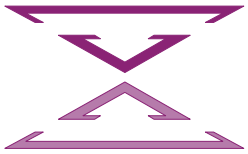


INFORMATION SCIENCE & COMPUTING

International Book Series Number 15

Knowledge – Dialogue – Solution

Supplement to
International Journal "Information Technologies and Knowledge" Volume 3 / 2009

ITHEA
SOFIA, 2009

Victor Gladun, Krassimir Markov, Vitalii Velychko, Alexey Volosyn, Krassimira Ivanova, Ilia Mitov (ed.)

Knowledge – Dialogue - Solution

International Book Series "INFORMATION SCIENCE & COMPUTING", Number 15

Supplement to the International Journal "INFORMATION TECHNOLOGIES & KNOWLEDGE" Volume 3 / 2009

Institute of Information Theories and Applications FOI ITHEA

Sofia, Bulgaria, 2009

This issue contains a collection of papers in the fields of Natural Language Processing, Pattern Recognition and Decision Making as well as other important areas of Artificial Intelligence.

Papers in this issue are selected from the XV-th International Conference "Knowledge-Dialogue-Solution", Kyiv, Ukraine – a part of the Joint International Events of Informatics "ITA 2009", Autumn Session.

International Book Series "INFORMATION SCIENCE & COMPUTING", Number 15
Supplement to the International Journal "INFORMATION TECHNOLOGIES & KNOWLEDGE" Volume 3, 2009

Edited by **Institute of Information Theories and Applications FOI ITHEA**, Bulgaria,
in collaboration with
- **V.M.Glushkov Institute of Cybernetics of NAS**, Ukraine,
- **Institute of Mathematics and Informatics, BAS**, Bulgaria,
- **Institute of Information Technologies, BAS**, Bulgaria.

Publisher: Institute of Information Theories and Applications FOI ITHEA, Sofia, 1000, P.O.B. 775, Bulgaria.
Издател: Институт по информационни теории и приложения ФОИ ИТЕА, София, 1000, п.к. 775, България
www.ithea.org, www.foibg.com, e-mail: info@foibg.com

General Sponsor: **Consortium FOI Bulgaria** (www.foibg.com).

Printed in Bulgaria

ISSN 1313-0455 (printed) **ISSN 1313-048X (online)** **ISSN 1313-0501 (CD/DVD)**

PREFACE

The scope of the International Book Series “Information Science and Computing” (**IBS ISC**) covers the area of Informatics and Computer Science. It is aimed to support growing collaboration between scientists from all over the world. IBS ISC is official publisher of the works of the members of the ITHEA International Scientific Society.

The official languages of the IBS ISC are English and Russian.

IBS ISC welcomes scientific papers and books connected with any information theory or its application.

IBS ISC rules for preparing the manuscripts are compulsory.

The rules for the papers and books for IBS ISC are given on www.foibg.com/ibsisc.

The camera-ready copyes of the papers should be received by ITHEA Submition System http://ita.ithea.org.

The camera-ready copyes of the books should be received by e-mail: info@foibg.com.

Responsibility for papers and books published in IBS ISC belongs to authors.

The Number 15 of the IBS ISC contains collection of papers from the fields of Knowledge Processing, Natural Language Interface, and Decision Making. Papers are peer reviewed and are selected from the XV-th International Conference “Knowledge-Dialogue-Decision”, (KDS) Kiev, Ukraine, a part of the Joint International Events of Informatics "ITA 2009" – autumn session.

ITA 2009 has been organized by

ITHEA International Scientific Society

in collaboration with:

- Association of Developers and Users of Intelligent Systems (Ukraine)
- V.M.Glushkov Institute of Cybernetics of National Academy of Sciences of Ukraine
- Taras Shevchenko National University of Kiev (Ukraine)
- Institute of Information Theories and Applications FOI ITHEA
- International Journal “Information Theories and Applications”
- International Journal "Information Technologies and Knowledge"
- Institute of Mathematics and Informatics, BAS (Bulgaria)
- Institute of Information Technologies, BAS (Bulgaria)
- Institute of Mathematics of SD RAN (Russia)
- Dorodnicyn Computing Centre of the Russian Academy of Sciences
- Universidad Politecnica de Madrid (Spain)
- BenGurion University (Israel)
- Rzeszow University of Technology (Poland)
- University of Calgary (Canada)
- University of Hasselt (Belgium)
- Kharkiv National University of Radio Electronics (Ukraine)
- Astrakhan State Technical University (Russia)
- Varna Free University “Chernorizets Hrabar” (Bulgaria)
- National Laboratory of Computer Virology, BAS (Bulgaria)
- Uzhgorod National University (Ukraine)

The main ITA 2009 events were:

KDS XVth International Conference "Knowledge - Dialogue – Solution"

i.Tech Seventh International Conference "Information Research and Applications"

MeL Fourth International Conference "Modern (e-) Learning"

INFOS Second International Conference "Intelligent Information and Engineering Systems"

CFDM International Conference "Classification, Forecasting, Data Mining"

GIT Seventh International Workshop on General Information Theory

ISSI Third International Summer School on Informatics

The XV-th anniversary of the International Conference "Knowledge-Dialogue-Decision" (KDS) is a remarkable event of ITA 2009.

KDS was established as result of the scientific co-operation organized within 1986-1992 by two international workgroups (IWG) researching the problems of data bases and artificial intelligence. As a result of tight relation between these problems in 1990 in Budapest appeared the International scientific Working Group of "Data Base Intellectualization" (IWGDBI) integrating the possibilities of databases with the creative process support tools. Heads of the IWGDBI ware V.Gladun (Ukraine) and R.Kirkova (Bulgaria).

In 1992 the "Association of Developers and Users of Intelligent Systems" (ASPIS) had been established. ASPIS was chosen to be the main organizer of the international conferences KDS. In the same time the "Institute of information theories and application FOI ITHEA" (ITHEA) started working in Bulgaria. During the years, in collaboration with ITHEA, KDS had been organized in Ukraine, Poland, Russia, and Bulgaria. From 2008 KDS is organized as two connected ITHEA ISS events: in Bulgaria (summer) and in Ukraine (autumn).

More information about ITA 2009 International Conferences is given at the www.ithea.org.

The great success of ITHEA International Journals, International Book Series and International Conferences belongs to the whole of the ITHEA International Scientific Society.

We express our thanks to all authors, editors and collaborators who had developed and supported the International Book Series "Information Science and Computing".

General Sponsor of IBS ISC is the **Consortium FOI Bulgaria** (www.foibg.com).

Kiev-Sofia, October 2009 *V. Gladun, Kr. Markov, V. Velychko, A. Volosyn, Kr. Ivanova, I. Mitov*

TABLE OF CONTENTS

Computing

Decision Making

Economics Decision Support Systems

Philosophy and Methodology of Informatics

INDEX OF AUTHORS

Intelligent NL Processing

СЕМАНТИЧЕСКИЙ АНАЛИЗ ТЕКСТОВ ЕСТЕСТВЕННОГО ЯЗЫКА: ЦЕЛИ И СРЕДСТВА

Леонид Святогор, Виктор Гладун

Аннотация: *В данной работе предлагается расширенное толкование понятия «текст естественного языка» и предлагается схема полного освоения его семантического ресурса за счёт «компьютерного понимания» и диалога. Указываются средства достижения указанной цели в процессе семантической обработки текстов – использование трёхуровневой онтологии для извлечения из текста онтологического смысла, а также ввод обратной связи для дополнительного уточнения в диалоге содержания дискурса.*

Ключевые слова*: семантический анализ текста, онтология, смысл, диалог.*

ACM Classification Keywords: *1.2.7 Natural Language Processing - Text analysis*

Conference*: The paper is selected from XV*th *International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.*

Введение

В начале и в конце семантического анализа естественно-языковых текстов стоит Слово*.* Методы анализа разнообразны и зависят от решаемой в прикладной области задачи, и существует не одно направление обработки текстовой информации. В условном разделении можно выделить методы семантической обработки текстов, которые нацелены на «лингвистические преобразования», например – *перевод на иностранный язык и обратно; краткий пересказ; конспектирование; тезисное представление; аннотирование* и на решение других прагматических задач. С другой стороны, у исследователей искусственного интеллекта интерес к тексту лежит в области «извлечения знаний» – *классификация сообщений, ответы на вопросы, контекстный перевод и понимание дискурсов* [Sowa, 2002]. Здесь применяются методы концептуального анализа. При этом можно заметить оформление двух проблем: **(а)** синтез систем представления знаний – онтологий и **(б)** разработка систем семантического анализа и машинного «понимания» текстов при помощи онтологий.

Проблема **(а)** решается широким фронтом; из последних, практически успешных работ можно указать на исследование [Палагин и др., 2009], где из корпуса профессиональных текстов автоматически извлекается подструктура знаний в одном из разделов предметной области (ПрО) «Материаловедение». Для синтеза онтологии используются формально-логические и синтаксические средства анализа.

Следует заметить, что «конкурирующим» подходом может служить разработка структур знаний при помощи экспертов и инженеров по знаниям. В этом случае готовой базой для разработчиков онтологий

служат учебники, свежие публикации и другие пособия по описанию ПрО [Поспелов, 1988; Гаврилова и Хорошевский, 2001; Ной и МакГиннесс, 2001].

В проблеме **(б)** наш подход состоит в следующем [Гладун и др., 2008; Святогор и Гладун, 2009].

Если *описание* ситуации, изложенной в тексте, может быть достигнуто чисто лингвистическими средствами, то *понимание* ситуации возможно за рамками лингвистического ресурса текста – мобилизацией когнитивных усилий человека и его индивидуальных знаний. Например, как отмечает Г.С. Поспелов, связное восприятие текста возможно лишь при его понимании.

Аналогично тому, как человеческое понимание рождается при согласовании внешней информации с его ментальной (когнитивной) моделью мира, «компьютерное понимание» может быть достигнуто отображением информации на определённую и формально-заданную систему знаний. Проще говоря, чтобы «понимать» что-то, надо его «узнавать». В машинной обработке текстовой информации роль памяти человека выполняет компьютерная система формальной репрезентации знаний – онтология: именно она позволяет совместить анализ текста с его компьютерным «пониманием». Процедурно это достигается достаточно просто: необходимо найти проекцию текста на компьютерную онтологию.

Говорить о «понимании информации» можно лишь в контексте окружающего её знания.

Конкретная задача искусственного интеллекта состоит в следующем. Задан текст ЕЯ, или сообщение. То, «о чём говорится», можно назвать *темой*, *содержанием*, *интенцией сообщения* или *коммуникативным смыслом*; этот смысл требуется из предъявленного текста ***извлечь*** [Мальковский, 1985; Штерн, 1998, стр.145].

Проверка «качества понимания», или релевантности текста извлечённому смыслу, происходит за рамками онтологии, например – экспертной оценкой смысла или по результату принятого решения.

Актуальность и цель работы

Машинное понимание языка является парадигмой искусственного интеллекта.

Кардинальным вопросом остаётся один: что мы хотим получить от текста?

Мы хотим получить его формальный (компьютерный) смысл. Для этого рассматривается структура Системы семантического анализа ЕЯ текстов, в которой должны быть предусмотрены и объединены базовые процедуры: грамматический *анализ*, взаимодействие текста с *онтологией*, получение результата – формального *понимания* текста через онтологию и, наконец, – уточнение его *смысла*. В конечном итоге, предлагаемая технология семантического анализа преследует цель – добиться лучшего взаимного понимания автора текста и его потребителя через компьютер, общую базу знаний и родной язык.

Взгляд на семантический ресурс ЕЯ текста

Потенциальные возможности текстового документа намного выше тех, что мы используем, и для того, чтобы «понимать» текст, нужно вначале выяснить его потенциальные возможности.

Текст рассматривается не только как вместилище информации – данных, фактов и знаний, которые требуется из него извлечь. *Он представляет собой языковое, информационное и культурное явление,* которое актуально для данного периода существования социума и может быть востребовано потомками. Отсюда следует, что текстовый материал изначально, априори «погружён» в некоторую общечеловеческую систему накопления и интерпретации знаний, в которой он сам был *порождён*. С другой стороны, текст генерируется как индивидуально, так и коллективно и может быть *востребован* также индивидуально или коллективно. Это означает, что содержание («семантическое наполнение»)

материала часто является *многоплановым*, и каждый план имеет свою глубину изложения. Выполняя свою коммуникационную функцию, *текст обязан быть понимаемым пользователем, то есть – должен* ***отображаться в базе знаний потребителя*** и взаимодействовать с ней.

Семантический компонент текста ЕЯ давно был зафиксирован лингвистами, которые определили предложение как выражающее законченную мысль. Какую мысль? Как её кратко и неискажённо сформулировать? Поиски ответа продолжаются в рамках когнитивной лингвистики [Штерн, стр.129], психологии [Балл, 2006], искусственного интеллекта [Поспелов, 1988]. В произвольном дискурсе мысль облачена в лексическую оболочку в соответствии с правилами грамматики, и в ряде случаев эту оболочку надо «сбросить». В интересующих нас случаях документ или сообщение несёт актуальную информацию или стабильное знание. ***Концентрированное выражение знания мы называем смыслом***. Это концентрированное знание, или смысл, надо извлечь, чтобы потом с ним оперировать. Следовательно, текст нужно хранить и беречь, поскольку он содержит определённый авторский замысел, интеллектуальный ресурс и в социогуманистическом плане является продолжением баз знаний.

Признавая ёмкий семантический ресурс текста, мы приходим к выводу, что во многих приложениях его потребительская ценность быстро не исчерпывается. Извлечение смысла вряд-ли является одноразовой операцией коммуникации. Можно надеяться на то, что раскрыть содержательный ресурс текста полностью (если это возможно!) удастся многократным к нему обращением. В этом состоит ***активная функция*** документа, которая используется в некоторых приложениях для глубинного семантического анализа [Поспелов, 1988; Штерн, стр.70].

Какие выводы следуют из расширенного толкования понятия «Текст»?

Решению задачи полного раскрытия семантического ресурса текста, на наш взгляд, способствует такая Система семантического анализа ЕЯ текстов (Система), которая удовлетворяет следующим требованиям:

Первое. Партнёры интеллектуального общения вместе с текстом погружены в единую компьютерную среду онтологического знания.

Второе. Предварительная лингвистическая обработка исходного текста (морфологический, синтаксический и семантический анализ предложений) необходима для снятия «лексической оболочки» и выделения термов, несущих содержательную нагрузку**.**

Третье. Результатом компьютерного семантического анализа связного текста должен быть формальный или адаптированный текст ЕЯ, который выражает его смысловое содержание.

Четвёртое. Система должна обеспечивать самоконтроль авторского намерения – насколько адекватно он выражает свои мысли.

Пятое. Система должна многократно активизировать текст с целью более глубокого проникновения в смысл сообщения.

В результате самого общего взгляда на желаемые качества Системы семантического анализа можно сделать вывод, что потенциальные возможности текста реализуются при помощи двух механизмов: ***анализа через онтологию и активного диалога.***

Средства обеспечения интеллектуальных функций Системы семантического анализа

Первое требование – взаимопонимание партнёров коммуникации – обеспечивается единой системой представления общих и профессиональных знаний, накопленных в социуме. В качестве контекстной среды общения предлагается формальная семантическая сеть – ***иерархическая трёхуровневая***

онтология (см. ниже), сформулированная в работе [Гладун и др., 2008], которая может быть расширена и дополнена спектром любых предметно-ориентированных онтологий [Поспелов, 1988; Гаврилова и Хорошевский, 2001; Палагин и др., 2009].

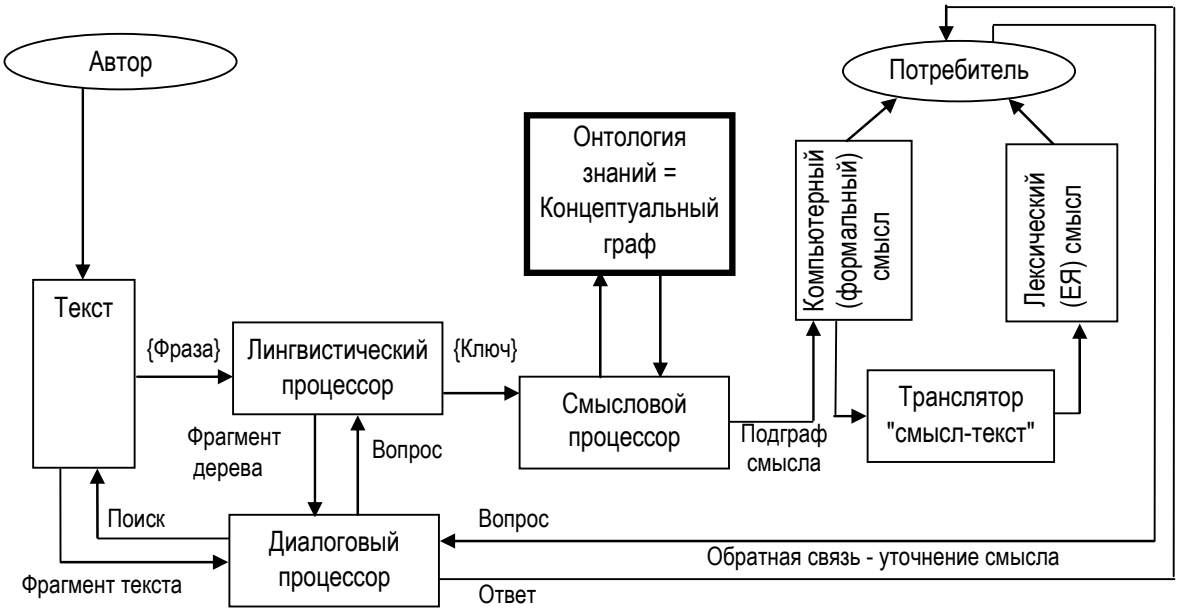
Второй тезис – предварительный разбор текста – выполняется ***лингвистическим процессором***, ориентированным на семантический анализ обычной текстовой информации [Поспелов, 1988; Штерн, стр.261]. Например, в лингвистической модели «Смысл–Текст» используются пять уровней репрезентации выражения, включая синтаксический, морфологический и семантический уровни [Мельчук, 1999]. В самом простом случае от лингвистического процессора требуются: построить дерево синтаксического разбора, выделить ядерные конструкции предложений, построить отношения, определить «значимые» лексические группы, в частности – ключевые слова текста [Палагин и др., 2009].

Третье условие означает, что из текста необходимо извлечь его смысловое содержание. Смысл связного текста, согласно его *определению* в работе [Святогор и Гладун, 2009], формализуется через онтологию – как совокупность подграфов концептуального графа. Задача выявления смысла в некотором текстовом фрагменте (и в целом тексте) возлагается на ***смысловой процессор***. Ниже приводится разъяснение онтологического смысла и кратко описывается автоматическая процедура его получения.

В четвёртом требовании предусмотрена сервисная возможность ***коррекции текста его автором***. При желании он сравнивает результат автоматического выделения смысла со своими внутренними намерениями.

Наконец, для более глубокого раскрытия смысла, уточнения фактов и других данных (*пятое условие*) предусмотрен режим диалога, который реализуется на естественном языке ***диалоговым процессором***. В процессе диалога смысл может существенно измениться, что, в свою очередь, может служить поводом для корректировки онтологии (активность текста).

Как смысловой, так и диалоговый процессоры выполняют интеллектуальную миссию, привлекая внелингвистические знания. К ним следует добавить вспомогательный ***транслятор «смысл – текст»,*** он (в необходимых случаях) поможет человеку более содержательно истолковать формальный подграф смысла.



Блок-схема Системы семантического анализа ЕЯ текстов

На рисунке показана блок-схема Системы семантического анализа ЕЯ текстов, в которой взаимодействуют указанные выше функциональные компоненты.

После установления структуры Системы дадим краткое описание её составных частей, имеющих – в контексте данной работы – принципиальное значение.

Иерархическая трёхуровневая онтология внешнего мира

В общем случае в искусственном интеллекте понятие «картина мира» появляется как синоним понятий «модель мира», «модель предметной области» [Штерн, стр.156]. В области *Knowledge representation* исследователями предложено немало онтологических систем репрезентации концептуальных знаний о мире (известные онтологии Дж. Совы, Микрокосмос, РуТЕЗ* и другие). Обладая мощным философским и лексическим потенциалом, общие онтологии скорее относятся к области гносеологических моделей, чем к системам, пригодным для семантического анализа текстов живого языка: взаимодействие онтологий с реальным текстом, так же как и результаты онтологической работы, остаются не определёнными.

В предлагаемой структуре Системы семантического анализа ЕЯ текста онтологии отводится ведущая роль. Как система отображения общественного интеллекта она позволяет интерпретировать текстовую информацию на языке общих знаний и объединяет тройку «автор – текст – потребитель» в единую интеллектуальную среду. Любое явление может быть понято и интерпретировано *только в контексте общепринятого и стабильного знания*.

Базисом Системы служит новая иерархическая трёхуровневая онтология – **ИО*3** [Гладун и др., 2008]. Она отличается двумя особенностями: **(а)** сетевая структура даёт принципиальную возможность объединить – *в рамках единой конструкции* – знания высшего уровня абстракции, общедоступные (повседневные и актуальные) знания среднего уровня и профессиональные знания нижнего уровня; **(б)** одновременно она ориентирована на работу с конкретными текстами. Кроме того, показано, что **(в)** результатом извлечения из текста знаний должен быть *«онтологический смысл».* Этот смысл поддаётся строгой формализации и компьютерной обработке [Святогор и Гладун, 2009].

***Иерархическая трёхуровневая онтология* ИО*3** представляет собой семантическую сеть в форме концептуального ориентированного графа.

*«**Концептуальный граф** – это способ семантической (понятийной) репрезентации ситуаций и знаний в моделях понимания естественного языка, принятия решений, рассуждений и т.д.»* [Штерн, стр.195]. Узлами онтологического графа (*онто-графа*) служат лексические единицы – слова естественного языка, которые трактуются как ***категории*** и ***понятия***. *«Целью категоризации является пояснение нового через известное и структурирование мира при помощи обобщений»* [там же, стр.160]. Понятием служит *«денотат некоторой сущности или явления, который кодируется языковым знаком»*. Категории и понятия называют ***концептами,*** которые в лексической семантике выступают в роли элементов «понятийного метаязыка» [там же, стр.191]. По образному определению Дж. Совы *«существует мост между языком и приложением* (ПрО). *Он не нуждается в полной информации о языке и приложении, но он должен содержать крючки (зацепки), которые привязывают языково-зависимые слова к языково-зависимой грамматике и к языково-независимым, но зависимым от приложений, концептуальным структурам»* [Sowa, 2002].

***Онтология* ИО*3** организована как «пирамида концептуальных знаний». Концепты обладают разной степенью обобщения. Наиболее абстрактные категории образуют верхний уровень онтологии; в соответствии с парадигмой академика В.И. Вернадского о биосфере и ноосфере, это – *Материя, Вещество, Жизнь, Разум*… Концепты среднего уровня образуют описательный континуум знаний. Они

раскрывают значения категорий верхнего уровня через более употребительную в актуальной деятельности общества лексику, например: *Время*, *Движение, Порядок, Человек, Общество, Организация, Развитие, Управление, Транспорт, Биология, Борьба за существование* и другие. На нижнем уровне пирамиды знаний располагаются концепты двух типов: часть из них обозначают обиходные понятия повседневной жизни, привычные объекты и ситуации (*комната, ложка, верх*…), другие концентрируются вокруг профессиональных знаний ПрО (*концепт, отношение, онтология*…). Множество концептов ПрО может быть пустым.

Узлы онто-графа соединяются формальными и неформальными (*ассоциативными*) связями. Ориентация связей в графе направлена сверху вниз – от концептов более высокого уровня обобщения к концептам, которые их характеризуют.

Контакт текста с онтологией происходит на нижнем уровне **ИО*3;** соответствующая процедура описана ниже.

Почему разработке «хорошей онтологии» придаётся столь большое значение? Ответ можно найти в принципиально важном утверждении Р. Шенка**: *«Метаязыком для внутренней смысловой репрезентации текстов является граф концептуальных зависимостей, который отображает смысловую структуру ситуации»*** [Шенк, 1980].

Фантом «компьютерного смысла»

Понятие «смысл ситуации» является очень ёмким и трактуется в разных дисциплинах по-разному. Первоначально понятия «смысл» и «мысль» имели ментальный характер – это показывает семантическая близость слов ***«смысл»*** и ***«со-мысль»***. К формализации данного понятия первыми подошли лингвисты: модель И. Мельчука «Смысл–Текст» признаётся классикой лингвистической теории.

В обширной лингвистической литературе имеется много толкований понятия ***«смысл».*** Смысл соотносится с целями коммуникации [Штерн, стр.185]; трактуется как «структура ситуации» [там же, стр.197]; связывается с категорией «понимания» текста [там же, стр.283] и т.д. Смысл может быть описан совокупностью ***денотатов***, освобождённых от эмоциональных, модальных, прагматических, стилистических и других оттенков [там же, стр.81]. Недостаток подобных определений – в их чрезвычайно слабой формализации.

С другой стороны, потребности автоматизации процессов семантического анализа ЕЯ текстов стимулируют развитие вычислительной лингвистики и компьютерных моделей [Демьянков,1985; Мальковский, 1985], а также компьютерную поддержку этих процессов [Zaboleeva–Zotova & Orlova, 2009]. В конечном итоге методы вычислительной лингвистики сводятся к морфологической интерпретации, операциям над синтаксисом и семантикой, а в практическом плане решают частные задачи. Выделить именно ***продукт семантической обработки*** – универсальный и стандартизованный результат, который затем можно использовать в совершенно разных приложениях, – такая цель не преследуется.

Фантом «компьютерного смысла текста» ускользает.

Вместе с тем потребности автоматического анализа ЕЯ текстов настоятельно требуют ответа на вышеприведенный кардинальный вопрос: что может служить универсальным продуктом семантической обработки текста? Разделяя позицию Р. Шенка, мы в качестве такого универсального продукта вводим в рассмотрение ***«формализованный смысл».***

На этом пути может материализоваться и оформиться «компьютерный смысл».

Онтологический смысл

На помощь в определении понятия «смысл» приходит онтология **ИО*3**.

Идея состоит в том, что если «пропустить» текст через онтологию, которая является структурой знаний, то на выходе получим ***концентрированное знание***, которое коррелирует с текстом. Концептуальный фильтр онтологии даст на выходе концептуальный, или *онтологический смысл*.

В работе [Святогор и Гладун, 2009] приводится формальное определение «онтологического смысла»; здесь оно повторяется тезисно.

Задан концептуальный онтологический граф **ИО*3.**

Элементарный смысл определяется как пара соединённых соседних узлов онтологического графа. Связи не обязательно именуются, они могут лишь фиксировать факт некоторого взаимодействия двух слов (например, *ворона–птица, пассажир–самолёт, развитие–прогресс*). Онто-граф состоит из множества связанных между собою элементарных смыслов, которые вступают в дозволенные комбинации. Связная часть онто-графа, соединяющая два удалённых узла, образует ***подграф;*** при изменении в нём стрелок на противоположные (снизу – вверх) получается *цепочка подграфа*.

Цепочка связанных элементарных смыслов, которая начинается в некотором «активном» узле и заканчивается в вершине онтологии, образует *онтологическую цепочку активного узла*. Цепочка, выделяемая активным узлом на онтологическом графе, трактуется как *смысловая траектория* и называется ***онтологическим смыслом активного слова.***

Комментарий. Процесс возбуждения смысловой траектории начинается с того, что в предложении из ядерной конструкции выделяется некоторое «ключевое слово». Если оно присутствует в онтологическом графе, то активное слово возбуждает соседний концепт, возбуждение передаётся дальше на высшие уровни онтологии – вплоть до вершины пирамиды. Результатом процесса является цепочка, то есть –

дискретная упорядоченная последовательность взаимосвязанных концептов; она является формальным онтологическим смыслом входного слова в заданной «картине знаний».

Пример. В ядре предложения выделены ключи: **ворона** и **сыр**. Для них будут построены соответствующие концептуальные цепочки: 1) **ворона** – *Птица – Полёт – Движение – Биосфера – Жизнь – Материя* и 2) **сыр** – *Еда – Жизнь – Материя.*

Связи пока-что не интерпретируются – в данной репрезентации онтологического смысла они не имеют значения: необходимо и достаточно зафиксировать *только связь пары объектов*. Далее, поскольку ключи находятся внутри одного предложения, то автоматически будет построена связь, ранее в онтологии отсутствовавшая: **ворона – сыр**. В итоге, линейный формат онтологического смысла будет следующий:

ворона (***птица, полёт, движение, биосфера, жизнь, материя***) **–**

– сыр (***еда, жизнь, материя***).

Семантическое пояснение суммарной цепочки, в случае необходимости, может быть дано позже, а связи могут быть интерпретированы *повторным обращением* к тексту (ворона ***имеет*** сыр).

В зависимости от целей семантического анализа цепочки можно укорачивать за счёт абстрактных категорий. Кроме того, цепочки могут быть как линейными, так и разветвлёнными, то есть – с присоединением узлов, примыкающих к траектории.

Результатом полного просмотра текста является множество – «пучок смысловых траекторий», который можно трактовать как «семантический портрет текста».

Онтологический смысл может быть ***целью и результатом*** семантического анализа ЕЯ текста благодаря таким свойствам:

– ключевые слова в смысловой цепочке извлекаются непосредственно из текста;

– эти слова помещаются в контекст общих знаний, которые организованы как концептуальная смысловая среда (онтология);

– множество смысловых цепочек даёт краткое, дискретное и формализованное описание текста (фрагмента текста) – «семантический портрет текста» в терминах общих знаний.

Функцию выделения в тексте значимого слова выполняет лингвистический процессор**.** Функция выделения в онтологическом графе смысловой траектории возлагается на смысловой процессор**.** Функцию адаптированного представления смысла выполняет транслятор «смысл – текст» (см. рисунок).

Онтологический смысл, извлечённый из текстового документа компьютерной системой, становится ***элементом Базы знаний***, которая доступна всем партнёрам коммуникации.

Понимание связного текста

Определив формально онтологический смысл, мы можем в прагматическом плане говорить о «компьютерном понимании текста». Под этим будем подразумевать, что: **(а)** компьютер выявляет «замысел автора», интерпретирует его внутри собственной системы знаний и **(б)** производит над смыслом определённые операции, в том числе – даёт ему естественно-языковую грамматическую интерпретацию [Штерн, стр. 60; Мальковский, 1985; Демьянков, 1985].

Предложенный аппарат компьютерного понимания включает триаду ***«Онтология – Текст – Смысл».*** Графические структуры триады непосредственно поддерживают процесс понимание текста, потому что семантическая сеть в явном виде способна отвечать на основные вопросы понимания, например: «***Что связывает*** *кошку с мышкой?*», «***Что общее*** *существует между человеком и птицей*?», «***Чем отличается*** *человек от птицы?*». Подчеркнём, что ответы даются не в текстовом лексиконе, а на концептуальном уровне общего знания – на метаязыке онтологии.

В то же время говорить о понимании авторского замысла, имея дело с многоплановым и лексически-избыточным текстовым документом, довольно сложно.

Как указывалось, онтологический смысл, который является продуктом семантического анализа полного текста, реализуется в виде пучка траекторий, мощность которого зависит от длины текста. Траектории активизируются ключевыми словами и «накрывают» весь текст дискретно, но не хаотично. Последовательность развития ситуации сохраняется.

Траектории упорядочены по времени их появления, они семантически взаимосвязаны одной фразой, абзацем, разделом. Отдельные ключевые слова вступают во взаимодействие через общие концепты более высокого уровня. Отдельные траектории пересекаются и частично сливаются, причём их концепты пересекаются в разных комбинациях и с разной частотой. Этот сложный механизм структурно и схематично отражает, в принципе, всю семантическую сложность и ***связность текста***, все повороты его тематики. «Неучтёнными» остаются лишь детали, подробности, числовые данные и т.п. Однако уточнение деталей принципиально выходит за рамки выявления смысла и требует другой технологии.

Фундаментальная ценность механизма выявления онтологического смысла кроется в том, что он, создавая графический портрет текста и описывая его метаязыком онтологии, позволяет человеку сложить хотя и самое общее и схематичное, но вполне адекватное представление о ситуации, дать ему концентрированную информацию, возбудить целенаправленные вопросы, отсеять лишние гипотезы.

В итоге потребитель получает определённую ясность – в чём состоит суть сообщения.

Что касается человеческого понимания онтологического смысла, формат которого непривычен для (современного) человека, то для его преобразования в грамматическую языковую форму предусмотрен, как указано выше, специальный транслятор «смысл – текст». Принципы трансляции разработаны в лингвистической модели «смысл–текст» [Мельчук, 1999].

Раскрытие смысла в диалоге

«Текст обогащает смысл» (эта формула принадлежит С. Васильеву, 1988 г).

Как было показано выше, активная функция текста проявляется в том, что он является, в принципе, неисчерпаемым источником интереса, причём при повторном чтении возможно не только переосмысливание дискурса, но даже изменение онтологии **(*Sic!*).** Особенно это относится к учебным материалам и высокохудожественным произведениям. Поэтому необходимо обеспечить многократный доступ пользователя к первоисточнику для более полного раскрытия его ресурса. Это возможно в режиме диалога.

В теории репрезентации знаний используют понятие ***«страты знаний»*** [Гаврилова и Хорошевский, 2001]. Стратификация знаний производится по типу их анализа, при этом различают: *Зачем* – знания*, Кто* – знания*, Что* – знания, *Как* – знания, *Где* – знания, *Почему* – знания и т.д.

Если в упрощённом виде использовать эту методологию для организации диалога, то следует модернизировать лингвистический процессор. В процессе построения дерева синтаксического разбора предложения связи (стрелки) между членами предложения должны быть проиндексированы вопросами, например, такими: *кто, что-делает, где, когда, какой, сколько, зачем, каким-образом* и т.д. Такая индексация создаёт для диалога лингвистическую базу. Вопросы к тексту, сформулированные пользователем, активизируют соответствующие группы слов, которые формируют ответ и тем самым раскрывают «глубинные падежи» ситуации [Поспелов, 1988; Мальковский, 1985; Штерн, стр. 71].

Заключение

Предлагаемый взгляд на общие ресурсы текста ЕЯ, на задачи его семантического анализа, способы и результат обработки приводит к единой триаде: *«Онтология – Текст – Смысл»*. Задачей семантического анализа текста полагается извлечение концентрированного знания, релевантного замыслу автора. Платформой извлечения знания служит онтология **ИО*3** – концептуальная система репрезентации общих знаний о мире и предметных областях. Результатом взаимодействия текста с онтологией является онтологический смысл – множество взаимосвязанных подграфов онтологического графа.

Онтологический смысл извлекается из онтологического графа «смысловым процессором» и интерпретируется с помощью транслятора «смысл – текст».

Для более глубокого изучения содержания документа используется «диалоговый процессор», который исследует дерево синтаксического разбора предложения и по заданному вопросу находит в тексте фрагмент, служащий конкретным ответом на вопрос пользователя.

Компьютерное понимание текста достигается за счёт: 1) погружения текста в единую среду знаний – онтологию, 2) формального представления смысла в памяти компьютера и 3) возможности операций над онтологическим смыслом.

Возможные применения

Предлагаемую новую информационную технологию можно использовать для формирования Баз данных, архивирования электронных документов, их индексирования, классификации и поиска в Интернет. В виртуальных лабораториях возможно на её основе создавать интеллектуальные банки данных, работающие в единой среде знаний.

Данная технология ориентирована на автоматическое извлечение ***метаданных*** из текстовых документов. При соответствующей доработке она может служить в системах автоматического реферирования научных публикаций, а в перспективе – для осмысленного интерпретирования мультимедийных документов.

Благодарности

Работа опубликована при финансовой поддержке проекта **ITHEA XXI** Института информационных теорий и приложений FOI ITHEA Болгария www.ithea.org и Ассоциации создателей и пользователей интеллектуальных систем ADUIS Украина www.aduis.com.ua.

Литература

[Sowa, 2002] J.F. Sowa. Concepts in Lexicon: Introduction. Architectures for Intelligent Systems. – IBM Systems Journal, vol. 41, No. 3, 2002. – pp. 331 – 349.

[Zaboleeva–Zotova & Orlova, 2009] A. Zaboleeva–Zotova, Y. Orlova. Computer Support of Semantic Text Analysis of a Technical Specification on Designing Software. – International Book Series, Number 9. Intelligent Processing. Supplement to the International Journal "Information Technologies & Knowledge" Volume 3 / 2009. – ITHEA, Sofia, 2009. – p. 29.

[Балл, 2008] Психология в рациогуманистической перспективе. Избранные работы. – К.: «Основа», 2006. – 401 с.

[Гаврилова и Хорошевский., 2001] Гаврилова Т.А., Хорошевский В.Ф. Базы знаний интеллектуальных систем. – СПб.: Питер, 2001. – 384 с.

[Гладун и др., 2008] В. Гладун, В. Величко, Л. Святогор. Структурирование онтологии ассоциаций для конспектирования естественно–языковых текстов – International Book Series, Number 2. Advanced Research in Artificial Intelligence. Supplement to the International Journal "Information Technologies & Knowledge" Volume 2 / 2008. – ITHEA, Sofia, 2008. – p. 153.

[Демьянков, 1985] Демьянков В.З. Основы теории интерпретации и её приложения в вычислительной лингвистике. – М.: Изд-во Моск. ун-та, 1985. – 76 с.

[Мальковский, 1985] М.Г. Мальковский. Диалог с системой искусственного интеллекта. – М.: Изд-во Моск. ун-та, 1985. – 213 с.

[Мельчук, 1999] Мельчук И.А. Опыт теории лингвистических моделей «Смысл – Текст». – М.: Школа «Языки русской культуры», 1999. – 346 с.

[Ной и МакГиннесс] Natalya F. Noy, Deborah L. McGuinness: http://protege.stanford.edu/publications/ontology_development/ontology101.html

[Палагин и др., 2009] А. Палагин, С. Крывый, В. Величко, Н. Петренко. К анализу естественно-языковых объектов – International Book Series, Number 9. Intelligent Processing. Supplement to the International Journal "Information Technologies & Knowledge" Volume 3 / 2009. – ITHEA, Sofia, 2009. – p. 36.

[Поспелов, 1988] Г.С. Поспелов. Искусственный интеллект – основа новой информационной технологии. – М.: Наука, 1988. – 279 с.

[Святогор и Гладун., 2009] Л. Святогор, В. Гладун. Определение понятия «Смысл» через онтологию. Семантический анализ текстов естественного языка. – International Book Series, Number 9. Intelligent Processing. Supplement to the International Journal "Information Technologies & Knowledge" Volume 3 / 2009. – ITHEA, Sofia, 2009. – p.53.

[Шенк, 1980] Шенк Р. Обработка концептуальной информации. Пер. с англ. – М.: Энергия, 1980.

[Штерн, 1998] І.Б. Штерн. Вибрані топіки та лексікон сучасної лінгвістики. Енциклопедичний словник. – К.: «АртЕк», 1998. – 335 с.

Информация об авторах

Святогор Леонид Александрович – *Ин-т кибернетики им. В.М. Глушкова НАН Украины, Киев-187 ГСП, 03680, просп. акад. Глушкова, 40, e-mai: aduis@rambler.ru*

Гладун Виктор Поликарпович – *Ин-т кибернетики им. В.М. Глушкова НАН Украины, Киев-187 ГСП, 03680, просп. акад. Глушкова, 40, e-mai: aduis@rambler.ru*

БАЗОВАЯ ОНТОЛОГИЯ РАСПРЕДЕЛЕННОЙ ВИРТУАЛЬНОЙ ЛАБОРАТОРИИ ПРОЕКТИРОВАНИЯ СЕНСОРНЫХ СИСТЕМ

Александр Палагин, Владимир Романов, Крассимир Марков, Виталий Величко, Игорь Галелюка, Крассимира Иванова, Петер Станчев, Илия Митов, Милена Станева

***Аннотация**: В статье рассмотрен алгоритм построения онтологии виртуальной лаборатории автоматизированного проектирования. Сформулированы требования, согласно которым разработана онтология. Приведены фрагмент глоссария и пример реализации разработанной онтологии специальными программными средствами.*

***Ключевые слова**: Виртуальная лаборатория; автоматизация проектирования; онтология.*

***ACM Classification Keywords**: J.6 Computer-Aided Engineering – Computer-Aided Design (CAD); K.4.3 Organizational Impacts – Computer-Supported Collaborative Work.*

***Conference**: The paper is selected from XVth International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.*

Введение

Описанная в [Palagin et al, 2009] компьютерная технология разработки сенсорных систем с помощью виртуальной лаборатории автоматизированного проектирования (virtual laboratories of computer-aided design – VLCAD) позволяет специалистам различных предметных областей, таких как химия, биология, биохимия, физика самостоятельно проверить возможность создания измерительного устройства и спроектировать новый прибор вплоть до разработки конструкторской документации. VLCAD создается на базе формализованного представления теоретических знаний, принципов организации, методов и средств автоматизированного проектирования и тестирования информационно-измерительных систем и приборов с использованием методологии системной интеграции [Палагин и Кургаев, 2003]. Для формализованного описания сложных систем, структурирования и представления знаний о некоторой предметной области в машинной форме все более широко используются онтологии. Онтология, как правило, описывает иерархию концептов предметной области и существенные свойства каждого концепта с помощью механизма "атрибут – значение". Связи между концептами могут быть описаны с помощью дополнительных логических утверждений. Эффективность использования онтологий особенно проявляется в таких наукоемких исследовательских областях, как техника представления и управления знаниями, моделирование объектов и процессов, проектирование баз данных, информационная интеграция и обнаружение знаний [Гладун, 1994]. Рассмотрим онтологическое представление VLCAD.

Построение онтологии виртуальной лаборатории автоматизированного проектирования

Под онтологией в данной работе будем понимать кортеж множеств $O = \langle X, R, F \rangle$, где X – конечное множество концептов (понятий) заданной предметной области — виртуальная лаборатория автоматизированного проектирования, R – конечное множество отношений между концептами X, F – конечное множество функций интерпретации, заданных на множествах X и/или R. В базовой онтологии примем, что множество F тождественно множеству аксиом A, представляющих истинные высказывания о

соответствующих понятиях *X*. По цели создания данная онтология относится к прикладной онтологии объектов VLCAD без определения отношений с онтологией верхнего уровня (метаонтологией). Общими требованиями к разработке онтологии являются:

1) ясность – онтология должна эффективно передавать содержание введенных концептов;

2) согласованность – все понятия и определения онтологии должны быть логически не противоречивы;

3) расширяемость – онтология должна допускать возможность расширять словарь терминов без необходимости ревизии существующих понятий.

4) структурированность – простота для понимания и поиска понятий, что достигается применением системологического подхода к анализу предметной области [Соловьева, 1999].

Процесс построения онтологии состоит из ряда этапов. При этом эти этапы выполняются не последовательно, а итерационно в зависимости от полноты и точности накопленных данных. Основные этапы построения онтологии включают: определение целей создания и использования онтологии, перечня вопросов, на которые позволяет дать ответ онтология; рассмотрение вариантов повторного использования онтологии; построение списка понятий, терминов предметной области и их свойств; определение классов объектов предметной области и их иерархии; определение свойств классов – слотов; задание ограничений и аксиом, построение диаграмм неиерархических (функциональных) отношений; ввод экземпляров объектов онтологии.

Цели построения онтологии VLCAD можно разделить на 2 группы – цели разработчика лаборатории и цели разработчика электронного устройства. Для разработчика лаборатории основной целью построения онтологии является формальное описание и классификация элементов виртуальной лаборатории и связей между ними для оценки функциональности текущего состояния VLCAD и планирования дальнейшей разработки лаборатории. С точки зрения разработчика электронного устройства основной целью онтологии является активная помощь пользователю при создании сенсорной системы с общением на ограниченном естественном языке. Ответ системы должен представлять собой описание процедур, методов и компонентов виртуальной лаборатории, которые позволяют решить задачу, поставленную в запросе пользователя. Для базовой онтологии VLCAD было принято ограничить функциональность следующим:

— предоставление информации, как о полной, так и о частичной структуре лаборатории в текстовом и графическом виде;

— предоставление информации о зависимостях между отдельными компонентами виртуальной лаборатории по запросу пользователя, включая входные и выходные данные для каждого компонента.

Для заданной предметной области создан глоссарий, который включает термины VLCAD и их естественно-языковые описания. Всего описано 62 термина, а фрагмент глоссария приведен в Таблице 1. В качестве языка для формального представления онтологии принят язык OWL (Web Ontology Language), а для непосредственного построения онтологии использовался редактор Protégé-OWL. Язык OWL поддерживает эффективное представление онтологий в терминах классов и свойств, обеспечение простых логических проверок целостности онтологии и связывание онтологий друг с другом. Логическая модель позволяет использовать механизм рассуждений, с помощью которого можно проверять последовательность построения утверждений и определений, корректно поддерживать иерархию концептов [Horridge et al, 2004].

Таблица 1

Термин	Описание термина
Базы данных	Файл (или несколько файлов) данных, которые хранятся в ПК или на носителях данных и содержат информацию, необходимую для проектирования устройств. Условно делятся на библиотеки.
Библиотека типовых функциональных решений	Является частью **Баз данных**. Содержит типовые функциональные решения для реализации функций (или подсистем) проектируемого устройства
Библиотека систем сбора данных	Является подклассом **Библиотеки типовых функциональных решений**. Содержит варианты реализаций (экземпляры) систем сбора данных, которые предназначены для получения измерительной информации и ее первичной обработки

На основе глоссария была построена классификационная схема с использованием таксономических отношений "ПОДКЛАСС-НАДКЛАСС" (subclass-of), "ЧАСТЬ-ЦЕЛОЕ" и "КЛАСС-ЭКЗЕМПЛЯР" (isInstanceOf). Схема иерархии классов VLCAD приведена на рисунке 1

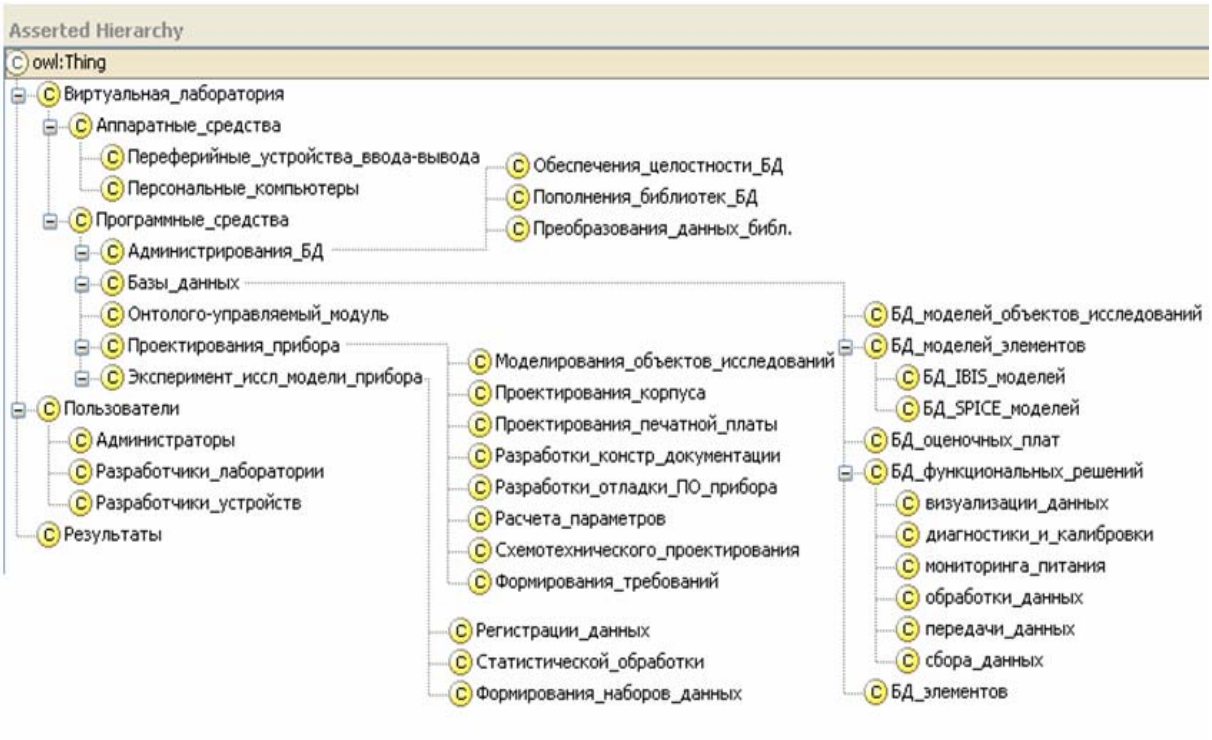


Рисунок 1. Классификационная схема виртуальной лаборатории.

Данная схема позволяет получить информацию о структуре лаборатории и состоянии разработки отдельных функциональных модулей, для чего используется добавленное свойство типа Annotation Datatype Property – *реализация компонента.*

Для описания функциональных отношений между отдельными компонентами виртуальной лаборатории заданы следующие объектные свойства (Object properties): *обеспечивают работу, сохраняются на,*

формируют ↔ является результатом, используются, управляет. Выбранные отношения позволяют находить необходимые пользователю зависимости между компонентами лаборатории. Функциональные отношения в редакторе Protégé-OWL опишем через ограничения, задаваемые для классов. Фрагмент схемы функциональных отношений приведен на рисунке 2.

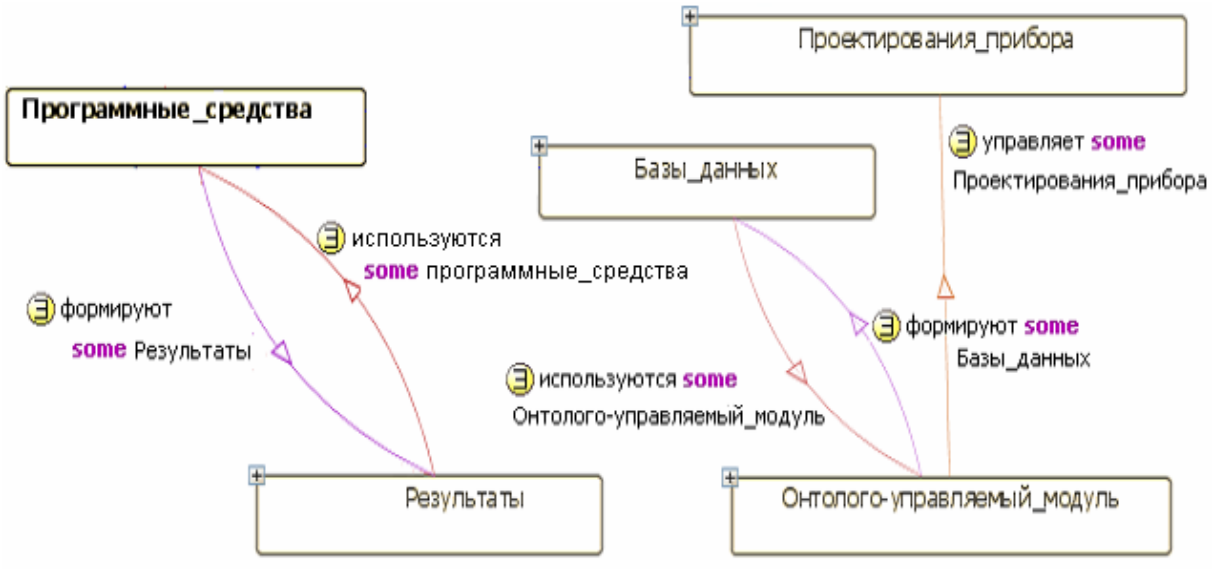


Рисунок 2. Фрагмент схемы функциональных отношений между классами VLCAD.

После определения функциональных отношений в онтологию добавляются экземпляры объектов.

Выводы

1. Разработана базовая онтология распределенной виртуальной лаборатории проектирования сенсорных систем, которая содержит основные понятия, описывающие предметную область VLCAD и виртуальных методов проектирования, а также отношения, семантически значимые для этой предметной области.
2. На следующем этапе данная онтология будет дополнена экземплярами объектов. VLCAD получит развитие в направлении предоставления пользователю контекстно-зависимой помощи как по компонентам виртуальной лаборатории, так и по процессам и методам проектирования. Общение с пользователем будет организовано на естественном языке в рамках выбранной предметной области.

Благодарности

Работа выполнена при финансовой поддержке Министерства образования и науки Украины в рамках совместного Украинско-Болгарского проекта № **145 / 23.02.2009** «Разработка распределенных виртуальных лабораторий на основе прогрессивных методов доступа для поддержки проектирования сенсорных систем» и Болгарского национального научного фонда в рамках совместного Болгарско-Украинского проекта **D 002-331 / 19.12.2008** с тем же названием.

Литература

[Horridge et al, 2004] Matthew Horridge, Holger Knublauch, Alan Rector, Robert Stevens, Chris Wroe. A Practical Guide To Building OWL Ontologies Using The Prot´eg´e-OWL Plugin and CO-ODE Tools Edition 1.0. The University Of Manchester. – 2004. – 117 p.

[Palagin et al, 2009] Oleksandr Palagin, Volodymyr Romanov, Krassimir Markov, Vitalii Velychko, Peter Stanchev, Igor Galelyuka, Krassimira Ivanova, Ilia Mitov. Developing of Distributed Virtual Laboratories for Smart Sensor System Design Based on Multi-dimensional Access Method. // International Book Series “INFORMATION SCIENCE & COMPUTING”, Number 8 FOI ITHEA, Sofia, Bulgaria. - 2009. – pp. 155-161.

[Гладун, 1994] Гладун В. П. Процессы формирования новых знаний / В. П. Гладун. – София: СД "Педагог 6", 1994. – 397 с.

[Палагин и Кургаев, 2003] Палагин А.В., Кургаев А.Ф. Проблемная ориентация в развитии компьютерных архитектур // Кибернетика и системный анализ. – 2003. – № 4. – С. 167–180.

[Соловьева, 1999] Соловьева Е.А. Естественная классификация: системологические основы. ХТУРЕ Харьков. – 1999. – 222 с.

Информация об авторах

Александр Палагин *– заместитель директора, Институт кибернетики имени В.М. Глушкова Национальной академии наук Украины, Академик Национальной академии наук Украины, доктор технических наук, профессор; проспект Академика Глушкова, 40, Киев-187, 03680, Украина; e-mail: palagin_a@ukr.net*

Владимир Романов *– заведующий отделом, Институт кибернетики имени В.М. Глушкова Национальной академии наук Украины, доктор технических наук; проспект Академика Глушкова, 40, Киев-187, 03680, Украина; e-mail: dept230@insyg.kiev.ua, VRomanov@i.ua*

Крассимир Марков *– старший научный сотрудник, Институт математики и информатики Болгарской академии наук; ул. "Академик Г.Бончев", 8, София-1113, Болгария; e-mail: markov@foibg.com*

Виталий Величко *– докторант, Институт кибернетики имени В.М. Глушкова Национальной академии наук Украины, кандидат технических наук, доцент; проспект Академика Глушкова, 40, Киев-187, 03680, Украина; e-mail: velychko@aduis.com.ua*

Игорь Галелюка *– научный сотрудник, Институт кибернетики имени В.М. Глушкова Национальной академии наук Украины, кандидат технических наук; проспект Академика Глушкова, 40, Киев-187, 03680, Украина; e-mail: galib@gala.net*

Крассимира Иванова *– научный сотрудник, Институт математики и информатики Болгарской академии наук; ул. Академика Г.Бонтчева, 8, София-1113, Болгария; e-mail: ivanova@foibg.com*

Петер Станчев *– профессор, Kettering University, Flint, MI, 48504, USA / Институт математики и информатики Болгарской академии наук; ул. "Академик Г.Бончев", 8, София-1113, Болгария; e-mail: pstanche@kettering.edu*

Илия Митов *– аспирант, Институт математики и информатики Болгарской академии наук; ул. "Академик Г.Бончев", 8, София-1113, Болгария; e-mail: mitov@foibg.com*

Милена Станева *– научный сотрудник, Институт математики и информатики Болгарской академии наук; ул. "Академик Г.Бончев", 8, София-1113, Болгария; e-mail: mstaneva@math.bas.bg*

АВТОМАТИЗИРОВАННОЕ СОЗДАНИЕ ТЕЗАУРУСА ТЕРМИНОВ ПРЕДМЕТНОЙ ОБЛАСТИ ДЛЯ ЛОКАЛЬНЫХ ПОИСКОВЫХ СИСТЕМ

Виталий Величко, Павел Волошин, Светлана Свитла

***Аннотация**: В статье рассмотрен метод автоматизированного создания тезауруса терминов предметной области на основе синтактико-семантического анализа естественно-языковых текстов для повышения релевантности поиска в полнотекстовых локальных поисковых системах. Использование предложенного метода позволяет сократить затраты времени на составление и редактирование тезауруса.*

***Ключевые слова**: локальные полнотекстовые поисковые системы, тезаурус терминов, синтактико-семантический анализ.*

***ACM Classification Keywords**: I.2.7 Natural Language Processing - Text analysis*

***Conference**: The paper is selected from XVth International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.*

Введение

Количество электронных документов, которые использует в своей ежедневной деятельности современная компания, стремительно возрастает. При этом данные хранятся в различных хранилищах, каждое из которых имеет собственную структуру (базы данных, информационные порталы, электронные библиотеки и т.д.) либо хранилище документов вообще неструктурированно (файлы на жестком диске пользователя).

Поэтому для обеспечения жизнедеятельности крупных государственных структур и частных корпораций необходимым условием является использование локальных поисковых систем для осуществления поиска по внутренним информационным ресурсам.

Одними из основных требований к подобным системам являются:

- обязательная полнотекстовая индексация всех информационных ресурсов, в которых осуществляется поиск, независимо от типов файлов и структуры хранения данных;
- наличие лингвистического процессора для выделения лексем, который позволяет осуществлять поиск по всем падежным формам искомого слова или словосочетания, что особенно важно для флективных языков, в частности, русского и украинского языка;
- упорядочивание результатов поиска по релевантности найденных документов.

На сегодняшний день в Украине наибольшее распространение получили локальные поисковые системы, такие как META, Google Desktop Search, Yandex.Server. В тоже время, все большее распространение и популярность приобретает программный продукт компании Microsoft — "Microsoft Office SharePoint Server 2007", который является средством организации коллективной работы с корпоративными данными и используется для решения чрезвычайно широкого спектра задач [1]:

- управление информационным порталом и бизнес-данными;
- поиск по данным предприятия;
- коллективная работа с документами;

- бизнес-аналитика;
- автоматизация деловых процессов.

Поиск по данным предприятия, осуществляемый Microsoft Office SharePoint Server 2007, не только удовлетворяет всем вышеперечисленным основным требованиям к локальным поисковым системам, но и имеет такие преимущества как: легкость администрирования; настройка страницы вывода результатов поиска; поиск не только по содержимому, но и по свойствам документа; возможность подключения списков расширения и замещения для управления поисковым запросом и т.д.

Список замещения определяет шаблон, который заменяется в поисковом запросе одним или несколькими словами – подстановками. Например, "*рост*" шаблон, а "*возрастание*" - подстановка. При вводе поискового запроса "*рост*" поисковый механизм отобразит результаты поиска только для поискового запроса "*возрастание*". Результаты поиска для слова "рост" отображаться не будут. Список расширения – это группа подстановок, которые являются синонимами. Запросы, содержащие слова из множества подстановок, расширяются путем включения всех других подстановок. Например, можно создать список расширения, где следующие подстановки являются синонимами: "*возрастание*", "*наращивание*", "*расширение*". При вводе поискового запроса "*возрастание*" поисковый механизм отобразит результаты поиска также для слов "*наращивание*" и "*расширение*".

Однако, все вышеперечисленные поисковые системы не имеют встроенных списков терминов (тезаурусов), учитывающих синонимию понятий предметной области. Для получения наиболее полных результатов поиска, пользователям в поисковых запросов приходится постоянно перечислять все синонимы введенного термина (например, *Россия* – это и *Российская Федерация*, и *РФ*).

При наличии тезауруса терминов предметной области, пользователю в поисковом запросе достаточно ввести только один термин. Если в тезаурусе есть список синонимов к введенному слову, то в результатах поиска будут присутствовать как документы, которые содержат слово, введенное пользователем, так и документы, содержащие слова-синонимы.

К сожалению, из-за отсутствия формализованных словарей терминов для конкретных предметных областей, автоматическое создание тезауруса невозможно. Ручное составление тезауруса является весьма трудоемкой задачей, так как требует экспертного анализа значительного количества документов организации (корпорации) для выделения списка терминов предметной области, при этом достаточно трудно оценить полноту полученного списка. Для решения такой задачи необходимо использовать автоматизированное создание списка терминов предметной области.

Принципы автоматического выбора терминов

Для построения понятийного аппарата из текстов предметной области используется поиск и выделение субстантивных именных словосочетаний, выражаемых схемой: согласуемое слово + существительное. В этой модели существительное является главным словом, а согласуемое слово — зависимым и может выражаться как прилагательным, так и существительным [2]. Словосочетания могут включать в свой состав также предлоги и сочинительные союзы. Количество слов в именных словосочетаниях колеблется от двух до пятнадцати и в среднем составляет три слова [3]. В работе [2] приводится 18 шаблонов именных словосочетаний, используемых для выделения терминов предметной области. В русском языке синтаксическая структура терминов предметной области более чем в 90 процентов случаев соответствует следующим пяти шаблонам: одиночные существительные, прилагательные, и сокращения; существительное + существительное в родительном падеже; прилагательное + существительное;

прилагательное + прилагательное + существительное; существительное + прилагательное + существительное в родительном падеже [4].

Вместе с тем существуют сложные словосочетания, используемые для обозначения понятий и терминов, состоящих из трех и более значимых слов. Выражение понятий и терминов словосочетаниями в пять и более слов, с использованием союзов и предлогов встречается редко, особенно такими словосочетаниями, в которых части речи не чередуются (например, прилагательное + прилагательное + прилагательное + существительное + существительное в родительном падеже). Эксперименты по выделению терминов показали, что в украинском языке для предметной области "экономика" целесообразно увеличить количество слов в синтаксической структуре именных словосочетаний до пяти. Словосочетания длиной пять и более слов используются в наименованиях организаций, в определении понятий относящихся к финансово-экономической сфере деятельности организаций. Шаблоны именных словосочетаний, используемых для поиска терминов, приведены в Таблице 1.

Таблица 1

№	Структура шаблона	Пример термина
1	Аббревиатура	СНГ; МВФ
2	Существительное	Система; государство
3	Существительное + существительное_в_родительном_падеже	Президент Украины; Глава государства
4	Прилагательное + существительное	Конституционный Суд; экономический рост
5	Существительное + существительное_в_родительном_падеже + существительное_в_родительном_падеже	Министр финансов Украины; Указ Президента Украины
6	Прилагательное + существительное + существительное_в_родительном_падеже	Финансовая система государства; Верховная Рада Украины; Бюджетный кодекс Украины
7	Существительное + прилагательное_в_родительном_падеже + существительное_в_родительном_падеже	Органы исполнительной власти; девальвация национальной валюты
8	Прилагательное + прилагательное + существительное	Среднемесячная заработная плата
9	Существительное + существительное_в_родительном_падеже + существительное_в_родительном_падеже + существительное_в_родительном_падеже	Постановление Кабинета Министров Украины; архив Министерства обороны Украины
10	Прилагательное + существительное + существительное_в_родительном_падеже + существительное_в_родительном_падеже	Национальная программа иммунопрофилактики населения
11	Существительное + прилагательное_в_родительном_падеже + существительное_в_родительном_падеже + существительное_в_родительном_падеже	Средства Государственного бюджета Украины; Председатель Верховной Рады Украины

№	Структура шаблона	Пример термина
12	Прилагательное + прилагательное + существительное + существительное_в_родительном_падеже	Государственная налоговая администрация Украины
13	Существительное + существительное_в_родительном_падеже + прилагательное_в_родительном_падеже + существительное_в_родительном_падеже	Система использования бюджетных средств; орган управления государственной службой
14	Прилагательное + существительное + прилагательное_в_родительном_падеже + существительное_в_родительном_падеже	Центральные органы исполнительной власти
15	Существительное + прилагательное_в_родительном_падеже + прилагательное_в_родительном_падеже + существительное_в_родительном_падеже	Зона чрезвычайной экологической ситуации; последствия мирового финансового кризиса
16	Прилагательное + прилагательное + прилагательное + существительное	Независимый государственный финансовый контроль
17	Прилагательное + существительное + существительное_в_родительном_падеже + существительное_в_родительном_падеже + существительное_в_родительном_падеже	Финансово-экономическая деятельность Министерства обороны Украины
18	Существительное + прилагательное_в_родительном_падеже + существительное_в_родительном_падеже + существительное_в_родительном_падеже + существительное_в_родительном_падеже	Комитет финансового контроля Министерства финансов
19	Прилагательное + прилагательное + существительное + существительное_в_родительном_падеже + существительное_в_родительном_падеже	Обесцененные денежные сбережения граждан Украины
20	Существительное + существительное_в_родительном_падеже + прилагательное_в_родительном_падеже + существительное_в_родительном_падеже + существительное_в_родительном_падеже	Фонд содействия молодежному жилищному строительству; нормы бюджетного законодательства Украины
21	Прилагательное + существительное + прилагательное_в_родительном_падеже + существительное_в_родительном_падеже + существительное_в_родительном_падеже	Специальный фонд государственного бюджета Украины
22	Существительное + прилагательное_в_родительном_падеже + прилагательное_в_родительном_падеже + существительное_в_родительном_падеже + существительное_в_родительном_падеже	Управление государственными финансовыми ресурсами Украины

№	Структура шаблона	Пример термина
23	Существительное + существительное_в_родительном_падеже + существительное_в_родительном_падеже + прилагательное_в_родительном_падеже + существительное_в_родительном_падеже	Источник финансирования дефицита Государственного бюджета; реформа системы контроля государственных финансов
24	Прилагательное + прилагательное + прилагательное + существительное + существительное_в_родительном_падеже	Государственный внебюджетный Пенсионный фонд Украины
25	Прилагательное + существительное + существительное_в_родительном_падеже + прилагательное_в_родительном_падеже + существительное_в_родительном_падеже	Главный распорядитель средств Государственного бюджета
26	Существительное + прилагательное_в_родительном_падеже + существительное_в_родительном_падеже + прилагательное_в_родительном_падеже + существительное_в_родительном_падеже	Контроль экономической деятельности коммерческих организаций
27	Прилагательное + прилагательное + существительное + прилагательное_в_родительном_падеже + существительное_в_родительном_падеже	Международный аудиторский контроль финансовой деятельности
28	Существительное + существительное_в_родительном_падеже + прилагательное_в_родительном_падеже + прилагательное_в_родительном_падеже + существительное_в_родительном_падеже	Система управления государственными финансовыми ресурсами
29	Прилагательное + существительное + прилагательное_в_родительном_падеже + прилагательное_в_родительном_падеже + существительное_в_родительном_падеже	Государственное регулирование внешней экономической деятельности
30	Существительное + прилагательное_в_родительном_падеже + прилагательное_в_родительном_падеже + прилагательное_в_родительном_падеже + существительное_в_родительном_падеже	Орган внешнего государственного финансового контроля
31	Существительное + существительное_в_родительном_падеже + существительное_в_родительном_падеже + существительное_в_родительном_падеже + существительное_в_родительном_падеже	Программа деятельности Кабинета Министров Украины

Автоматическое выделение однословных и многословных терминов, кроме шаблонов, использует результаты синтактико-семантического анализа текста. Распознание поверхностных семантических отношений осуществляется с помощью анализа флексий полнозначных слов, учитывая предлоги и союзы, без предварительного полного грамматического разбора и построения синтаксических отношений, которые используется в традиционной грамматике [5].

Процедура выделения терминов из текста включает два основных этапа.

На первом этапе происходит непосредственный поиск в тексте слов и словосочетаний – кандидатов в термины. В качестве однословных терминов выбираются существительные и аббревиатуры. Многословные термины формируются с помощью определенных типов отношений между словами предложения, путем постепенного присоединения слов к однословному термину-существительному. Для терминов – именных словосочетаний используются следующие основные типы отношений между словами: *объектное*, *принадлежность* (между двумя существительными), *определительное* (между прилагательным и существительным), *однородные слова* (между двумя существительными или двумя прилагательными). Выделенные группы слов проверяются на соответствие заданным шаблонам, приведенным в Таблице 1. Порядок расположения в предложении слов, образующих термин, может точно не соответствовать заданному шаблону, но обязательным условием выделения термина является соответствие отношений между словами определенным типам отношений. Это позволяет, например, из предложения "*Построение онтологии указанной предметной области*" выделить термин "*онтология предметной области*".

На втором этапе список кандидатов в термины фильтруется: учитывается значимость выделенных словосочетаний (приближенность в дереве разбора к подлежащему или сказуемому предложения) и частота, с которой они встречаются в тексте.

Приведенные принципы автоматического построения списка возможных терминов реализованы для украинского и русского языков в программе "Конспект".

Построение тезауруса терминов

Полученный предварительный список терминов редактируется вручную с помощью утилиты – редактора тезауруса терминов предметной области (общий вид окна редактора изображен на Рис. 1.)

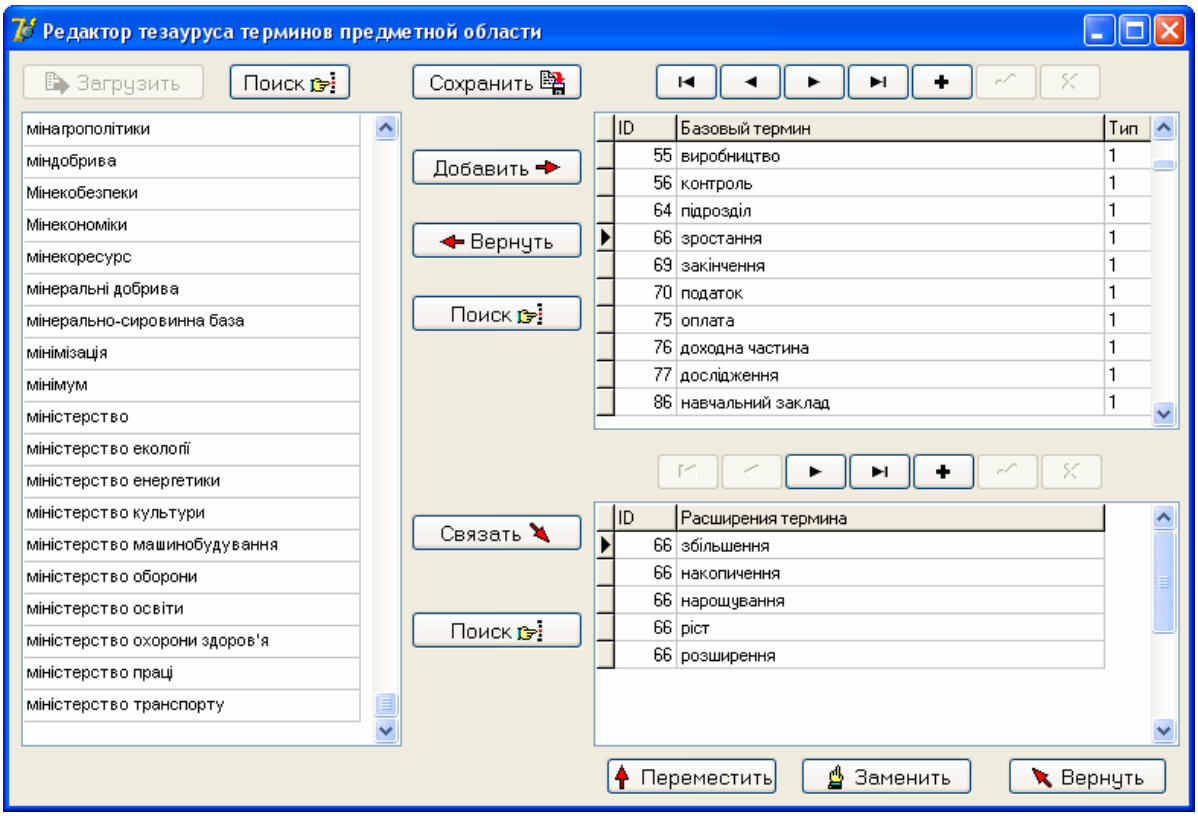


Рисунок 1. Главная форма редактора тезауруса терминов предметной области

Входными данными для утилиты является список терминов, сформированный программой "Конспект". Эксперт-аналитик вручную связывает термины, являющиеся синонимами для заданной предметной области (см. Рис.1). Полученные кортежи синонимов терминов предметной области сохраняются в XML-файл заданной структуры, который может использоваться поисковой системой среды Microsoft Office SharePoint Server 2007 в качестве тезауруса (списка расширений). В общем виде процесс автоматизированного построения тезауруса терминов предметной области изображен на Рис. 2.

Рисунок 2. Схема процесса автоматизированного построения тезауруса терминов

Эксперименты

Рассмотренный метод автоматизированного создания тезауруса терминов предметной области был использован для обработки текстов на украинском языке, относящихся к финансово-экономической сфере деятельности организаций. Из 172 различных текстовых документов за 2000-2008 года общим объемом примерно 3000 страниц текста без форматирования (примерно 12 Мб файлов формата txt) автоматически было выделено 26 860 слов и словосочетаний. Из сформированного списка для дальнейшего ручного редактирования терминов было оставлено 6,5 тыс. слов и словосочетаний, которые встречались в исходных текстов более трех раз. Термины, не имеющие синонимов, были исключены из тезауруса. В качестве базовых терминов было выбрано около 400 слов и словосочетаний, а в список расширений добавлено примерно 650 терминов. От общего количества терминов в тезаурусе однословные термины составили 76%, двухсловные – 21%, термины, состоящие из трех и более слов– 3%.

Выводы

В статье рассмотрена методика автоматизированного создания тезауруса терминов предметной области на основе синтактико-семантического анализа естественно-языковых текстов. Приведенные принципы автоматического построения списка возможных терминов реализованы для украинского и русского языков в программе "Конспект". Использование автоматизированного метода построения тезауруса терминов предметной области с помощью системы анализа естественно-языковых текстов позволяет значительно сократить затраты времени на составление и редактирование тезауруса. Дальнейшими направлениями исследований является совершенствование алгоритма выделения именных словосочетаний для поиска терминов произвольной длины на основе определенных поверхностных семантических отношений без явного задания всех возможных шаблонов структуры терминов, выделение некоторых типов глагольных словосочетаний и использование их как синонимов к именным.

Благодарности

Работа опубликована при финансовой поддержке проекта **ITHEA XXI** Института информационных теорий и приложений FOI ITHEA Болгария www.ithea.org и Ассоциации создателей и пользователей интеллектуальных систем ADUIS Украина www.aduis.com.ua.

Литература

1. Ноэл М., Спенс К. Microsoft SharePoint 2007 Полное руководство. Издательство: Вильямс, 2008 г. 832 с.
2. Найханова Л.В. Основные аспекты построения онтологий верхнего уровня и предметной области // В сборнике научных статей “Интернет-порталы: содержание и технологии”. Выпуск 3. / Редкол.: А.Н. Тихонов (пред.) и др.; ФГУ ГНИИ ИТТ “Информатика”. – М.: Просвещение, 2005. – С. 452-479.
3. Белоногов Г.Г., Кузнецов Б.А. Языковые средства автоматизированных информационных систем. М.: Наука, 1983.
4. B. Dobrov, N. Loukachevitch, O. Nevzorova. The technology of new domains’ ontologies development // Proceedings of the X-th International Conference “Knowledge-Dialogue-Solution” (KDS’2003).- Varna, Bulgaria.-2003.- pp.283-290.
5. Палагін О.В., Світла С.Ю., Петренко М.Г., Величко В.Ю. Про один підхід до аналізу та розуміння природномовних об’єктів. Комп’ютерні засоби, мережі та системи. -2008, №7. с.128-137.

Информация об авторах

Виталий Величко *– Институт кибернетики имени В.М. Глушкова Национальной академии наук Украины, кандидат технических наук, доцент; проспект Академика Глушкова, 40, Киев-187, 03680, Украина; e-mail: velychko@aduis.com.ua*

Павел Волошин *– ЗАО "Софтлайн", 03142, г. Киев, ул. В.Стуса, 35/37, аналитик компьютерных систем, e-mail: pavelv@unicyb.kiev.ua.*

Светлана Свитла *– Институт кибернетики имени В.М. Глушкова Национальной академии наук Украины, научный сотрудник; проспект Академика Глушкова, 40, Киев-187, 03680, Украина; e-mail:ssve@i.ua*

METHODS OF SYNTHESIZING REVERSIBLE SPATIAL MULTIVALUED STRUCTURES OF LANGUAGE SYSTEMS

Grigoriy Chetverikov, Irina Leshchinskaya, Irina Vechirskaya

Abstract: *The basic construction concepts of multivalued intellectual systems (MIS), which are adequate to primal problems of person activity and using hybrid tools with many-valued coding are considered. The concepts are agreed with the dialectic laws opened by a man and their manifestations in problems connected with creation of identification systems prediction and recognition of imagery in which the interactive operational mode is a main part of the whole complex of intellectual properties. The law of unity and struggle of contrasts changes and alternation of coding indications of messages about objects in neurons of a brain - from space to temporal and from two-place to multivalued.*

Keywords*: multivalued intellectual system, language systems, parallelism (spatial), AFP (algebra of finite predicates), AFP-structures, knowledge base, multiplevalued logic, multistate reversible element.*

ACM Classification Keywords*: I.2.4. Knowledge Represetation Formalisms and Methods.*

Conference*: The paper is selected from XVth International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.*

Introduction

Developing and improving computer facilities underline the process of automatizing mental activity, which was the starting - point of emergence of concept of artificial intelligence. However, the successes in the field of intellectualizing computer machines are insignificant especially if one compares the achieved results with anticipated ones and forecast. The orientation on the attain of qualitatively new technologies of information processing manifest itself in attempts to realize systems of artificial intelligence (AI) on Neumann computers. Therefore, new requirement of the technology of information processing are caused by need for solving problems which are badly formalized and the availability of user who is not a professional programmer. (1) Thus, we came to realizing one of the variants of developing AI systems - this is the way of analyzing modeling and synthesizing a natural language intelligent interface by means multiple-valued logical systems, in particular by the algebra finite predicates as well as the theory of multiple-valued structures and coding. Since the advent of computers facilities research has been carried out and realization at the level of engineering solutions multi-valued structures and coding in view of high information saturation of their signals has been conducted. Structures of data processing means, which are conducted on the basis of multiple-valued logical elements and modules with appropriate links, are called multiple-valued ones. All the objects, which are described by finite structural alphabet: elements, modules, structures, system of computer, measuring and control facilities and natural language information tools are classified among such structures.

At present there exist a great number of uncoordinated approaches and methods of building and applying multiple-valued structures, however, their systematization and classification are not available (i.e. any kind of an ordered system of realization means). At the same time the optimum design and technical realization of computer machines on the base of multi-valued structures are impossible without simultaneous development of entirely new (nontraditional) kinds of mathematical models and their research for various models of operation and interpretation of the modeling results. All this has resulted in a critical situation, which is caused by the absence of the integral theory of constructed highly effective multi-valued structures of spatial type. The analysis done shows

that the problem of developing the generalized theory of building highly effective multiple-valued computer structures and coding for language systems can be solved only within the class of intuitive and constructivist theories [1-4].

In particular, in works the accent on the concept of neuro -physiologic and neuro - cybernetic aspects of alive brain mechanisms is made. It is connected with the following natural neuron structures from nervous cells - neurons, essentially are highly effective recognizing systems and, for this reason, is of interest not only for doctors and physiologists, but also for the experts designing artificial intelligence systems. However direct transfer of research results of neuro- physiologists in engineering practice is now impossible because of a lack of an appropriate bioelectronic technology and an element basis, that has led to development and creation of a set of varieties of artificial neurons realized on the elements of the impulse technology.

As the corollary, non-adequacy of used principles of coding and element basis to simulated processes entails a redundancy, complication and non evidence of used mathematical and engineering means of transformations, loss of a micro level of parallelism in handling expected fast acting and flexibility of restructuring without essential modifications of architecture and connections.

1. Choosing the Body of Mathematics

The availability of algebra of finite predicates (AFP) provides an interesting opportunity of realizing a transition from algebraic description of information processes to their description in the form of equation in the language of given algebra and the equations specify relations between its variables [5,6]. All the variables in the equation possess equal rights and any of them can be both independent and dependent ones. The presence of equations and their advantage over algorithms consist in the fact that there appears an opportunity to calculate the reaction of the system even in case of the incomplete definiteness of initial information, whereas an incompletely developed algorithm is unable to operate. One should note that by means of AFP-structures which realize appropriate finite predicates. The given approach is similar to the process of constructing combinational circuits by the formula of the algebra of logic. Depending on the level of functional and structural realization we have AFP-structures of the first, second and third level 6].

The algebra of finite predicates is used as the body of mathematics of the research. We treat AFP as the one which is represented by the set M of all the predicates U^m. Let T be a set of all relations on U^m, Q be a set of all predicates on U^m. The relation T and predicate Q are called corresponding to each other, if for any $x_1, x_2, \ldots, x_m$ we have:

$$Q(x_1, x_2, \ldots, x_m) = \begin{cases} 0, if\ x_1, x_2, \ldots, x_m \notin T; \\ 1, if\ x_1, x_2, \ldots, x_m \in T. \end{cases} \quad (1)$$

In accord with (1) there can be a transition from the arbitrary relation T to predicate Q corresponding to the said relation T. The predicate Q which is found by the expression is called the characteristic function of relation T.

The condition of the form:

$$a(x_i) = x_i^{\,a} = \begin{cases} 0, if\ a \neq x_i; \\ 1, if\ a = x_i. \end{cases} \quad (2)$$

is called predicate of recognizing an object $a \in U$ of variable $x_i, i = \overline{(l, m)}$.

The predicate $a(x_i)$ should be considered as the predicate $a(x_1, x_2, \ldots, x_i, \ldots, x_m)$ from $P \subseteq Q$, whose all arguments, except x_i, are negligible. We will replace the expression in the form $a(x_i)$, where $i = \overline{l, m}$, $a \in U$

by x_i^a (here a is called an exponent of the variable x_i. Thus, the set T and basic elements x_i^a ($i = \overline{l,m}$, $a \in U$) and basic operations: disjunction, conjunction and negation is called the algebra of finite predicates over M. Eliminating the operation of negation out of the basis of the given algebra enables to obtain the so called disjunction and conjunction algebra of predicates (DCAP). Its completeness is proven [5]. Thus, the given algebra is considered as an instrument of research but not as its subject.

2. Formalizing the Concept of Unification of Spatial Multiple-valued Structures

A concept of unifying (reducing to uniformity and indissoluble interaction) two-digit and multi-valued means of processing appropriate (symbolic) data semi digital in a natural language. The present approach is based upon a single methodological and special purpose principle by applying the proposed methods of the theory of intelligence [7] for mathematical description and appropriate formalization of the concept of unifying input/output data [8] and their intermediate transformation [9] an appropriate AFP-structure of the third sort (Fig.1) [5].

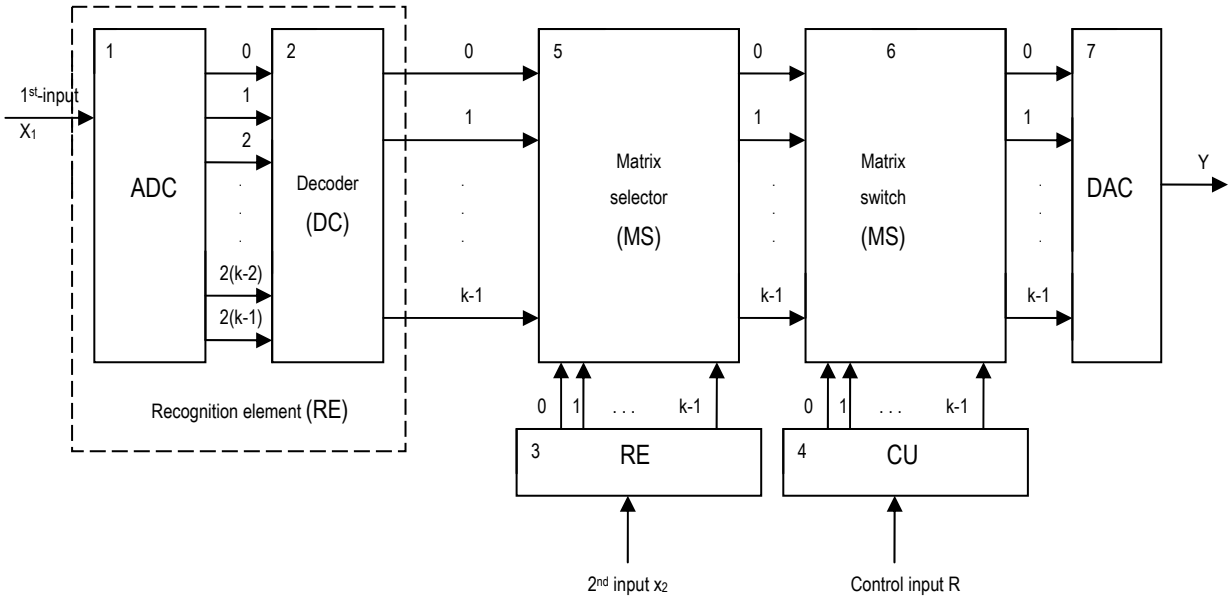


Fig.1. Two-input AFP-structure of the third sort

AFP-structure of the third sort is proposed on the basis of the sdf architectural studies presented in the works [5-8], as well as proceeding from the need for structurizing problems in developing the intuitive and constructivistical theory of constructing multiple-valued structures of spatial type for language systems.

AFP-structure of the third sort based upon a two-input universal multivalued functional converter [5] includes the following components: a recognition element (RE) n-valued variable which is formed by a parallel ADC together with a spatial DC, a MS and MSW, a control unit, a parallel digital-to-analogue converter (key switch). After describing the logic of operation of these components by the appropriate equations of the algebra of finite predicates, we will obtain their mathematical models. The use of the concept of unification and the given algebra will ensure boundary parallelism and uniformity of the structure as a whole. Obtaining analytical relations of input /output variables of component will make it possible to formalize and automatize synthesis procedure of multiple-valued structures of spatial type [5,6].

An RE of k-valued variable realizes the recognition predicate - the basic element of the algebra of finite predicates. The AFP is specified on the set M of all n-place k-valued predicates, i.e. the function of the form

$y = f(x_1, x_2, \ldots, x_n)$, where $x_1, x_2, \ldots, x_n$ are letter variables specified on alphabet of letters $A = \{a_1, a_2, \ldots, a_n\}$, $y \in \{0,1\}$ is a logic variable, i.e.

$$x_i^{a_j} = \begin{cases} 0, if\ a_j \neq x_i; \\ 1, if\ a_j = x_i, (i = \overline{1,n},\ j = \overline{1,k}). \end{cases}$$

As the recognition element has k -levels ($0 \div (k-1)$), the number of output recognition signals for a spatial set of the RE will constitute: $k : a_1, a_2, \ldots, a_{k-1}$. Thus, paralleling the process at the level of the boundary speed of response of appropriate transformation is provided already at the AFP-structure input. The logic of operation of DC particularly decoder 2 (DC 2) is described by a set of the AFP equation of the following form:

$$\begin{cases} b_{1,2,3}^0 = \overline{y_1}; \\ b_{1,2,3}^1 = y_1 y_2; \\ \ldots\ \ldots\ \ldots\ \ldots \\ b_{1,2,3}^{k-1} = y_{k-1}; \end{cases}$$

The logic of operation of MS 3 is described by a system of the AFP equations of the following form:

$$\begin{cases} g_{00} = b_1^0 \,\&\, b_2^0;\ g_{01} = b_1^0 \,\&\, b_2^1;\ g_{0(k-1)} = b_1^0 \,\&\, b_2^{k-1}; \\ g_{10} = b_1^1 \,\&\, b_2^0;\ g_{11} = b_1^1 \,\&\, b_2^1;\ g_{1(k-1)} = b_1^1 \,\&\, b_2^{k-1}; \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ g_{(k-1)0} = b_1^{(k-1)} \,\&\, b_2^0;\ g_{(k-1)1} = b_1^{(k-1)} \,\&\, b_2^1;\ g_{(k-1)(k-1)} = b_1^{(k-1)} \,\&\, b_2^{k-1}. \end{cases}$$

where g_{ij} is the output signals of matrix selector, which take the values out of the set $E_2 \in \{0,1\}$, $i, j \in \{0,1,\ldots,k-1\}$. The logic of operation of the space switch in the AFP-structure of the first sort is described by the following system of the AFP equation:

$$\begin{cases} g_{00}^0 \vee g_{01}^1 \vee \ldots \vee g_{0(k-1)}^{(k-1)} = q_0, \\ g_{10}^0 \vee g_{11}^1 \vee \ldots \vee g_{1(k-1)}^{(k-1)} = q_1, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ g_{(k-1)0}^0 \vee g_{(k-1)1}^1 \vee \ldots \vee g_{(k-1)(k-1)}^{(k-1)} = q_{(k-1)}. \end{cases}$$

Correspondingly, the switch input signals are formed at the expense of the wire OR for signals with the same indexes but with different commutation level:

$$\begin{cases} z_0 = r_0 \vee s_0 \vee q_0, \\ z_1 = r_1 \vee s_1 \vee q_1, \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ z_{(k-1)} = r_{(k-1)} \vee s_{(k-1)} \vee q_{(k-1)}. \end{cases}$$

The final result of the universal conversion can be formally represented in the form of the following operator picture:

$$F(z_i) = \max_{i=0}^{k-1} \left(\min_{i=0}^{k-1} z_i t_i \right),$$

where $i = \overline{1,k-1}, (t_0, t_1, \ldots, t_{k-1})$ are sets of signals of adjusting (selecting) the output functions of the universal AFP-structure of the third sort. Thus, the aim of this approach is achieved by decompiling multiple-valued

hardware means (AFP-structures of the third sort) into multiple-valued and two-value discrete and analogous subunits, especially in the part of their intermediate spatial information transformation.

The research has shown that the application of traditional methods of combinational synthesis in functionally complete bases as disjunction (conjunction) normal forms to multiple-valued structures of spatial type is ineffective from the point of view of retaining the properties of uniformity and parallelism of structural formations [3,5]. There is a need for seeking objects of research which are the most natural and closest to the inner logic of functioning for a natural language particulary of corresponding structures a variety of algebraic and logical means of modeling and new methods of synthesis of corresponding structures [5,9].

In problem of information processing in a natural language (for instance, in a Slavonic one) one needs to recognize and process not less than 33 letters already at the level of phonetics. Consequently, beginning already with problems of phonetical level of language information processing, value increase is an inevitable problem of further research. It is evident that considering such an approach to the investigation of problems of creating and constructing AFP-structures of the third sort on the base of universal multiple-valued functional converters requires expanding from the point of view of increasing the values of a structural alphabet ($k > 3$)

As the converter has a property of universality, the power of a set of functions which can be realized by a one-input AFP-structure of the third sort makes up the value $N = k^k$. Increase the digits of the structural alphabet (the number of analogue-to-digital converter parallel stages out of comparator series).

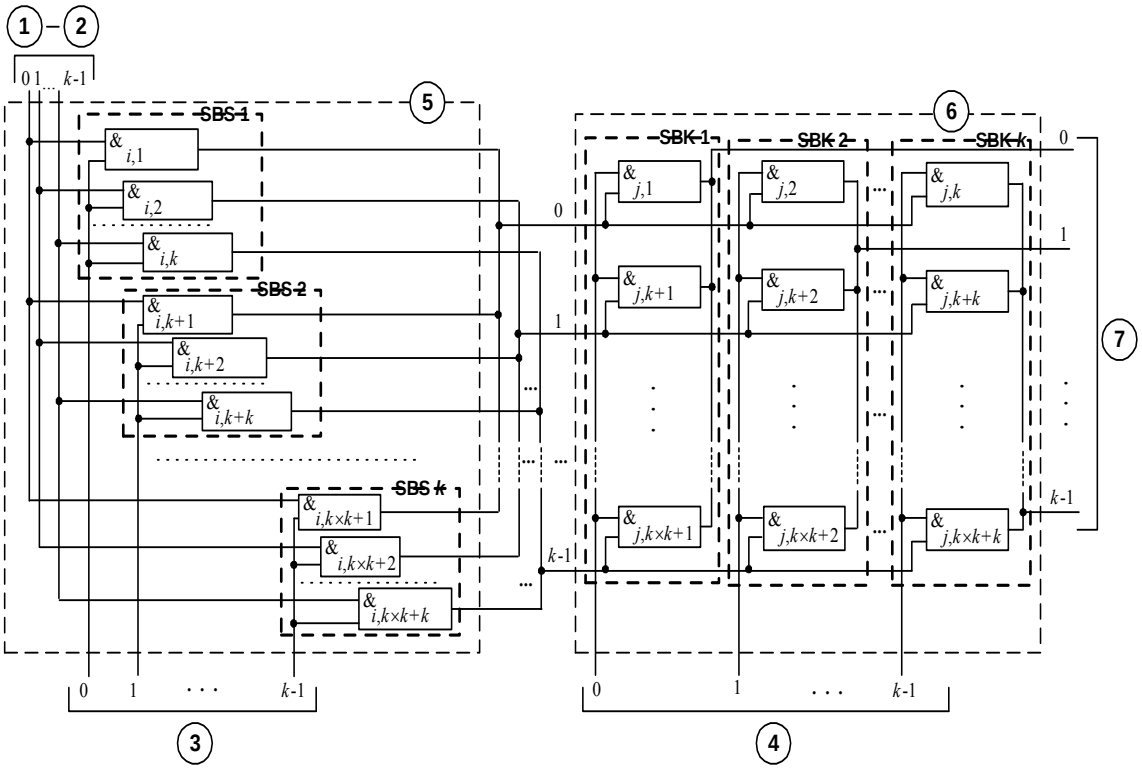


Fig.2. Structure and logic of realization of combinationed of matrix circuits of: a) a selector b) a switch

With a joint reference voltage divider from 3 to 10 makes it possible to increase the functional potentialities with respect to realizing a set of logical functions of one variable to 1010/33 times. Thus, applying two-input AFP-structure, of the third sort, as well as constructing a space decoder, a matrix selector and switch on conjunction elements enable to ensure uniformity and homogeneity of its inner structure, as well as to increase speed of

response at the expense of the boundary parallelism of the structure. The AFP-structure uses logical, but not computing methods of intermediate transformation with the use of the concept of unifying 2 digit and k-valued coding, which ensures the simplification of structure of intermediate subunits of a matrix selector and a switch (Fig.2)

3. Solution Process of the Task of Hypothetically Connected Subscribers

This part of article is devoted to building of formalization methods of the relation with linear logical transformation. It is a main tools for realization of logic network which focused on parallel information processing and its program realization.

Let variables x_i, $i = 1,2,...,12,...$ - are the telephone numbers of city Kharkiv and Kharkiv area. The task consists of finding all telephone numbers of subscribers, with which can be connected subscribers numbers x_1, x_2, x_4, x_7, x_9. The numbers set of subscribers, on which entrance rings are fixed will designate y_j.

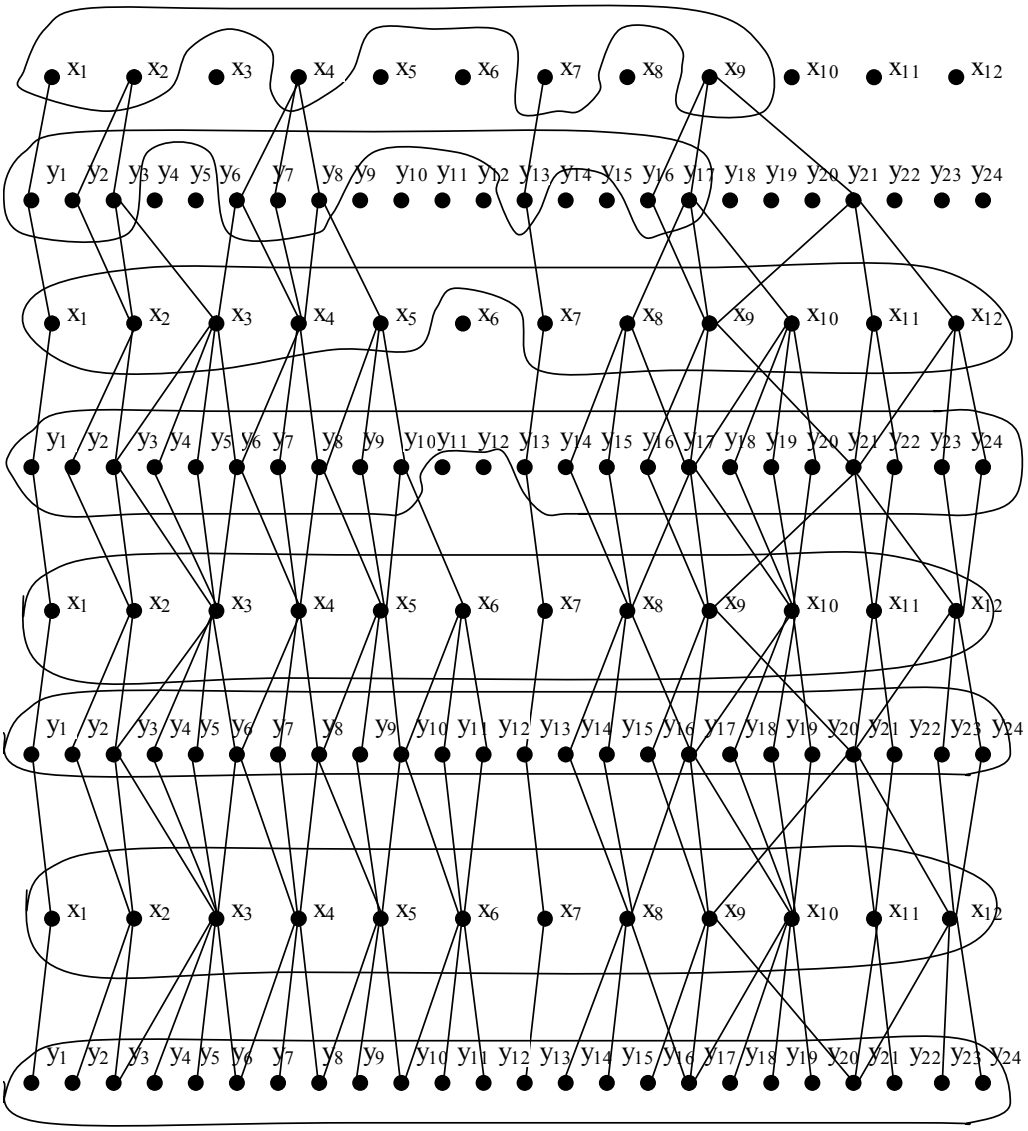


Fig.3. Finding of the hypothetically connected subscribers

On Fig.3 evidently, that subscribers, x_1, x_2, x_4, x_7, x_9 hypothetically connected with subscribers $y_1 - y_{24}$, a decision is found for 3 steps. The using the developed method of finding the degree of linear logical transformation allow to minimize time to search decisions of this task.

Variable x_i, $i = 1,2,...,12,...$ - are the telephone numbers of Kharkiv and Kharkiv area. It is necessary to find the telephone numbers of subscribers, with which can be connected subscribers with numbers $x_1 = 0572230508$, $x_2 = 0572945376$, $x_4 = 0577124387$, $x_7 = 0577774004$, $x_9 = 0577711691$. The numbers set of subscribers, where rings acted is presented in [10].

Thus, the followings numbers of subscribers Kharkiv and Kharkiv area were found:

$x_3 = 0577153256$, $x_5 = 0577356578$, $x_6 = 0572995633$, $x_8 = 0577332376$, $x_{10} = 0572278745$, $x_{11} = 0577126534$, $x_{12} = 0572937694$.

Conclusion

Thus, the above listed results make it possible to make the following important conclusion: using new algebraic and logical means of modeling of natural language constructions in the form of a system of equations based on the AFP-language and explicit way of specifying a finite alphabet operator which underlies method of solving these equations, ensures realization the property of reversibility of AFP-structures and a wide paralleling of symbolical information processing. Fundamental research of the algebraic and logical structure of an natural language as well as algebraic and logical means of its modeling in the form of AFP-structures of the first, second and the third sort permits to come close to the solution of the important scientific problem: attaing qualitatively new technologies of symbolical information processing on the basis of the concept of unification and methods of synthesizing reversible spatial multivalued structures of language systems. The way of method of n-th power of linear logical transformation's adaptation in complex accounts computer-based system of integrated data-processing system of telecommunications enterprise is adduced.

Acknowledgements

The paper is published with financial support by the project ITHEA XXI of the Institute of Information Theories and Applications FOI ITHEA Bulgaria www.ithea.org and the Association of Developers and Users of Intelligent Systems ADUIS Ukraine www.aduis.com.ua.

Bibliography

[1] Bondarenko M., Chetverikov G., Karpukhin A. Analiz problemi sozdaniya novich tekhnicheskich sredstv dlya realizatsii lingvisticheskogo interfeisa. Proc of the 10thInternational Conference KDS – 2003, Varna, Bulgaria, June 16-26, 2003, pp.78-92.

[2] Bondarenko M., Chetverikov G., Deyneko Zh. Application of a numerically – analytical method for simulation of non-linear resonant circuits. 10th International Conference Mixed Design Of Integrated Circuits And System (MIXDES 2003), Lodz, Poland, 26-29 June 2003, pp.131-133.

[3] Bondarenko M., Chetverikov G., Leshchinsky V. Synthesis Methods of Multiple-valued Structures of Bilogical Networks // Proc. Of the 12th International Conf. "Mixed design of integrated circuits and systems" (MIXDES 2005), Krakow (Polska), 2005. – pp. 201-204.

[4] Bondarenko M., Chetverikov G., Karpukhin A. Structural Synthesis of Universal Multiple-Valued Structures of Artificial Intelligence Systems // Proc. Of the 9th World Multi-Conference in Systemics,Cybernetics and Informatics (WMSCI 2005). – Orlando, Florida (USA), 2005. Vol.VII. – pp. 127-130.

[5] M.F. Bondarenko, Z.D. Konoplyanko, G.G. Chetverikov. Theory fundamentals of multiple-valued structures and coding in artificial intelligence systems.-Kharkiv: Factor-druk, 2003. – 336 p.

[6] M.F. Bondarenko, Z.D. Konopljanko, G.G. Chetverikov. Osnovy teorii synteza nadshvydkodiuchikh structur movnykh sistem shtuchnogo intellectu, Monografia. – K.: IZMN, 1997. – 386 p.

[7] Bondarenko M.F., Lyahovets S.V., Karpukhin A.V., Chetverikov G.G. Sintez shvidkodiuchikh structur lingvistichnich ob'ekhtiv., Proc. of the 9th International Conference KDS – 2001, St.Peterburg, Russia, 2001, pp. 121–129.

[8] Bondarenko M.F., Bavykin V.N., Revenchuk I.A., Chetverikov G.G. Modeling of universal multiple-valued structures of artificial intelligence systems, Proc. of the 6th International Workshop "MIXDES'99", Krakow, Poland, 17–19 June 1999, pp. 131–133.

[9] Bondarenko M., Chetverikov G., Vechirskaya I. Multiple-valued structures of intellectual systems Proc of the International Conference KDS – 2008, Varna, Bulgaria, June 21-28, 2008, pp. 53-59.

[10] Vechirskaya I. Reshenie zadachi nahojdeniay gipoteyicheski svyazanyh objektov. Pravo i bezpeka. Kharkiv: KHNUVS, 2009.- pp. 268-274.

Authors' Information

Irina Leshchinskaya *– Researcher; State National University of Radio-Electronics P.O. Box 14, Lenin's avenue, Kharkov, 61166, Ukraine, irina.leshchinsky@gmail.com*

Irina Vechirskaya *– Researcher; State National University of Radio-Electronics P.O. Box 14, Lenin's avenue, Kharkov, 61166, Ukraine, ira_se@list.ru*

Grigoriy Chetverikov *- Professor, Doctor of Technical Sciences, State National University of Radio-Electronics P.O. Box 14, Lenin's avenue, Kharkov, 61166, Ukraine, chetvergg@kture.kharkov.ua*

COMPUTER TECHNOLOGY FOR SIGN LANGUAGE MODELLING

**Iurii Krak, Iurii Kryvonos, Olexander Barmak,
Anton Ternov, Bohdan Trotsenko**

***Abstract**: The work suggests a complex informational technology for sign language modeling (for Ukrainian sign language implementation). The technology and corresponding mathematical model for formalization of sign language primitives, dactyl alphabet and face mimics primitives have been suggested. The corresponding software applications have been created which allow gestures as well as face mimics storage and reproduction. The results have been analyzed and the roadmap for further research has been created*

***Keywords**: Sign language modeling, Computer technology, Face mimics modeling, Ukrainian sign language*

***ACM Classification Keywords**: I.2.8 Problem Solving, Control Methods, and Search H.1.1 Systems and Information*

***Conference**: The paper is selected from XV[th] International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.*

Introduction

Significant limits of existing means for sign language reproduction give a reason to create more powerful algorithms which would allow to create computer-bases systems for sign language tutoring with aim to facilitate communication with deaf people and people having hearing disabilities. The authors suggest [1,2] a concept of informational technology of non-verbal communications with deaf people. Using a virtual 3d human model, the complex informational technology includes functionality to reproduce sign language gestures, gestures of dactyl alphabet, proper face mimics during pronunciation.

Implementation of the feature to generate pronunciation animation for a custom gesture requires having proper informational and mathematical models. Therefore, the following problem statement has been formulated:

1) To create informational and mathematical models for the formal sign language morpheme description and for the face mimics synthesis;
2) Using the created models, to create a technology along with corresponding software for the gesture capturing, storage and reproduction, with support for the face mimics.

Model of sign language morphemes fixation

Process of sign language reproduction on a 3d human model can be regarded as animation of the corresponding frequency of different skeleton states.

The skeleton human model is a simplified correspondent of the human skeleton. It can be formalized as the hierarchical structure of kinematic pairs, which reflect the basic human bones.

Modern 3d animation software (Poser, 3D Studio Max) are able to generate animation using virtual skeleton and information on angles change. Therefore, for the formal description of a gesture we can use sets which reflect the simplified human skeleton and changes in time of angles between bones: $H = \{H_i : H_i = \{k, d_i, M_i \in M\}\}$ – the simplified human skeleton (bones hierarchy) and change of angles in time, where H_i – i-th bone of the skeleton ($i = 0, \cdots, N-1$, N – amount of bones in the skeleton); k – index of the parental bone; $d_i = [x_i, y_i, z_i]^T$ –

coordinate of the bone's ending point using the coordinate system connected with the parental bone; $M = \{M_i : M_i = \{order_i, \theta_i\}\}$ – the angles and order of angle application for the each bone.

For the storage of the gesture using this formalism we suggest to use the BVH[3] file format. The format was introduced by Biovision company exactly for the storage of movement. The BVH is one of the best formats for this kind of storage, with its sole disadvantage being the absence of the complete definition of the starting pose; yet this disadvantage is irrelevant to our problem. File in BVH format consists of two parts. The header contains the hierarchy of the model and the initial pose of the skeleton and description of the animatory part. The primary reason for using this file format is its simplicity and suitable for the our problem. It's very convenient that it is recognized by primary software on the market. The figure 1 shows a virtual human skeleton that is used for the formal description of the sign language reproduction.

Technology for gesture obtaining and storage

The history of 3d animation is much more than a decade old, thus the progress is essential. This work benefits from the Motion Capture technology for modeling a virtual human, which is designed for the sign language reproduction. Systems using Motion Capture appeared on beginning of 90ies of the past century. The problem of 3d animation was among the most important ones before the technology appeared. The underlying principle of the Motion Caption is quite simple: the real-world human plays the role for the virtual character. Typical Motion Capture implementation contains a number of signal emitters which are tightened to the real-world human. The information about the placement of the emitters is perceived by some detectors and thereon registered by a computer. Later on the information is processed by software, which reconstructs movements for the virtual character.

The primary disadvantage of existing Motion Capture systems is the relatively large cost of the equipments and the service for the data processing. Authors suggest much simpler implementation of the technology. Given as granted that for the sign language digitalization a relatively small amount of movements should be captured, the following is suggested:

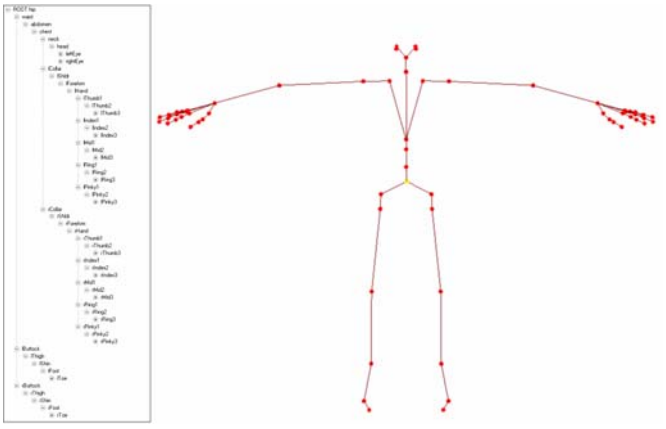


Figure 1.Virtual human skeleton

In order to obtain a set of angles which reflect the changes of bones relatively to the initial state of the skeleton the following technological scheme is suggested:

1) Using 3 cameras a real-world human performing a custom gesture is recorded. The cameras are set on the same focus distance (L) from the object of the recording and thus fixate 3 projections: frontal, left and right (see figure 2). States of the skeleton characterize N frames of the recording (using 15 frames per second).
2) For the fixation of i-th state ($i = 1, \cdots, N$) of the skeleton we use the proper frames from the recordings (see figure 3). The skeleton is then set to correspond to the image (see figure 4).
3) Let's combine points of skeleton projection with the corresponding image of the real-world human (for the frontal, left and right frames), we obtain new coordinates: $\left(x_i^{front}, y_i^{front}, -1\right)$ – for the frontal, $\left(x_i^{left}, y_i^{180^\circ}, z_i^{left}\right)$ – for the left projection and $\left(x_i^{right}, y_i^{180^\circ}, z_i^{right}\right)$ – for the right projection. Where:

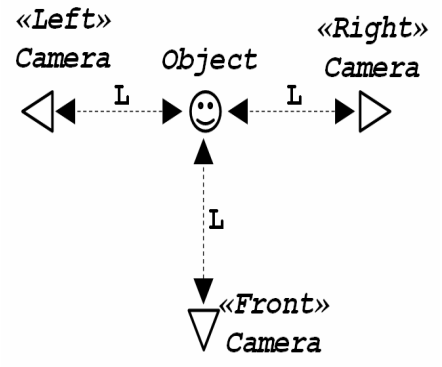
$$x_i^{front} = x_i^{180^\circ} + off_i^{front_x},\ y_i^{front} = y_i^{180^\circ} + off_i^{front_y},\ x_i^{left} = x_i^* + off_i^{left_x},\ x_i^{right} = x_i^* + off_i^{right_x},\ z_i^{right} = z_i^* + off_i^{right_z} \quad (1)$$

The shifts off_i obtained this way we will apply for all the frames which correspond to the gesture. For the more exact values off_i it is recommended for the first frame to make the real-world human stay in a T-position (or close to it).

4) For the j-th ($j = 1, \cdots, N$) frame on the pair of images (frontal and left or frontal and right) we consider points of bones connections which have actually moved ($\left(x_{i,new}^{front}, y_{i,new}^{front}\right)$ and $\left(x_{i,new}^{left\ or\ right}, y_{i,new}^{left\ or\ right}\right)$) and then computer their 3d coordinates (relatively to the center of the skeleton):

$\left(x_i^{new}, y_i^{new}, z_i^{new}\right) = \left(x_i^{old}, y_i^{old}, z_i^{old}\right)$,

where $x_i^{old} = x_{i,new}^{front} - off_{i,x}^{front}$, $y_i^{old} = y_{i,new}^{front} - off_{i,y}^{front}$, $z_i^{old} = x_{i,new}^{left\ or\ right} - off_{i,x}^{left\ or\ right}$ (2)



Firgure 2. Placement of camera during recording

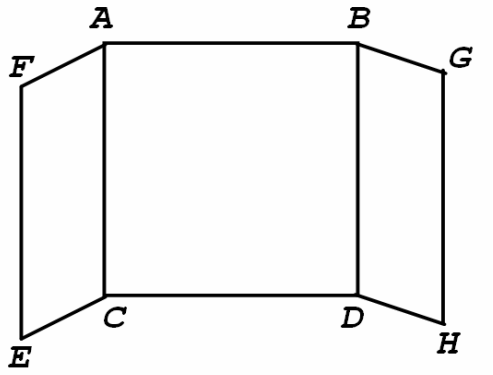


Figure 3. Planes for frames projection

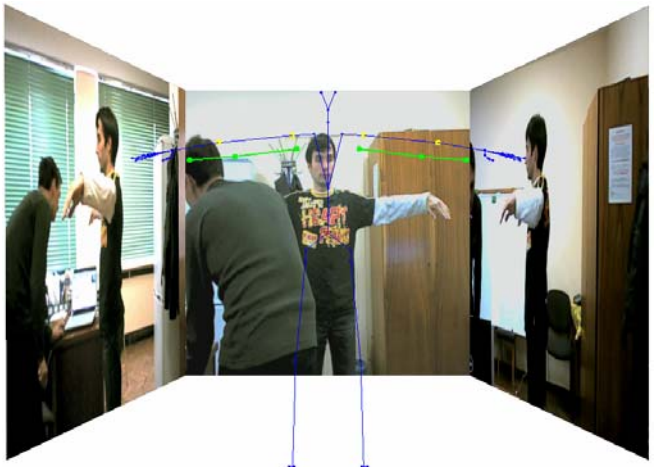


Figure 4. Working scene example for the angles detection

Then, given the 3d coordinates of the new location of the bones connections we compute the Euler angles, which is the equivalent information of the points coordinates in combination with the previous intersection location.

Consider coordinate system XYZ and two vectors in it: $r_1 = (x_1, y_1, z_1)$ and $r_2 = (x_2, y_2, z_2)$ (see figure 5). What are the Euler angles $\varphi_x, \varphi_y, \varphi_z$ for the rotation of X, Y, Z which will combine r_1 into r_2. For this:

1) Let's build a rotation matrix T, which transforms vector r_1 into vector r_2:

$r_1 = Tr_2$, where $T = \begin{pmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{pmatrix} = \begin{pmatrix} \cos(\alpha_{X^\wedge X'}) & \cos(\alpha_{X^\wedge Y'}) & \cos(\alpha_{X^\wedge Z'}) \\ \cos(\alpha_{Y^\wedge X'}) & \cos(\alpha_{Y^\wedge Y'}) & \cos(\alpha_{Y^\wedge Z'}) \\ \cos(\alpha_{Z^\wedge X'}) & \cos(\alpha_{Z^\wedge Y'}) & \cos(\alpha_{Z^\wedge Z'}) \end{pmatrix}$.

Given that the vectors are normalized, we obtain:

$$t_{11} = x_1x_2 + y_1y_2 + z_1z_2. \quad (3)$$

$$t_{21} = -y_1x_2 + \frac{x_1y_2}{\sqrt{x_1^2 + y_1^2}} + \frac{y_1z_1z_2}{\sqrt{x_1^2 + y_1^2}}. \quad (4)$$

$$t_{31} = z_1x_2 + z_2\sqrt{x_1^2 + y_1^2}. \quad (5)$$

$$t_{12} = -x_1y_2 + \frac{y_1x_2}{\sqrt{x_2^2 + y_2^2}} + \frac{z_1y_2z_2}{\sqrt{x_2^2 + y_2^2}}. \quad (6)$$

$$t_{22} = y_1y_2 + \frac{x_1x_2 + y_1z_1y_2z_2}{\sqrt{x_1^2 + y_1^2}\sqrt{x_2^2 + y_2^2}}. \quad (7)$$

$$t_{32} = -z_1y_2 + \frac{y_2z_2\sqrt{x_1^2 + y_1^2}}{\sqrt{x_2^2 + y_2^2}}. \quad (8)$$

$$t_{13} = x_1z_2 + z_1\sqrt{x_2^2 + y_2^2}. \quad (9)$$

$$t_{23} = -y_1z_2 + \frac{y_1z_1\sqrt{x_2^2 + y_2^2}}{\sqrt{x_1^2 + y_1^2}}. \quad (10)$$

$$t_{33} = z_1z_2 + \sqrt{x_1^2 + y_1^2}\sqrt{x_2^2 + y_2^2}. \quad (11)$$

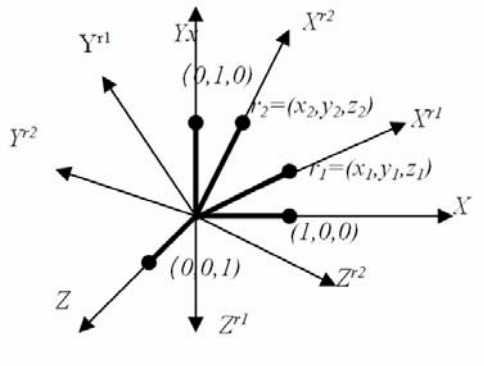


Figure 5. Coordinate systems: $XYZ, X^{r_1}Y^{r_1}Z^{r_1}, X^{r_2}Y^{r_2}Z^{r_2}$

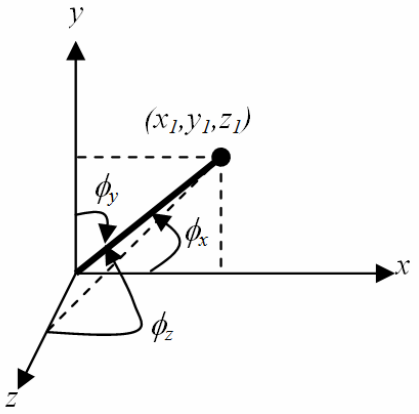


Figure 6. Directing angles for the rotation

2) We then compute Euler angles $\varphi_x, \varphi_y, \varphi_z$. It should be underlined, that the multiplication of matrices is not commutative, and thus 6 cases of multiplication precedence should be considered. Therefore, we compute angles for the following 6 orders of multiplication usage: ZYX, YXZ, YZX, XYZ, XZY, ZXY.

a) Multiplication order ZYX, ($c() = \cos(), s() = \sin()$):

$$\begin{pmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{pmatrix} = \begin{pmatrix} c(\varphi_y)c(\varphi_z) & c(\varphi_z)s(\varphi_x)s(\varphi_y) - c(\varphi_x)s(\varphi_z) & c(\varphi_x)c(\varphi_z)s(\varphi_y) + s(\varphi_x)s(\varphi_z) \\ c(\varphi_y)s(\varphi_z) & c(\varphi_x)c(\varphi_z) + s(\varphi_x)s(\varphi_y)s(\varphi_z) & -c(\varphi_z)s(\varphi_x) + c(\varphi_x)s(\varphi_y)s(\varphi_z) \\ -s(\varphi_z) & c(\varphi_y)s(\varphi_x) & c(\varphi_x)c(\varphi_y) \end{pmatrix} \quad (12)$$

Then

$$\varphi_z = arctg\left(\frac{t_{21}}{t_{11}}\right) + n\pi \ . \quad (13)$$

$$\varphi_y = arctg\left(\frac{-t_{31}}{\sqrt{t_{11}^2 + t_{21}^2}}\right) + n\pi \ . \tag{14}$$

$$\varphi_x = arctg\left(\frac{t_{32}}{t_{33}}\right) + n\pi \ . \tag{15}$$

b) The angles for the other orders are computed in a similar way.

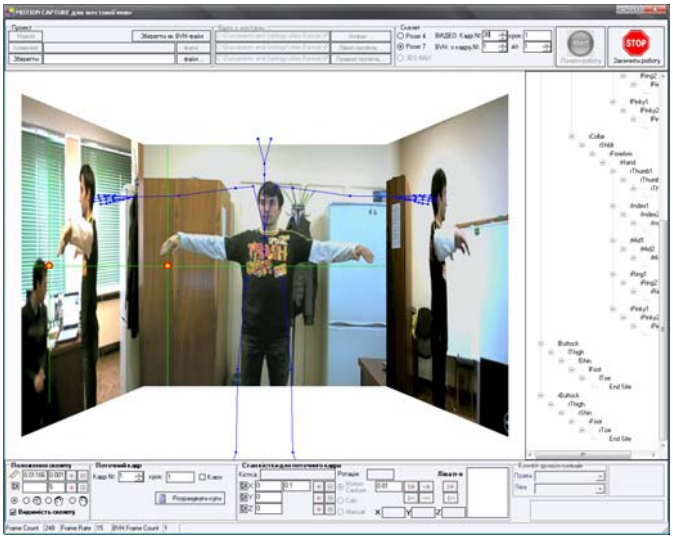


Figure 7. The application which implements motion capture technology

Model for face mimics synthesis for the pronunciation process and a technology for visemes computation.

Face mimics play a role of an additional information channel in the sign language. In order to support this additional channel and make gesture reproduction look more real, special research work has been conducted and then face visemes have been introduced. From the numerous researches [4,5] we conclude that visual alphabet contains no more than 6-15 elements. The actual number depends on the required level of natural look.

Using results of the researches on face mimics and visuals mentioned above a set of visemes has been creates for 6 vocals and 32 consonants (see table 1), as required for the Ukrainian language phonetics. Phonemes have been classified into 15 groups, excluding additional group for the idle state (no pronunciation). We define viseme as a position of lips which is naturally observed during pronunciation of a particular phoneme.

For each group of phonemes (i.e. for each viseme) frames of face mimics were computer. These frames are further used for pronunciation animation (see figure 8).

TABLE 1.

viseme	*phoneme*	*viseme*	*phoneme*	*viseme*	*phoneme*	*viseme*	*phoneme*
1	а	5	і, и	9	п, б, м	13	р
2	е	6	й	10	в, ф	14	л', р'
3	о	7	ш, ж, ч, дж	11	т, д, н, л	15	т', д', н'
4	у	8	к, г, х, ґ	12	с, з, ц, дз	16	**idle state**

3d gesture animation and face mimics reproduction model and its implementation

For the process of gesture and face mimics animation synthesis the following formal description is suggested. The formal description using proper set of parameters and algorithms. The 3d virtual human model which implements the animation and face mimics has such attributes:

Viseme	"у"	"а"	"тднл"	"л'р'"	"сзцдз"	"пбм"	с/сп
МВО							

Figure 8. Sample visemes for Ukrainian sign language

$V = \{V_i : V_i = \{x, y, z\}\}$ – set of triangles' vertices for the triangulation of the 3d human model;

$N = \{N_i : N_i = \{x, y, z\}\}$ – set of normals for the vertices; $T = \{T_i : T_i = \{u, v\}\}$ – set of texture coordinates for the vertices; $V^{ind} = \{V_i^{ind} : V_i^{ind} = \{k_1, k_2, k_3\}\}$ – set of indices which specify the order of triangles construction; $I = \{I_i : I_i = \{img\}\}$ – set of pictures for the textures

For the skeleton-based animation modeling it is necessary and sufficient to compute the new values of vertices (V). To reach this goal we suggest to using so-called skinning method. Skinning can be defined as an algorithm of computing new values of vertices as a weighted sum of points which belong to different bones of the skeleton. The model of skeleton-based animation can be formalized as follows:

$MH = \{MH_i : HM_i = \{k, \{l_1, \cdots l_m\}, d_i, Glb_i, Order_i\}\}$– description of the simplified human skeleton (bones hierarchy) for the implementation of the skeleton-based animation, where MH_i – i-th bone of the skeleton ($i = 0, \cdots, N-1$, N – number of bones in the skeleton); k – index of the parental bone; $\{l_1 \cdots l_m\}$ – set of children's indices, $d_i = [x_i, y_i, z_i]^T$ – coordinates of the ending point of the bone in the coordinate system which center is in the beginning of the bone; Glb – vector which is used for determination of the bone's coordinates in the global coordinate system; $Order_i$ – order of rotations.

$\text{Skin} = \{Skin_i : Skin_i = \{(IndexVertex_1, Weight_1), \cdots\}\}$ – set of vertices which influence the current point.

For each vertex v skinning is computed as follows:

$$v_j^{'} = \sum_{i=0}^{N} \{(v_j * IBM_{H_i} * JM_{H_i}) * JW_{H_i}\}, \quad (16)$$

where: n - number of bones related to the vertex v;

IBM_{H_i} - inverted bind-pose matrix for the bone H_i; JM_{H_i} - moment matrix for the bone H_i; JW_{H_i} - weight coefficient for the influence of the bone H_i on the vertex v.

For the modeling of animation process we use so-called morphing method. It is a way of smooth change of properties from one state to another. The states in between are called key states. Intermediate states are computed based upon key states using weight coefficients. Thus it is a key part for the animation. The morphing process for face mimics is described as follows:

Face mimics is built for a virtual 3d human model is built using a segmented (or targeted) morphing of the head in a common state and the head showing the most distinct face mimics.

The formula for relative morphing of M basic morphs is:

$$V' = V + \sum_{m=1}^{M} w_m \cdot TM_m \,, \tag{17}$$

where w_m - weight coefficients, V – mesh of the basic model without morphing, TM_m - the mesh of the distinct face mimics. The result is a linear combination of meshes.

Method of segmented morphing allows to form several facial expressions based of a relatively small amount of morphs. It enables animation as well.

Additional advantage of segmented morphing is a feature to animate jaws independently of eyes, lips and emotional expression.

For the 3d reproduction of gestures and face mimics an application has been created which implements the skeleton-based (for the gesture) and morphing animation. The application benefits from OpenGL 3d API and shows 3d human (see figure 9) using the attributes mentioned above. The application uses morphing and skinning to show animation (see figure 10).

Figure 9. 3D Virtual Human

Figure 10. Frames of animation for the gesture "you"

For testing the correctness of gestures appropriate software has been created (see figure 11):

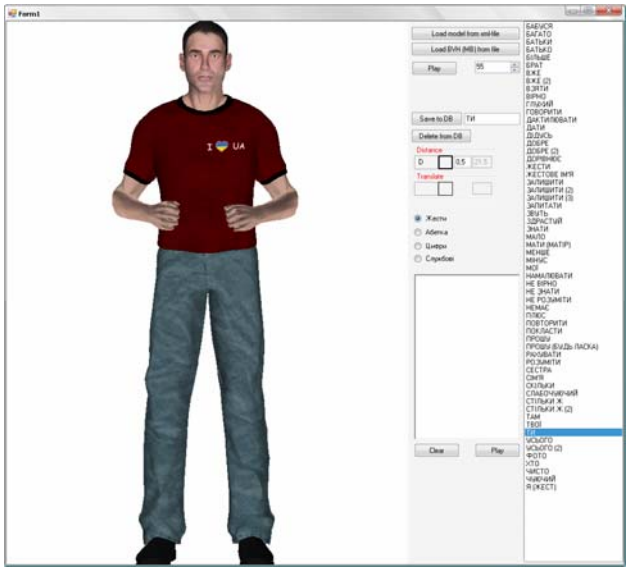


Figure 11. Software for gesture reproduction testing

Implementation of Ukrainian sign language

A specialized software has been created to implement the Ukrainian sign language. The software uses the same methodology as in Ukrainian schools for deaf children. Those underlying teaching materials are issued by the Ministry of Education of Ukraine [6] and are targeted at a beginner's level.

The functionality included in software has 3 informational categories (topics, words and sentences) and viewport with virtual human who shows gestures. The primary information category is "topics". It contains main methodological information for each lesson: points to make, skills to teach (or to learn), necessary information for comprehension and explanations. It also contains a list of gestures relative to the topic and sample sentences with the gestures (see figure 13).

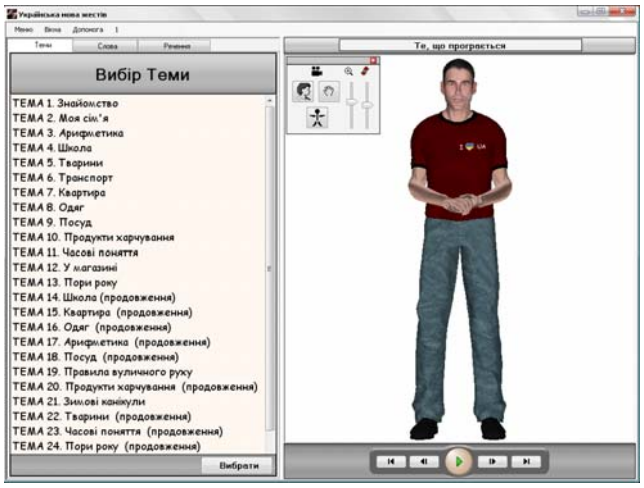


Figure 12. Software application "Ukrainian Sign Language"

The categories "Words" (see figure 14) and "Sentences" are secondary. They contain all the gestures and samples of sentences respectively.

Gesture reproduction viewport has a special meaning. First of all, because of the feature to demonstrate custom gesture's dynamics, that is, to show a gesture frame-by-frame. Because the gestures were digitized using recordings of native speakers and assuming that the software will be widely spread, it gives a rise to a standard of the sign language. The feature of showing a gesture frame-by-frame is a true competitor to a live teacher who needs to demonstrate gestures very slow and with lots of explanations during the teaching process. Using a single program for teaching sign language means that all the deaf children on all the territory of Ukraine will know the same gestures and thus communicate more effectively. This is how a standard of the sign language can emerge.

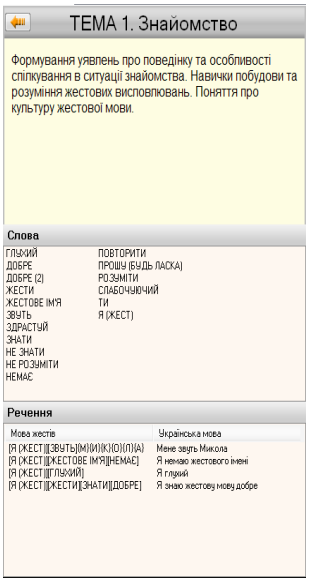


Figure 13. "Topic" category

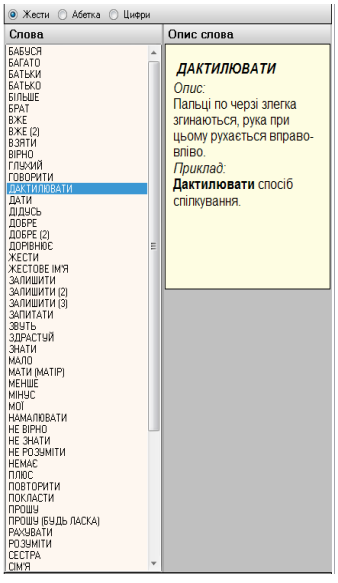


Figure 14. "Words" category "Meeting new people"

Conclusions

A set of gestures has been digitized using motion capture technology and material with recorded gestures. Gesture reproduction on a virtual human showed the technology to be able to repent naturally looking gestures, all very close to the video they have been digitized from.

Appropriate software has been created, which contains several lessons of the Ukrainian sign language. The lessons have been taken from the materials for schools for the deaf children [6].

The created software is a candidate to represent a standard of the Ukrainian sign language. It can solve the problem of differences between gestures with the same meaning, which are inevitably present as a result of teaching from different speakers.

Further development is aimed on improvement of the suggested technology:

- taking into account natural limits for each connection of bones in order to have a virtual model capable of intelligent controlling;
- fill the database with the major set of gestures of the Ukrainian sign language, that is, to create a standard of the language;
- create a mean for semantic connection of sentences in the spoken language to sentences in the sign language.

Acknowledgment

We are very grateful to Michael Rodda a Chartered Psychologist, Professor Emeritus at the University of Alberta, and Foreing Member of the Academy of Pedagogical Sciences of Ukraine for the support of our research. It is very important for us.

Bibliography

[1] Kryvonos Yu.G., Krak Yu.V., Barmak O.V., Ternov A.S. “Information technology of nonverbal communication with deaf people”, Artificial Intelligence, no 3, 2008, p. 325-331. (on Ukrainian)

[2] Krak Yu.V., Barmak O.V., Gandha O.S., Ternov A.S., Shatkovskii N.N. “Computer system for communication with deaf people”. Advanced Studies in Software and Knowledge Engineering, International Book Series “INFORMATION SCIENCE & COMPUTING”, Number 4, Supplement to the International Journal “INFORMATION TECHNOLOGIES & KNOWLEDGE”, Volume 2, 2008, p.161-165.

[3] Lander J. “Working with Motion Capture File Formats. Game Developer”. Miller Freeman Inc. USA., 1998

[4] Bilodid J.K. “Modern Ukrainian language”. Kyiv, Naukova dumka, 1969. (on Ukrainian)

[5] Darris Dobbs and Bill Fleming. “Animating Facial Features and Expressions”. Charles River Media, 1999.

[6] Hryschenko E.S., Stiopkin V.V. “Ukrainian sign language for primary school”. Kyiv, Bohdana, 2004. (on Ukrainian)

Authors' Information

Iurii Krak – *the senior scientist, Institute of Cybernetics of National Academy of Science of the Ukraine, address: 40 Glushkov ave., Kiev, Ukraine, 03680; e-mail: krak@unicyb.kiev.ua*

Iurii Kryvonos – *the Deputy Director, Institute of Cybernetics of National Academy of Science of the Ukraine, address: 40 Glushkov ave., Kiev, Ukraine, 03680*

Olexander Barmak – *the senior scientist, Institute of Cybernetics of National Academy of Science of the Ukraine, address: 40 Glushkov ave., Kiev, Ukraine, 03680;e-mail: barmak@svitonline.com*

Anton Ternov – *the junior scientist, Institute of Cybernetics of National Academy of Science of the Ukraine, address: 40 Glushkov ave., Kiev, Ukraine, 03680*

Bohdan Trotsenko – *the junior scientist, Institute of Cybernetics of National Academy of Science of the Ukraine, address: 40 Glushkov ave., Kiev, Ukraine, 03680*

Knowledge Engineering

THEORETIC-EXPERIMENTAL MULTICRITERIA METHOD FOR NEURAL NETWORK CLASSIFIERS' ARCHITECTURE

Albert Voronin, Yuriy Ziatdinov, Anna Antonyuk

Abstract: *The problem state and multicriteria optimization procedure of neural network classifier's architecture is considered. The scalar convolution of criteria with nonlinear trade-off scheme is offered as a goal function. The search methods of optimization with discrete arguments are used. The neural network classifier of texts as an example is given.*

Keywords: *multicriteria optimization, neural nets, classifier.*

ACM Classification Keywords: *H.1 Models and Principles – H.1.1 – Systems and Information Theory; H.4.2 – Types of Systems; C.1.3 Other Architecture Styles – Neural nets*

Conference: *The paper is selected from XVth International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.*

Introduction

The important version of artificial neural nets are neural network classifiers. They are used for technological and medical diagnostics, different kind of information sources classification. In general case the structure of q-layer feedforward neural network classifier is represented in Figure 1.

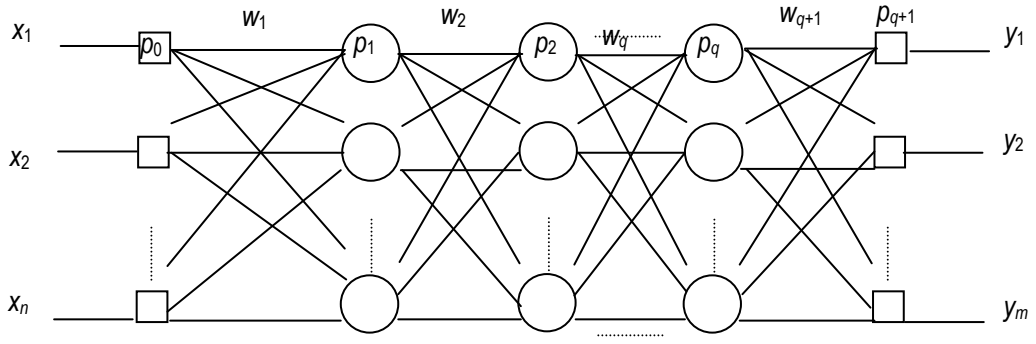


Fig. 1

Here $x_1, x_2, ..., x_n$ are features of classification object which compose the input vector $x = \{x_i\}_{i=1}^{n}; p_0 = n$ is the number of neural elements in the receptor layer; $p_1, p_2, ..., p_q$ mean number of neurons in each of hidden layers; $p_{q+1} = m$ is the number of neurons in the output layer (number of classes); $y = \{y_k\}_{k=1}^{m}$ is output vector of neural network which defines belonging of classification object to one of m classes; $w_1, w_2, ..., w_q, w_{q+1}$ - is the vector of synaptic weights of a neural network.

Let's present some necessary items from the neural networks theory [1-3]. Artificial neural network is a set of neural elements and connections between them. Each neuron has a group of synapses - unidirectional input links connected to the outputs of others neurons. Each synapse is characterized by its weight (determined at learning of a network). Neuron has current state defined as a weighted sum of its inputs: $s = \sum_{i=1}^{n} w_i x_i$. The output of neuron is a function of its state which is called the activation function: $y = f(s)$. Activation or inhibition signal through the axons (output connection of the neuron) goes to the following neuronal synapses. Activation functions can be of threshold or continuous kinds (bipolar sigmoid, gaussian, etc.). The set of all neurons of artificial neural network is divided into subsets, called layers. Layer is a set of neurons on which in every time tact enter the parallel signals received from other neurons of the network [2]. The output of the classifier is a vector of activation functions $y = \{y_k\}_{k=1}^{m}$. The index j *for* which output has maximum activity, i.e. $\max_{k \in [1,m]} y_k = y_j$, corresponds to the index of classification object class.

The number of input layer neurons is determined by the dimension of input attributes vector and can not be edited. Similarly, the number of neurons of output layer $p_{q+1} = m$ is determined by the number of classes on which the space of characters is divided, and is also a constant value. The number of processing (hidden) layers q and the number of neurons in each of them represents the conception of neural network architecture [1] and can serve as arguments (independent variables) in process of its optimization.

In this paper, we reduce the research to the case where value q is fixed and specified. Then optimization arguments of neural classifier architecture are the number of neurons in every processing layer which compose the vector of independent variables $p = \{p_j\}_{j=1}^{q}$. The quality of neural classifier operating depends on choice of architecture we do.

The problem consists in choice of such architecture when neural classifier has the best operating properties in given conditions.

Problem formulation

In general, the problem can be formally represented by the problem

$$p^* = \arg \underset{p \in P}{extr}\, Y(p), \tag{1}$$

where $Y(p)$ is objective function; $\underset{p\in P}{extr}$ is an operator of objective extremalization function with arguments p; P is admissible domain of independent variables.

Let us make additional particular assumptions for constructive problem solution. Let associate each property of neural classifier with a quantitative characteristic $f(p)$ *which has a sense of* quality operating criterion. One of such criteria is the probability of classification error. Let us define this criterion experimentally and approximately present it as the number of classification errors $e(p)$ divided into the general, large enough number of tests N:

$$f_1(p) = \frac{e(p)}{N}. \quad (2)$$

It is supposed that with growth of neurons number in processing layers within some reasonable limits classification accuracy increases, and the value of this criterion decreases. Maximum permissible network error value should be known from physical considerations and is given as a restriction $f_1(p) \le A_1$.

The second criterion characterizes time that is needed to train the neural network with current architecture p. There is a strong correlation between that time and the total number of classifier neurons in hidden layers. So, let represent this criterion in the form of

$$f_2(p) = \sum_{k=1}^{q} p_k. \quad (3)$$

It should be noted that this criterion also characterizes signal passing time through the neural network from input to output. Criterion value increases with growth of neurons number. The maximum permissible value of the second criterion is defined by acceptable neural network training time and is presented as the restriction $f_2(p) \le A_2$.

There are other criteria for different properties of the neural classifier characteristic. In this paper we will confine ourselves to presenting just two main criteria, remembering that the proposed method allows including the other classifier properties.

Admissible domain of optimization arguments is given by the parallelepiped restriction $P = \{p | 0 \le p_k \le P_u, k \in [1, P_u], u \in [1, q]\}$, where P_u is the maximum number of neurons in u-th layer.

As the both of included in view criteria should be minimized (the smaller criterion, the better corresponding property of classifier), so objective function extremalization operator becomes $\underset{p\in P}{extr} = \min_{p\in P}$.

Thus, the both of criteria are contradictory, nonnegative, subject to minimization and limited. There are all the grounds to use the scalar convolution of criteria with nonlinear trade-off scheme as an objective function [4]. In unified version such convolution is represented by the formula

$$Y(p) = Y[f(p)] = \frac{A_1}{A_1 - f_1(p)} + \frac{A_2}{A_2 - f_2(p)}. \quad (4)$$

where $f(p) = \{f_r(p)\}_{r=1}^{r=2}$ is two-dimensional vector of particular criteria. Considering (2), (3) and (4), the optimization problem of neural classifier architecture for expression (1) is transformed to

$$p^* = \arg\min_{p \in P} \left[\frac{A_1}{A_1 - e(p)/N} + \frac{A_2}{A_2 - \sum_{k=1}^{q} p_k} \right]. \quad (5)$$

It is easy to see that dependence $e(p)$ in formula (5) a priori is unknown and subject to experimental determination.

Method of solution

Among the optimization problems, there are such of them where arguments can acquire only discrete values according to their physical nature. Discrete values can always be reduced to an integer by special normalization. Such problems are considerably more difficult then continuous multicriteria problems and for their solutions should be applied other approaches [5].

The set of acceptable discrete values may be infinite, finite, or even consisted of only two values, 0 and 1, for example. In the first case the problem degenerates into a continuous optimization problem. To solve it the efficient and formalized algorithms and software are proposed in [4]. In the last case, the integer programming with Boolean variables, with specific methods occurs (logic synthesis of finite automaton, Rvachev's functions, etc.). From our standpoint, the most interesting and substantial is the case when the set of acceptable discrete values is not so large that problem degenerates into a continuous one, but also is not so small that problem can be solved by simple enumeration. Just this kind is the problem of nonlinear discrete (integer) programming.

Methods of discrete programming don't have such unity as methods of variational calculus have, and in most cases represent a set of particular techniques suitable for solution of particular problems. But their urgency requires their development and improvement because, as a rule, the most of important applied problems are reduced to the problems of partially or completely discrete programming. The complexity of solving discrete (integer) programming problems increases in the case of multicriteria problem.

In the case when components of possible multicriteria problems solutions can acquire only discrete values $p_k^{(P_u)}, k \in [1, P_u], u \in [1, q]$, the scalar convolution of criteria with nonlinear trade-off scheme $Y(p)$ is the lattice function defined on the discrete set P. Optimization of lattice objective function build on nonlinear trade-off scheme is reduced to the nonlinear programming problem with discrete (integer) arguments, which solution, as noted above, is difficult enough.

To solve this problem, we assume that under the discrete set P there is an auxiliary area of continuous arguments $p_c \in P_c$, which contains all discrete points $p_k^{(P_u)}$ and all continuous space between them. In the area P_c the continuous function $Y(p_c)$ is defined, which coincides with the lattice function $Y(p)$ in points $p_k^{(P_u)}$.

This assumption allows obtaining analytical solution, if in the expression (5) the dependence $e(p)$ is specified in the regression model form, for example.

Then it is possible to use the necessary condition of the function's minimum: $\frac{\partial Y(p_c)}{\partial p_c} = 0$. The solution of this equations system gives the trade-off - optimal continuous point $p_c *$. The last step of the algorithm is searching on P the discrete point nearest to $p_c *$, which will be the required discrete solution $p *$. Unfortunately, in our case specifying of the analytic dependencies is very difficult or even impossible.

As a basic we consider the case when functions $e(p)$ and therefore $Y(p)$ are unknown, but it is possible to define $Y(p)$ function's values at the points $p_k^{(P_u)}$ by measurement or calculation. Then we can organize a nature or calculating experiment, which in result will realize the search movement to the required trade-off - optimal discrete point $p *$.

There are different approaches to the search procedure organization, which should give the sequence of improving solutions. One of them is the discrete analogue of the Nelder-Mead simplex-planning method (the method of deformed polyhedron) [4]. This is the modification of gradient methods, which is very often and successfully applied in practice. The second is the non-local (dual) approach [4], which is often more efficient than the gradient methods.

As the search procedures use local or nonlocal models of continuous function $Y(p_c)$, so for called above variants exist the general necessity of searching the discrete point p_d on P which is the nearest to the continuous solution p_c at the current or final iteration. If the number of hidden layers q is not great, then the solution of this problem is not difficult (a simple rounding to the integer value). With multilayer classifiers we recommend to use the following algorithmic technique. At the point p_c is placed the center of hypersphere, which diameter increases from zero until the surface of the sphere will not touch the nearest discrete point, which thereby is identified as p_d. There are various software implementation of this algorithm. There are neural classifiers of various types and purposes.

Multicriteria optimization of neural network texts classifier

As an example, let us consider in general terms the optimization problem of neural network text classifier architecture. Text-classification system [3] consists of two main parts: the frequency analyzer with system dictionary and neural network classifier itself (Fig. 2).

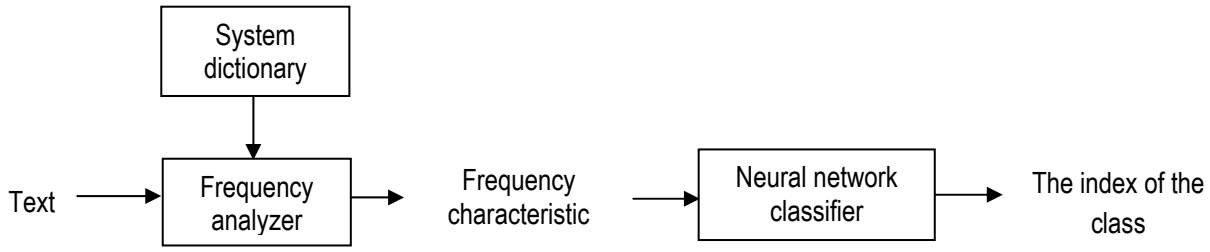


Fig.2

The text enters the input of the system, the output conforms the index of subject to which the text refers (business, politics, medicine, sport, spam, etc.). Before we proceed to optimize the neural network classifier architecture, it is necessary to perform the following steps:

1. Identifying m classes, with which the system will operate.

2. Selecting suitable training texts $t_k, k \in [1,m]$ and verifying (test) texts $t_l, l \in [1,L], L \geq m$.

3. From the set of training texts in special way words $v_i, i \in [1,n]$ are extracted and the system dictionary V is formed.

4. The frequency analyzer determines for each word v_i from the system dictionary V its frequency of occurrence x_i in the given text t_k. Frequency characteristic is the vector $x = \{x_i\}_{i=1}^{n}$ of text attributes, which dimension is equal to the number of words in the system dictionary $v_i \in V$.

After obtaining of the training texts frequency analyzer results, we can begin to train the neural classifier with some architecture $p = \{p_j\}_{j=1}^{q}$. The training process of neural network consists in specifying of such its weighting coefficients $w_1, w_2, ..., w_q, w_{q+1}$ where the maximum network error at the training texts for this architecture does not exceed the maximum permissible value. Specific learning algorithms are not considered.

Now we can proceed directly to the vector optimization procedure. To optimize the neural network classifier architecture let us use the search method of simplex-planning. Suppose, for specificity that the number of processing layers q = 2. Then the idea of the method in continuous form can be illustrated by Fig.3.

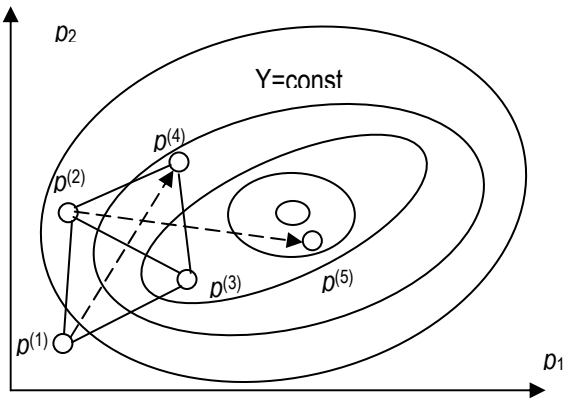


Fig.3

On the arguments plane $p_1 - p_2$ in some starting area we construct the initial regular simplex, which in two-dimensional case is represented by isosceles triangle with apexes $p^{(1)}, p^{(2)}, p^{(3)}$. For each of three simplex architectures we realize the process of the classifier learning and feed to inputs a series of test texts t_l. In each simplex apex we determine the number of classification errors $e^{(1)}, e^{(2)}, e^{(3)}$ with total number of tests $N = L$. By formula (2) we get criteria $f_1^{(1)}, f_1^{(2)}, f_1^{(3)}$. By formula (3) criteria

$f_2^{(1)}, f_2^{(2)}, f_2^{(3)}$ are defined. Formula (4) that serves in this case, as not objective, but as the evaluation function for our example is given by

$$Y(p_1, p_2) = \frac{A_1}{A_1 - e(p_1, p_2)/L} + \frac{A_2}{A_2 - p_1 - p_2} \qquad (6)$$

It provides the values of the scalar convolutions $Y^{(1)}, Y^{(2)}, Y^{(3)}$ for start simplex architectures. Comparing these values between each other, we find that one of them, $Y^{(1)}$, for example, is bigger (i.e. worse) than others. Most probably it could be predicated that the architecture $p^{(4)}$ obtained by mirror reflection of the worst in base simplex point $p^{(1)}$ relatively to the center of opposite bound, would be better. By all calculations for architecture $p^{(4)}$ let us form the new simplex with apexes $p^{(2)}$, $p^{(3)}$ and $p^{(4)}$. Comparing the values $Y^{(2)}, Y^{(3)}, Y^{(4)}$, we find that one of the points, $p^{(2)}$, for example, is worse than others in the second simplex. By reflecting of this point relatively to the center of the second simplex opposite bound, we get the architecture $p^{(5)}$, etc., until we get the architecture p^* that corresponds to the objective function minimum.

This is only the illustration of the simplex-planning method idea. In fact, this method in the Nelder-Mead modifying provides simplexes adaptation to the topography of the objective function by means of the polyhedrons deformation, it has well-developed algorithms and software. In addition, we must not forget that we have the case of optimization with integer arguments, that dictates the necessity of nearest discrete solution p_d searching for every obtained continuous solution p_c.

The second, non-local search method is rather complicated in realization, but it is usually more effective [4, 5]. The method is based on the iterative construction of the non-local model $Y(p)$ «floated» together with the system of changing base points and refined by experimental results. The set of control points is compressed and constricted to the required extremum point (« shagreen leather »). At each iteration at the same time and interdependent is realized as our conception about the objective function at extremum region improvement, so the definition of such arguments extremum estimation, which is adequate to the level of these representations at the given iteration. Therefore, the non-local optimization method refers to the dual class and can be named as the method of dual programming.

Both of search methods provide the series of experiments execution. Obtained herewith experimental data can be used to build analytical regression models of particular criterion $f_1(p) = e(p)/L$. Using these models, we can realize not search, but analytical vector optimization of the other same type neural network classifiers architecture. If it will prove to be difficult, then a search procedure is executed, and with not nature, but computational experiment, that is much easier.

Solving the problem of the regression models construction, we must specify the type of approximating dependence, known accurate within unknown regression coefficients. Analysis of the problem leads to the assumption that with sufficient for practice accuracy, we can limit ourselves by the linear regression:

$$f_1(p_1, p_2) \approx (a_1 p_1 + a_2 p_2)/L, \qquad (7)$$

where a_1, a_2 are regression coefficients, determined from experimental data by least-squares procedure. A linear regression model is checked for the adequacy by mathematical statistics methods. If it is necessary, the model can be complicated.

Considered methods provide the start of search procedure from the architecture, which in decision-maker's opinion is close enough to the optimal point. If in the search procedure occurs increasing of neurons number in processing layers, the neural networks theory [1] characterizes this approach as the constructive. If the start number of neurons is superfluous the approach is as called destructive (The Rodin principle: to model a sculpture, you need to take a block of marble and remove from it unnecessary).

The implementation of stated in the paper stages of vector optimization provides such neural network classifier architecture, which systematically correlates contradictory criteria of its functioning effectiveness.

Acknowledgements

The paper is published with financial support by the project ITHEA XXI of the Institute of Information Theories and Applications FOI ITHEA Bulgaria www.ithea.org and the Association of Developers and Users of Intelligent Systems ADUIS Ukraine www.aduis.com.ua.

Bibliography

1. E.V. Bodyansky, O.G. Rudenko, Artificial neural networks: architectures, training and application [in Russian], TELETEH, Kharkov: (2004).
2. V.A. Golovko, Neural networks: training, organization and application [in Russian], IPRZHR, Moscow (2001).
3. V.S. Borisov, "Self-training texts classifier in natural language", Cybernetics and system analysis, No. 3., 169-176 (2007).
4. A.N.Voronin, J.K. Ziatdinov, A.I. Kozlov, Vector optimization of dynamical systems [in Russian], Tehnika, Kiev (1999).
5. A.N.Voronin, P.D Mosorin., A.G Yasinsky, "Multicriteria problems of optimization with discrete arguments", in: Automation, 2000. International conference of automatic management: Works. -Vol.1, 75-78, Lvov (2000).

Authors' Information

Albert N. Voronin *- National Aviation University, faculty of Computer Information Technologies, Dr.Sci.Eng., professor, 03058, Kiev-58, Kosmonavt Komarov avenue, 1, Ukraine, e-mail: alnv@voliacable.com*

Yuriy K. Ziatdinov *- National Aviation University, faculty of Computer Information Technologies, Dr.Sci.Eng., professor, Head of Chair; 03058, Kiev-58, Kosmonavt Komarov avenue, 1, Ukraine, e-mail: oberst@nau.edu.ua*

Anna A. Antonyuk *- National Aviation University, faculty of Computer Information Technologies, post graduate student, 03058, Kiev-58, Kosmonavt Komarov avenue, 1, Ukraine, e-mail: niuriel@mail.ru*

AN INVESTIGATION OF PROBLEM OF ARTIFICIAL NEURON NETWORKS APPLICABILITY FOR SOLUTION OF PROBLEM OF FORECASTING OF GAS DISPERSION IN ATMOSPHERE

Igor Shostak, Valentina Davidenko

Abstract. *The problems of artificial neuron networks (ANN) applicability for forecasting of hazardous pollutants (HP) discharge into the atmosphere are considered in this article. It is spoken in detail about existing approaches' and methods' advantages and disadvantages. The fundamental scheme of ANN usage for modelling is offered. The scope of its applicability is described.*

Keywords*: emergency, artificial neuron networks, air pollution, gas dispersion.*

ACM Classification Keywords*: H.3.4 Systems and Software*

Conference*: The paper is selected from XV[th] International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.*

Introduction

The problem of gas dispersion modelling is of interest not only in science, but also in practical life, especially if this problem has deal with emergency discharge of HP into the atmosphere due to emergency situation at transportation time or in permanent placements (stores, chemical plants and so on). These situations in turn increase probability of HP affection of civilian population or possible damage of special resources.

The problem of provision of emergency services with precise information about forecasting of HP emission in the atmosphere for efficient salvage operations.

Problem Statement

In terms of mathematical description the problem of gas dispersion modelling comes to evaluation of function (1) with known initial parameters.

$$F(t_n, \{Met_n ... Met_0\}, \{x_n, y_n, z_n\}, \{t_0, Met, S, \{x_0, y_0, z_0\}\}) = Con_n, \quad (1)$$

where t_n – interest time at which we want to know gas concentration Con_n;

$\{x_n, y_n, z_n\}$ – location coordinates of interest place in which we want to know gas concentration Con_n;

$\{Met_n ... Met_0\}$ – description of meteorological parameters in a period *[0..n]* time units;

$$\{Met_i = \{Met_{ij}\}, \quad (2)$$

where Met_{ij} – meteorological parameters which are get at time moment *i* from meteorological station *j*;

$$Met_{ij} = \{Temp_{ij}, Humid_{ij}, WS, WD\}, \; j=1,k, \quad (3)$$

where $Temp_{ij}$ – set of temperatures measured from different heights;

$Humid_{ij}$ – relative air humidity;

WS – set of wind speeds measured on different heights;

WD – set of wind directions measured on different heights;

k – amount of meteorological stations;

S – emission characteristic parameter:

$$S = \{Type, Vol\_All, Press_0, Temp_0, Vol\_Ex\}, \quad (4)$$

where *Type* – gas's type;

Vol_All – total gas volume in the source;

Vol_Ex – released gas volume ($Vol\_Ex \le Vol\_All$);

$Press_0$ – gas pressure inside the source at the release moment;

$Temp_0$ – gas temperature inside the source at the release moment.

However the problem of evaluation of gas concentration in some fixed location with certain radius is greatly more actual in the real practice.

The problem of modelling for short-term periods is considered. So having these conditions we can account meteorological parameters to be constant for further calculations.

Analysis of Existing Approaches

Existing models and approaches which describe gas dispersion in the atmosphere don't have a satisfactory accuracy when its deal with open space. And they are almost useless for forecasting and further building of the most precise picture of HP dispersion into the urban regions.

In some cases when gas dispersion process is described by physical laws expressed through mathematical formulas it becomes impossible to take into account all physical and statistical laws. And it leads to simplify to that view which can be programmed (e.g. Navier-Stokes equations) but accuracy is extremely decreased. In other cases Gaussian plume model and its modifications are used but it give to us only a very approximate picture what is going on, because shape of potential gas dispersion region is determined as ideal regular curve.

The Lagrange and Eurlian model-based algorithms give more precise results. The main idea of these methods consists in division of gas emission into many elementary boxes with individual characteristics. But these models don't sufficiently take into account complexity of that processes which occur at dispersion time, especially in cases concerning with complex obstacles (e.g. 3D location between arbitrarily situated buildings).

Task Decision

So assigned task has following characteristics:

- nonlinearity;
- accounting of huge count of factors (meteorological conditions, geometrical parameters of landscape and different objects);
- dynamics of process and parameter variability.

ANN can theoretically cope with this task (as follows from problem definition). And it is one of all ANN's purposes [1]. It is also declared in [2] about possibility of application of artificial intelligence techniques, especially about ANN for decision of similar tasks. It is necessary to formulate the task in terms of ANN and properly prepare data for possibility of ANN application.

Figure 1 shows scheme of gas dispersion in 2D flat ground which includes following peculiarities:

- the area is divided into conventional square regions - the cells;

- coordinate system is usually relative with center in the locate of gas emission source with information S (4);
- meteorological stations with information Met (2).

It is obvious that all meteorological parameters are not known exactly at any point of space under consideration. In order to determine approximate parameter value an interpolation of existing data is used.

- gas concentration level is determined in all corners of space cells, i.e. any cell is characterized by concentration data C:

$$C_{ij} = \{M_{ij}, Con_{ij}\}, \text{ i=1,n, j=1,m,} \quad (5)$$

where M_{ij} – interpolated environment parameters at time *i* for cell *j*;

Con_{ij} – concentration level at time *i* in cell *j*;

m – current amount of cells;

n – amount of time units.

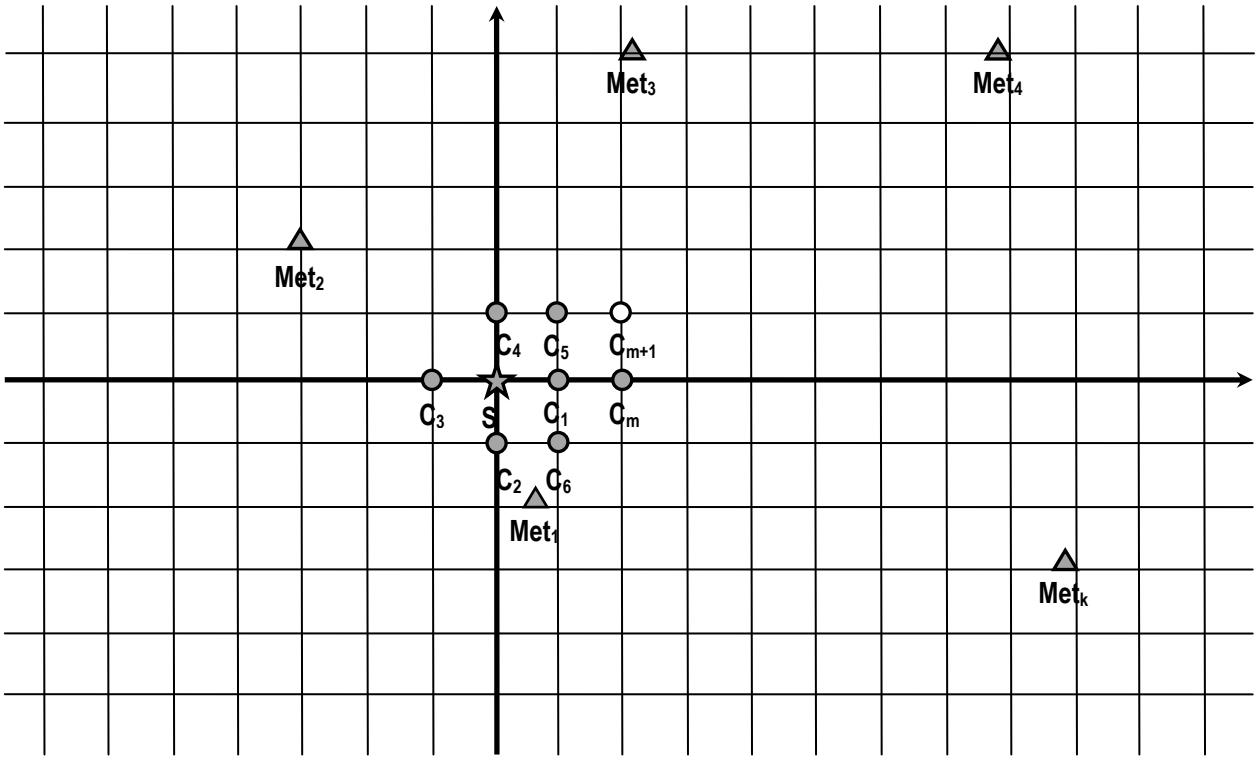


Fig. 1. Gas dispersion general scheme

– gas emission source;

– meteorological station;

– locations with determined concentration;

– location with undetermined concentration.

The algorithm of determining of gas concentration in any point of space has similar nature of graph breadth-first search algorithm, i.e.:

1. At first concentrations in the cells which are the nearest to the release source are determined.

2. A concentration in the cell is determined by using concentration data from neighbor cells and meteorological conditions interpolated from parameters measured in the weather stations (Fig.2). This step is performed by means of ANN.

The ANN's structure is multi-layer feed-forward perceptron. The learning process is implemented on data sampling which is received from different experiments. The measurements of meteorological parameters and gas concentration level are made in different places of the space when experiments are carried out [3].

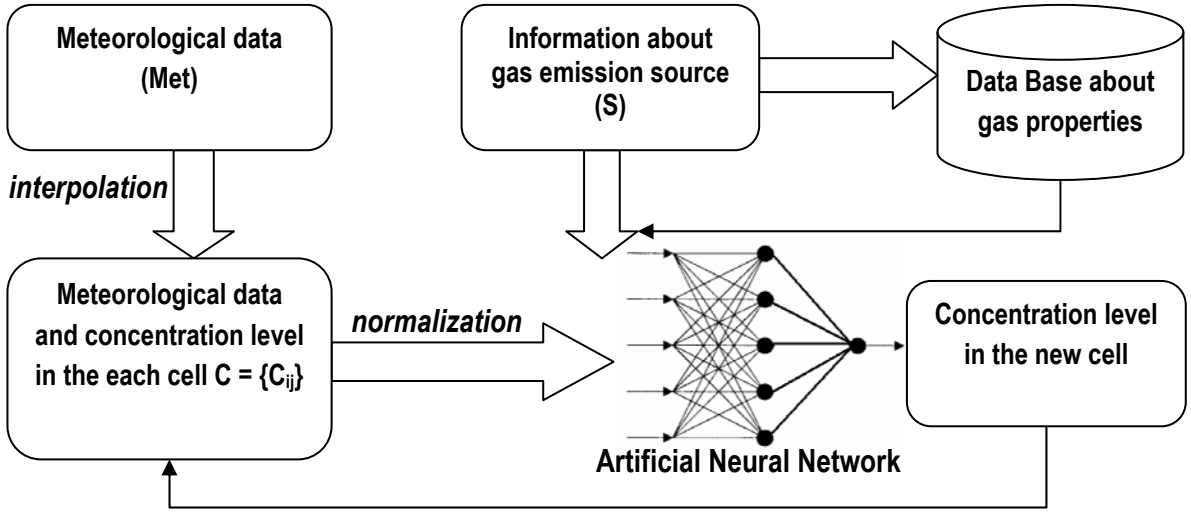


Fig. 2. General scheme of algorithm of computation of concentration in the cell

The advantages of proposed approach:

- accounting of many factors with different natures in one system;
- an ability to find decisions in nonlinear dependences.

The disadvantages of proposed approach:

- this approah needs a lot of time to analyse whole picture;
- there is an ANN's teaching problem: it is hard to receive precise information about different data.

Proposed approach testing

Proposed approach was tested on base of data from open source files [5].The Kincaid field experiments were performed in 1980-1981. This investigation was a part of the EPRI Plume Model Validation and Development Project. A numerous reports described these experiments are comprehensive source of information about gas dispersion process. The Kincaid power plant is situated in Illinois, USA and is surrounded by flat farmland with some lakes. The power plant has a 187 m stack with a diameter of 9 m. During the experiment, SF6-gas was released from the stack. The most meteorological measurements were taken from a "Central Site" located around 650 m east of the Kincaid plant. The radiosonde data supplied on diskette are routine data from the station Peoria, 120 km north of the source.

So as stated above it is necessary to represent research area in the grid format. This performance claims two stages:

1. To choose a step – grid cell size. In fact this is distance in kilometers.
2. To calculate average concentration in each cell.

A step parameter has been chosen according to the following rule: the amount of cells with nonzero concentration must be maximal and the amount of cells with zero concentration because this would make ANN teaching easier. As a result of the test carried out on base of data extracted on May 7, 1980 at 11 a.m. is Tab.1.

Tab. 1. The "useful" cells amount dependence on step parameter

Step size, km	An amount of cells with average concentration $0<con_i<10$	An amount of cells with average concentration $10<=con_i<100$	An amount of cells with average concentration $100<=con_i<1000$	Total amount of cells with nonzero average concentration
1	7	9	9	25
0.5	6	13	15	34
0.1	7	19	21	47
0.05	7	19	21	47
0.01	7	19	21	47

So, optimal step size is 0.1 km.

After preliminary described above the set of data which specifies concentration distribution for each cell of investigated area at certain time is generated. The grid with 0.1 km step for situation on May 7, 1980 at 10 a.m. is depictured on Fig.3. This figure shows that cells with nonzero concentration is not situated consecutive but have disordered arrangement. Gas monitors were located rarely and this explains previous fact, but it doesn't mean that unmeasured concentration is equal to zero. But lack of information about precise data doesn't allow to use these cells for analysis. It is offered the following decision of this problem: not only neighbouring cells concentrations have to be included into the learning sample for ANN but also distances between cells and analysis cell (is bold encircled in Fig.3) have to be included too.

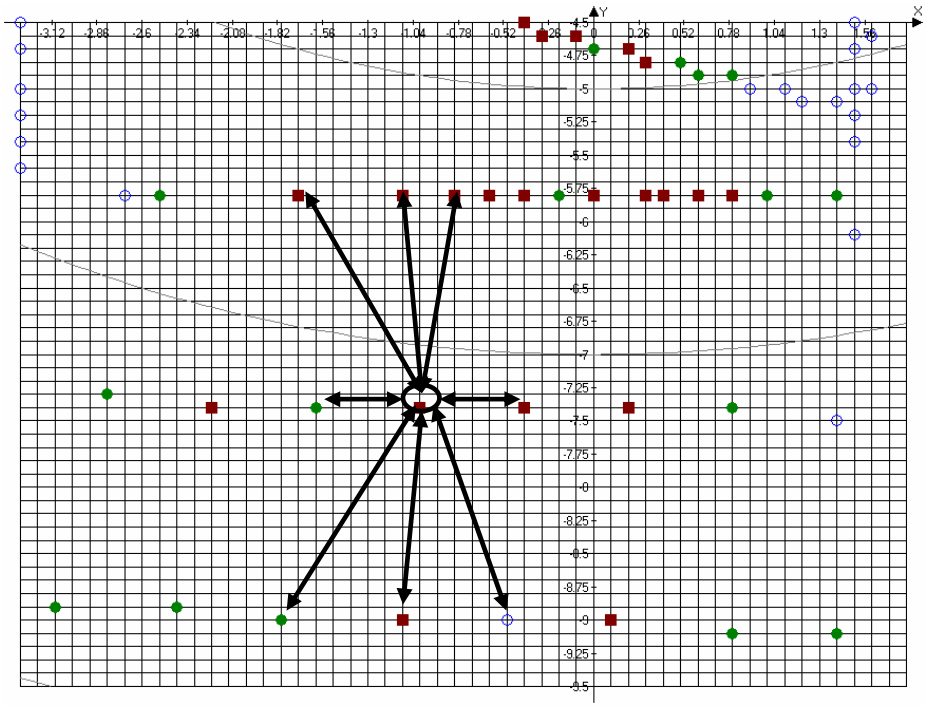


Fig.3. SF6 gas concentration distribution on May 7, 1980 at 10 a.m.

empty circles – cells with average concentration $0< con_j <10$;

filled circles – cells with average concentration $10<= con_j <100$;

filled squares – cells with average concentration $100<= con_j <1000$.

The analysis of eight nearest cells is performed for each cell taking into account distances. So, the learning sample LernSamp has following structure (6):

$$LearnSamp = \{learn_i\}, \text{ i = 1..k,} \quad (6)$$

where $learn_i$ – training pattern;

k – amount of training patterns.

$$learn_i = \{Con_i, Dist_i, WS_i, WD_i, Temp_i\}, \quad (7)$$

where $Con_i = \{con_j\}$ – concentration levels in the eight nearest cells, j=1..8;

$Dist_i = \{dist_j\}$ – distances between eight nearest cells and analysis cell, j=1..8; $WS_i, WD_i, Temp_i$ is described above (3).

As stated above ANN is multi-layer feed-forward perceptron. As a result of tests carried out network's structure was determined: first (input) layer consists of 20 neurons (19 of input data + 1 addition neuron); first hidden layer consists of 4 neurons, second hidden layer – 6 neurons. Output layer consists of single element which output result determines a grade of membership to classes with low (<10), middle (<100) or high (<1000) concentrations. Hyperbolic tangent function (8) is chosen to be activation function for ANN testing.

$$\varphi(v) = a \cdot th(b \cdot v), (a, b) > 0 \quad (8)$$

As a result of ANN's teaching the average learning error was minimized to 0.007 value (Fig.4).

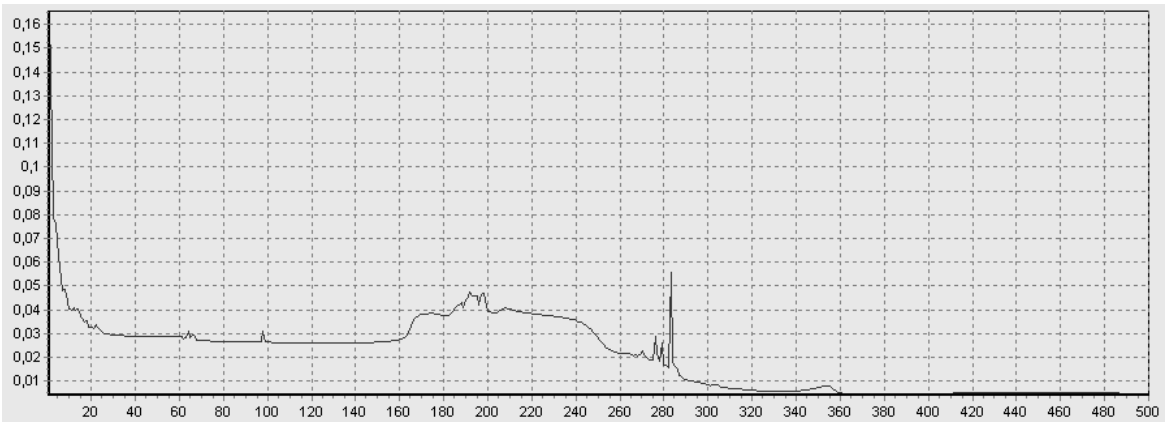


Fig. 4. The average learning error – learning epoch curve

Conclusion

Proposed approach can be applied for gas dispersion modelling in case of accidental releases of hazardous pollutants. Unfortunately, there is no research works about modelling such situations in Ukraine. It leads to impossibility to estimate model efficiency in circumstances of certain climatic zone. Nevertheless estimations made on basis of data [3] confirm acceptability and applicability of chosen approach of concentration evaluation. For the further development of this approach it is necessary:

- to carry out additional research of dependence of effective ANN teaching from meteorological parameters;
- to give a qualitative and quantitative assessments of the algorithm by means of data received experimentally.

Acknowledgements

The paper is published with financial support by the project ITHEA XXI of the Institute of Information Theories and Applications FOI ITHEA Bulgaria www.ithea.org and the Association of Developers and Users of Intelligent Systems ADUIS Ukraine www.aduis.com.ua.

Bibliography

[1] Simon Haykin Neural Networks: A Comprehensive Foundation, 2nd edition – Upper Saddle River, NJ, USA: Prentice Hall PTR, 1998. – 842p.

[2] Sven-Erik Gryning, Ekaterina Batcharova Air pollution modeling and its application XIII – 2000. – 815p.

[3] Thompson, R. S. (1993): Building Amplification Factors for Sources Near Buildings – A Wind-Tunnel Study. Atmospheric Environment Part A – General Topics 27, 2313-2325.

[4] M.J. Borysiewicz, M.A. Borysiewicz Atmospheric dispersion modelling for emergency management CoE MANHAZ, Institute of Atomic Energy

[5] Olesen, H.R.2005: User's Guide to the Model Validation Kit. National Environmental Research Institute, Denmark. 72pp. – Research Notes from NERI no. 226.

Authors' Information

Valentina Davidenko *–post-graduate student, National Aerospace University named after N.E. Zhukovsky "KhAI", 61070, Chkalova str. 17, Kharkov, Ukraine; e-mail: Valyuxa@ukr.net*

Igor Shostak *– professor in Software Engeneering depr., National Aerospace University named after N.E. Zhukovsky "KhAI", 61070, Chkalova str. 17, Kharkov, Ukraine; e-mail: iv_shostak@rambler.ru*

СЕГМЕНТАЦИЯ АНОМАЛЬНЫХ ОБЛАСТЕЙ НА МУЛЬТИСПЕКТРАЛЬНЫХ ИЗОБРАЖЕНИЯХ ШЕЙКИ МАТКИ

Катерина Малышевская

Аннотация: *В работе представлена сегментация мультиспектральных изображений шейки матки. Представлена сеть Кохонена для сегментации изображений и коротко изложена ее суть. Приведены результаты работы, сделан анализ полученных результатов и выводы относительно проведенной работы.*

Ключевые слова*: сегментация, сеть Кохонена, шейка матки, диагностика.*

ACM Classification Keywords*: I.5.1 Pattern Recognition - Neural nets*

Аннотация: *В работе представлена сегментация мультиспектральных изображений шейки матки. Представлена сеть Кохонена для сегментации изображений и коротко изложена ее суть. Приведены результаты работы, сделан анализ полученных результатов и выводы относительно проведенной работы.*

Conference*: The paper is selected from XVth International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.*

Введение

В данной работе рассмотрена возможность сегментации мультиспектрального изображения шейки матки. Такая задача продиктована необходимостью ранней диагностики заболевания с использованием компьютерной системы, которая поможет врачу определить области с большим риском возникновения раковой трансформации ткани. Данная система базируется на утверждении, что оптические свойства здоровой клетки отличаются от свойств больной и это отличие более выражено, чем отличия клеток разных людей. В качестве исходных данных в данной работе мы имеем информацию о 206 пациентках, прошедших диагностику в клинике при помощи новой оптической системы, внедренной в медицинском университете Аризоны (США). Параллельно врач ставил диагноз пациенткам путем классификации типов тканей, взятых на биопсию. Участки, из которых была сделана биопсия, были обозначены на снимке, и результаты биопсии были сопоставлены с указанными участками на изображении [Schoonmaker J. 2007]. Данные о каждой пациентке хранятся в отдельном файле формата MATLAB

Постановка задачи

Существуют следующие аномальные области: SEA (squamous epithelial abnormalities) - доброкачественные изменения плоского эпителия, койлоциты без изменений, позволяющих предположить CIN, Squamous cell changes - изменения плоского эпителия без четких признаков опухоли, CIN-I - дисплазия легкой степени, CIN-II - дисплазия умеренной степени, CIN-III - интраэпителиальная неоплазия тяжелой степени, понятие объединяет тяжелую дисплазию и внутриэпителиальный рак (CIS - carcinoma in situ), рак, подозрительный на инвазию, инвазивный плоскоклеточный рак. [Koss L.G. 1989] Для создания системы, которая поможет распознавать эти аномальные области, на первом этапе необходимо провести сегментацию изображения. Далее, для того чтобы определить тип ткани в каждом сегменте, необходимо классифицировать каждый сегмент используя базу данных содержащую образцы различных тканей полученных по результатам биопсий. В работе [Schoonmaker J. 2007] была проведена классификация

типов тканей по пикселям. Формулы для классификации подбирались эмпирическим путем. Но с увеличением базы пациентов подобранные таким образом формулы могут работать некорректно. Задачей этой работы является автоматизировать работу системы распознавания аномальных областей и приспособить ее для работы с новыми изображениями. Для сегментации изображений используются Карты Кохонена. Использование этой сети обусловлено желанием избежать влияния человеческого фактора при работе с изображениями.

Исходные данные

Технология, с помощью которой были получены электронные изображения тканей, называется «Квадропризменное апертурное разделение» (“Quad-prism Aperture-Splitting”, QPAS). Данная технология способна одновременно разделять свет отдельных длин волн и/или поляризаций, фильтровать каждый и затем перепроецировать на отдельные квадранты электронного оптического сенсора. QPAS оснащена четырьмя наборами фильтров, которые собирают:

1. Четыре диапазона отраженного поляризованного света с поляризатором, параллельным источнику света.
2. Четыре диапазона отраженного поляризованного света с поляризатором, перпендикулярным источнику света.
3. Восемь диапазонов флуоресценции, используя источник с длиной волны 365 мм (два набора фильтров).

Таким образом, мы собираем 16 каналов данных, которые и будут использоваться для классификации тканей.

Методы

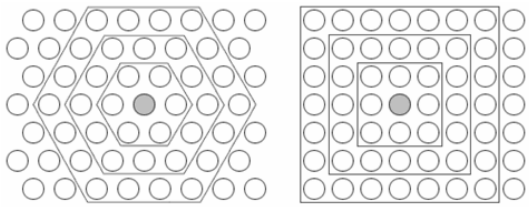
Для сегментации изображений была разработана программа, в которой использовались готовые библиотеки. Приложения с усеченной Гауссовой и колоколообразной функциями соседства были разработаны самостоятельно для проверки их эффективности.

Карты Кохонена Сети, называемые картами Кохонена, - это нейронные сети которые используют неконтролируемое обучение. Обучающее множество состоит лишь из значений входных переменных, в процессе обучения нет сравнивания выходов нейронов с эталонными значениями. Можно сказать, что такая сеть учится понимать структуру данных. Идея сети Кохонена принадлежит финскому ученому Тойво Кохонену (1982 год). Основной принцип работы сетей - введение в правило обучения нейрона информации относительно его расположения [Зайченко Ю.П. 2004].

Самоорганизующаяся сеть подразумевает использование упорядоченной структуры нейронов. Обычно используются одно и двумерные сетки. При этом каждый нейрон представляет собой n-мерный вектор-столбец $w = [w_1, w_2, ..., w_n]^T$, где n определяется размерностью исходного пространства (размерностью входных векторов), w_i - вес i-го нейрона. Применение одно и двумерных сеток связано с тем, что возникают проблемы при отображении пространственных структур большей размерности.

Обычно нейроны располагаются в узлах двумерной сетки с прямоугольными или шестиугольными ячейками. При этом, как было сказано выше, нейроны также взаимодействуют друг с другом. Величина этого взаимодействия определяется расстоянием между нейронами на карте. На рисунке 1 дан пример расстояния для шестиугольной и четырехугольной сеток.

При этом легко заметить, что для шестиугольной сетки расстояние между нейронами больше совпадает с евклидовым расстоянием, чем для четырехугольной сетки.



а) б)

Рис.1. Расстояние между нейронами на карте для шестиугольной (а) и четырехугольной (б) сеток.

При этом количество нейронов в сетке определяет степень детализации результата работы алгоритма, и в конечном счете от этого зависит точность обобщающей способности карты.

Начальная инициализация карты. При реализации алгоритма сети заранее задается конфигурация сетки (прямоугольная или шестиугольная), а также количество нейронов в сети. Некоторые источники рекомендуют использовать максимально возможное количество нейронов в карте. При этом начальный радиус обучения (neighborhood в англоязычной литературе) в значительной степени влияет на способность обобщения при помощи полученной карты. В случае, когда количество узлов карты превышает количество примеров в обучающей выборке, то успех использования алгоритма в большой степени зависит от подходящего выбора начального радиуса обучения. Однако, в случае, когда размер карты составляет десятки тысяч нейронов, время, требуемое на обучение карты, обычно бывает слишком велико для решения практических задач. Таким образом необходимо достигать допустимого компромисса при выборе количества узлов.

Перед началом обучения карты необходимо проинициализировать весовые коэффициенты нейронов. Удачно выбранный способ инициализации может существенно ускорить обучение и привести к получению более качественных результатов. Существуют три способа инициирования начальных весов.

Инициализация случайными значениями, когда всем весам даются малые случайные величины.

Инициализация примерами, когда в качестве начальных значений задаются значения случайно выбранных примеров из обучающей выборки

Линейная инициализация. В этом случае веса инициируются значениями векторов, линейно упорядоченных вдоль линейного подпространства, проходящего между двумя главных собственными векторами исходного набора данных. Собственные вектора могут быть найдены, например, при помощи процедуры Грама-Шмидта.

Обучение. Обучение состоит из последовательности коррекций векторов, представляющих собой нейроны. На каждом шаге обучения из исходного набора данных случайно выбирается один из векторов, а затем производится поиск наиболее похожего на него вектора коэффициентов нейронов. При этом выбирается нейрон-победитель, который наиболее похож на вектор входов. Под похожестью в данной задаче понимается расстояние между векторами, обычно вычисляемое в евклидовом пространстве.

Таким образом, если обозначить нейрон-победитель как c, то получим $|x - w_c| = \min_i \{\|x - w_i\|\}$, w_i – вес i-го нейрона.

После того, как найден нейрон-победитель производится корректировка весов нейросети. При этом вектор, описывающий нейрон-победитель и вектора, описывающие его соседей в сетке, перемещаются в направлении входного вектора.

При этом для модификации весовых коэффициентов используется формула:

$$w_i(t+1) = w_i(t) + h_{ci}(t) * [x(t) - w(t)]$$

где t обозначает номер эпохи (дискретное время). При этом вектор x(t) выбирается случайно из обучающей выборки на итерации t, w_i – вес i-го нейрона. Функция h(t) называется функцией соседства нейронов. Эта функция представляет собой невозрастающую функцию от времени и расстояния между

нейроном-победителем и соседними нейронами в сетке. Эта функция разбивается на две части: собственно функцию расстояния и функции скорости обучения от времени. где r определяет положение нейрона в сетке.

Обычно применяется такие функции: Гауссова функция $h(d,t)=e^{-\frac{d^2}{2\sigma^2(t)}}$, также могут использоваться усеченная Гауссова $h(d,t)=e^{-\frac{d^2}{2\sigma^2(t)}}1(\sigma(t)-d)$ и колоколообразная $h(d,t)=1(\sigma(t)-d)$, $\sigma(t)$ – радиус функции соседства в период времени t, a d – растояние между нейронами сетки. [Kohonen T. 1997].

При этом, как показали дальнейшие эксперименты, лучший результат получается при использовании Гауссовой функции расстояния. При этом является убывающей функцией от времени. Часто эту величину называют радиусом обучения, который выбирается достаточно большим на начальном этапе обучения и постепенно уменьшается так, что в конечном итоге обучается один нейрон-победитель. Наиболее часто используется функция, линейно убывающая от времени.

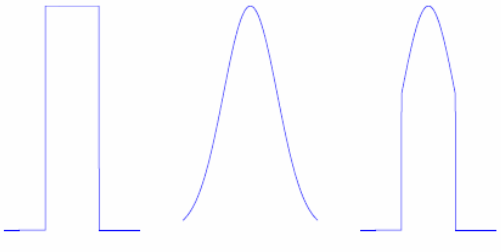


Рис. 2

Рассмотрим теперь функцию скорости обучения a(t). Эта функция также представляет собой функцию, убывающую от времени. Наиболее часто используются два варианта этой функции: линейная и обратно пропорциональная времени – вида $a(t)=\frac{A}{t+B}$ где A и B это константы, t - время. Применение этой функции приводит к тому, что все вектора из обучающей выборки вносят примерно равный вклад в результат обучения.

Обучение состоит из двух основных фаз: на первоначальном этапе выбирается достаточно большое значение скорости обучения и радиуса обучение, что позволяет расположить вектора нейронов в соответствии с распределением примеров в выборке, а затем производится точная подстройка весов, когда значения параметров скорости обучения много меньше начальных. В случае использования линейной инициализации первоначальный этап грубой подстройки может быть пропущен [Стариков А. 2000].

Применение алгоритма. Данную методику можно использовать для поиска и анализа закономерностей в исходных данных. При этом, после того, как нейроны размещены на карте, полученная карта может быть отображена разными цветами.

Результаты сегментации

Ниже приведены результаты сегментации используя различные функции. Различные сегменты выделены разными оттенками серого.

По изображениям видно, что применение всех функций дает приблизительно одинаковый результат, поэтому в дальнейшем будет использоваться только Гауссова функция принадлежности, так как это наиболее стандартная функция.

Далее представлены результаты сегментации на различное количество кластеров 10 (Рис. 4), 20 (Рис. 5) и 50 (Рис. 6) соответственно:

Изображения были разбиты на разное количество сегментов для дальнейшей проверки по базе типов тканей. Оптимальное количество сегментов можно будет определить при дальнейшей работе, хотя уже на данном этапе видно, что 20 сегментов отображают те области, которых не видно при 10 сегментах (отличия обведены для наглядности на рис. 4 и рис. 5). И как видно на рис. 5 и рис. 6, что 20 и 50

сегментов практически не отличаются, хотя проверка 50 сегментов по базе типов тканей может занять значительно больше времени.

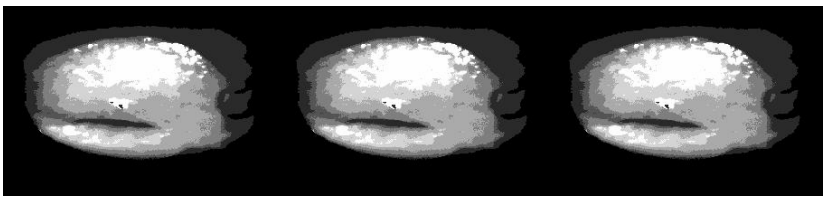


Рис. 3 Сегментация изображения, применяя Гауссову, усеченную Гауссову и колоколообразную функции

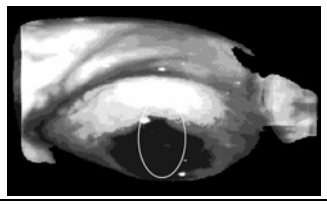


Рис. 4. 10 кластеров

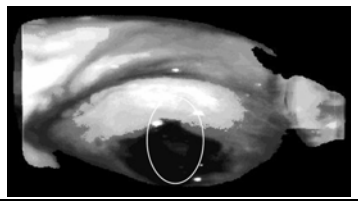


Рис. 5. 20 кластеров

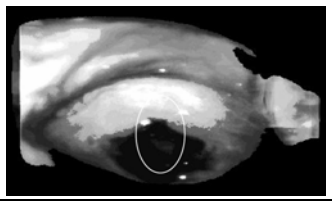


Рис. 6. 50 кластеров

Выводы

- Использование сети Кохонена обусловлено намерением избежать влияния человеческого фактора при сегментации изображения.
- По результатам экспериментов видно, что использование всех функций соседства дает одинаковый результат.
- Были проведены эксперименты по разбиению на разные количества сегментов и уже на данном этапе работы видно, что использование 10 сегментов может быть недостаточным, так как не полностью отображает картину состояния органа.

Благодарности

Работа опубликована при финансовой поддержке проекта **ITHEA XXI** Института информационных теорий и приложений FOI ITHEA Болгария www.ithea.org и Ассоциации создателей и пользователей интеллектуальных систем ADUIS Украина www.aduis.com.ua.

Библиография

[Kohonen T. 1997] "Self-Organizing Maps"(2-nd edition), Springer, 1997.

[Koss L.G. 1989] Koss L.G. The complex test for cervical cancer detection. In: The Journal of the American Medical Association. – 1989

[Schoonmaker J. 2007] Schoonmaker J. et al. Automatic Segmentation of Uterine Cervix for in vivo Localization and Identification of Cervical Interaepithelial Neoplasia // Apogen Technologies 2007

[Зайченко Ю.П. 2004] Зайченко Ю.П. Основи проектування інтелектуальних систем. Навчальний посібник – К.: Видавничий Дім «Слово», 2004

[Стариков А. 2000] BaseGroup 2000.

Информация об авторе

Катерина Малышевская *– аспирант Национальный Технический Университет Украины «Киевский Политехнический Институт», Киев, Украина, e-mail: volovik_katya@yahoo.com*

Mathematical Foundation of Artificial Intelligence

ЛИНЕЙНЫЕ СТРУКТУРЫ В ПРИКЛАДНЫХ ЗАДАЧАХ И КОНСТРУКТИВНЫЕ МЕТОДЫ ИХ ОПИСАНИЯ

Николай Кириченко, Владимир Донченко

Аннотация: *Рассмотрены свойства основных линейных структур в евклидовых пространствах, развиты конструктивные способы их описания и построения на основе систематического развития и применения аппарата псевдообращения по Муру - Пенроузу. Важность приведённых результатов проиллюстрирована на широком спектре прикладных задач: от регрессионного анализа и задач прогноза до теории оптимального управления.*

Ключевые слова*: Псевдообращение по Муру - Пенроузу, сингулярное представление матрицы, метод наименьших квадратов, линейная регрессия, системы оптимального управления, прогноз, кластеризация, искусственные нейронные сети.*

ACM Classification Keywords*: G.3 Probability and statistics, G.1.6. Numerical analysis: Optimization; G.2.m. Discrete mathematics: miscellaneous.*

Conference*: The paper is selected from XVth International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.*

Вступление

Как отмечалось в работе [Донченко,2009], апелляция к «структуре объекта» ассоциируется с представлением об объекте, как чём-то едином, составленном из взаимодействующих между собою частей. Как правило, «структура» и «связи» между частями объекта, рассматриваемого как единое целое, употребляются как реализующие одно и тоже представление об объекте с тем дополнением, что «структура» – это совокупность частей объекта плюс «связи» между ними. Среди важнейших математических структур, отмеченных в упомянутой выше работе, особое место занимают «линейные структуры». К ним можно отнести линейные пространства и подпространства, гиперплоскости а также – линейные операторы и функционалы. Среди линейных структур по богатству возможностей использования связей занимает евклидово пространство: конечномерное линейное плюс скалярное произведение. Богатство свойств линейных структур: и в варианте линейных пространств и подпространств, и в варианте операторов соответствующего вида, – в математическом моделировании объектов трудно переоценить. Это касается как абстрактных математических, так и исследований прикладного характера. В полной мере сказанное выше относится, в частности, к алгебре, регрессионному анализу, теории случайных процессов, теории дифференциальных и интегральных уравнений, систем оптимального управления, прикладным задачам классификации, прогноза и т.д. Важную роль в прикладных исследованиях играют конструктивные методы описания соответствующих

объектов. В том, что касается линейных операторов и линейных функционалов, вопрос конструктивности решается построением матриц соответствующих объектов, а для операций – использованием операций матричной алгебры. В том, что касается подпространств, порождённых теми или иными совокупностями векторов, дело обстоит сложнее. Их конструктивное описание можно получить, связав указанные объекты с пространством значений подходящей матрицы, которое в свою очередь описывают подходящим ортогональным проектором. Именно этот подход развивается ниже. Отметим, что ортогональные проекторы играют важную роль в исчерпывающем исследовании систем линейных алгебраических уравнений (СЛАУ). Принципиально важны они также в постановке и решении важных оптимизационных задач с квадратическим функционалом качества в евклидовых пространствах, в том числе – в построении наилучших квадратических приближений правой части СЛАУ значениями левой, когда СЛАУ несовместна. Такие наилучшие приближения называют также псевдорешениями. Конструктивное описание ортогональных проекторов в связи с естественными подпространствами линейного оператора прямо определяется псевдообращением по Муру - Пенроузу[Moore,1920], [Penrose, 1955](см. также, например, [Алберт, 1977]). Отметим, также, что важную роль в конструктивном решении прикладных задач с использованием линейных структур играет сингулярное представление (его называют также сингулярным разложением или SVD - представлением) матрицы в специфической записи в виде взвешенной суммы тензорных произведений специального набора пар векторов. Ниже рассматриваются основные свойства линейных структур, основные особенности и возможности (“tips and tricks”) их конструктивного описания, а также – использования в для конструктивного решения важнейших прикладных задач прогноза, кластеризации и классификации, в других областях.

Евклидовы пространства и подпространства: базовые “tips and tricks”

В дальнейшем, говоря об евклидовом пространстве будем иметь в виду множество конечных числовых последовательностей одной и той же длины n, записанных в столбик с покоординатными операциями сложения и умножения на скаляр и суммой покоординатных произведений в качестве скалярного произведения. Стандартным образом, Именно такой вариант евклидового пространства будем стандартным образом обозначать через R^n, а его элементы – через $a = \begin{pmatrix} a_1 \\ \cdots \\ a_n \end{pmatrix}$. Стандартные ортонормированные базисы, составленные из векторов с единственной единичной компонентой (остальные – нули) на месте с соответствующим номером будут обозначаться для R^m и R^n соответственно через $e(j) \in R^m, j = \overline{1,m}, e_{(i)} \in R^n, i = \overline{i,n}$. Оператор A из R^n, в R^m: $A : R^n \to R^m$, в ортонормированных базисах $e(j) \in R^m, j = \overline{1,m}, e_{(i)} \in R^n, i = \overline{i,n}$ будем отождествлять с $m \times n$ - матрицей $A = (a_{ij})$ этого оператора. Для матрицы $A = (a_{ij})$ будем использовать также блочное представление по столбцам (столбцовое) и строкам(строчное):

$$A = \begin{pmatrix} a_{(1)}^T \\ \cdots \\ a_{(m)}^T \end{pmatrix} = (a(1) \vdots \ldots \vdots a(n)), a_{(i)}^T \in R^n, i = \overline{1,m}, a(j) \in R^m, j = \overline{1,n}.$$

Линейное пространство всех $m \times n$ матриц будем обозначать $R^{m \times n}$.

Линейное подпространство, порождённое системой векторов $c_k \in R^p, k = \overline{1,K}$ будет обозначаться через $L(c_k, k = \overline{1,K}) \equiv L(c_1 \ldots, c_K)$, а линейное подпространство значений линейного оператора $A : R^n \to R^m$ – через L_A.

Первым из набора “tips and tricks”является утверждение о том, что

1. “tips and tricks”: $L_A = L(a(1), \ldots, a(n))$.

Таким образом, линейное подпространство, порождённое набором векторов, совпадает с подпространством значений матрицы, составленной из векторов набора, как из столбцов.

2. “tips and tricks”: для элементов столбцового и строчного представления матрицы $A \in R^{m \times n}$ справедливы соотношения

$$a(j) = Ae(j), j = \overline{1,n},$$

$$a_{(i)}^T = e_{(i)}^T A, i = \overline{1,m}.$$

3. “tips and tricks”: для произведения произвольных матриц B, C со столбовым и строчным представлением $B = (b(1) \vdots \ldots \vdots b(r)), b(j) \in R^m, j = \overline{1,r}, C = \begin{pmatrix} c_{(1)}^T \\ \cdots \\ c_{(r)}^T \end{pmatrix}, c_{(r)}, c(i) \in R^n, i = \overline{1,r}$ соответственно и диагональной матрицы $\Lambda = diag(\lambda_1, \ldots, \lambda_r)$ справедливо соотношение

$$B\Lambda C = \sum_{i=1}^{r} \lambda_i b(i) c_{(i)}^T.$$

Важной составляющей аппарата конструктивного описания и использования линейных структур является понятие ортогонального проектора, которое полностью отвечает стандартному геометрическому представлению об ортогональном проектировании. Общей, основой эффективного использования ортогональных проекторов является наличие двух эквивалентных определений таких проекторов и возможности их конструктивного построения в связи с линейными подпространствами через псевдообращение.

4. “tips and tricks” («геометрическое определение ортогонального проектора»): для разложения $R^p = L + L^\perp$ в прямую сумму ортогональных подпространств ортогональным проектором P_L на линейное подпространство $L \subseteq R^p$ называется оператор, определяемый соотношением $P_L x = P_L(x_L + x_{L^\perp}) = x_L$,

где

$$x = x_L + x_{L^\perp}, x_L \in L, x_{L^\perp} \in L^\perp \quad (1)$$

– однозначное представление произвольного вектора $x \in R^p$ по двум составляющим ортогональной суммы. Очевидным образом оператор ортогонального проектирования является линейным оператором.

5. “tips and tricks”:разложение (1) произвольного вектора $x \in R^p$ в силу симметричности относительно ортогональных слагаемых определяет одновременно два ортогональных проектора: $P_L, P_{L^\perp}$ с очевидным соотношением

$P_L + P_{L^\perp} = E_p$,

где E_p – единичная матрица соответствующей размерности.

6. “tips and tricks”: для ортогонального проектора P_L на подпространство L оператор $Z_L \equiv E_p - P_L$ является ортогональным проектором на ортогональное дополнение $L^\perp$ к $L : Z_L \equiv E_p - P_L = P_{L^\perp}$.

7. “tips and tricks“(абстрактное определение ортогонального проектора): для того, чтобы линейный оператор $P : R^p \to R^p$, был оператором ортогонального проектирования необходимо и достаточно, чтобы он был идемпотентным симметричным оператором. Линейное пространство L_P, на которое совершается ортогональное проектирование в соответствии с «геометрическим определением» описывается одним из двух соотношений:

$L_P = \{x : x = Pu, u \in R^p\} = \{x : x = Px, x \in R^p\}$.

8. “tips and tricks“(сингулярное или SVD- представление произвольной $A \in R^{m\times n}$): для произвольной $A \in R^{m\times n}$ ранга $r \le \min(m,n)$ справедливо следующее представление матрицы в виде взвешенной суммы тензорных произведений

$$A = \sum_{i=1}^{r} \lambda_i u_i v_i^T, \qquad (2)$$

где:

- $\lambda_1^2 \ge \ldots \ge \lambda_r^2 > 0$ общий набор ненулевых собственных чисел матриц AA^T, A^TA.
- $u_i \in R^m, i = \overline{j,r}$ - ортонормированный набор собственных векторов матрицы AA^T, отвечающих ненулевым собственным числам: $AA^T u_i = \lambda_i^2, \lambda_i^2 > 0, i = \overline{1,r}, u_i^T u_j = \delta_{ij}, i \ne j$;
- $v_i \in R^n, i = \overline{j,r}$ - ортонормированный набор собственных векторов матрицы A^TA, отвечающих ненулевым собственным числам: $A^TAv_i = \lambda_i^2 v_i, \lambda_i^2 > 0,, i = \overline{1,r}, v_i^T v_j = \delta_{ij}, i \ne j$.

9. “tips and tricks“(определение псевдообращения через SVD - представление матрицы): для произвольной $A \in R^{m\times n}$ с SVD - представлением (2) псевдообратная к ней, A^+, определяется соотношением

$$A^+ = \sum_{i=1}^{r} \lambda_i^{-1} v_i u_i^T : R^m \to R^n. \qquad (3)$$

Псевдообращение в дальнейшем будет обозначаться аббревиатурой ПдО.

Заметим, что SVD - определение ПдО (соотношение (3)) позволяет легко установить, что ПдО коммутирует с транспонированием, а также ряд других полезных соотношений, в частности, что $A^T(A^T)^+ = A^+A$.

10. “tips and tricks“: ортогональные проекторы на L_A, L_{A^T} определяется соотношением

$$AA^+ = \sum_{i=1}^{r} u_i u_i^T,$$

$$A^T(A^T)^+ = A^+A = \sum_{i=1}^{r} v_i v_i^T$$

11. “tips and tricks“:операторы $Z(A), Z(A^T)$, определяемые соотношениями

$$Z(A) = E_n - A^+A,$$

$$Z(A^T) = E_m - A^{T^+}A^T = E_m - AA^+,$$

являются операторами ортогонального проектирования на $L_A^{\perp}$, $L_{A^T}^{\perp}$

Важность последних соотношений определяются тем, что $L_{A^T}^{\perp}$ является множеством нулей оператора A.

12. “tips and tricks“:подпространство $L_{A^T}^{\perp}$ является ядром $KerA$ (множеством нулей) оператора A:

$$L_{A^T}^{\perp} = KerA = Z(A)R^n.$$

13. “tips and tricks“: для совместности СЛАУ $Ax = y$ необходимо и достаточно, чтобы $y^T Z(A^T)y = 0$. В этом случае A^+y является наименьшим по норме решением. Оно ортогонально к $KerA$, а множество всех решений Ω_y описывается соотношением

$$\Omega_y = A^+y + Z(A)R^n = \{x : x = A^+y + Z(A)v, v \in R^n\}. \quad (4)$$

14. “tips and tricks“: если СЛАУ $Ax = y$ несовместна, т.е $y^T Z(A^T)y > 0$ множество, определяемое соотношением (4) описывает совокупность всей наилучших квадратических приближений правой части значениями левой:

$$\Omega_y = A^+y + Z(A)R^n = \underset{x \in R^n}{Arg\min} \| Ax - y \|^2. \quad (5)$$

Значение невязки для любого наилучшего квадратического приближения составляет $y^T Z(A^T)y$.

15. “tips and tricks“: для того, что матричное уравнение $AX = Y, A \in R^{m \times n} X \in R^{n \times N}, Y \in R^{m \times N}$ имело корни необходимо и достаточно, чтобы $trY^T Z(A^T)Y = 0$. В этом случае множество Ω_Y определяется соотношением

$$\Omega_Y = \{X : X = A^+Y + Z(A)V, V \in R^{n \times N}\}. \quad (6)$$

16. “tips and tricks“: для линейной зависимости вектора $d \in R^m$ от столбцов матрицы $A \in R^{m \times n}$ необходимо и достаточно, чтобы выполнялось соотношение

$$d^T Z(A^T)d = 0. \quad (7)$$

Ссылка на векторы--элементы блочного представления матрицы A нисколько не ограничивают возможности применения утверждения этого пункта для определения линейной зависимости того или иного вектора от фиксированного набора векторов. Для использования результата в общем случае необходимо и достаточно составить матрицу, в которой векторы набора являются столбцами, и дополнительно использовать п.1“tips and tricks“.

Отметим также, что условием линейной независимости строки $a^T, a \in R^n$ от строк матрицы $A \in R^{m \times n}$ является условие

$$a^T Z(A)a = 0. \qquad (8)$$

17. “tips and tricks“(прямые формулы Гревиля): ПдО произвольной матрицы $A \in R^{m\times n}$, дополненной строкой $a^T \in R^n$, определяется элементами $P \in R^{n\times m}, q \in R^n$ блочного представления ПдО $\begin{pmatrix} A \\ a^T \end{pmatrix}^+ = (P \vdots q)$ расширенной матрицы:

$$q = \begin{cases} \dfrac{Z(A)a}{a^T Z(A)a}, a^T Z(A)a > 0(íåç.) \\ \dfrac{R(A)a}{1 + a^T R(A)a}, a^T Z(A)a = 0(çàâ.) \end{cases}, \qquad (9)$$

$$P = (E_{ô} - qa^T)A^+,$$

где $R(A) = A^+ A^{+T}$.

Первая строка в (9) отвечает случаю линейной независимости строки - расширения от строк матрицы A, второй – линейной зависимости.

С учётом коммутирования транспонирования с ПдО, прямые формулы Гревиля очевидным образом переписываются для варианта расширения матрицы столбцом.

18. “tips and tricks“(обратные формулы Гревиля): для блочного представления ПдО расширенной матрицы $\begin{pmatrix} A \\ a^T \end{pmatrix}^+ = (P \vdots q)$ ПдО матрицы A определяется соотношениями

$$A^+ = \begin{cases} (E - \dfrac{qq^T}{||q||^2})P, a^T q = 1(íåçàâ.) \\ (E + \dfrac{qa^T}{1 - a^T q})P, a^T q \neq 1(çàâ.)) \end{cases}. \qquad (10)$$

Условие в первой строке в соотношениях (10) отвечает линейной независимости строки, которая удаляется, а второй – зависимости от остальных строк матрицы A. Справедливость этих условий непосредственно вытекает из п.17 “tips and tricks“, и используется, когда дополнительно известна ПдО для расширенной матрицы.

19. “tips and tricks“: квадрат расстояния $\rho^2(a, L_A)$ вектора $a \in R^m$ от подпространства L_A определяется соотношением

$$\rho^2(a, L_A) = \min_{y\in R^m} ||a - y||^2 = a^T Z(A^T)a.$$

20. “tips and tricks“: квадрат расстояния $\rho^2(a, \Gamma(b, L_A))$ вектора $a \in R^m$ от гиперплоскости $\Gamma(b, L_A) = b + L_A$ определяется соотношением

$$\rho^2(a, \Gamma(b, L_A)) = \min_{y\in\Gamma(b,L_A)} ||a - y||^2 = (a - b)^T Z(A^T)(a - b).$$

Отметим, что привязка подпространств в пп.19, 20 к множеству значений оператора A не ограничивает сферу применимости результатов. С помощью п.1 ‘tips and tricks’ они очевидным образом распространяются на ситуацию, когда подпространство порождается заданной конечной совокупностью векторов.

Формулы аналитического возмущения ПдО описывают ПдО возмущённой матрицы, когда возмущение имеет вид аддитивной добавки $ab^T, a \in R^m, b \in R^n$. В работе [Кириченко,1997] исчерпывающим образом исследованы варианты представления $(A+ab^T)^+$, которые, как оказывается, определяются тем, являются ли линейно зависимыми компоненты возмущения: a, b^T от, соответственно, столбцов и строк матрицы A, а также сохранением или падением ранга возмущенной матрицей, когда одновременно a, b^T линейно зависимы в связи с матрицей A. И условия линейной независимости и условие падения ранга носят аналитический характер. В п.16 “tips and tricks” представлены условия линейной зависимости. Условие сохранения ранга представлено следующим пунктом.

21. “tips and tricks“:ранги матриц A и $A+ab^T$ одинаковы: $rank(A+ab^T)=rankA$, тогда и только тогда, когда $b^TA^+a \neq -1$.Ранг возмущённой матрицы падает, когда $b^TA^+a=-1$.

Принимая во внимание громоздкость соответствующих формулировок, ниже приведен один из вариантов утверждения о виде ПдО для возмущённой матрицы. С полным вариантом утверждения можно ознакомиться в уже упомянутой работе [Кириченко,1997].

22. “tips and tricks“(аналитические формулы возмущения ПдО матриц, фрагмент): если компоненты возмущения: a, b^T линейно не зависимы от, соответственно, столбцов и строк матрицы A, т.е. $a^TZ(A^T)a>0, b^TZ(A)b>0$, то

$$(A+ab^T)^+ = A^+ - \frac{A^+aa^TZ(A^T)}{a^TZ(A^T)a} - \frac{Z(A)bb^TA^+}{b^TZ(A)b} + Z(A)ba^TZ(A^T)\frac{1+b^TA^+a}{a^TZ(A^T)ab^TZ(A)b}.$$

Применения “tips and tricks“- свойств ПдО: линейная регрессия, скалярные наблюдения

Применение ПдО в линейной регрессии определяется тем, что МНК – оценка $\hat{\beta}$ (оценка метода наименьших квадратов) неизвестного параметра $\beta \in R^p$ линейной регрессии $y=\sum_{j=0}^{p-1}\beta_j f_j(x)+\varepsilon$ на основе наблюдений $(x_i, y_i), x_i \in R^m, y_i \in R^1, i=\overline{1,n}$ определяется решением оптимизационной задачи

$$\hat{\beta} \in \underset{\beta \in R^p}{Arg\min} \| X\beta - Y \|^2, \quad (11)$$

в которой X - матрица плана, а Y - вектор – столбец с компонентами $y_i \in R^1, i=\overline{1,n}$ – вектор наблюдений. В соответствии с соотношением (4) п.13 “tips and tricks“ общее решении задачи (11) определяется соотношением

$$\hat{\beta} = X^+Y + Z(X)v, v \in R^p * \quad (12)$$

со свободным параметром $v \in R^p$.

Решение задачи МНК - оценивания в виде (12) полностью согласуется с классическим решением уравнения Гаусса – Маркова в виде

$$\hat{\beta} = (X^TX)^{-1}X^TY,$$

Поскольку в случае полного столбцового ранга матрицы плана $X^{+} = (X^{T}X)^{-1}X^{T}Y$ а $Z(X) = 0$. В этом случае классическом случае множество МНК – оценок в (12) является одно элементным.

Применения “tips and tricks“- свойств ПдО: задача терминального управления

Под задачей терминального управления для линейноЙ динамической системы с дискретным временем

$$x(k+1) = A(k)x(k) + b(k)u(k),$$

$$x(0) = x_{(0)},$$

где

$x(k) \in \mathrm{R}^{\mathrm{n}}$, $\mathrm{u(k)} \in \mathrm{R}^{1}$, $A(k) \in \mathrm{R}^{\mathrm{n}\times\mathrm{n}}$, $b(k) \in \mathrm{R}^{\mathrm{n}}$, $k = \overline{0,N}$, имеют в виду задачу выбора такого управления $u(k)$, $k = \overline{0,N}$, которое позволяет вывести фазовую траекторию в момент $N+1$ на уровень $\mathrm{x}_{(1)}$ или, если это невозможно, выбором того же управления минимизировать отклонение $\left\| x(N+1) - x_{(1)} \right\|^{2}$.

Принципиальным результатом для исследования задачи терминального управления является теорема редукции, позволяющая свести задачу терминального управления к СЛАУ.

Теорема (теорема редукции). Задача терминального управления является эквивалентной СЛАУ

$$W(N+1)u = x_{(1)} - A(N-1)A(N-2)\ldots A(0)x_{(0)},$$

в которой вектор $u \in R^{N+1}$ – объединенный вектор управления, а матрица $W(N+1)$ является блочной:

$$W(N+1) = \left(W(N+1,0) \vdots W(N+1),1) \vdots \ldots \vdots W(N+1,N)\right),$$

с блоками $W(N+1),k), k = \overline{0,N}$, определяемыми соотношениями

$$W(N+1,k) = A(N)A(N-1)\ldots A(k+1)b(k), k = \overline{0,N-1},$$

$$W(N+1,N) = b(N).$$

Теорема редукции позволяет исчерпывающим образом исследовать задачу терминального управления с помощью пп.13,14 “tips and tricks”.

Следует добавить, что аналогичным образом с помощью ПдО удаётся исчерпывающим образом исследовать задачу терминального наблюдения в том числе в случае ошибок и шумов(см., например, [Кириченко, Донченко, 2005])

Применения “tips and tricks“- свойств ПдО: кластеризация

ПдО расширяет возможности кластеризации, позволяя эффективно погружать классифицируемые объекты в подходящие подпространства или гиперплоскости. П.1 “tips and tricks” дает возможность связывать подпространство, порождённое набором векторов, с подходящей матрицей. Если объект связывается с гиперплоскостью, то её смещение – это, как правило, среднее по векторам порождающей совокупности, а подпространство – это подпространство значений матрицы, построенной из центрированных средним векторов порождающей совокупности, как из столбцов. Результаты пп..19, 20 “tips and tricks” обеспечивают возможность конструктивного вычисления расстояний от объектов(подпространств или гиперплоскостей), ассоциируемых с порождающей совокупностью. Применение стандартных рекуррентных последовательно уточняемых разбиений с расстояними

соответствия из п.19 или п. 20 “tips and tricks” придаёт процедуре кластеризации необходимой завершённости. С подробностями можно ознакомиться, например в работах, например, [Кириченко, Донченко, 2007], [Кириченко, Донченко, 2008]).

Применения “tips and tricks“- свойств ПдО: RFT- функциональные сети

Искусственные нейронные сети являются стандартным технологическим инструментом исследования в задачах прогноза, классификации и кластеризации. В сущности, они представляют собой графическое изображения: графы суперпозиций, стандартизованных функциональных элементов. В этом смысле их с полным правом можно назвать функциональными сетями. Стандартизация функциональных элементов(нейронов) проявляется в том, что они реализуют скалярную функцию векторного аргумента как суперпозицию линейного функционала и скалярной функции скалярного аргумента:

$$y = F(w^T x), w, x \in R^m, F : R^1 \to R^1 .$$

Упомянутая стандартизация – унифицированность может проявляться и в выборе внешней функции: F, которая называется функцией инициализации нейрона. Она может быть фиксированной или принимать значения из конечного набора функций.

Заметим, что, как правило, фиксированной является и структура сети: количество стандартных функциональных элементов нейронов и способ их соединения: топология сети.

Возможности функциональных сетей можно значительно расширить, если 1) придать большую функциональную универсальность и обеспечить адаптивность в построении каждого из стандартизованных элементов; 2) обеспечить более гибкие возможности в соединении элементов: большую свободу в формировании топологии сети; 3) гарантировать адаптивное построение структуры всей сети в целом. Последнее может быть конструктивно реализовано в ходе выполнения последовательных шагов наращивания сети, имеющих рекуррентный характер.

Реализация такой программы для задачи прогноза - восстановления функции, представленной своими значениями $(x_i, y_i), x_i \in R^n, y_i \in R^m, i = \overline{1,n}$ состоит в построении стандартного функционального элемента(RFT - преобразователя) в виде

$$y = A_+ \Psi(Cx), \qquad (13)$$

в котором матрица C – матрица предварительного преобразования вектора признаков x, Ψ – нелинейное покоординатное преобразование измененного вектора признаков, A_+ - матрица МНК оценивания на выборке $(x_i, y_i), x_i \in R^n, y_i \in R^m, i = \overline{1,n}$.

Эффективность преобразователя (13) в прогнозе зависимости может контролироваться по невязке и по тестовой выборке.

В общем адаптивная процедура построения функциональной сети развивает идею МГУА А.Г.Ивахненко

Заключение

В работе проанализированы важные в прикладном отношении аспекты использования линейных структур в рамках евклидовых пространств в решении прикладных задач математического моделирования. В числе других рассмотрены конструктивные способы порождения подпространств и гиперплоскостей, а также ортогональных проекторов, связанных с указанными объектами. Упомянутая конструктивность обеспечивается применением псевдообращения по Муру – Пенроузу(ПдО), а также новыми результатами в этой области. Важность и эффективность использования приведённых результатов проиллюстрирована

на широком спектре задач, включающем линейную регрессию, в том числе в векторную, теорию оптимального управления, кластеризацию, прогноз и функциональные сети, являющиеся обобщение искусственных нейронных сетей.

Благодарности

Работа опубликована при финансовой поддержке проекта **ITHEA XXI** Института информационных теорий и приложений FOI ITHEA Болгария www.ithea.org и Ассоциации создателей и пользователей интеллектуальных систем ADUIS Украина www.aduis.com.ua.

Литература

[Moore,1920] Moore E.H. On the reciprocal of the general algebraic matrix // Bulletin of the American Mathematical Society. – 26, 1920. – P.394 -395.

[Penrose, 1955] Penrose R. A generalized inverse for matrices // Proceedings of the Cambridge Philosophical Society 51, 1955. – P.406-413.

[Алберт, 1977] Алберт А. Регрессия, псевдоинверсия, рекуррентное оценивание. – М.: Наука. – 1977.– 305 с.

[Донченко. 2009] Донченко В.С. Неопределённость и математические структуры в прикладных исследованиях/ Human aspects of Artificial Intelligence Inernational Book Series Information science & Computing.– Number 12.– Supplement to International Journal “Information technologies and Knowledge”. –Volume 3.–2009.– P. 9-18.

[Кириченко, 1997] Кириченко Н.Ф. Аналитическое представление псевдообратных матриц //Киб. и СА.- №2. –1997.– С.98-122.

[Кириченко, Донченко, 2005] Кириченко М.Ф., Донченко В.С. Задача термінального спостереження динамічної системи: множинність розв’язків та оптимізація//Журнал обчислювальної та прикладної математики. – 2005. –№5– С.63-78.

[Кириченко, Донченко, 2007] Кириченко Н.Ф., Донченко. В.С. Псевдообращение в задачах кластеризации// Киб. и СА.- №4, 2007– С.98-122.

[Кириченко, Донченко, 2008] В.С Кириченко Н.Ф. Донченко В.С. Гиперплоскости в «множествах и расстояниях соответствия»: кластеризация / Artificial Intelligence and Decision Making.– International book series “Information Science&Computing”, Number 7.–Supplement to the International Journal “Information Technologies and Knowledge” V.2/2008 FOI ITHEA, Sofia 2008.– P. 25-36.

Информация об авторах

Николай Ф. Кириченко *– Профессор Институт кибернетики НАН Украины;*

Владимир С. Донченко – *Профессор; Киевский национальный университет имени Тараса Шевченко, факультет кибернетики, Украина, e-mail: voldon@unicyb.kiev.ua*

СЛОЖНОСТНО–СТРУКТУРНЫЙ ПОДХОД К ИССЛЕДОВАНИЮ ОБЛАСТИ ПРИМЕНИМОСТИ АЛГОРИТМА PROFIT

Татьяна Ступина, Иван Кулаков

***Аннотация**: В работе представлена формальная постановка сложностно - структурного подхода к исследованию области применимости алгоритма PROFIT (PROfile Forward and Inverse Tomographic modeling). Основная идея состоит в определении кластера моделей среды (разреза геологического строения земной коры), представленных изображениями, которые с допустимой погрешностью восстанавливаются алгоритмом. Метрика в пространстве изображений задается специальным образом. Она учитывает изменение поля скоростей и аномальные включения, количественный и качественный состав которых отражает этапы усложнения моделей сред.*

***Ключевые слова**: сейсмическое профилирование, томография, прямое моделирование, синтетическая модель.*

***ACM Classification**: G1.10. Mathematics of Computing - Applications*

***Conference**: The paper is selected from XVth International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.*

Введение

Подход к обработке экспериментальных данных зависит от специфики решаемой задачи в конкретной прикладной области и конечной цели, которая ставится в задаче. В различных областях знаний, целью которых является обнаружение причинно-следственных связей какого-либо процесса или явления, необходимо решать обратную задачу, которая, как правило, является неустойчивой как относительно входных данных, так и самого оператора инверсии. Для оценки качества полученного решения также необходимо уметь решать и прямую задачу. На каждом этапе возникают свои трудности, с которыми необходимо справляться, используя, предложенные к обработке данные и основываясь на выбранной модели. Оценить же качество выбранной модели независимо от качества обрабатываемых данных возможно лишь теоретически при известной «истинной» модели [Вапник, 1984].

В данной работе рассматривается алгоритм PROFIT прямого и обратного моделирования профильных данных в сейсмической томографии [Кулаков, 2007]. Преимущества данного алгоритма при решении задач на реальных и синтетических данных подробно отражены в работах И.Ю. Кулакова. В работе представлены основные этапы этого алгоритма, подчеркивающие многопараметрическую сложность и специфику решаемой задачи. Важной компонентой предложенного алгоритма является возможность построения реконструированной модели геологической среды, т.е. глубинного разреза, что является достаточно интересным и важным для задач геологической интерпретации в целях изучения строения земной коры и поиска полезных ископаемых. Для того чтобы оценить множество правдоподобных, относительно экспертных знаний о моделях сред, реконструкций предлагается сложностно - структурный подход.

В работе предложена формальная постановка сложностно - структурного подхода к исследованию области применимости алгоритма PROFIT. Основная идея состоит в определении кластера моделей среды, представленных изображениями глубинных разрезов, которые с допустимой погрешностью восстанавливаются алгоритмом. Элементы кластера формируются в метрическом пространстве

изображений относительно некоторого экспертного шаблона (модель, допустимая по смыслу с точки зрения строения земной коры в данном регионе). Другими словами, задается критерий, определяющий допустимую реконструированную модель среды, построенную алгоритмом PROFIT для данного географического региона и фиксированной системы наблюдений.

Основные понятия

В общем виде задача сейсмической томографии [Нолет, 1990] заключается в определении поля скоростей $v(r)$ по множеству измерений времени на поверхности

$$T_i = \int_{S_i} \frac{ds}{v(r)},\ i = 1,...,N.$$

Задача является достаточно сложной, поскольку неизвестная функция $v(r)$ в неявном виде присутствует в определении лучевой траектории S_i. Поэтому задача решается в линеаризованной постановке применением принципа Ферма. Обозначим время, предсказываемое начальной (референтной, опорной) моделью

$$T_i^\circ = \int_{S_i} \frac{ds}{v_\circ(r)}. \quad (1)$$

Где S_i° - траектория луча в начальной модели, тогда время задержки определяется следующим образом

$$\delta T_i = T_i - T_i^\circ = \int_{S_i} \frac{ds}{v} - \int_{S_i^\circ} \frac{ds}{v_\circ} \approx \int_{S_i^\circ} \left(\frac{1}{v} - \frac{1}{v_\circ} \right) ds,$$

$$\delta T_i = -\int_{S_i^\circ} \frac{\delta v(r)}{v_\circ^2(r)} ds. \quad (2)$$

Где $\delta v(r) = v(r) - v_\circ(r)$. Здесь была произведена замена неизвестной лучевой траектории на траекторию для референтной модели, поэтому в прямую задачу внесена поправка второго порядка. Обратная же задача может быть плохо обусловлена и нет уверенности, что ошибка второго порядка не приведёт к большой ошибке в решении. На практике очень часто система наблюдений не отвечает идеальным условиям, когда невозможно обеспечить достаточную лучевую плотность, поэтому обратная задача решается численными методами. С этой целью определённым образом задается параметризация по базису $h_j(r)$

$$\delta v(r) = \sum_{j=1}^{M} \gamma_j h_j(r).$$

Тогда уравнение временных задержек перепишется в следующем виде

$$\delta T_i = \sum_{j=1}^{M} \left(-\int_{S_i^\circ} \frac{1}{v_\circ^2(r)} \gamma_j h_j(r) ds \right) = \sum_j A_{ij} \gamma_j.$$

Где A_{ij} - матрица первых производных по параметрам, $A_{ij} = \frac{\partial T_i}{\partial \gamma_j} = -\int_{S_i^\circ} \frac{h_j(r)}{v_\circ^2(r)} ds$.

Таким образом, задача сводится к решению СЛАУ

$$A\gamma = \delta T \,. \tag{3}$$

В общем случае такая система является плохо обусловленной, поэтому при её решении применяются методы регуляризации и итеративные подходы обращения сильно разреженных матриц большого размера. В алгоритме PROFIT реализованы одни из таких методов, позволяющие за достаточно приемлемое время получать хорошее решения.

Принцип работы алгоритма PROFIT

Программа PROFIT состоит из двух основных компонент. Первая содержит алгоритмы томографической инверсии и может быть использована и в качестве отдельного и независимого программного кода для обработки сейсмических рефрагированных данных. Вторая компонента подключает к исполнению совместно прямое и обратное моделирование и предназначена для построения реконструированной модели среды. Код программы реализован на языке Фортран-90 и разработан под операционной системе Widows. Исходные и исполняемые коды программы находятся в свободном доступе.

Основная структура алгоритма томографической инверсии в программе основывается на общепризнанных вычислительных методах компьютерной томографии [Нолет, 1990]. Вычисления производятся итеративно в соответствии со следующими шагами:

1. Трассировка лучей в 2 D скоростной модели (стартовая модель для первой итерации или пересчитанная скоростная модель после предыдущих итераций);
2. Построение сеточной параметризации (только для первой итерации);
3. Вычисление матрицы и обращение;
4. Обновление скоростной модели. Переход на первый шаг.

Более подробно опишем наиболее важные моменты для каждого из этих шагов.

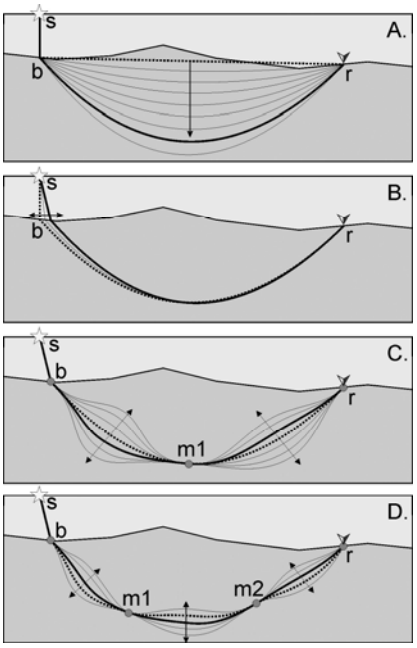


Рис. 1. Иллюстрация лучевого трассирования

Лучевое трассирование. Лучевое трассирование, применяемое в представляемом алгоритме, основывается на принципе Ферма и состоит в нахождении пути, обеспечивающим минимальное время между источником и приёмником. Эта идея является основой для расчета изгиба лучевого трассирования, которая широко применялась в последние десятилетия и внедрялась как основа в различных алгоритмах томографии локальных землетрясений, так и сейсмического моделирования.

В представленном алгоритме реализована альтернативная версия алгоритма расчёта изгиба, которая схематично показана на рис. 1. Поиск пути с минимальным временем состоит из последовательно выполняемых нескольких шагов вычисления изгиба луча *B*, первоначальная траектория которого описывается тригонометрической функцией

На первом шаге (Фигура A) точка на морском дне (точка *b*) располагается сразу под источником *S*. Мы начинаем с прямой линии между точкой *b* и приёмником *r*, потом последовательно изменяем её до получения минимального времени прохода луча. В первом приближении отклонение *A*(*s*) относительно начального прямого пути вычисляется в соответствии со следующей формулой:

$$A(s) = B\cos\left(\pi\frac{s-{}^{D}\!/_{2}}{D}\right), \tag{4}$$

где *B* – вычисленное значение изгиба, *s* - расстояние вдоль начального пути, *D* – длина начального пути. Значение параметра *B* настраивается таким образом, чтобы кривая S_i обеспечивала минимальное значение интеграла (1), т.е. $B^* = \arg\min_B T_i(B)$, где $v(s)$ – распределение скорости вдоль луча S_i.

На втором шаге (Фигура B), мы передвигаем точку *b* до получения минимального значения интеграла (1). В случае наземных наблюдений этот шаг пропускается, поскольку точки *S* и *b* имеют одинаковые координаты.

На третьем шаге дальнейшее отклонение от пути между *b* и *r* вычисляется итеративно с использованием следующей формулы:

$$A(s) = \frac{B}{2}\cos\left(2\pi\frac{s-{}^{(D_2-D_1)}\!/_{2}}{(D_2-D_1)}\right)+\frac{1}{2} \tag{5}$$

где D_1 и D_2 соответствуют длинам вдоль путей начала и конца текущего сегмента.

На первой итерации вычисление значения изгиба производится для всего сегмента *b*-*r*, представленном на Фигуре A, по формуле (4). На второй итерации (Фигура 1) путь делится на два равных сегмента с длинами *b* - *m*1 и *m*1-*r* соответственно и для каждого из них пересчитывается длина изгиба в соответствии с формулой (5). После определения минимального времени, весь путь делится на три части (Фигура D), и пересчет значений изгиба осуществляется для трёх сегментов *b*-*m*1, *m*1-*m*2, и *m*2-*r*. Последовательно эта процедура повторяется для четырёх, пяти, шести и более частей. Процедура заканчивается, когда длина сегмента становится меньше заданной величины.

Параметризация. В программе PROFIT 2D распределение скоростей задается по аналогии с ранее разработанным для 3D томографической инверсии подходом и реализованным в программе LOTOS-07 [Кулаков, 2007]. Скоростные аномалии (неоднородности) интерполируются билинейно между фиксированными узлами сетки $\{h_j(r)\}_{j=1,M}$, $r=(x,z)$, которые задаются на множестве вертикальных линий с фиксированным шагом Δx вдоль профиля. Вдоль каждой линии по направлению в глубину (по переменной z) вычисляется значение плотности лучей $\rho(S^j)$ как сумма длин лучей в единичном объёме. И тогда узлы j, $j=1,...,M$, распределяются в соответствии с лучевой плотностью. В вертикальном направлении, во избежание чрезмерных узловых флуктуаций, задаётся минимальный шаг Δz между узлами. В областях с пониженной лучевой плотностью расстояние между узлами наибольшее. Узлы не устанавливаются в областях с лучевой плотностью меньше заданного значения (например, 0.1 от среднего). Хотелось бы отметить, узловые шаги в вертикальном и горизонтальном направлении не равны между собой, поскольку мы полагаем различное разрешение по вертикали и горизонтали. Узлы сетки устанавливаются только в первой итерации в соответствии с распределением лучей, протрассированных в стартовой модели. На последующих итерациях скоростные изменения Δv_j пересчитываются в этих же узлах.

При построении узлового распределения по лучевой плотности сеть можно определённым образом адоптировать в соответствии с некоторыми особенностями данных и пересчитать для выделенных изменений в лучевой плотности. Распределение лучей и плотность являются индивидуальными

параметрами для каждой области рефрагированных данных, зависящие не только от инструментальной настройки, но и от композиции и геометрии геологической поверхности и среды.

Отметим также, что хотя малый размер (порядок) сетки обеспечивает обнаружение скоростных аномалий, расположенных в изучаемой области в тоже время это приводит к увеличению размерности матрицы обращения.

Вычисление матрицы и обращение. Матрица первых производных вычисляется на основе лучевых траекторий S_i, которые определяются с помощью лучевого трассирования по 2D модели. Каждый элемент A_{ij} матрицы A в уравнении (3) равен временному изменению Δt_i вдоль S_i луча за единичное скоростное изменение Δv_j в *j* - ом узле параметризации. Элементы матрицы определяются численно.

Обращение матрицы A осуществляется с помощью итеративного LSQR алгоритма [Слуис, Ворст, 1987], который достаточно быстро сходится при хорошо подобранных параметрах регуляризации Am (амплитуда демпфирования) и Sm (весовой параметр сглаживания скоростных вариаций в соседних узлах параметризационной сетки). Итерационно решаем систему уравнений:

$$\begin{pmatrix} \mathbf{A} \\ am\ \mathbf{I} \\ sm\ \mathbf{C} \end{pmatrix} d\gamma = \begin{pmatrix} dT \\ 0 \\ 0 \end{pmatrix}. \qquad (6)$$

Где $\mathbf{I}$ – диагональная единичная матрица размера M, $\mathbf{C}$ - прямоугольная матрица, каждая линия которой состоит из двух единичных элементов противоположного знака, являющихся соседними узлами в параметризационной сетке.

Оптимальные значения этих параметров зависят от многих факторов. Например, при увеличении размера данных параметр демпфирования должен увеличиваться, в то время как при увеличении числа узлов вследствие измельчения разбиения параметр демпфирования должен быть уменьшен. В случае же большого уровня шума в данных параметр демпфирования должен быть достаточно весомым для сохранения устойчивости решения. Процедура нахождения коэффициентов демпфирования до сих пор не формализована. К тому же соотношение между числом параметров, лучей и значениями коэффициентов амплитуд и сглаживания является нелинейным. Например, при удвоении числа лучей соответствующая амплитуда решения получается увеличением коэффициента демпфирования до 1.2. Для каждой сети наблюдений эти значения определяются индивидуально. В качестве основного критерием при нахождении весовых коэффициентов демпфирования используется оценка среднеквадратического среднего (RMS) на каждой итерации. Но он не является абсолютно объективным, поскольку не обеспечивает устойчивости решения. Альтернативным подходом, реализованным в программе PROFIT, к определению оптимальных значений параметров демпфирования является синтетическое моделирование на тестовых примерах с аномалиями фиксированного размера. Например, тесты с шахматной доской.

Скоростные аномалии, полученные после инверсии, вычисляются по регулярной сети и добавляются в скоростную модель, полученную на предыдущей итерации.

Формализация подхода к исследованию области применимости алгоритма

Для того чтобы оценить качество решения, получаемое алгоритмом, часто прибегают к тестированию на синтетических, тестовых или реальных данных. Выше представленный алгоритм имеет в своём коде компоненты генерации синтетических данных по моделям сред (решение прямой задачи). Обладая такой

инструментом, появляется возможность формально описать класс задач (множество геологических разрезов земной коры) хорошо решаемых данным алгоритмом.

Обозначим через $G =< (L^d, v_\circ^d, \kappa) >$, $d \in N$, графическое цветное изображение поля скоростей модели среды с описаниями: L^d - *d – ый «*выпуклый» контур, $v_\circ^d$ - скорость среды внутри контура, κ - количество связанных выпуклых контуров. Связанными будем называть контуры, имеющие общую границу. Пусть G^* - изображение синтетической модели среды, G^m - реконструированное изображение среды. Реконструированное изображение строится пользователем по томографическому изображению поля скоростей, являющегося решением алгоритма. При подборе параметров регуляризации и минимизации значения RMS, естественно предполагается, что синтетическая модель является неизвестной.

Зададим множество $\widetilde{G}(G^*)$ допустимых моделей G^m относительно синтетической G^* следующим образом

$$\widetilde{G}(G^*) = \left\{ G^m : g(G^m, G^*) \le \delta^* , \left\| T - T^\circ \right\| \le \varepsilon \right\}, \tag{7}$$

Где $g(G^m, G^*)$ - «метрическая» разность изображений, построенная на основе симметрической разности топологий, ограниченных контурами L^d изображений и учитывающая допустимый диапазон изменения скорости $\Delta v_\circ^d$ внутри контура; $\left\| T - T^\circ \right\|$ - RMS оценка качества решения при построении томографического изображения алгоритмом.

Определим множество (кластер) $\Im$ как область применимости алгоритма следующим образом

$$\Im = \left\{ G^* : \exists G \;\; g(G, G^*) \le \delta , \left\| T - T^\circ \right\| \le \varepsilon \right\}. \tag{8}$$

Если мощность этого множества достаточно большая, то можно говорить о большой области применимости алгоритма. Каждый элемент этого множества представляет собой существенно отличающиеся друг от друга изображения сред, например зоны субдукции и соляные купола.

Для того чтобы ранжировать изображения формализуем понятие сложности геологического изображения и зададим её парой $\mu = (\mu_1, \mu_2)$, $\mu_1 = d$, $\mu_2 = 1 - \min\limits_d V(L^d)$, $V(L^d)$ - нормированный объём (площадь) топологического множества, ограниченного контуром L^d.

В качестве примера на рис. 2,3,4 приведем изображения синтетической модели соляных куполов, являющиеся результатами работы алгоритма PROFIT.

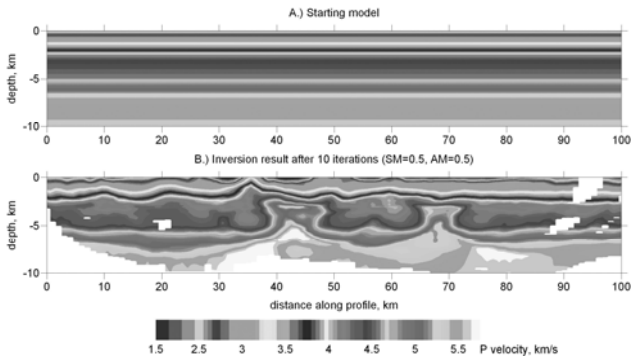


Рис. 2. Фигура В является томографическим изображением, представляющим результат инверсии с оптимально подобранными параметрами по временам, полученным по синтетической модели.

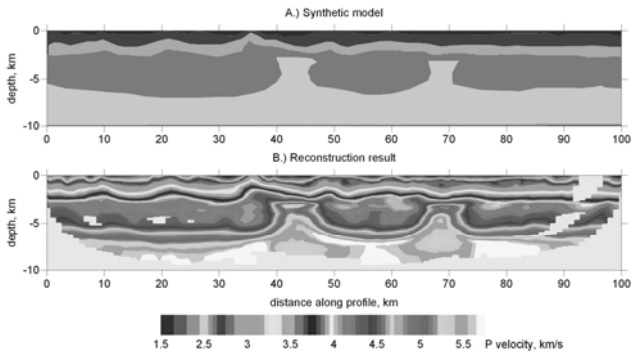


Рис. 3. Фигура В является томографическим изображением, представляющим результат инверсии по временам реконструированного изображения фигуры А.

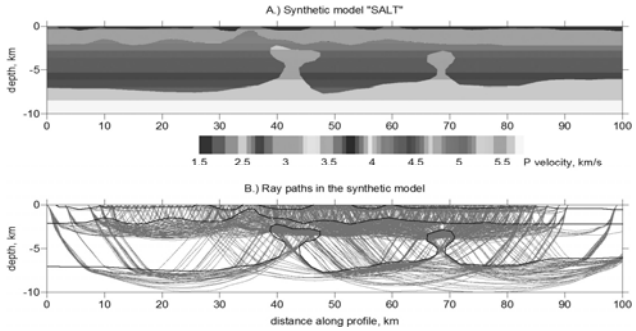


Рис. 4. Фигура А является синтетической моделью. Фигура В демонстрирует траектории лучевого трассирования.

Заключение

В работе предложена формальная постановка сложностно - структурного подхода к исследованию области применимости алгоритма PROFIT. Это позволяет количественно оценить возможности алгоритма относительно числа («мощность» множества $\mathfrak{I}$) и сложности (μ) реконструируемых геологических изображений. Отметим, что к настоящему моменту при интерпретации результатов томографической инверсии в основном используется экспертная оценка «качества» получаемого изображения. Нами предпринята попытка некоторой формализации оценки. Более того такой подход может быть использован и при сравнении томографических изображений с некоторыми поправками на размытые границы контура. В заключении также отметим, что несмотря на то, что при решении прямой задачи и вычислении времен по синтетической модели применялся один метод лучевого трассирования, построение лучей происходит абсолютно независимо. В будущем в целях наиболее объективного оценивания качества планируется оценивать качество алгоритма по синтетическим моделям с временами, полученными другими методами.

Благодарности

Работа опубликована при финансовой поддержке проекта **ITHEA XXI** Института информационных теорий и приложений FOI ITHEA Болгария www.ithea.org и Ассоциации создателей и пользователей интеллектуальных систем ADUIS Украина www.aduis.com.ua.

Библиография

[Вапник, 1984]. В.Н. Вапник. Алгоритмы и программы восстановления зависимостей. Москва: Наука, 1984.

[Кулаков, 2007] И.Ю. Кулаков. Геодинамические процессы в коре и верхней мантии земли по результатам региональной и локальной сейсмотомографии. Диссертация на соискание ученой степени доктора геолого-минералогических наук.ИГМ СО РАН, Новосибирск, 2007.

[Нолет, 1990] Г. Нолет. Сейсмическая томография. Москва: Мир, 1990.

[Слуис, Ворст, 1987] А. ван дер Слуис, Х. А. ван дер Ворст. Численное решение больших разреженных линейных алгебраических систем, возникающих в задачах томографии. В монографии [Г. Нолет, 1990].

Информация об авторах

Татьяна Ступина *– Институт нефтегазовой геологии и геофизики СО РАН, проспект Коптюга,3, Новосибирск, 630090, Россия; e-mail: stupinata@ipgg.nsc.ru*

Иван Кулаков *– Институт нефтегазовой геологии и геофизики СО РАН, проспект Коптюга,3, Новосибирск, 630090, Россия; e-mail: KoulakovIY@ipgg.nsc.ru*

ЭФФЕКТИВНОСТЬ БАЙЕСОВСКИХ ПРОЦЕДУР РАСПОЗНАВАНИЯ

Александра Вагис, Анатолий Гупал

Аннотация: *В основе байесовского подхода лежит определение таких структур описания объектов, для которых возможно построение эффективных (оптимальных) процедур распознавания. В настоящее время эти вопросы решены для дискретных объектов с независимыми признаками и объектов, которые описываются цепями Маркова. Получены детерминированные оценки погрешности байесовской процедуры распознавания в зависимости от размеров классов (обучающих выборок), количества признаков и количества значений признаков.*

Ключевые слова: *Байесовская процедура; погрешность процедуры; цепь Маркова; обучающая выборка.*

ACM Classification Keywords: *H.4.2 Information Systems Applications: Types of Systems: Decision Support.*

Conference*: The paper is selected from XV[th] International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.*

Введение

Задача распознавания рассматривается с точки зрения минимизации функционала среднего риска [Вапник, 1979].Такая постановка является обобщением классических задач, решаемых на основе метода наименьших квадратов, в том смысле, что наблюдению объекта может соответствовать не одно, а несколько состояний объектов. Она на порядок сложнее любой NP-полной задачи [Гэри, Джонсон, 1982]. При решении задач распознавания определяющим моментом является структура описания объекта. Если она известна, то не представляет труда по обучающей выборке построить байесовские процедуры, как это проделано для независимых признаков и цепей Маркова [Гупал, 1995], [Гупал, Вагис, 2001]. В дискретном случае показано, что оценка снизу сложности класса задач распознавания совпадает с верхней оценкой погрешности с точностью до абсолютной константы, т.е. байесовская процедура распознавания является субоптимальной. Количество классов и число значений признаков линейным образом входят в оценку погрешности байесовской процедуры распознавания. Таким образом, проблема минимизации среднего риска решена для дискретных объектов с независимыми признаками: построены оптимальные оценки погрешности процедур распознавания в зависимости от входных параметров задачи. Исследование эффективности байесовской процедуры распознавания на цепях Маркова основано на исследовании свойств оценок переходных вероятностей. Для нестационарных цепей Маркова получены асимптотические оценки погрешности байесовской процедуры распознавания, которые аналогичны оценкам для независимых признаков.

В монографии [Гупал, Сергиенко, 2008], по сути, впервые проведено обоснование индуктивного подхода с точки зрения построения эффективных процедур распознавания. Сделан вывод о том, что современную математику следует считать объединением дедуктивного и индуктивного подходов. Сфера применения дедуктивного подхода – получение точных решений в задачах малой размерности; область применения индуктивного подхода – решение задач распознавания и прогнозирования большой размерности на основе информации в обучающих выборках. В [Гупал, Сергиенко, 2008] излагаются прикладные аспекты развиваемой теории. Проведенные численные расчеты на основе информации из Всемирного банка белковых структур подтвердили высокую эффективность байесовских процедур распознавания (на моделях цепей Маркова различных порядков) в задачах распознавания вторичной структуры белков. Обсуждаются вопросы прогнозирования курсов акций ценных бумаг, валют, диагностики глиом головного мозга и т.д.

Постановка задачи распознавания в булевом случае

Пусть имеются множество объектов X, множество ответов Y (состояний объектов). Совокупность пар $X^l = (x^i, y_i)_{i=1}^l$ называется обучающей выборкой. Элементы множества X – это не реальные объекты, а лишь их описания, содержащие доступную часть информации об объектах. Невозможно исчерпывающим образом охарактеризовать, скажем, человека, геологический район, производственное предприятие или экономику страны. Поэтому одному и тому же описанию x могут соответствовать различные объекты, а, значит, и некоторое множество ответов (состояний объектов). Предполагается, что на множестве $X \times Y$ существует неизвестное вероятностное распределение $P(x, y)$, согласно которому сгенерирована выборка пар $X^l = (x^i, y_i)_{i=1}^l$, *x* – булев вектор $x = (x_1, x_2, ..., x_n)$ размерности *n*, состояние *y* принимает два значения нуль и единица. Для функции распознавания $A(x)$ определяется качество обучения в виде функционала

$$I(A) = \sum_{x \in X} \sum_{y=0}^{1} (A(x) - y)^2 P(x, y). \quad (1)$$

Минимизация среднего риска (1) является обобщением классических задач, решаемых на основе метода наименьших квадратов, т.е. когда наблюдению $x = (x_1, x_2, ..., x_n)$ соответствует не одно, а несколько состояний объектов (исходов экспериментов). В.Н. Вапник и А. Я. Червоненкис были одними из первых, кто придал задачам распознавания строгую математическую трактовку. В дальнейшем задача минимизации среднего риска (1) привлекла внимание многих ученых. Для булевых векторов $x = (x_1, x_2, ..., x_n)$ число различных функций $A(x) \in \{0,1\}$ равно 2^{2^n}. С математической точки зрения проблема минимизации среднего риска (1) сложнее любой NP-полной задачи, кроме того, проверку решения нельзя провести за полином операций, поскольку усреднение в (1) выполняется по всем 2^n векторам $x = (x_1, x_2, ..., x_n)$.

Предполагается, что наблюдаемый объект описывается вектором $x_1, x_2, ..., x_n$, f, где $x_1, x_2, ..., x_n$ – признаки (измерения) объекта, f – состояние объекта. Пусть проведено *m* наблюдений над объектами и в каждом случае было зафиксировано состояние объекта. Имеем обучающую выборку $V = (V_0, V_1, V_2)$ следующего вида: первая часть V_0 – булева матрица размерности $m_0 \times n$, где m_0 – число строк. Каждая строка представляет собой вектор $x = (x_1, x_2, ..., x_n, f)$, который выбран в соответствии с распределением P при условии $f = 0$. Вторая часть V_1 – булева матрица размерности $m_1 \times n$, где m_1 – число строк. Каждая строка матрицы – вектор x, который выбран на основе распределения P при условии $f = 1$. Последняя часть V_2 – булев вектор размерности m_2. Каждая компонента этого вектора – наблюдаемое значение состояния f, которое выбирается в соответствии с распределением P, можно считать, что m_2 = *m*= $m_0 + m_1$. ***Индуктивный шаг.*** Требуется построить такую процедуру индуктивного вывода, которая по измерениям $x_1, x_2, ..., x_n$ любого следующего объекта и обучающей выборке $V = (V_0, V_1, V_2)$ определяет состояние f объекта.

Погрешность процедуры. Сложность класса задач. Полагаем, что процесс определения состояния f объекта по известным значениям вектора входа $x = (x_1, x_2, ..., x_n)$ проводится с помощью функции $A(x)$, т.е. *f* = $A(x)$. Так как множество функций $A(x)$ конечно, то среди них существует наилучшая функция $A^*(x)$, такая, что $P(x, A^*(x)) = \max(P(x,0), P(x,1))$. Легко заметить, что функция $A^*(x)$ минимизирует средний риск (1).

Погрешность функции $A(x)$ – усредненная величина

$$\upsilon(A, P) = \sum_x P(x, A^*(x)) - P(x, A(x)). \quad (2)$$

т.е. $\upsilon(A^*,P)\equiv 0$. Погрешность (2) в отличие от методов минимизации эмпирического риска указывает на отклонение погрешности функции $A(x)$ от минимума среднего риска, поэтому для практических расчетов с ней удобнее работать. Индуктивная процедура распознавания Q строит по данной выборке $V=(V_0,V_1,V_2)$ и вектору x функцию $A(x)=Q(V,x)$. Погрешность процедуры Q на распределении *P* – усредненная величина:

$$\upsilon(Q,P)=\sum_{V\in W}\upsilon(A,P)P_1(V)\,. \qquad (3)$$

Усреднение (3) – принципиальный момент в теории оценивания погрешности процедур распознавания, заключающийся в усреднении погрешности функции по множеству всех обучающих выборок, имеющих заданные размеры классов. В результате такого усреднения получается детерминированная оценка погрешности байесовской процедуры распознавания для всего класса задач, учитывающего совокупность всевозможных распределений вероятностей. Заметим, что в методах минимизации эмпирического риска аналогичные оценки носят асимптотический и вероятностный характер, в связи с чем, их неудобно использовать на практике [Вапник, 1979]. Операция усреднения по своей сути выполняет функцию контроля: формула (3) оценивает качество работы процедуры на всевозможных новых объектах, не входящих в состав обучающей выборки.

Класс задач $C\equiv C(m_0,m_1,m_2,n)$ – совокупность всевозможных распределений вероятностей P, детерминированные числа m_0,m_1,m_2,n – вход задачи, они определяют размеры выборки. Погрешностью процедуры распознавания Q на классе C называется число $\upsilon(Q,C)=\sup_{P\in C}\upsilon(Q,P)$. Сложность класса C определяется величиной $\mu(C)=\inf_{Q}\upsilon(Q,C)=\inf_{Q}\sup_{p\in C}\upsilon(Q,P)$. В конечном итоге нужно построить такую процедуру распознавания Q, для которой число $\upsilon(Q,C)$ мало отличается от числа $\mu(C)$.

Байесовская процедура распознавания

Нами выбран прямой подход к построению методов распознавания: необходимо определить такие структуры описания объектов, для которых можно построить эффективные и даже оптимальные процедуры распознавания. Этот подход имеет очевидное преимущество перед другими подходами: вместо использования достаточно сложных процедур минимизации эмпирического риска или других функционалов качества (в специальном классе параметрических функций) используется простое байесовское правило классификации, которое используется для решения задач большой размерности. Известно,что оптимальным правилом классификации является байесовское правило, которое максимизирует апостериорную вероятность распределения классов при заданном наборе признаков объекта. Однако проблема заключается в том, что распределение вероятностей неизвестно, известны лишь эмпирические данные относительно значений признаков объектов и их состояний, которые составляют обучающую выборку. Построенное правило классификации должно хорошо работать за пределами обучающей выборки, т.е. обладать обобщающей способностью. Необходимо также, чтобы оценка погрешности процедуры достаточно быстро уменьшалась при увеличении размеров классов объектов, составляющих обучающую выборку.

Пусть $d=(d_1,d_2,...,d_n)$ – булев вектор. Считаем, что распределения P из класса C при каждом d удовлетворяют условию $P(x=d)=\prod_{j=1}^{n}P(x_j=d_j|f=i),\ i=0,1$, что означает независимость признаков x_j для каждого класса объектов. Рассматриваются случайные величины $\xi(d,i)$, которые зависят от d и i как от параметров,:

$$\xi(d,i)=(k(i)/m_2)\prod_{j=1}^{n}(k(d_j,i)/m_i)\,,\quad i=0,1\,. \qquad (4)$$

Здесь $k(d_j,i)$ – количество значений, равных d_j, j-го признака в j-м столбце матрицы V_i; $k(i)$ – количество значений целевого признака, равных i, в векторе V_2. Тогда функция распознавания определяется формулой

$$A(d)=\begin{cases}0, \text{ если } \xi(d,0)\ge\xi(d,1),\\ 1, \text{ если } \xi(d,0)<\xi(d,1).\end{cases} \quad (5)$$

Процедуру распознавания, определяемую соотношениями (4), (5), обозначим Q_B. Заметим, что величины $\xi(d,i)/(\xi(d,0)+\xi(d,1))$ представляют собой приближенные значения вероятностей $P(f=i|x_1=d_1,x_2=d_2,...,x_n=d_n)$, вычисленных по теореме Байеса. При подсчете оценок этих вероятностей существенным образом используется информация относительно размеров классов в обучающей выборке, поэтому эти значения входят в приводимые ниже оценки погрешности байесовской процедуры. Для байесовской процедуры проводить разбиение на обучающую и контрольную выборки не нужно, поскольку при подсчете погрешности учитывается работа процедуры на всем множестве обучающих выборок. Байесовская процедура (4), (5) строится программным образом, найти аналитический вид функции распознавания (5) невозможно. На основе неравенства Чебышева выведена оценка сверху погрешности байесовской процедуры на классе *C* [Гупал, 1995].

Теорема 1. *Существует абсолютная константа $\alpha<\infty$, такая, что справедливо неравенство*

$$\upsilon(Q_B,C)\le a\sqrt{\frac{n}{\min(m_0,m_1)}+\frac{1}{m_2}}. \quad (6)$$

При условии $\min(m_0,m_1)\ge 2n$ оценка сверху погрешности байесовской процедуры задается квадратным корнем, в противном случае (для малых выборок) она не превосходит единицы. Из доказательства теоремы вытекает, что абсолютная константа $a=4\sqrt{2}$. При выводе этой теоремы не нужно требовать, как в методах минимизации эмпирического риска, равномерную сходимость частот к их вероятностям, поскольку в силу усреднения (3) математическое ожидание и дисперсии частот в формуле (4) совпадает с вероятностями и дисперсиями бернуллиевских случайных величин. Показано, что если в обучающей выборке отсутствует один из классов, т.е. $\min(m_0,m_1)=0$, то любая процедура, в том числе и байесовская, работает плохо и ее погрешность строго положительна.

В теории минимизации эмпирического риска [Вапник, 1979] отсутствуют нижние оценки погрешности процедур, поэтому нельзя сделать вывод относительно эффективности предложенного подхода. Для доказательства оптимальности байесовской процедуры распознавания необходимо показать, что никакая другая процедура распознавания не может «работать лучше», чем байесовская процедура. В монографии [Гупал, Сергиенко, 2008] с помощью нетривиальных контрпримеров построены такие «плохие» распределения вероятностней, для которых нижняя оценка погрешности произвольной процедуры распознавания отличается от верхней оценки погрешности байесовской процедуры на абсолютную константу, т.е. байесовская процедура распознавания является субоптимальной.

Теорема 2. *Существует абсолютная константа $a_1>0$ такая, что справедливо следующее: каковы бы ни были целые числа m_0, m_1, m_2, натуральное число n, и процедура распознавания Q, существует такое распределение вероятностей P из класса C, что выполняется неравенство*

$$\upsilon(Q,P)\ge a_1\sqrt{\frac{n}{\min(m_0,m_1)}+\frac{1}{m_2}}. \quad (7)$$

Получен также интересный результат относительно того, что байесовская процедура распознавания в булевом случае эквивалентна процедуре распознавания, построенной на использовании отделяющей гиперплоскости.

Байесовская процедура. Дискретный случай

С точки зрения получения оценок погрешностей дискретный случай сложнее, чем булев случай, так как необходимо дополнительно учесть такие параметры, как число классов и количество значений признаков. Построение оценок в дискретном случае – нетривиальная задача, поскольку заранее неизвестно как количество классов, количество признаков и число значений признаков будут представлены в оценке погрешности байесовской процедуры распознавания.

Пусть имеется конечное множество *X* описаний объектов $b = (x_1, x_2, ..., x_n, f)$; здесь n – натуральное число, $x_j \in \{0,1,...,g-1\}$, $j = 1,2,...,n$; $f \in \{0,1,...,h-1\}$; g – число значений признаков, h – количество классов (состояний), $g \geq 2$, $h \geq 2$. В обучающей выборке V количества объектов различных классов заданы; более того, на практике они часто определяются заранее. Поэтому считаем, что выборка *V* состоит из $h+1$ части, $V = (V_0, V_1, ..., V_h)$. Пусть $m_0, m_1, ..., m_h$ – целые неотрицательные детерминированные числа. Часть V_i представляет собой целочисленную матрицу размерностью $m_i \times n$. Каждая строка этой матрицы является наблюдаемым значением вектора $x = (x_1, x_2, ..., x_n)$, описывающего объект класса i, который выбран из множества *X* в соответствии с распределением вероятностей P при условии $f = i$. Последняя часть V_h представляет собой целочисленный вектор размерностью m_h. Каждая компонента этого вектора есть наблюдаемое значение целевого признака f, которое выбирается в соответствии с распределением P. Иными словами, вероятность того, что j–я компонента вектора V_h принимает значение i, равна $P(f = i)$. Все строки матрицы V_i, $i = 0,1,...,h-1$ и компоненты вектора V_h являются независимыми случайными элементами. Считаем $m_0 \leq m_1 \leq ... m_{h-1}$; выполнение этого неравенства всегда можно обеспечить путем перенумерации классов объектов.

Байесовская процедура распознавания определяется на основе формулы

$$\xi(d, i) = \frac{k(i)}{m_h} \prod_{j=1}^{n} \frac{k(d_j, i)}{m_i},\ i \in \{0,1,...,h-1\}. \qquad (8)$$

Здесь $k(d_j, i)$ – количество значений, равных d_j, j-го признака в j– м столбце матрицы V_i; $k(i)$ – количество значений целевого признака f, равных i, в векторе V_h. Значение функции распознавания $A(d)$ равно минимальному числу s из множества $\{0,1,...,h-1\}$ такому, что

$$\xi(d, s) \geq \xi(d, i),\ i \in \{0,1,...,h-1\}. \qquad (9)$$

Приведем оценку сверху погрешности процедуры Q_B на классе C [Белецкий, 2006].

Теорема 3. *Существует абсолютная константа* $a < \infty$ *такая, что справедливо неравенство*

$$\upsilon(Q_B, C) \leq a\sqrt{\frac{gn}{m_0} + \frac{h}{m_h}}. \qquad (10)$$

Нижняя оценка сложности класса задач отличается от верхней оценки (10) на абсолютную константу. Таким образом, проблема минимизации среднего риска решена для дискретных объектов с независимыми признаками: построены оптимальные оценки погрешности процедур распознавания в зависимости от входных параметров задачи.

Байесовская процедура распознавания на цепях Маркова

Процедуры распознавания для независимых признаков имеют ограниченную область практических приложений. Большой интерес представляет развитие байесовских процедур распознавания на случай зависимых испытаний, связанных в цепь Маркова. Для того чтобы исследовать эффективность процедур распознавания на цепях Маркова и обосновать применение этих процедур на практике, необходимо изучить поведение оценок переходных вероятностей.

Рассмотрим ситуацию, когда $x_1, x_2, ..., x_n$ образуют последовательность зависимых случайных величин, связанных в цепь Маркова. Для модели цепи Маркова первого порядка вероятность цепочки $x_1, x_2, ..., x_n$ задается соотношением

$$P(x_1, x_2, ..., x_n | f) = P(x_1 | f) \times P(x_2 | x_1, f) \times ... \times P(x_n | x_{n-1}, f). \quad (11)$$

Здесь $P(x_1 | f)$ – начальное распределение вероятностей величины x_1, $P(x_k | x_{k-1}, f)$, $k = 2,..., n$, – нестационарные переходные вероятности. Как и в дискретном случае полагаем, что величины $x_j \in \{0,1,..., g-1\}$, $j = 1,2,..., n$; $f \in \{0,1,..., h-1\}$; g – число значений признаков, h – число классов.

Начальное распределение вероятностей $P(x_1 = d_1 | f = i)$ аппроксимируется частотой $\frac{k(d_1, i)}{m_1}$, где $k(d_1, i)$ – количество значений, равных d_1 в первом столбце матрицы V_i. В байесовской процедуре распознавания используются оценки переходных вероятностей, построенных в виде частот

$$\hat{p}(x_j = d_j | x_{j-1} = d_{j-1}, i) = \frac{k((d_{j-1}, d_j), i)}{k(d_{j-1}, i)}, j = 2,..., n \quad (12)$$

Здесь $k((d_{j-1}, d_j), i)$ – число пар (d_{j-1}, d_j) в строках матрицы V_i, $k(d_{j-1}, i)$ – число значений, равных d_{j-1} в (j–1)-м столбце матрицы V_i, $i \in \{0,1,..., h-1\}$.

В монографии [Гупал, Сергиенко, 2008] исследовались свойства оценок переходных вероятностей (12). На основе этих результатов получены асимптотические оценки погрешности байесовской процедуры распознавания, которые аналогичны оценкам для независимых признаков.

Индуктивный подход в математике

В монографии [Гупал, Сергиенко, 2008] излагаются основные принципы дедуктивного и индуктивного подходов в математике. Дедуктивно-аксиоматический подход используется в искусственной среде, созданной человеком, где основным требованием является получение точного решения и строгое доказательство теорем. Дедуктивный вывод проводится на основе лишь одного класса – истинных утверждений, ложные утверждения не используются. Отсюда следует неполнота формальных систем.

Индуктивный подход имеет непосредственное отношение к естественным наукам, поскольку объекты и явления природы изучаются на основе измерений и опытных данных. В индуктивном подходе на основе конечного числа наблюдений и исходов предыдущих экспериментов невозможно точно предсказать исход следующего эксперимента. Принципы индуктивных вычислений кардинально отличаются от дедуктивных принципов. Индуктивные процедуры формируют приближенное решение задачи, оценка погрешности которого (как мы видели для байесовских процедур) зависит от входных параметров задачи, причем она стремительно убывает при увеличении объемов обучающих выборок. В математике и образовании индуктивный подход широко не использовался, поскольку до последнего времени он не имел убедительного обоснования.

Дедуктивные вычисления проводятся на основе одного класса – истинных утверждений. Индуктивные процедуры строятся на основе информации, которая представлена в обучающих выборках. Причем, обучающие выборки содержат информацию относительно конечного числа классов наблюдаемых

объектов (по крайней мере, не менее двух). Поэтому для них удается построить эффективные и даже оптимальные процедуры распознавания.

В монографии, по сути, впервые проведено обоснование индуктивного подхода с точки зрения построения эффективных процедур распознавания. Сделан вывод о том, что современную математику следует считать объединением дедуктивного и индуктивного подходов. Сфера применения дедуктивного подхода – получение точных решений в задачах малой размерности; область применения индуктивного подхода – решение задач распознавания и прогнозирования большой размерности на основе информации в обучающих выборках. На наш взгляд, эти выводы будут полезны математикам, изучающим проблемы построения оснований современных математических дисциплин.

Заключение

Опыт решения задач в области компьютерного материаловедения (наплавочные материалы, качественные стали, сложные материалы) показал высокую достоверность методов индуктивного вывода. Разработанные методы продемонстрировали свою эффективность при решении сложных биологических задач, в частности, при моделировании процесса селекции сельскохозяйственных культур и прогнозировании белковых соединений. На основе байесовских процедур исследуются диагностические способы определения степени злокачественных опухолей головного мозга. Можно сделать вывод о том, что индуктивные процедуры имеют серьезные перспективы развития, поскольку процессы, протекающие в живой природе, выполняются на основе индуктивных вычислений.

Благодарности

Работа опубликована при финансовой поддержке проекта **ITHEA XXI** Института информационных теорий и приложений FOI ITHEA Болгария www.ithea.org и Ассоциации создателей и пользователей интеллектуальных систем ADUIS Украина www.aduis.com.ua.

Библиография

[Белецкий, 2006] Белецкий Б.А., Вагис А.А., Васильев С.В., Гупал Н.А. Сложность байесовской процедуры индуктивного вывода "Проблемы управления и информатики", 2006, №6. С.55-70.

[Вапник, 1979] Вапник В.Н. Восстановление зависимостей по эмпирическим данным. – М: Наука, 1979.– 448 с.

[Гупал, 1995] Гупал А.М., Пашко С.В., Сергиенко И.В. Эффективность байесовской процедуры распознавания. "Кибернетика и системный анализ", 1995, №4. С.76-89.

[Гупал, Вагис, 2001] Гупал А.М., Вагис А.А. Статистическое оценивание марковской процедуры распознавания. "Проблемы управления и информатики", 2001, №2. С.62-71.

[Гупал, Сергиенко, 2008] Гупал А.М., Сергиенко И.В. Оптимальные процедуры распознавания.– Киев: Наукова думка, 2008.– 232 с.

[Гэри, Джонсон, 1982] Гэри М., Джонсон Д. Вычислительные машины и труднорешаемые задачи. – М: Мир, 1982.– 416 с.

Сведения об авторах

Вагис Александра Анатольевна *– Институт кибернетики им. В.М. Глушкова НАН Украины, к.ф.-м.н., с.н.с., Киев, Украина. E-mail: valex_ic@mail.ru*

Гупал Анатолий Михайлович *– Институт кибернетики им. В.М. Глушкова НАН Украины, д.ф.-м.н., чл.-к. НАН Украины, зав. отд., Киев, Украина. E-mail: gupal_anatol@mail.ru*

ПОВЫШЕНИЕ ТОЧНОСТИ РЕШЕНИЯ ОБРАТНОЙ ЗАДАЧИ С ИСПОЛЬЗОВАНИЕМ СЛУЧАЙНЫХ ПРОЕКЦИЙ

Елена Ревунова, Дмитрий Рачковский

***Аннотация**: Приводятся результаты исследования разрабатываемого подхода к решению дискретной некорректной обратной задачи с использованием методов псевдообращения и случайных проекций. В известных обратных задачах Филипса и Барта получена относительная ошибка восстановления сигнала не хуже, чем для методов регуляризации Тихонова.*

***Ключевые слова**: обратная задача, устойчивое решение, случайные проекции, регуляризация Тихонова*

***ACM Classification Keywords**: I.5.4 Signal processing, I.6 Simulation and Modeling, G.1.9 Integral Equations*

***Conference**: The paper is selected from XV[th] International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.*

Введение

Решение обратной задачи актуально для целого ряда приложений. К ним относятся интерпретация данных физических экспериментов в задачах исследования различных полей и сред, а также многие другие задачи. В практических приложениях требуется решать дискретную обратную задачу вида:

$$\mathbf{\Phi x}=\mathbf{y}, \tag{1}$$

где матрица $\mathbf{\Phi}\in\Re^{N\times N}$ и вектор $\mathbf{y}\in\Re^{N}$ известны и получены в результате оцифровки методом Галеркина уравнения Фредгольма первого рода [1, 2, 3, 4]; требуется оценить вектора сигнала $\mathbf{x}\in\Re^{N}$.

В случае, когда $\mathbf{y}$ содержит аддитивный шум, $\mathbf{\Phi}$ имеет высокое число обусловленности $||\mathbf{\Phi}^{-1}||\,||\mathbf{\Phi}||$, ряд сингулярных чисел $\mathbf{\Phi}$ плавно спадает к нулю, – задачу оценки $\mathbf{x}$ называют [1] дискретной некорректной обратной задачей. Такие свойства $\mathbf{\Phi}$ характерны для задач оптики, спектрометрии, электроразведки.

Решение дискретной некорректной обратной задачи как задачи наименьших квадратов

$$\mathbf{x}'=\mathrm{argmin}_{\mathbf{x}}\,||\mathbf{y}-\mathbf{\Phi x}||_2 \tag{2}$$

в виде решения минимальной нормы [5] на основе псевдообращения

$$\mathbf{x}'=\mathbf{\Phi}^{+}\mathbf{y} \tag{3}$$

является неустойчивым [1, 3]. Признаком неустойчивости является то, что малым изменениям вектора $\mathbf{y}$ соответствуют большие изменения решения $\mathbf{x}'$; при этом велико значение ошибки решения.

Для преодоления неустойчивости и, соответственно, повышения точности решения в условиях шума используют подход регуляризации [1, 2, 3, 4], состоящий в учете априорной информации – о том, что l_2-норма решения $||\mathbf{x}||_2$ мала [5]; либо что решение является разреженным [6], т.е. $||\mathbf{x}||_0$ мало; либо о точности аппроксимации [6].

При наличии априорной информации о малости l_2-нормы решения для получения устойчивого решения используют решение минимальной нормы (3) либо регуляризацию Тихонова [1]. В случае регуляризации Тихонова задачу формулируют следующим образом

$$\mathbf{x}'=\mathrm{argmin}_{\mathbf{x}}\,(\,||\mathbf{y}-\mathbf{\Phi x}||_2+\lambda||\mathbf{x}||_2\,), \tag{4}$$

где λ – параметр регуляризации.

Недостатком, присущим методам решения дискретных некорректных обратных задач на основе регуляризации Тихонова, является необходимость подбора λ-параметра регуляризации, от правильности которого в значительной мере зависит устойчивость решения. Методы подбора параметра регуляризации имеют высокую вычислительную сложность. Методы на основе псевдообращения не требуют подбора параметров, однако они не обеспечивают [1, 3] устойчивости и точности решения. Поэтому востребованными являются подходы к устойчивому решению дискретной некорректной обратной задачи на основе псевдообращения, но с точностью на уровне регуляризации Тихонова.

В данной работе приводятся результаты исследования разрабатываемого нами подхода к решению дискретной некорректной обратной задачи с использованием методов псевдообращения и случайных проекций на двух известных обратных задачах - Филипса и Барта [7], [8].

Решение дискретной некорректной обратной задачи

Нами предлагается подход к решению дискретной некорректной обратной задачи, использующий аналогию с современными работами в области Compressed Sensing ("сжатого зондирования") [9], [10], [11]. Здесь для высокоточного восстановление $\mathbf{x}$ по $\mathbf{Rx}$ в качестве проектора $\mathbf{R}$ используют матрицу с элементами, сформированными реализациями случайной величины с нормальным законом распределения [11], [12]. Такого рода случайные проекционные матрицы с $k<N$ используются также в теории [11], [13] и практике [14] вложений векторных пространств (vector space embeddings) для сокращения размерности векторов с целью ускорения оценки их сходства.

Для решения обратной задачи с использованием проекционного подхода умножим обе части исходного уравнения (1) на матрицу $\mathbf{R}\in\Re^{k\times N}$, $k\le N$, элементы которой – реализации случайной величины с нормальным распределением, нулевым средним и единичной дисперсией. Число столбцов N матрицы $\mathbf{R}$ фиксировано и определяется размерностью исходной матрицы $\mathbf{\Phi}$: число строк k может выбираться произвольно. Получаем уравнение

$$\mathbf{R\Phi x}=\mathbf{Ry},\ \mathbf{R\Phi}=\mathbf{A},\ \mathbf{A}\in\Re^{k\times N},\ \mathbf{Ry}=\mathbf{b},\ \mathbf{b}\in\Re^{k}. \tag{5}$$

Восстановление сигнала методом наименьших квадратов

$$\mathbf{x}'=\operatorname{argmin}_{\mathbf{x}}\ ||\mathbf{Ry}-\mathbf{R\Phi x}||_2. \tag{6}$$

Тогда восстановление сигнала $\mathbf{x}$ на основе псевдообращения получим как

$$\mathbf{x}'=(\mathbf{R\Phi})^{+}\mathbf{Ry},\ \mathbf{x}'=(\mathbf{A})^{+}\mathbf{b}, \tag{7}$$

либо как

$$\mathbf{x}'=((\mathbf{R\Phi})^{\mathrm{T}}\mathbf{R\Phi})^{+}(\mathbf{R\Phi})^{\mathrm{T}}\mathbf{Ry},\ \mathbf{x}'=(\mathbf{A}^{\mathrm{T}}\mathbf{A})^{+}\mathbf{A}^{\mathrm{T}}\mathbf{b}. \tag{8}$$

Восстановление сигнала методом регуляризации Тихонова получим как

$$\mathbf{x}'=\operatorname{argmin}_{\mathbf{x}}\ (\ ||\mathbf{Ry}-\mathbf{R\Phi x}||_2+\lambda||\mathbf{x}||_2\ ). \tag{9}$$

Точность решения обратной задачи будем оценивать с помощью относительной ошибки d восстановления истинного сигнала $\mathbf{x}$, вычисляемой как

$$d=||\mathbf{x}-\mathbf{x}'||/||\mathbf{x}||=||\Delta\mathbf{x}||/||\mathbf{x}||, \tag{10}$$

где $\Delta\mathbf{x}$ – ошибка решения, $\mathbf{x}'$ – вектор восстановленного сигнала.

Экспериментально исследуем поведение зависимости относительной ошибки восстановления сигнала от уровня аддитивного собственного шума в векторе сигнала $\mathbf{y}$ (1).

Экспериментальное исследование

Исследовались примеры дискретных некорректных обратных задач Филипса и Барта [7], [8].

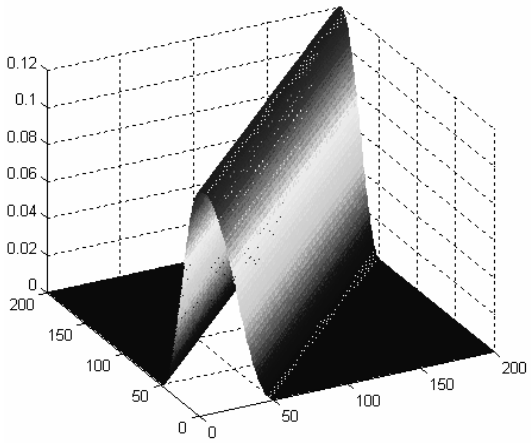


Рис. Матрица **Ф** оцифровки ядра в задаче Филипса

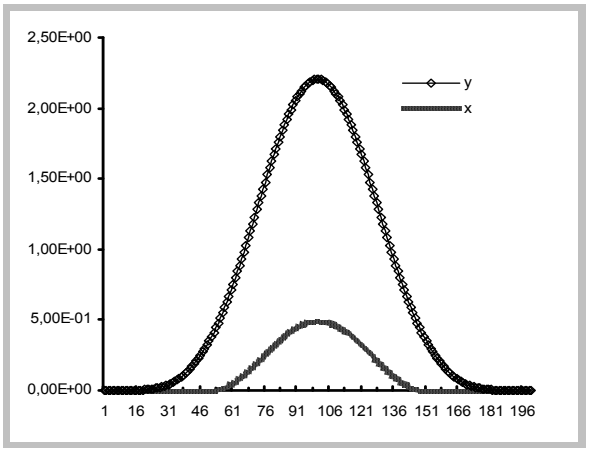


Рис. Дискретно заданные сигнал **x** правая части **y** в задаче Филипса

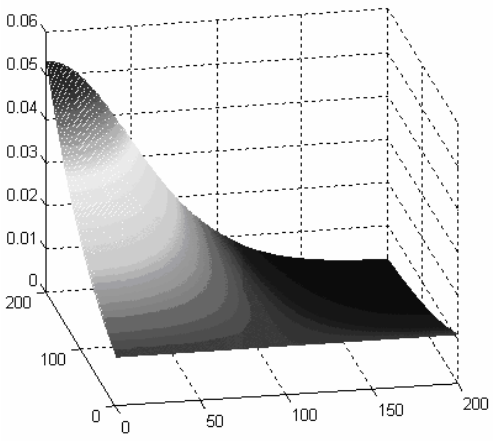


Рис. Матрица **Ф** оцифровки ядра в задаче Барта

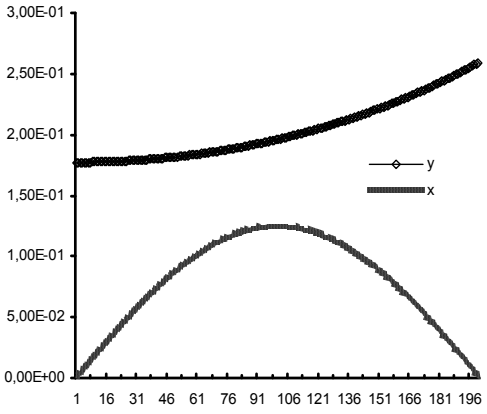


Рис. Дискретно заданные сигнал **x** и правая часть **y** в задаче Барта

В обеих задачах матрица **Ф**, полученная при оцифровке ядра, имела размерность 200×200, высокое число обусловленности ($||\mathbf{\Phi}^{-1}||\ ||\mathbf{\Phi}|| >> 1$), и ряд сингулярных чисел, плавно спадающий к нулю. Вектор правой части уравнения (2) после оцифровки искажался аддитивным шумом ε с нормальным распределением. Исследовалось поведение зависимости относительной ошибки восстановления сигнала от уровня аддитивного собственного шума в сигнале **y**. Сравнивались зависимости $d(\varepsilon)$ для методов на основе псевдообращения и регуляризации без применения проецирования с аналогичными зависимостями после применения проецирования. Размерность k для псевдообращения с проецированием выбиралась экспериментально по минимуму ошибки d.

Тестовая задача Филипса. Для задачи Филипса матрица Ф получается из аналитически заданной функции ядра $K(s,t) = \phi(s-t)$, $\phi(x) = 1+\cos(\pi x/3)$ при $|x|<3$, $\phi(x) = 0$ при $|x|\geq 3$, **y** из $g(s) = (6-|s|)(1+0.5\cos(\pi s/3)) + 9/2\ \pi \sin(\pi|s|/3)$ путем оцифровки по методу Галеркина [1].

Результаты зависимости относительной ошибки восстановления сигнала от уровня шума без умножения на матрицу-проектор (*pinv1*, *pinv2*, *reg1*, *reg2*, *reg3*) и результаты для экспериментов с умножением на матрицу **R** (*pinv1r*, *pinv2r*, *reg1r*, *reg2r*, *reg3r*) приведены на рис. 1 (расшифровка названий методов приведена далее).

Среди методов решения обратной задачи без проецирования наибольшую ошибку дают методы на основе псевдообращения матрицы **Φ**. При уровне шума *nl* 1e-7 относительная ошибка *d* (10) для *pinv1* составляет 0.36, а для *pinv2* – 0.14. При уровнях шума с 1e-5 и 1e-6 (для *pinv1* и *pinv2* соответственно) относительная ошибка *d* превышает 1.0, что свидетельствует о значительном искажении формы восстанавливаемого сигнала. Метод *pinv2* во всем исследованном диапазоне шума дает ошибку меньшую, чем *pinv1*.

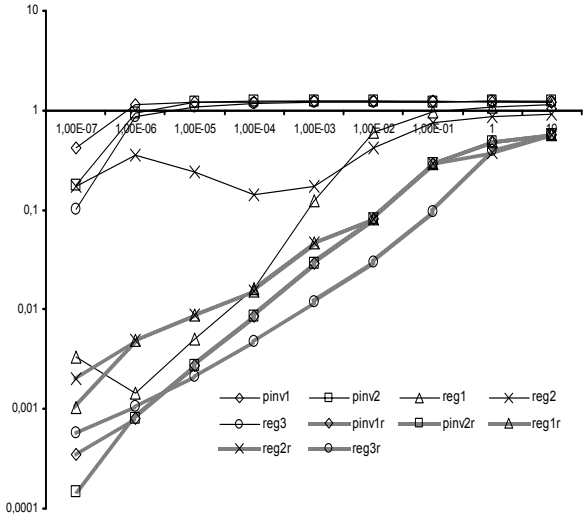


Рис. 1. Зависимость относительной ошибки от уровня шума

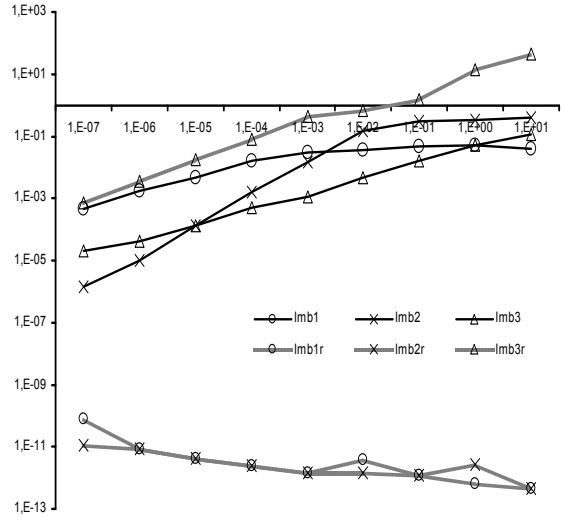


Рис. 2. Зависимость параметра регуляризации от уровня шума

Для регуляризации Тихонова с подбором параметра регуляризации по методу *L*-кривой *reg2* в диапазоне шума от 1e-7 до 1e-3 значение ошибки *d* изменяется в пределах от 0.14 до 0.35; при дальнейшем возрастании уровня шума ошибка *d* растет. Высокая относительная ошибка *d* этого метода связана с тем, что получаемые им значения параметра регуляризации λ не оптимальны.

Наименьшие значения ошибки *d* демонстрирует метод основе регуляризации Тихонова с подбором λ по методу кроссвалидации *reg1*. В диапазоне шумов 1e-7 – 1e-3 метод обеспечивает наименьшую ошибку среди всех исследованных нами методов регуляризации без проецирования. Низкие значения ошибки в указанном диапазоне шумов обеспечиваются благодаря удачному подбору параметра регуляризации.

Зависимость ошибки *d* от уровня шума, полученная методом регуляризации Тихонова с подбором λ по методу обобщенной невязки *reg3*, аналогична зависимости для методов на основе псевдоинверсии: при уровнях шума 1e-7–1e-6 она составляет 0.1 – 0.86; а при уровнях шума выше 1e-5 *d*>1. Такое поведение относительной ошибки связано с тем, что выбор λ методом обобщенной невязки осуществляется неверно (рис.2). Видно, что значения параметра регуляризации *lmb3* для этого метода далеки от значений параметра регуляризации *lmb1*, полученных методом *reg1*, демонстрирующим лучшую точность.

Рассмотрим результаты методов решения обратной задачи *с проецированием*. При оптимальном *k* (в зависимости от шума *k* менялось от *k*=37 при *nl*=1e-7 до *k*=2 при *nl*=1e1) все они обеспечивают более устойчивое восстановление сигнала и более низкий уровень относительной ошибки, чем соответствующие методы без применения проецирования. Только для регуляризации с выбором λ по методу кросс-валидации ошибка после проецирования *reg1r* и при уровнях шума 1e-6–1e-4 приближается к значениям ошибки до проецирования *reg1*. При дальнейшем нарастании уровня шума ошибка для *reg1r* становится намного меньшей, чем для *reg1*.

Среди методов регуляризации с проецированием *reg1r*, *reg2r*, *reg3r* наименьшую относительную ошибку *d* демонстрирует метод с выбором λ по обобщенной невязке *reg3r*. Как видно из рис. 2, только для *reg3r* параметр регуляризации *lmb3r* растет с ростом шума, в то время как для методов *reg1r*, *reg2r* значения

параметра регуляризации *lmb1r*, *lmb2r* имеют тенденцию к убыванию. Это свидетельствует о неправильном выборе параметра регуляризации для *reg1r*, *reg2r*, так как известно [1], что его значение должно возрастать с ростом уровня шума.

Отметим, что после проецирования методы, основанные на псевдообращении матрицы, обеспечивают относительную ошибку не хуже, чем методы регуляризации. Учитывая меньшие вычислительные затраты методов *pinv1r*, *pinv2r*, предпочтительно их использование.

Тестовая задача Барта. Для задачи Барта матрица **Ф** получается из аналитически заданной функции ядра $K(s,t) = \exp(s \cos t)$, **y** из $g(s) = 2 \sin s / s$ путем оцифровки по методу Галеркина.

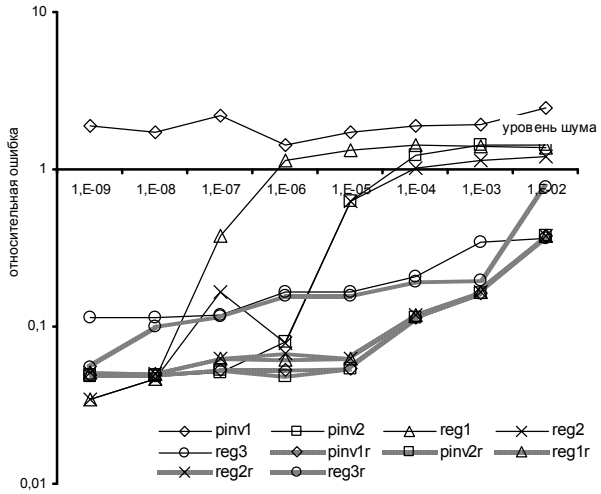


Рис. 3 Зависимость относительной ошибки восстановления сигнала *d* от уровня шума

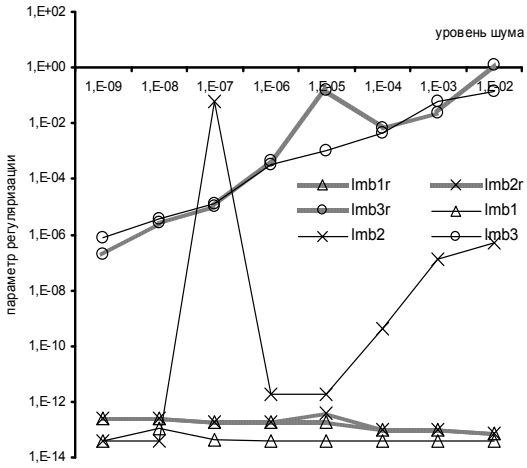


Рис. 4 Зависимость параметра регуляризации *lmb* от уровня шума

На рис.3 приведены экспериментально полученные зависимости относительной ошибки восстановления сигнала *d* от уровня шума без умножения на матрицу-проектор **R** для методов *pinv1*, *pinv2*, *reg1*, *reg2*, *reg3* и результаты для экспериментов с умножением на **R**: *pinv1r*, *pinv2r*, *reg1r*, *reg2r*, *reg3r*. В диапазоне собственных шумов 1e-8,...,1e-3 относительная ошибка восстановления сигнала для методов с умножением на матрицу **R** меньше относительной ошибки для методов без умножения.

После проецирования, при оптимальном *k* (в зависимости от уровня шума, *k* менялось от *k*=7 при *nl*=1e-9 до *k*=3 при *nl*=1e-2) относительная ошибка восстановления сигнала *d* для методов на основе псевдообращения матрицы во всем исследованном диапазоне шумов становится меньше ошибки для методов регуляризации.

Вывод

Для дискретных некорректных обратных задач Филипса и Барта, полученных в результате оцифровки по методу Галеркина ядра и правой части уравнения Фредгольма первого рода, при решении методами псевдообращения с использованием проецирования снижается относительная ошибка восстановления сигнала. При этом, при оптимальном *k*, точность результатов не хуже, чем для методов на основе регуляризации Тихонова без проецирования. Это делает использование методов на основе псевдообращения матрицы предпочтительным в силу их большей устойчивости (проявляющейся в плавном изменении относительной ошибки восстановления сигнала с ростом шума); точности восстановления сигнала **x**, сравнимой или лучшей, чем у методов регуляризации Тихонова; и меньших вычислительных затрат. Направлением дальнейших исследований являются экспериментальные и теоретические методы вычислительно эффективного априорного выбора оптимального *k*.

Благодарности

Работа опубликована при финансовой поддержке проекта **ITHEA XXI** Института информационных теорий и приложений FOI ITHEA Болгария www.ithea.org и Ассоциации создателей и пользователей интеллектуальных систем ADUIS Украина www.aduis.com.ua.

Литература

1. *Hansen P.C.* Rank-deficient and discrete ill-posed problems. Numerical Aspects of Linear Inversion. – SIAM, Philadelphia, 1998. – 247 p.
2. *Тихонов А.Н., Арсенин В.Я.* Методы решения некорректных задач. – М.: Наука, 1979. – 285 с.
3. *Морозов В.А.* Регулярные методы решения некорректно поставленных задач. – М.: Наука, 1987. – 239 с.
4. *Engl H.W., Hanke M., Neubaer A.* Regularization of inverse problems. Kluwer Academic Publishers, Dordrecht, 2000. – 321 p.
5. *Demmel J.W.* Applied Numerical Linear Algebra. – SIAM, Philadelphia, 1997. – 419 p.
6. *Donoho D., Elad M., Temlyakov V.* Stable Recovery of Sparse Overcomplete Representations in the Presence of Noise // IEEE Trans. Information Theory. – 2006. – №52. – P. 6-18.
7. *Baart M.L.* The use of auto-corellation for pseudo-rank determination in noisy ill-conditioned least-squared problems // IMA J.Numer.Anal. – 1982. – № 2. – P. 241-247.
8. *Phillips D.L.* A technique for the numerical solution of integral equation of the first kind // J.ACM. – 1962. – № 9. – P. 84-97.
9. *Candès E.J.* Compressive sampling // Proc. International Congress of Mathematics. – 2006. – №3. – P.1433-1452.
10. *Candès E.J., Romberg J., Tao T.* Stable signal recovery from incomplete and inaccurate measurements // Comm. Pure Appl. Math. – 2005. – № 59. – P. 1207-1223.
11. *Baraniuk R., Davenport M., DeVore R., Wakin M.* A Simple Proof of the Restricted Isometry Property for Random Matrices // Constructive Approximation. – 2008. – V.28, №3. – P. 253-263.
12. *Rudelson M., Veshynin R.* Sparse Reconstruction by Convex Relaxation: Fourier and Gaussian Measurements. CISS 2006 (40th Annual Conference on Information Sciences and Systems). – P. 207-212.
13. *Johnson W.B., Lindenstrauss J.* Extensions of Lipschitz mappings into a Hilbert space // Contemporary Mathematics. – 1984. – №26. – P.189-206.
14. *Мисуно И.С., Рачковский Д.А., Слипченко С.В.* Векторные и распределенные представления, отражающие меру семантической связи слов // Математичні машини і системи. – 2005. – № 3. – С. 50-67.

Информация об авторах

Елена Г. Ревунова *– научный сотрудник; Международный научно-учебный центр информационных технологий и систем НАН и МОН Украины, пр. Акад. Глушкова 40, Киев-03680, Украина; e-mail: helab@i.com.ua*

Дмитрий А. Рачковский *– ведущий научный сотрудник; Международный научно-учебный центр информационных технологий и систем НАН и МОН Украины, пр. Акад. Глушкова 40, Киев-03680, Украина; e-mail: dar@infrm.kiev.ua*

МЕТОДЫ И СРЕДСТВА СЕГМЕНТАЦИИ ПОЛЬЗОВАТЕЛЕЙ WEB-САЙТОВ

Дмитрий Ночевнов

***Аннотация**: В статье проанализирован характер данных о пользователях в сети WWW и метрики сайтов с точки зрения возможности их использования для сегментации посетителей Web-сайтов, предложена классификация таких метрик, рассмотрены основные пути и средства сбора и визуализации сведений о пользователях. Дан сравнительный анализ основных методов сегментации "с учителем" и "без учителя", а также средств сегментации различной сложности, в том числе Google Analytics, Web Mining for Clementine. Рассмотрены основные показатели качества сегментации и предложен адаптированный вариант процесса Cross-Industry Standard Process for Data Mining для сегментации.*

***Ключевые слова**: Web user segmentation, Web usage mining, Internet marketing.*

***ACM Classification Keywords**: H.2.8 Database Applications - Data mining. J.1 Administrative data processing - Marketing.*

***Conference**: The paper is selected from XVth International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.*

Введение

В Интернете представлены Web-сайты различных типов - для электронной коммерции, для выявления потенциальных потребителей, службы работы с покупателями, основанные на рекламе информационные сайты, информационные сайты, основанные на подписке и брендовые сайты. Их пользователи выступают в роли потребителей товаров и услуг, информация о которых размещена на сайте, и поэтому для исследования этих пользователей есть смысл использовать соответствующие маркетинговые методы и модели, наработанные на протяжении многих десятилетий. Однако особенности поведения человека в Интернете, характер собираемых о нём данных и природа самой сети заставляют маркетологов и предпринимателей искать более эффективные пути исследования посетителей Web-сайтов, более точного анализа и прогнозирования их поведения с целью повышения эффективности планирования бизнеса и продвижения товаров и услуг через WWW. Одним из методов таких исследования является сегментация пользователей.

1. Определение и назначение сегментации пользователей Web-сайтов

Сегментация пользователей (другое название **Profile Mining**) заключается в идентификации и анализе отдельных групп пользователей со схожей активностью, потребностями, желаниями, характеристиками, отличающихся устойчивыми признаками и реакцией на предложения, с целью определения размера и значения маркетинговых сегментов [Токарев, 705].

Сферами применения сегментации пользователей сайтов являются:

1) оптимизация производительности сайта за счёт выявления и исправления возникающих у посетителей из различных частей аудитории сложностей в работе с сайтом, учёта их предпочтений [Burby, 2007];

2) построение рекомендательных систем [Markov, 2007] и персонализация Web-страниц;

3) создание целевых страниц для различных сегментов посетителей с целью улучшения конверсии сайта [Burby, 2007];

4) в целевом маркетинге [Weinstein, 2004] для выявления и описания различных характеристик групп

пользователей, и применения к ним маркетинговых стратегий и стратегий разработки продуктов с целью повышения продаж, лучшего выявления, понимания и покрытия целевых рынков;

5) для более точного вычисления значений ключевых индикаторов продуктивности бизнеса (KPI) благодаря учёту сегментов и, соответственно, выявления реального прогресса организации в достижении собственных бизнес-целей [Burby, 2007].

С учётом двух основных составляющих сегментации [Mason, 2006] - способа сегментации и базы, относительно которой выполняется сегментация - были определены следующие ***цели данного исследования*:**

1) выполнить обзор базы сегментации пользователей сайтов и источников статистических данных;

2) дать сравнительный анализ основных методов сегментации;

3) выделить показатели качества сегментации;

4) определить основные этапы сегментации;

5) выполнить обзор существующих средств сегментации.

Определим сначала перечень данных о посетителях, которые могут быть собраны в базе сегментации, а также используемые для этого метрики.

2. База сегментации пользователей Web-сайтов и источники статистических данных

2.1. Характер данных о пользователях Web-сайтов

Во время работы пользователя в WWW вместе с запросом и через счётчики посещений на сайт передаются первичные сведения о пользователе (*primary data*), такие как системный язык, адрес источника перехода на сайт, адресе компьютера и др. Эти сведения, вместе с информацией о сделанных пользователями действиях, могут быть записаны в специальные слабоструктурированные журналы сервера, и позже проанализированы. Кроме этого в профайлах пользователя в базе данных сайта, а также в журналах работы Web-приложений обычно сохраняются дополнительные данные о пользователе (*secondary data*) [Burby, 2007]: сведения из регистрационной формы, рейтинги продуктов, статей, последние покупки и т.п., которые позволяют установить явные или скрытые предпочтения пользователя.

В Web Usage Mining принята следующая терминология [Liu, 2008]:

- *просмотры страниц* (*pageview*) – обобщённое представление коллекции Web-объектов, отображаемых в Web-браузере пользователя во время выполнения отдельного действия (например, щелчка мышкой на рекламном объявлении, чтения статьи, просмотра товара и добавление его в корзину);
- *сессии* (*sessions*) – последовательность просмотров Web-страниц отдельным пользователем на протяжении одного визита. Сессии могут быть в дальнейшем обобщены путём выбора подмножества просмотров интересуемых страниц в пределах сессии;
- *эпизоды* (*episodes*) – подмножество последовательностей сессий, обобщающих семантически или функционально связанные просмотры страниц.

Функцию сохранения журналов запросов поддерживает большинство Web-серверов. Счётчики посещений, вместе со средствами просмотра отчётов о собранных данных, предоставляются многими Web-службами, в том числе Google Analytics, Coremetrics, W3Counter, SpyLog, Bigmir)Net и т.п.

2.2. Классификация переменных сегментации пользователей сайтов

В области Web-аналитики и Web Mining при сборе данных о пользователях используются специальные *метрики сайтов*. Они составляют базу сегментации и могут быть соотнесены с некоторыми маркетинговыми *переменными сегментации*. Например, такой поведенческой переменной, как частота покупок, может быть поставлена в соответствие метрика "количество сделанных пользователем покупок на сайте", а язык потребителя можно считать совпадающим с системным языком, указанным в его HTTP-

запросах.

Рассмотрим подробнее метрики сайтов, используемые в области Web-аналитики, с точки зрения возможности их использования в качестве переменных сегментации пользователей.

В маркетинге при составлении схемы сегментации потребителей обычно используют две группы переменных [Kotler, 2006]:

1) *описательные характеристики*:

- географические: страна, район, область проживания и т.п.;
- демографические: возраст, пол, семейное положение, доходы, социальный класс и т.п.;
- психографические: тип личности и т.п.;

2) *характеристики поведения*, например способы использования продукта или торговой марки.

Эти группы переменных можно дополнить ещё тремя группами метрик сайтов:

1) *сведения, передаваемые с компьютера пользователя автоматически во время посещения сайта* ("*технические признаки пользователя*" согласно [Пелещишин, 2007]), включающие в себя:

- *данные о компьютере, передаваемые через поля заголовка HTTP-запроса*: характеристики программного обеспечения, системный язык, источник перехода на сайт, поисковый запрос пользователя, который привёл на сайт или страницу, географическое расположение провайдера, Cookies и т.п.
- данные о компьютере, которые могут быть прочитаны из Web-браузера с помощью счётчиков посещений (встроенных в Web-страницы JavaScript-программ): характеристики монитора, история просмотров страниц в текущем сеансе работы браузера и др.;

2) *дополнительная информация с сайта* [Liu, 2008]: ключевые слова просмотренного содержимого и атрибуты интересуемых продуктов или услуг;

3) *обобщённая Интернет-статистика:*

- глобальная и региональная Интернет-статистика, которую можно найти на сайтах W3Counter, Bigmir)net, SpyLog и др.;
- метрики отраслевой статистики (benchmarking), включающие в себя сведения о посетителях сайтов в зависимости от их отраслевой принадлежности и предоставляемые такими Web-службами, как Google Ad Planner, Google Trends, Google Benchmarking, Coremetrics, ClickZ Stats, Fireclick и др.

Только небольшую часть *описательных характеристик* пользователей можно хотя бы приблизительно определить автоматически по косвенным показателям, таким как системный язык, место расположения Интернет-провайдера пользователя, предпочитаемые товарам, временя и суммы покупок. Дополнительную информацию могут дать комментарии пользователей и указанные ими рейтинги статей и продуктов. Точное определение описательных характеристик возможно только из открытых источников в Интернете или на самом сайте через форму регистрации пользователя, различные анкеты и голосования.

Большая часть принятых в маркетинге *поведенческих характеристик*, таких как *ожидаемые выгоды от покупки*, *степень верности продукту*, *степень готовности купить продукт*, *отношение к продукту* тоже с трудом поддаются автоматическому измерению и могут быть определены только косвенно или же самым пользователем. В тоже время из первичных данных о посещении сайта можно значительно точнее, чем с помощью традиционного маркетингового анкетирования, рассчитать *частоту покупок* и *пользовательский статус* (новичок, бывший, потенциальный, или постоянный пользователь). Кроме этого ассоциацией *Web Analytics Association* [Cutroni, 2008] предложено несколько десятков дополнительных метрик, учитывающих посещаемые пользователем страницы и выполняемые им действия, внутренние и внешние источники перехода на сайт, длительность визитов, среднее количество просмотров страниц за визит, количество посещений целевых страниц и целевых действий (покупок, подписок, кликов на рекламные ссылки и т.п.). В частично обновлённом и дополненном виде эти метрики

поддерживаются многими Web-аналитическими порталами, среди которых Google Analytics, Coremetrics и др.

Большинство Web-метрик имеют не Гауссовский закон распределения [Clifton, 2008], а случайный. Из-за этого полученные без сегментации значения могут быть неверными. Причина этого в том, что собранная Интернет-статистика включает в себя одновременно как сведения о новых пользователях, так и о тех, которые вернулись на сайт, кто только ознакамливается с продуктами, о покупателях, сотрудниках и конкурентах фирмы. Каждая из этих групп может, например, посещать сайт в разное время и день недели. В такой ситуации разбиение пользователей на группы и вычисление для каждой из них собственного значения среднего времени посещения будет целесообразнее.

3. Сравнительный анализ основных подходов к сегментации пользователей сайтов

Рассмотрим существующие на сегодняшний день основные подходы к сегментации, принятые в Web Usage Mining [Markov, 2007] для анализа пользователей, и методы сегментации потребителей, используемые в маркетинговых исследованиях [Weinstein, 2004]. В зависимости от способа разбиения на сегменты, их можно разделить на методы сегментации "с учителем", "с подкреплением" и "без учителя".

3.1. Методы сегментации "с учителем"

Для методов сегментации "с учителем" характерна частичная предопределённость целевых метрик и наличие обучающих образцов [Mason, 2006], [Markov, 2007]. Получаемые сегменты основаны на некоторой предопределённой классификации, которая может быть простой (например, пол мужской/женский) или более сложной (например, "Первые пользователи с отменённой покупкой телевизора"). Разбиение выполняется на основе гипотезы об интересности, важности и значительности будущих сегментов.

Большинство существующих на сегодня средств анализа пользователей сайтов, такие как Google Analytics, Coremetrics, предлагают детерминированные подходы к сегментации с использованием статистических методов [Liu, 2008]. При этом данные могут обобщаться в предопределённые блоки вроде дней посещений, отдельных сессий, посетителей или доменов, из которых поступил запрос пользователя.

Как правило, методы обучения "с учителем" выполняются за 3 шага [Markov, 2007]: 1) составляется тренировочный набор данных с предварительно классифицированными значениями целевых переменных в дополнении к независимым переменным; 2) путём обучения с помощью тренировочного набора данных создаётся и проверяется модель классификации; 3) полученная модель применяется для классификации новых пользователей.

К сегментации "с учителем" относят [Weinstein, 2004], [Markov, 2007], [Liu, 2008] следующие методы:

- частично или полностью предопределённые ассоциативные правила;
- деревья решений;
- наивный байесовский классификатор;
- факторный, дискриминантный и регрессионный статистический анализ;
- метод опорных векторов (Support Vector Machines);
- автоматический определитель взаимодействия по критерию хи-квадрат (Chi-Square Automatic Interaction Detection);
- многомерное шкалирование;
- совместный анализ (Conjoint Analysis);
- моделирование структурными уравнениями (Structural Equation Modeling).

Ассоциативные правила особенно распространены в рекомендательных системах и позволяют найти Web-страницы, которые вместе посещаются, и группы продуктов, которые вместе покупаются [Liu, 2008]. Они представляют только локальный шаблон для некоторых записей и переменных сегментации, и не

могут считаться всеобщими и рассматриваться в качестве строгой модели [Markov, 2007]. Пример ассоциативного правила: "если время просмотра страницы Default.html небольшое, то длительность сессии тоже будет небольшой с достоверностью 80,825% и подтверждением в 22,989% записей".

Факторный анализ применяется к большому количеству переменных с целью их сокращения до ключевых факторов, что позволит лучше понимать маркетинговую ситуацию [Weinstein, 2004]. В маркетинге применяется две его основных разновидности: 1) *R-факторный анализ*, сокращающий массив данных путём поиска подобий в значениях переменных; 2) *Q-анализ*, ищущий группы подобных потребителей.

Множественный регрессионный анализ полезен для анализа ассоциаций между переменными сегментирования [Малхорта, 2002]. Более точную классификацию даёт *метод опорных векторов* за счёт выявления нелинейных зависимостей между переменными сегментирования [Markov, 2007]. Его недостатком является вычислительная сложность обучающего алгоритма.

Автоматический определитель взаимодействия по критерию хи-квадрат – наиболее общий классификатор. Он категоризирует все независимые непрерывные и дискретные целые переменные по подобию и определяет результирующие категории для переменных в виде целых групп [Weinstein, 2004].

Дискриминантный анализ используется для исследования разницы между сегментами или предсказания возможности членства в группах, например, сравнения поклонников марки с противниками, опытных пользователей с новичками и т.п. Он выполняется с помощью специальных уравнений, называемых *дискриминантными* функциями [Weinstein, 2004], и используется для анализа данных в том случае, когда зависимая переменная категориальная, а независимые переменные – интервальные [Малхорта, 2002].

Многомерное шкалирование – это маркетинговый метод графического представления атрибутов продуктов, основанного на восприятии и предпочтениях пользователей. Целью метода является идентификация рыночных сегментов с близкими потребностями и взглядами относительно продуктов [Weinstein, 2004].

Ещё одним используемым в маркетинге методом является *совместный анализ* (Conjoint Analysis) – основанный на измерений вклада различных атрибутов продукта в принятие решения о покупке. Он моделирует предпочтения или реакции пользователей в терминах набора атрибутов товара. Затем эти предпочтения ранжируются, оцениваются и группируются в однородные сегменты [Weinstein, 2004].

Моделирование структурными уравнениями (Structural Equation Modeling) - это подход к моделированию, раскрывающий связи между множеством наблюдаемых переменных в терминах скрытых переменных [Weinstein, 2004].

Частным случаем сегментации "с учителем" является *сегментация "с подкреплением"* (*reinforcement learning*), когда для каждого прецедента имеется пара «ситуация, принятое решение». К этому виду относят, в частности, *эволюционное моделирование* [Снитюк, 2008].

3.2. Методы сегментации "без учителя"

Для сегментации "без учителя" [Weinstein, 2004], [Markov, 2007], [Mason, 2006], [Liu, 2008] применяют:

- кластерный анализ;

- ассоциативные правила;

- нейронные сети;

- разведовательный анализ данных (Exploratory Data Analysis).

Кластеризация обычно применяется первой во время анализа данных с отсутствующими предопределёнными значениями метрик [Mason, 2006]. При этом переменные не разделяют на зависимые и независимые, и проверяются взаимозависимые связи всего набора переменных [Малхорта, 2002]. Общая цель кластерного анализа: максимизировать подобие членов в пределах каждого кластера и максимизировать разницу между кластерами. Недостатком этого метода является опасность создания статистически правильных, но бессмысленных сегментов в случае неправильных начальных данных.

Нейронные сети - более мощный инструмент анализа, однако его сложнее настраивать и интерпретировать результаты по сравнению с кластерным анализом [Mason, 2006].

Также для предварительной подготовки данных о действиях пользователя на сайте может быть использован *разведывательный анализ данных* [Markov, 2007]. Этот вид статистического анализа позволяет выполнить пробную оценку набора данных, уменьшить его размерность, проверить взаимосвязи между переменными и выявить интересуемые подмножества записей журнала посещений. Результаты анализа отображаются в виде простых графиков и таблиц для поддержки принятия решения о выполнении более глубокого исследования с использованием специальных методов сегментации.

4. Показатели качества сегментации

Не все схемы сегментации полезны с точки зрения маркетинга. Согласно работам [Токарев, 705], [Weinstein, 2004], [Kotler, 2006] в общем случае сегменты должны быть:

- *измеряемыми*;

- *однородными в пределах сегмента* и *разнородными между сегментами*;

- *достаточно большими* и *прибыльными*; для этого они должны полностью охватывать однородную группу пользователей;

- *доступными*, то есть должна быть возможность доступа и обслуживания пользователей сегментов;

- *практичными*, то есть пригодными к использованию эффективными маркетинговыми программами привлечения и обслуживания потребителей;

- *стабильными*, то есть оставаться принципиально различимыми и по разному отвечать на разные комбинации маркетинговых элементов и программ.

5. Этапы сегментации пользователей Web-сайта

В различных областях, использующих сегментацию пользователей - Web usage mining, Web-аналитика и маркетинг - существуют свои методики её выполнения. Так, в маркетинге сегментация потребительских рынков выполняется в три этапа [Токарев, 705]: 1) выбор критериев (переменных) сегментации; 2) выбор метода сегментации; 3) выбор целевых сегментов. Этих шагов недостаточно для сегментации пользователей Web-сайтов. Во-первых, в данном случае при выборе метрик и методов необходимо учитывать область применения результатов сегментации: это может быть повышение эффективности работы сайта, его персонализация, или же уточнение потребительских сегментов. Во-вторых, первичные данные о пользователях, сохраняемые в журналах серверов, непригодны для непосредственного использования и нуждаются в дополнительной обработке. Чтобы их использовать, согласно [Jiawei, 2006], нужно:

1) предварительно их очистить от несущественной информации вроде загрузок изображений или же записей про посещение сайта Web-агентами, сжать и трансформировать в удобный для поиска и анализа важной и полезной информации;

2) из этих данных построить многомерный массив, где в качестве измерений будут использованы URL, время, IP-адреса, информация о содержании посещённых Web-страниц, дополнительные данные о пользователе из журналов Web-приложений, чтобы затем иметь возможность определить характеристики и последовательности действий пользователей, вычислить поведенческие метрики сайта и выполнить сегментацию.

Эти вопросы в значительной степени учтены в методике *Cross-Industry Standard Process for Data Mining* (CISP DM), которая предложена совместно компаниями SPSS, NCR, DaimlerChrysler и OHRA. В контексте данной методики [Markov, 2007] и с учётом необходимости дополнительной обработки первичных данных о пользователе [Liu, 2008], сегментация пользователей может выполняться как итеративная и адаптивная последовательность фаз:

1. Фаза определения бизнеса или постановки исследования.
2. Фаза сбора, анализа и выборки данных.
3. Предварительная очистка, объединение и интеграция данных.
4. Фаза моделирования и сегментации.
5. Фаза оценки результатов.
6. Фаза применения результатов.

Некоторые фазы сегментации могут зависеть от результатов предыдущих фаз. В свою очередь любая фаза может быть повторно выполнена с новыми условиями, если этого будет нужно для удовлетворительного выполнения последующих фаз. Например, в зависимости от поведения и характеристик модели сегментации может появиться необходимость вернуться к фазе подготовки данных для их дополнительной очистки перед фазой оценки результатов.

6. Средства сегментации

Инструменты сегментации можно разделить на предлагаемые в сети WWW и представленные на рынке программного обеспечения. Первые, как правило, призваны автоматизировать работу Web-аналитиков, и предлагают сегментацию для углублённого анализа аудитории сайта. Сфера их использования ограничена, так как у исследователя нет доступа непосредственно к статистическим данным, метрики уже разбиты на предопределённые категории и для сегментации применяются простые методы обучения "с учителем". К таким службам можно отнести бесплатный сервис Google Analytics, и коммерческие Coremetrics, Nedstat и др. Они имеют много общего: большое количество предопределённых сегментов, которые можно дополнить собственными, объединить и, определив граничные условия, использовать для фильтрации данных, собранных с помощью собственного счётчика посещений. Они так же позволяют создавать отчёт о характеристиках отфильтрованных пользователей за некоторый период, сравнивать их характеристики с другими группами посетителей.

При наличии массивов статистических данных о пользователях, можно выполнить сегментацию с помощью представленных на рынке коммерческих средств профессионального статистического анализа вроде Statistica, библиотеки "Web Mining for Clementine" SPSS, или попробовать самостоятельно выполнить расчёты с помощью библиотеки Machine learning framework в среде Mathematica от Wolfram Research, пакета анализа данных MS Excel. Также есть возможность разработать собственную программу сегментации на языке Java, используя специальную открытую коллекцию алгоритмов машинного обучения WEKA.

Выводы

Вместе с продолжающимся ростом и распространением электронной коммерции и других информационных систем, основанных на WWW, накапливаются и коллекции первичных слабоструктурированных данных об историях посещений и характеристиках пользователей, собираемых Web-сайтами во время ежедневных операций. Анализ подобной информации может помочь определить реальную информацию о клиентах, составить маркетинговые стратегии для продуктов и услуг, оценить эффективность компаний по продвижению товаров, повысить продуктивность Web-приложений, персонализировать содержание. Одним из видов такого анализа является сегментация пользователей Web-сайтов.

В статье проанализирован характер данных о пользователях WWW и метрики сайтов в контексте их использования для сегментации посетителей Web-сайтов, предложена классификация метрик,

рассмотрены основные пути и средства сбора и визуализации сведений о пользователях. Дан сравнительный анализ основных методов сегментации "с учителем" и "без учителя", а также средств сегментации различной сложности, в том числе Google Analytics, Web Mining for Clementine. Также рассмотрены основные показатели качества сегментации и предложен адаптированный вариант процесса CISP DM для сегментации.

Благодарности

Работа опубликована при финансовой поддержке проекта **ITHEA XXI** Института информационных теорий и приложений FOI ITHEA Болгария www.ithea.org и Ассоциации создателей и пользователей интеллектуальных систем ADUIS Украина www.aduis.com.ua.

Литература

[Burby, 2007] Burby Jason, Atchison Shane. Actionable Web Analytics Using Data To Make Smart Business Decisions. - Wiley Publishing Inc., 2007.

[Clifton, 2008] Clifton Brian. Advanced Web Metrics with Google Analytics. - Wiley Publishing, Inc., 2008.

[Cutroni, 2008] Cutroni Justin. Google Analytics Compliance with WAA Standard Metrics. - September 21, 2008 – Blog at http://www.epikone.com/blog/2008/09/21/google-analytics-compliance-with-waa-standard-metrics/.

[Jiawei, 2006] Han Jiawei, Kamber Micheline. Data Mining: Concepts and Techniques. Second Edition. - Elsevier Inc., 2006.

[Kotler, 2006] Kotler Philip, Keller Kevin. Marketing management. —Twelfth ed., 2006.

[Liu, 2008] Liu Bing. Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data. – Springer-Verlag Berlin Heidelberg, 2007. – 532p.

[Markov, 2007] Markov Zdravko, Larose Daniel T. Data Mining the Web: Uncovering Patterns in Web Content, Structure, and Usage. - Wiley, 2007.

[Mason, 2006] Mason Neil. Defining and Using Segmentation. Part 1, 2. - The ClickZ Network. – Jun 27, 2006 - Article at http://www.clickz.com/showPage.html?page=3615916.

[Weinstein, 2004] Weinstein Art. Handbook of Market Segmentation. Strategic Targeting for Business and Technology Firms. - The Haworth Press, 2004. - 626p.

[Малхорта, 2002] Малхорта Нэреш К. Маркетинговые исследования. Практическое руководство, 3-е издание.: Пер. с англ. — М.: Издательский дом "Вильяме", 2002. — 960 с.: ил. — Парал. тит. англ.

[Пелещишин, 2007] Пелещишин А.М. Позиціювання сайтів у глобальному інформаційному середовищі: Монографія. - Львів: Видавництво Національного університету "Львівська політехніка", 2007. - 260с.

[Снитюк, 2008] Снитюк В.Є. Прогнозування. Моделі. Методи. Алгоритми. - К.: "Маклаут", 2008. - 367с.

[Токарев, 705] Токарев Б.Е. Маркетинговые исследования: учебник. – М.: Экономистъ, 2007. – 624с.

Authors' Information

Дмитрий Ночевнов *– доцент кафедры информационных технологий проектирования Черкасского государственного технологического университета, бул.Шевченка, 460, г. Черкассы, Украина, 18006; e-mail: dmitry.ndp@gmail.com*

Computing

SAT-BASED METHOD OF VERIFICATION USING LOGARITHMIC ENCODING

Liudmila Cheremisinova, Dmitry Novikov

***Abstract**: The problem under discussion is to check whether a given combinational network realizes a system of incompletely specified Boolean functions. SAT-based procedure is discussed that formulates the overall problem as conventional conjunctive normal form (CNF) on the basis of encoding of multiple-output cubes the Boolean functions are specified on and checking whether the combinational network realizes them using a SAT solver. The novel method is proposed that speeds up the SAT-based procedure due to suggested efficient procedure of logarithmic encoding multiple-output cubes that allows reducing the number of variables to be additionally introduced into the CNF under construction.*

***Keywords**: design automation, verification, simulation, CNF satisfiability.*

***ACM Classification Keywords**: B.6.2 [Logic Design]: Reliability and Testing; G.4 [Mathematical Software]: Verification; B.6.2 [Reliability and Testing]: Error-checking.*

***Conference**: The paper is selected from XVth International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.*

Introduction

The role of combinational verification becomes more and more important with the rapid increase of the complexity of designs synthesized by modern CAD (computer-aided design) tools. Today, verification is a bottleneck in the overall VLSI design cycle as it consumes up to 70% of design effort [Drechsler, 2000], [Kuehlmann, 2002]. The objective of verification is to ensure that implemented and specified behaviors are the same. In a typical scenario, there are two structurally similar circuit implementations of the same design, and the problem is to prove their functional equivalence. In contrast to that, in the paper the verification task is examined for the case, when desired functionality of the system under design is incompletely specified. Such a case usually occurs on early stages of designing when assignments to primary inputs of designed device exist which will never arise during a normal mode of the device usage. Thus when hardware implementing this device, its outputs in response of these inputs may be arbitrary specified. In this case verification methodologies must consider only possible input scenarios to the design under verification and verify that every possible output signal of the implemented behavior has its intended (described in initial specification) value.

We consider the verification problem for the case, when 1) desired incompletely specified functionality is given in the form of a system of incompletely specified Boolean functions (ISF system); 2) functions of the system are specified on intervals (cubes) of values of Boolean input variables and these intervals are large enough; 3) the ISF system is implemented in the form of a combinational circuit. Efficient methods for equivalence checking were proposed [Drechsler, 2000], [Kuehlmann, 2002], [Kropf, 1999] most of them can be organized into two major categories: simulation and SAT based equivalence checking. Both these techniques have in our case some peculiarities following from incompletely specified functionality of one of the compared descriptions.

At present, logic simulation is the most widely used technique for ensuring the correctness of digital integrated circuits in industry because of its scalability and predictable run-time behavior. This technique is based on verifying a digital system by stimulating inputs of the circuit with binary signal values that propagate in the circuit leading to a corresponding activation of its outputs, whose values must be consistent with the expected ones. In our case a special type of simulation could be applied: guided simulation, when inputs are assigned based on certain information, provided by the design specification. This could reduce in some cases the search space of the simulator. However though the specification of the designed circuit with n inputs would be specified with a small number of multi-output cubes, the overall size of Boolean space covered by them can contain up to 2^n combinations of n input variables. That is why simulation is infeasible for state-of-the-art designs.

In a modern combinational equivalence checking flow based on formal verification approach, both circuits to be verified are transformed into a single circuit called a miter derived by combining the pairs of inputs with the same names and feeding the pairs of outputs with the same names into EXOR gates, which are ORed to produce the single output of the miter. The miter is a combinational circuit with the same inputs as the original circuit and there is constant 0 on its output if and only if the two original circuits produce identical output values under all possible input assignments. To check whether it takes place usually SAT solvers are used requiring their input to be in conjunctive normal form (CNF). This type of solvers can be applied to circuits by converting them into CNF form. Majority of SAT applications derived from circuit representation rely on some version of Tseitin transformation [Tseitin, 1983] for producing CNF of the circuit called as conventional CNF. A circuit-to-CNF conversion has a linear complexity and introduces as many variables as there are primary inputs and gates in the circuit.

In our case when we have a circuit and a system of incompletely specified Boolean functions we cannot construct a miter. The key idea proposed is how to organize testing of all multiple-output cubes whether they are realized by the given circuit. We present a novel SAT based verification method of testing if the given circuit implements all the multiple-output cubes representing the system of incompletely specified Boolean functions based on encoding its multiple-output cubes. The proposed method speeds up the SAT-based procedure due to suggested efficient procedure of logarithmic encoding multiple-output cubes that allows reducing the number of variables to be additionally introduced into the CNF under construction.

Basic definitions

A system $F(X) = \{ f_1(X), f_2(X), \ldots, f_m(X) \}$ (where $X = (x_1, x_2, \ldots, x_n)$) of incompletely specified Boolean functions is represented as mapping of n-dimensional Boolean space E^n into m-dimensional space $\{0,1,–\}^m$, i.e. $F(X)$: $E^n \rightarrow \{0,1,–\}^m$, where the symbol "–" denotes don't-care condition, such a function is not defined over the whole Boolean space E^n. In the case of ISF don't-care points in E^n may differ for different functions. A completely specified Boolean function $f(X)$ realizes an ISF $g(X)$ iff $f(X)$ can be derived from $g(X)$ by assigning either 0 or 1 to each don't-care point of E^n.

An ISF is specified by off-set, on-set and dc-set as subsets of E^n, i.e. sets U_f^0, U_f^1 and U_f^{ds} of cubes ($U_f^1 \cup U_f^0 \cup U_f^{ds} = E^n$). Let us specify a system $F(X)$ of ISFs as a set S_F of multiple-output cubes. A multiple-output cube $(\boldsymbol{u}, \boldsymbol{t}) \in S_F$ is a pair of row ternary vectors $\boldsymbol{u}$ and $\boldsymbol{t}$ (or conjunctions) of dimensions n and m that are called further as its input and output parts. The input part $\boldsymbol{u}$ represents a cube in E^n or a conjunction of some literals (variables $x_i \in X$ or its inversions). The output part $\boldsymbol{t}$ is a ternary vector of values of functions for the cube $\boldsymbol{u}$ or a conjunction of some literals $f_j \in F$. For each $f_j \in F$ the j-th entry t^j of $\boldsymbol{t}$ is 1 or 0 ($t^j = 1, 0$) if all the minterms of the cube $\boldsymbol{u}$ are in the on-set U_{fj}^1 or in the off-set U_{fj}^0 correspondingly; otherwise t^j is don't-care. Representation of a function with multiple-output cubes has the following distinctive feature: cubes $\boldsymbol{u}_i$ and $\boldsymbol{u}_j$ can intersect each other and don't-care value of an element t_i^j of the output part of $(\boldsymbol{u}_i, \boldsymbol{t}_i)$ means that either the function f_j is don't-care on the whole cube $\boldsymbol{u}_i$ or f_j does not take the same definite value on the whole interval $\boldsymbol{u}_i$, i.e. there exist at least two minterms covered by the cube $\boldsymbol{u}_i$ on which f_j has different definite values.

The system $F(X)$ of ISFs given by the set S_F of multiple-output cubes $(\boldsymbol{u}_i, \boldsymbol{t}_i)$ can be represented in matrix form by a pair of ternary matrices $\boldsymbol{U}$ and $\boldsymbol{T}$ of the same cardinalities. The matrix $\boldsymbol{U}$ contains input parts of multiple-output

cubes from S_F as its rows; similarly matrix $\boldsymbol{T}$ specifies output parts as its rows. For example, ISF system $F(X)$ specified by S_F= {($x_3 x_4 x_5$, f_1), ($x_1 x_2$, $f_1 \bar{f}_2$), ($\bar{x}_2 \bar{x}_3 \bar{x}_4$, $\bar{f}_1 f_2$), ($\bar{x}_2 x_3 \bar{x}_4$, $\bar{f}_2$), ($\bar{x}_2 x_4 \bar{x}_5$, f_2)} is represented as follows:

$$\boldsymbol{U} = \begin{array}{ccccc} x_1 & x_2 & x_3 & x_4 & x_5 \\ - & - & 1 & 1 & 1 \\ 1 & 1 & - & - & - \\ 0 & 0 & 0 & - & \\ - & 0 & 1 & 0 & - \\ - & 0 & - & 1 & 0 \end{array} \quad \boldsymbol{T} = \begin{array}{ccc} f_1 & f_2 & \\ 1 & - & 1 \\ 1 & 0 & 2 \\ 0 & 1 & 3 \\ - & 0 & 4 \\ - & 1 & 5 \end{array} \qquad (1)$$

A CNF represents a completely specified Boolean function as conjunction of one or more clauses, each being in its turn a disjunction of literals. CNF representation is popular among SAT algorithms because each clause must be satisfied (evaluated to 1) for the overall CNF to be satisfied. The SAT problem is concerned with finding a truth assignment of CNF literals, which simultaneously satisfies each of its member clauses. If such an assignment exists the CNF is referred to as satisfiable, and the assignment is known as a satisfying assignment.

Matrix representation of CNF formula C containing k clauses and p distinct variables is a ternary matrix $\boldsymbol{C}$ having k rows and p columns. The entry c_i^j of the matrix in the i-th row and the j-th column is 1, 0 or "–" depending on in what a form (x_j or $\bar{x}_j$) the variable x_j appears or does not appear in the i-th clause of C.

SAT-based verification of logical descriptions with functional indeterminacy

The conventional CNF of a combinational network specifies all combinations of signal values of its terminals that can take place when it functions. The procedure of derivation of conventional CNF is known, it associates a CNF formula with each network gate that captures the consistent assignments between gate inputs and its output. All such gate local CNFs are joined then in the overall network conventional CNF by using the conjunction operation. CNF for a gate representing a local function $y = f(z_1, z_2, \ldots, z_k)$ is based on defining and representing in a CNF form a new Boolean function $\varphi(y, f) = y \sim f(z_1, z_2, \ldots, z_k)$ [Kunz, 2002] that is true in the only case when both functions y and $f(z_1, z_2, \ldots, z_k)$ assume the same value.

When we have conventional CNF C of a combinational network and a system $F(X)$ of ISFs, the arguments and the functions of which correspond to primary inputs and outputs of the network, a problem under discussion is to check whether a given network implements the ISF system. It is true if it takes place for each multiple-output cube of the system. In terms of the network CNF this condition could be reformulated as follows. For every multiple-output cube $(\boldsymbol{u}_i, \boldsymbol{t}_i) \in S_F$ a value assignment satisfying the conjunction $\boldsymbol{u}_i \wedge \boldsymbol{t}_i$ must satisfy the network CNF. Or, this condition could be reformulated in the form more suitable for the task of SAT solving: a network realizes an ISF system $F(X)$ iff for every multiple-output cube $(\boldsymbol{u}_i, \boldsymbol{t}_i) \in S_F$ a value assignment contradicting to it and satisfying the conjunction $\boldsymbol{u}_i \wedge \bar{\boldsymbol{t}}_i$ is unsatisfying assignment for a network CNF. If $\boldsymbol{u}_i = x_1^i x_2^i \ldots x_{ni}^i$ and $\boldsymbol{t}_i = f_1^i f_2^i \ldots f_{mi}^i$ then the CNF P_i specifying the contradiction of the multiple-output cube $(\boldsymbol{u}_i, \boldsymbol{t}_i)$, called as the cube-prohibitive CNF, consists of the following $n_i + 1$ clauses regarding the set $X \cup F$ of variables:

$$P_i(X, F) = x_1^i x_2^i \ldots x_{ni}^i (\bar{f}_1^i \vee \bar{f}_2^i \vee \ldots \vee \bar{f}_{mi}^i).$$

For example, the cube-prohibitive CNF for the multiple-output cube $s_2 = (x_1 x_2, f_1 \bar{f}_2)$ of the mentioned ISF system $F(X)$ (1) contains three clauses: $P_2(X, F) = (x_1)(x_2)(\bar{f}_1 \vee f_2)$. And all cube-prohibitive CNFs for the ISF system are shown in the third column of Table 1.

Appending clauses of P_i to the network conventional CNF C results in CNF $C_{Pi} = C \wedge P_i$. It is not difficult to prove that CNF C_{Pi} is satisfiable iff the network does not realize the i-th multiple-output cube. As to the whole ISF system it is not realized by the network iff it does not implement at least one of its multiple-output cubes, i.e. if the following formula is satisfiable:

$$C_P = C \wedge P = C \wedge (P_1 \vee P_2 \vee \ldots \vee P_l), \qquad (2).$$

where P is the ISF system prohibitive formula that, in general, is not in the form of CNF.

Table 1

Encoding cube-prohibitive CNFs $P_i^k(X, F, W)$of the ISF system (1)

№	$P_i(X, F)$	cube-prohibitive CNFs $P_i(X, F)$										unary codes					logarithmic codes			logarithmic codes		
		x_1	x_2	x_3	x_4	x_5	z_1	z_2	z_3	f_1	f_2	w_1	w_2	w_3	w_4	w_5	w_1	w_2	w_3	w_1	w_2	w_3
1.	P_1	–	–	1	–	–	–	–	–	–	–	0	–	–	–	–	0	0	0	0	0	0
2.		–	–	–	1	–	–	–	–	–	–	0	–	–	–	–	0	0	0	0	0	0
3.		–	–	–	–	1	–	–	–	–	–	0	–	–	–	–	0	0	0	0	0	0
4.		–	–	–	–	–	–	–	–	0	–	0	–	–	–	–	0	0	0	0	0	0
5.	P_2	1	–	–	–	–	–	–	–	–	–	–	0	–	–	–	0	0	1	–	0	1
6.		–	1	–	–	–	–	–	–	–	–	–	0	–	–	–	0	0	1	–	0	1
7.		–	–	–	–	–	–	–	–	0	1	–	0	–	–	–	0	0	1	–	0	1
8.	P_3	–	0	–	–	–	–	–	–	–	–	–	–	0	–	–	0	1	0	–	1	0
9.		–	–	0	–	–	–	–	–	–	–	–	–	0	–	–	0	1	0	–	1	0
10.		–	–	–	0	–	–	–	–	–	–	–	–	0	–	–	0	1	0	–	1	0
11.		–	–	–	–	–	–	–	–	1	0	–	–	0	–	–	0	1	0	–	1	0
12.	P_4	–	0	–	–	–	–	–	–	–	–	–	–	–	0	–	0	1	1	–	1	1
13.		–	–	1	–	–	–	–	–	–	–	–	–	–	0	–	0	1	1	–	1	1
14.		–	–	–	0	–	–	–	–	–	–	–	–	–	0	–	0	1	1	–	1	1
15.		–	–	–	–	–	–	–	–	–	1	–	–	–	0	–	0	1	1	–	1	1
16.	P_5	–	1	–	–	–	–	–	–	–	–	–	–	–	–	0	1	0	0	1	–	–
17.		–	–	–	1	–	–	–	–	–	–	–	–	–	–	0	1	0	0	1	–	–
18.		–	–	–	–	1	–	–	–	–	–	–	–	–	–	0	1	0	0	1	–	–
19.		–	–	–	–	–	–	–	–	–	0	–	–	–	–	0	1	0	0	1	–	–
20.	$Q(W)$											1	1	1	1	1	1	1	–			
21.																	1	–	1			

Encoding of cube-prohibitive CNFs

Converting the formula $P = P_1 \vee P_2 \vee \ldots \vee P_l$ (from 2) into a CNF form could be done always, but that can be hard problem. We propose the method of construction of ISF system prohibitive CNF P that has linear complexity. It is based on encoding multiple-output cubes and their prohibitive CNFs using some coding variables $w_i \in W$. After encoding, cube-prohibitive CNFs $P_i(X, F)$ are transformed into encoded cube-prohibitive CNFs $P_i^k(X, F, W)$ and the formula (2) into

$$C^k{}_P = C \wedge (P_1{}^k \wedge P_2{}^k \wedge \ldots \wedge P_l{}^k) \wedge Q(W), \tag{3}$$

where $Q(W)$ provides that the CNF $C^k{}_P$ will be satisfiable iff at least one CNF $C_{Pi} = C \wedge P_i$ is satisfiable. From now on $Q(W)$ is called as alternative CNF.

When transforming the formula (2) to the CNF form (3) each cube-prohibitive CNF P_i is encoded by a code in the form of a clause $d_i = w_{i1}{}^{\sigma i1} \vee w_{i2}{}^{\sigma i2} \vee \ldots \vee w_{ir}{}^{\sigma ir}$ ($\sigma_{ir} \in \{0,1\}$ and $w_{ir}{}^1 = w_{ir}$, $w_{ir}{}^0 = \bar{w}_{ir}$) where $w_{ij} \in W$. As $P_i(X, F) = x_1{}^i x_2{}^i \ldots x^i_{ni} (\bar{f}_1{}^i \vee \bar{f}_2{}^i \vee \ldots \vee \bar{f}_{mi}{}^i)$ then

$$P_i^k(X, F, W) = (x_1{}^i \vee d_i) \ldots (x^i_{ni} \vee d_i) (f_1{}^i \vee \ldots \vee f_{mi}{}^i \vee d_i).$$

All used codes should differ from each other. To formulate the conditions the alternative CNF $Q(W)$ in (3) must satisfy for the chosen encoding of cube-prohibitive CNFs, let us denote by f_Q and f_{di} functions defined by $Q(W)$ and $d_i(W)$ and by $U_Q{}^1$ and $U_{di}{}^1$ – their on-sets that are subsets of the Boolean space E^r of all different minterms of Boolean variables w_i.

Assertion 1. Codes $d_1(W)$, $d_2(W)$, …, $d_l(W)$ of cube-prohibitive CNFs and the appropriate alternative CNF $Q(W)$ must satisfy the following conditions:

1) $(\bigwedge_i f_{di}) \wedge f_Q = 0$ or $(\bigcap_i M_{di}^1) \cap M_Q^1 = \varnothing$;

2) $(\bigwedge_{i \neq j} f_{di}) \wedge f_Q \neq 0$ or $(\bigcap_{i \neq j} M_{di}^1) \cap M_Q^1 \neq \varnothing$ for all $j \in \{1,2,\ldots,r\}$.

The first condition ensures the CNF $P(X, F, W) = (P_1^k \wedge P_2^k \wedge \ldots \wedge P_l^k) \wedge Q$ (the part of the CNF C^k_P) be unsatisfiable when the analyzed ISF system is realized by the network, i.e. when all cube-prohibitive CNFs $P_i(X, F)$ (not coded) are unsatisfiable with respect to the network CNF C. If the first condition takes place then there exists no value assignment of variables of the set W that can ensure all P_i^k and $Q(W)$ be true. The second condition ensures the CNF $P(X, F, W)$ be satisfiable when the analyzed ISF system is not realized by the circuit, i.e. there exists at least one multiple-output cube, for example the j-th one, that is not realized by it. Thus, a variable value assignment can be found satisfying the j-th cube-prohibitive CNF $P_j(X, F)$ (and thus $P_j^k(X, F, W)$ too). The fulfillment of the second condition guaranties that there exists at least one assignment of coding variables that ensures satisfiability of $Q(W)$ and all cube prohibitive CNFs P_i^k except the j-th one (that is satisfiable by the assumption).

Encoding by codes of unit length

The method of unary encoding has been proposed in [Cheremisinova, 2008], that introduces as many coding variables w_i as there exist multiple-output cubes in the ISF system specification. According to the method the following encoded cube-prohibitive CNF is generated for each multiple-output cube $(\boldsymbol{u}_i, \boldsymbol{t}_i) = (x_1^i x_2^i \ldots x_{ni}^i, f_1^i f_2^i \ldots f_{mi}^i)$:

$$P_i^k(X, F, W) = (x_1^i \vee w_i)(x_2^i \vee w_i) \ldots (x_{ni}^i \vee w_i)(\bar{f}_1^i \vee \ldots \vee \bar{f}_{mi}^i \vee w_i),$$

and the following alternative CNF Q satisfying the Assertion 1 is (see the forth column of Table 1):

$$Q(W) = \bar{w}_1 \vee \bar{w}_2 \vee \ldots \vee \bar{w}_l.$$

Such an alternative CNF Q satisfies the conditions of the Assertion 1 proposed above:

$$f_{di} = w_i,\ f_Q = \bar{w}_1 \vee \bar{w}_2 \vee \cdots \vee \bar{w}_l,\ \bigwedge_i f_{di} = w_1 w_2 \ldots w_l, \text{ so } (\bigwedge_i f_{di}) \wedge f_Q = 0;$$

$$\bigwedge_{i \neq j} f_{di} = w_1 w_2 \ldots w_{j-1} w_{j+1} \ldots w_l, \text{ so } (\bigwedge_{i \neq j} f_{di}) \wedge f_Q = w_1 w_2 \ldots w_{j-1} \bar{w}_j w_{j+1} \ldots w_l \neq 0.$$

In [Cheremisinova, 2008] it has been proven that an ISF system is realized by the network with functional description specified by CNF C iff the following CNF is unsatisfiable:

$$C \wedge (P_1^\kappa \wedge P_2^\kappa \wedge \ldots \wedge P_l^\kappa) \wedge (w_1 \vee w_2 \vee \ldots \vee w_l).$$

Encoding by codes of logarithmic length

The unary encoding of cube-prohibitive CNFs introduces too many additional variables into the CNF that is the input for a SAT solver. The considerable reduction of the number of coding variables and accordingly the reduction of the complexity of resulting CNF can be achieved when to encode cube-prohibitive CNFs with codes of logarithmic length. In that case the minimal number of coding variables could be introduced, it is $r = \lceil log_2 l \rceil$, where l is the number of multiple-output cubes. In this method cube-prohibitive CNFs are encoded with different elementary dizjunctions $d_i = w_{i1}^{\sigma_{i1}} \vee w_{i2}^{\sigma_{i2}} \vee \ldots \vee w_{ir}^{\sigma_{ir}}$ of size r. Accordingly, the following encoded cube-prohibitive CNF is generated for each multiple-output cube $(\boldsymbol{u}_i, \boldsymbol{t}_i) = (x_1^i x_2^i \ldots x_{ni}^i, f_1^i f_2^i \ldots f_{mi}^i)$:

$$P_i^k = (x_1^i \vee d_i)(x_2^i \vee d_i) \ldots (x_{ni}^i \vee d_i)(\bar{f}_1^i \vee \ldots \vee \bar{f}_{mi}^i \vee d_i),$$

An alternative CNF $Q(W)$ for the given encoding of cube-prohibitive CNFs could be found on the grounds of the following assertion, that follows from the Assertion 1.

Assertion 2. If cube-prohibitive CNFs are encoded with codes $d_1(W)$, $d_2(W)$, …, $d_l(W)$ then the appropriate alternative CNF $Q(W)$ is specified by the function $f_Q = \overline{\bigwedge_i f_{di}}$, where f_{di} is the function defined by $d_i(W)$.

Indeed, two conditions mentioned in the Assertion 1 are satisfied:

1) $(\bigwedge_i f_{di}) \wedge (\overline{\bigwedge_i f_{di}}) = 0$;

2) $(\bigwedge_{i\neq j} f_{di}) \wedge \overline{\bigwedge_i f_{di}} = (f_{d1} \dots f_{dj-1} f_{dj+1} \dots f_{dl}))\ (\bar{f}_{d1} \vee \dots \vee \bar{f}_{dj-1} \vee \bar{f}_{dj} \vee \bar{f}_{dj+1} \vee \dots \vee \bar{f}_{dl}) = f_{d1} \dots f_{dj-1}\ \bar{f}_{dj}\ f_{dj+1} \dots f_{dl} \neq 0.$

In conformity with the Assertion 2 the alternative CNF $Q(W)$ will be:

$$\begin{aligned} Q(W) &= \neg((w_{11}^{\sigma 11} \vee w_{12}^{\sigma 12} \vee \dots \vee w_{1r}^{\sigma 1r}) \wedge \dots \wedge (w_{l1}^{\sigma l1} \vee w_{l2}^{\sigma l2} \vee \dots \vee w_{lr}^{\sigma lr})) = \\ &= \neg(w_{11}^{\sigma 11} \vee w_{12}^{\sigma 12} \vee \dots \vee w_{1r}^{\sigma 1r}) \vee \dots \vee \neg(w_{l1}^{\sigma l1} \vee w_{l2}^{\sigma l2} \vee \dots \vee w_{lr}^{\sigma lr}) = \\ &= \bar{w}_{11}^{\sigma 11}\ \bar{w}_{12}^{\sigma 12} \dots\ \bar{w}_{1r}^{\sigma 1r} \vee \dots \vee\ \bar{w}_{l1}^{\sigma l1}\ \bar{w}_{l2}^{\sigma l2} \vee \dots \vee\ \bar{w}_{lr}^{\sigma lr}. \end{aligned} \quad (4)$$

Thus alternative CNF $Q(W)$ will consists of $2^l - l$ clauses that correspond to those combinations of variable w_i values that are not used for encoding of cube-prohibitive CNFs. For example, when using codes

$$\{(\bar{w}_1 \vee \bar{w}_2 \vee \bar{w}_3), (\bar{w}_1 \vee \bar{w}_2 \vee w_3), (\bar{w}_1 \vee w_2 \vee \bar{w}_3), (\bar{w}_1 \vee w_2 \vee w_3), (w_1 \vee \bar{w}_2 \vee \bar{w}_3)\}$$

for encoding five cube-prohibitive CNFs there are three free codes

$$\{(w_1 \vee \bar{w}_2 \vee w_3), (w_1 \vee w_2 \vee \bar{w}_3), (w_1 \vee w_2 \vee w_3)\}$$

that generate alternative CNF $Q(W) = (w_1 \vee w_2)(w_1 \vee w_3)$ (the fifth column of Table 1).

To simplify the alternative CNF $Q(W)$, it to have less clauses and literals, the task arises to choose such $2^l - l$ clauses that would be free from encoding of cube-prohibitive CNFs and could generate simpler CNF $Q(W)$ (if to apply the identical transformations to them).

The method of encoding by codes of logarithmic length

The idea of the proposed method of logarithmic encoding of cube-prohibitive CNFs is to take constant 1 as an alternative CNF $Q(W)$. So we should to select such codes for cube-prohibitive CNFs that 1) they satisfy the conditions of the Assertion 1 and 2) the on-sets of the functions f_{di} corresponding to them cover the whole Boolean space E^r:

1) $\bigwedge_i f_{di} = 0$ or $\bigcap_i M_{di}^1 = \varnothing$;

2) $\bigwedge_{i\neq j} f_{di} \neq 0$ or $\bigcap_{i\neq j} M_{di}^1 \neq \varnothing$ for all $j \in \{1,2,\dots,r\}$;

3) $\bigvee_i f_{di} = 1$ or $\bigcup_i M_{di}^1 = E^r$.

Let us call a clause containing the literals corresponding to all variables from W as a complete one, its size equals $r = |W|$. The function $f_{di}(W)$ specified by a complete clause $d_i(W) = w_{i1}^{\sigma i1} \vee w_{i2}^{\sigma i2} \vee \dots \vee w_{ir}^{\sigma ir}$ takes the value 0 on the only minterm that is $w_{i1}^{\bar{\sigma}i1} w_{i2}^{\bar{\sigma}i2} \dots w_{ir}^{\bar{\sigma}ir}$. And the function $f_{di}(W)$ specified by an arbitrary clause $d_i(W) = w_{i1}^{\sigma i1} \vee w_{i2}^{\sigma i2} \vee \dots \vee w_{ip}^{\sigma ip}$, $p < r$, takes value 0 on 2^{r-p} minterms. Let us denote the set of these minterms as $I(d_i)$.

Assertion 3. Clauses $d_1(W)$, $d_2(W)$, …, $d_l(W)$ may be chosen as codes of cube-prohibitive CNFs, allowing an alternative CNF $Q(W)$ =1, if they satisfy the following conditions:

1) $\bigcup_i I(d_i) = E^r$;

2) each $I(d_i)$ contains at least one minterm (concerned with the clause d_i) which enters into no other set $I(d_j)$.

These conditions follow from three ones sited above. From the Assertion 3 follows that any *i*-th minterm concerned with one of *I* codes of cube-prohibitive CNFs belongs to the only set $I(d_i)$ but each of $2^r - I$ free minterms belongs to one or more sets $I(d_i)$.

The following method of encoding *I* cube-prohibitive CNFs by codes of logarithmic length results from the Assertion 3.

1. Find $r = \lceil log_2 I \rceil$.

2. Take *I* different minterms from E^r as ones concerned with codes, they are accepted as initial codes.

3. Form the set *M* of free minterms of E^r that are concerned with no code.

4. For each initial code $d_i(W)$ having initially the size *r* find the maximal interval in the set $I(d_i) \cup M$. That could be done using Quine–McCluskey method [Zakrevskij, 2008]. The clause corresponding to the found maximal interval will be the code of the *i*-th cube-prohibitive CNF satisfying the second condition of the Assertion 3.

Let us consider the above example of ISF system consisting of five multiple-output cubes. Here three variants of encoding cube-prohibitive CNFs are shown (minterms concerned with codes are on the left from them, free minterms are below the set of concerned minterms):

	w_1 w_2 w_3	w_1 w_2 w_3	w_1 w_2 w_3	w_1 w_2 w_3	w_1 w_2 w_3	w_1 w_2 w_3
P_1	0 0 0	0 0 0	0 1 0	0 – 0	0 0 0	0 0 –
P_2	0 0 1	– 0 1	0 1 1	0 – 1	0 1 0	0 1 –
P_3	0 1 0	– 1 0	1 0 0	– 0 0	1 0 0	1 0 –
P_4	0 1 1	– 1 1	1 0 1	– 0 1	1 1 0	1 1 0
P_5	1 0 0	1 – –	1 1 0	1 1 –	1 1 1	– – 1
	1 0 1		0 0 0		0 0 1	
	1 1 0		0 0 1		0 1 1	
	1 1 1		1 1 1		1 0 1	

The first encoding variant is shown in the sixth column of Table 1 and the functions $f_{di}(W)$ realized by the clauses $d_i(W)$ corresponding to the codes of this variant are specified by the following truth table:

w_1	w_2	w_3	f_{d1}	f_{d2}	f_{d3}	f_{d4}	f_{d5}
0	0	0	1	1	1	0	0
0	0	1	1	1	0	1	0
0	1	0	1	0	1	1	0
0	1	1	1	1	1	1	0
1	0	0	1	1	1	0	1
1	0	1	1	1	0	1	1
1	1	0	1	0	1	1	1
1	1	1	0	1	1	1	1

One could compare cods shown in the fifth and the sixth columns of Table 1 that are received by simple logarithmic encoding and by the proposed improved method. It should note too that it is better to encode cube-prohibitive CNFs with more clauses by codes with less definite components.

The optimality criterion when searching for the encoding of cube-prohibitive CNFs is the total number of literals of all codes. It should be noted that from the point of view of the complexity of the encoding searched for, in some cases it is not indifferent which of the minterms will be concerned with some code and which are free. That can influence on the encoding complexity. The matter is each minterm is adjacent to as many minterms as its size is, and it can be merged only with each of them. Thus the upper bound of the number of don't cares in codes is equal to the sum of sizes of free minterms minus the number of pairs of adjacent free minterms. So it is desirable to choose $2^r - I$ free minterms to minimize the number of adjacent among them.

From the procedure of constructing codes of cube-prohibitive CNFs and the Assertion 3 follows that the proposed method can ensure finding optimal codes of logarithmic length. The more is the number of free minterms the less is the size of constructed codes. In the worth case when we have 2^r multiple-output cubes (and correspondingly 2^r cube-prohibitive CNFs) the number of free minterms equals to 0 and the codes are of size *r*.

Results of computer experiments

Two encoding methods have been implemented on C++ programming language: one based on unary encoding and the other based on the proposed here logarithmic encoding. Two similar programs of verification using these methods of encoding cube-prohibitive CNFs were compared on the same set of pseudo-random pairs of descriptions: ISF system and combinational network implementing it. MiniSat solver [MiniSat] has been used to check whether the CNF $C^{\kappa}_{P} = C \wedge (P_1^k \wedge P_2^k \wedge \ldots \wedge P_l^k)$ is satisfiable.

The experiments have shown that the considerable reduction of variables (when using logarithmic encoding) did not bring about substantial speedup of the solution of SAT problem. This fact could take place because of two antagonistic impacts on the efficiency of SAT-solver. On one hand, the reduction of variables should to speed up SAT-solver. For example, if verified ISF system consists of l = 32000 multiple-output cubes then it is necessary to introduce 15 coding variables using logarithmic encoding compared with 32000 using unary encoding. But on the other hand, the increment of the sizes of clauses in the case of logarithmic encoding (in comparison with the unary encoding) complicates SAT-solving.

Acknowledgements

The paper is published with financial support by the project ITHEA XXI of the Institute of Information Theories and Applications FOI ITHEA Bulgaria www.ithea.org and the Association of Developers and Users of Intelligent Systems ADUIS Ukraine www.aduis.com.ua.

Bibliography

[Cheremisinova, 2008] L. Cheremisinova, D. Novikov. SAT-Based Approach to Verification of Logical Descriptions with Functional Indeterminacy. In: Proc of 8th International Workshop on Boolean Problems, Freiberg: September 18–19, 2008, pp. 59–66.

[Drechsler, 2000] R. Drechsler. Formal Verication of Circuits. Kluwer Academic Publishers, 2000.

[Kropf, 1999] T. Kropf. Introduction to Formal Hardware Verification. Springer, 1999.

[Kuehlmann, 2002] A. Kuehlmann, A.J. van Eijk Cornelis: Combinational and Sequential Equivalence Checking. In: Logic synthesis and Verification. Ed. S.Hassoun, T.Sasao and R.K.Brayton. Kluwer Academic Publishers, 2002, pp. 343–372.

[Kunz, 2002] W. Kunz, J. Marques-Silva, S. Malik. SAT and ATPG: Algorithms for Boolean Decision Problems. In: Logic synthesis and Verification. Ed. S.Hassoun, T.Sasao and R.K.Brayton. Kluwer Academic Publishers, 2002, pp. 309–341.

[MiniSat] The MiniSat Page / http://minisat.se/MiniSat.html.

[Tseitin, 1983] G.C. Tseitin. On the Complexity of Derivation in Propositional Calculus. In: Studies in Constructive Mathematics and Mathematical Logic, part 2, 1968, pp. 115–125. Reprinted in J.Siekmann and G.Wrightson, eds., Automation of Reasoning, vol. 2, 1983, Springer-Verlag, pp. 466–483.

[Zakrevskij, 2008] A. Zakrevskij, Yu. Pottosin, L. Cheremisinova. Optimization in Boolean space. TUT Press, 2008, 241 p

Authors' Information

Liudmila Cheremisinova *– Principal Researcher, The United Institute of Informatics Problems of National Academy of Sciences of Belarus, Surganov str., 6, Minsk, 220012, Belarus, e-mail: cld@newman.bas-net.by*

Dmitry *Novikov – Post graduate student, The United Institute of Informatics Problems of National Academy of Sciences of Belarus, Surganov str., 6, Minsk, 220012, Belarus, e-mail: yakov_nov@tut.by*

MULTICRITERION PROBLEMS ON THE COMBINATORIAL SET OF POLYARRANGEMENTS

Natalia Semenova, Lyudmyla Kolechkina

***Abstract**: The multicriterion problem of discrete optimization on the feasible combinatorial set of polyarrangements is examined. Structural properties of feasible region and different types of efficient decisions are explored. On the basis of development of ideas of Euclidean combinatorial optimization and method of general criterion possible approaches for the solution of multicriterion combinatorial problem on the set of polyrrangements is developed and substantiated.*

***Keywords**: multicriterion optimization, discrete optimization, polyarrangements, Pareto-optimal solution, weakly and strongly efficient solutions, combinatorial set of polyarrangements.*

***ACM Classification Keywords**: G 2.1 Combinatorics (F2.2), G 1.6 Optimization*

***Conference**: The paper is selected from XVth International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.*

Introduction

The multicriteria problems of optimization on different sets continue to attract attentions of many researchers [1-10]. The models of discrete combinatorial optimization are widely used at the decision of important problems of the geometrical planning, economy, placing of objects, control process of treatment of information, acceptance of decisions and others. Lately in the area of research of different classes of combinatorial models, developments of new methods of their decision great attention is paind to the methods which are based on the use of structural properties of combinatorial sets [2, 8-15].

A new and actual problem which unites multicriterion of alternatives and feasible sets of decisions having different combinatorial characteristics is formulated and is explored in this work.

It is of common knowledge, most combinatorial optimization problems can be taken to the problems of the integer programming, but it is not always justified, as an opportunity account of combinatorial properties of problems [2] is lost here.

The systematic study of properties of Euclidean combinatorial sets and their research is described in many works Along with well - known Euclidean combinatorial sets of transpositions, placing, combinations, breaking up more complicated structures are polycombsnatorial sets are selected. Interest to such sets is conditioned by the different applied problems, as their certain number is well described by polycombinatorial constructions [12, 14].

t should be noted that problems of Euclidean combinatorial optimization on the polycombinatorial sets are combined with combinatorial polyhedrons and their properties which are the protuberant shells of such sets. The promoted interest to combinatorial and polycombinatorial configurations is conditioned by researches of the last years in the area of computer technologies at creation of modern algorithms and programs for solution of optimization problems. So polycombinatorial criteria is fated by the necessities of practice. The paper continues the studies of multicriterion problems over combinatorial and polycombinatorial sets presented in [8, 9]. The interrelation established between multicriterion problems over combinatorial sets and optimization problems over a continuous feasible set is used to study some structural properties of the feasible domain and to formulate and prove a number of theorems on the optimality conditions for different types of efficient solutions of the problems considered. We propose a polyhedral approach to solving vector combinatorial problems over a set of polyarrangements. It is based on the methods of principal criterion, cutting planes, and relaxation.

1.Problem statement. Basic definitions

The multicriterion problems are examined:

$$Z(\Phi, P_{qk}^{ns}(A,H)): \quad \max\left\{\Phi(a) \mid a \in P_{qk}^{ns}(A,H)\right\},$$

consisting of maximization of vectori criterion $\Phi(a)$ on the Euclidean combinatorial set of polyarrangements, where $\Phi(a)$ $\Phi_i : R^n \to R^1, i \in N_l$.

For statement of material we use the concept of multiset A, which is determined by foundation $S(A)$ and multiple of elements $k(a)$.

Let a multiset $A = \{a_1, a_2, ..., a_q\}$, its base $S(A) = \{e_1, e_2, ..., e_k\}$, where $e_j \in R^1 \ \forall j \in N_n$, and the multiplicity of elements $k(e_j) = r_j, j \in N_k$, $r_1 + r_2 + ... + r_k = q$, be given.

Take an arbitrary $n \in N_q$. Call the ordered n-selection from the multiset A as the set

$$a = \left(a_{i_1}, a_{i_2}, ..., a_{i_n}\right), \tag{1}$$

where $a_{i_j} \in A \ \ \forall i_j \in N_n, \ \forall j \in N_n, \ i_s \neq i_t$, if $s \neq t \ \ \forall s \in N_n$.

Let $P_{qk}^{n}(A)$ be a general combinatorial set of n-arrangements, induced by q>n elements from the multiset A, *k* its elements being different. Denote by $E(A)$ the image of the set $P_{qk}^{n}(A)$ mapped into R^n. Any point $x \in E(A)$ is such that its coordinates take different values from the multiset A of real numbers, i.e., $x = \left(x_1, x_2, ..., x_n\right)$, where $x_j = a_{i_j} \ \ a_{i_j} \in A \ \forall i, j \in N_n$.

Let us represent a set N_q as an ordered partition into *s* (where *s*<*q*) nonempty pairwise disjoint subsets $J_1, ..., J_s$, i.e., such that $J_i \cap J_j = \varnothing$, $J_i \neq \varnothing, J_j \neq \varnothing, \forall i, j \in N_s$, and an ordered decomposition of the number n into s terms $n_1, n_2, ..., n_s$, that satisfies the condition $1 \le n_i \le q_i$, $\forall i \in N_s, |J_i| = q_i$. Obviously, $q_1 + q_2 + ... + q_s = q$ and $n_1 + n_2 + ... + n_s = n$.

Denote by *H* a set of elements of the form $h = (h(1), ..., h(n)) = (h^1, ..., h^s)$, where $h(j) \in N_n$, $j \in N_n$, a, and h^i is an arbitrary permutation of elements of the set $J_i \ \forall i \in N_s$.

Let a submultiset A^i of the multiset A consist of the elements of A whose numbers belong to the set J_i : $A^i = \{a_1^i, ..., a_{n_i}^i\}$, $|J_i| = n_i$.

Definition 1. [12] A set

$$P_{qk}^{ns}(A,H) = \left\{(a_{h(1)}, ..., a_{h(n)}) \middle| a_{h(i)} \in A \ \forall i \in N_n, \forall h \in H\right\} \subset R^n$$

is called a general set of polyarrangements.

Without loss of generality, let us arrange elements of the multiset A in nondecreasing order: $a_1 \le a_2 \le \ldots \le a_n$. Obviously, this arrangement also remains for each submultiset $A^i \ \ i \in N_s$, of A.

2. Properties of Euclidean set of polyarrangements

The convex hull of a set of polyarrangements $P_{qk}^{n}(A)$ is a polyhedron of polyarrangements $\Pi_{qk}^{ns}(A,H) = \operatorname{conv} P_{qk}^{ns}(A,H)$, whose set of vertices consists of elements of the set of polyarrangements: $\operatorname{vert} \Pi_{qk}^{ns}(A,H) = P_{qk}^{ns}(A,H)$.

Theorem 1. A polyhedron of polyarrangements $\Pi_{qk}^{ns}(A,H)$ is defined by the set of all solutions of the following system of inequalities:

$$\begin{cases} \sum\limits_{j=1}^{n_i} x_j \le \sum\limits_{j=1}^{n_i} a_j^i, i \in N_s, \\ \sum\limits_{j=1}^{m_i} x_{\alpha_j} \ge \sum\limits_{j=1}^{m_i} a_j^i, \quad m_i \in N_{q_i-1}, \alpha_j \in J_i, \forall i \in N_s \end{cases} \tag{2}$$

$$\alpha_j \ne \alpha_t, \forall j \ne t, \forall j,t \in J_i.$$

Let us consider some properties of the polyhedron $\Pi_{qk}^{ns}(A,H)$ and its relationship with the general set of polyarrangements.

Obviously, s subsystems of linear inequalities describing polyhedra of arrangements $\Pi_{q_i k_i}^{n_i}(A^i)$, being convex combinations of the sets of arrangements $a_{h^i}, i \in N_s$, can be separated out from the system of linear inequalities (2).

Therefore,

$$\Pi_{q_i k_i}^{n_i}(A^i) = \left\{ x \in R^{n_i} \left| \sum_{j=1}^{n_i} x_j \le \sum_{j=1}^{n_i} a_{q_i - j}^i, \sum_{j=1}^{m_i} x_{\alpha_j} \ge \sum_{j=1}^{m_i} a_j^i \right. \right\},$$

$m_i \in N_{q_i-1}, \alpha_j \in J_i, \alpha_j \ne \alpha_t, \forall j \ne t, \forall j,t \in J_i, \forall i \in N_s.$

The product of polyhedra $M_1,...,M_s$, is known to be a set $\bigotimes\limits_{i=1}^{S} M_i = \left\{ x \in R^{d_1+...+d_s} \mid x = (x_1,...,x_s),\ x_i \in M_i \ \ \forall i \in N_s \right\}$, where M_i is an d_i-dimensional polyhedron.

Lemma. 1) The product of polyhedrons is a polyhedron;

2) $\dim(\bigotimes\limits_{i=1}^{S} M_i) = \sum\limits_{i=1}^{S} \dim M_i$, where $\dim M$ is dimension of set M;

3) k-measured the verges of polyhedron $\bigotimes\limits_{i=1}^{S} M_i$ form the set with the elements of kind $\bigotimes\limits_{i=1}^{s} F_i$, where $F_i - k_i$ is the measured verge of polyhedron M_i and $k_1 + ... + k_s = k$.

By Statement 3.2 [12],

$$\bigotimes_{i=1}^{S} \Pi_{q_i k_i}^{n_i}(A^i) = \left\{ x \in R^{d_1+...+d_s} \mid x = (x_1,...,x_s),\ x_i \in \Pi_{q_i k_i}^{n_i}(A^i) \ \ \forall i \in N_s \right\},$$

i.e., a point $x \in \overset{s}{\underset{i=1}{\otimes}} \Pi^{n_i}_{q_i k_i}(A^i)$ satisfies each of s subsystems of the system (2). Hence, we may state that if a_{h^i} is a vertex of the polyhedron $\Pi^{n_i}_{q_i k_i}(A^i)$, then $a(h) = \overset{s}{\underset{i=1}{\otimes}} a_{h^i}$, $a(h)=a_{h^1},...,a_{h^s}$, where $a(h) \in P^{ns}_{qk}(A,H)$.

Next theorems are just [14].

Theorem 2. $\Pi^{ns}_{qk}(A,H) = \overset{s}{\underset{i=1}{\otimes}} \Pi^{n_i}_{q_i k_i}(A^i)$.

Theorem 3. For n<q, the polyhedron of polyarrangements $\Pi^{ns}_{qk}(A,H)$ is combinatorially equivalent to the polyhedron of polypermutations $\Pi^{s}_{qk}(A,H)$ of dimension n.

The vertices of the polyhedron $\Pi^{ns}_{qk}(A,H)$ are elements of the set of polyarrangements $P^{ns}_{qk}(A,H)$.

Theorem 4. A vertex $a(h) \in \mathrm{vert}\,\Pi^{ns}_{qk}(A,H)$ is adjacent to a vertex $a(z) \in \mathrm{vert}\Pi^{ns}_{qk}(A,H)$ if and only if $a(z)$ can be formed from $a(h)$ by a permutation of two unequal components a^i_l и a^i_j, $j \in J_{q_i-1}$, $i \in N_s$.

Note that the total number *p* of linear inequalities appearing in the system (2) and describing the polyhedron of polyarrangements $\Pi^{ns}_{qk}(A,H)$ is very high. It can be reduced in some cases.

Statement 1. If only k out of n coordinates $x_j, i \in N_n$, of a point $x \in R^n$ are different, then the number of inequalities of the system that describe the convex polyhedron $\Pi^{ns}_{qk}(A,H)$ can be reduced by excluding $N = \sum_{i=1}^{s} N_i$ inequalities, where $N_i = 1 + q_i + \sum_{j=i+1}^{q_i} C^j_{q_i}$.

Proof. Let us call an aggregate of inequalities of the subsystem for a subset J_i, $i \in N_s$, of the system (2), having equal values m_i of the upper limit of summation, the m_i th group of inequalities of this subsystem. Each mi m_i th group includes $C^{m_i}_{q_i}$ inequalities. Hence, the total number of inequalities describing the polyhedron $\Pi^{n_i}_{q_i k_i}(A^i)$ is $p_i = \sum_{i=0}^{q_i} C^{m_i}_{q_i} = 2^{q_i}$, $i \in N_s$. Since there are k_i different coordinates a^i_j, $j \in J_i$, out of q_i, then some inequalities can be excluded from the ith subsystem of inequalities (2). In view of the condition $a_1 \le a_2 \le \ldots \le a_q$ for any $j \in N_{m_i-1}, m_i \le q_i, i \le N_s$, the following equality holds: $a^i_j = a^i_{j+1}$. Therefore, if the inequalities of the first group in the subsystem (2) hold, the inequalities of the second, third, … m_i th, $i \in N_s$, groups will also hold. Indeed, since $x_j \ge a^i_1$, $j \in J_i$, $i \in N_s$, the condition $\sum_{j=1}^{m_i} x_{\alpha_j} \ge m_i a^i_1$ is satisfied for any $m_i \in N_n$. Hence, the inequalities of the second, third, …, m_i th, $i \in N_s$, groups can be excluded from each subsystem of system (2) describing the polyhedron of polyarrangements $\Pi^{ns}_{qk}(A,H)$, and the total number of inequalities in the m_i th subsystem will be $N_i = 1 + q_i + \sum_{j=i+1}^{q_i} C^j_{q_i}$. If the set of numbers

$\left(a_1^i, a_2^i, ..., a_n^i\right)$ possesses the property $a_j^i = a_{j+1}^i \ \ \forall j \in N_{n_i-1} \setminus N_{n_i-m_i}$, $i \in N_s$, the reasoning may be similar. Then it will suffice to leave only the inequalities of the first, second, ..., $(m_i - j)$ th group in the subsystem of the system (2). Therefore, $N = \sum_{i=1}^{s} N_i$ inequalities can be excluded from the system (2).

Let a set of polyarrangements $P_{qk}^{ns}(A,H)$ be mapped into the Euclidean space R^n and let us formulate the problem $Z(F,X)$ of maximizing a vector criterion $F(x)$ on a feasible set X:

$$Z(F,X): \max\{F(x) \mid x \in X\}.$$

To each point a $a \in P_{qk}^{ns}(A,H)$, there corresponds a point $x \in X$ such that $F(x) = \Phi(a)$, where $F(x) = (f_1(x), f_2(x), ..., f_l(x))$, $f_i : R^n \to R^1, i \in N_l$, X is a nonempty set defined as follows: $X = \operatorname{vert} \Pi_{qk}^{ns}(A,H)$, where $\Pi_{qk}^{ns}(A,H) = \operatorname{conv} P_{qk}^{ns}(A,H) = \Pi$. Let the problem $Z(F,X)$ contain also convex constraints that form a convex closed set $D \subset R^n$ of the form $D = \{x \in R^n \mid Bx \le d\}$. Then the feasible set $X = \operatorname{vert} \Pi_{qk}^{ns}(A,H) \bigcap D$.

In multicriterion optimization problems, the traditional concept of optimality is replaced with Pareto optimality (efficiency), weak efficiency (Slater optimality), and strong efficiency (Smale ptimality). Thus, by solutions of the problem $Z(F,X)$ we will mean elements of the following sets: $P(F,X)$ of efficient (Pareto optimal) solutions, $Sl(F,X)$ of weakly efficient (Slater optimal), and $Sm(F,X)$ of strongly efficient (Smale optimal) solutions. According to [4–7], the following statements are true for each feasible $x \in X$:

$$x \in Sl(F,X) \Leftrightarrow \{y \in X \mid F(y) > F(x)\} = \varnothing \quad (3)$$

$$x \in P(F,X) \Leftrightarrow \{y \in X \mid F(y) \ge F(x), F(y) \ne F(x)\} = \varnothing \quad (4)$$

$$x \in Sm(F,X) \Leftrightarrow \{y \in X \mid y \ne x, F(y) \ge F(x)\} = \varnothing. \quad (5)$$

$$Sm(F,X) \subset P(F,X) \subset Sl(F,X). \quad (6)$$

Since $|X| < \infty$, the set $P(F,X) \ne \varnothing$ and is externally stable [15].

3. Structural properties and optimality conditions of different sets of efficient solutions

Theorem 5. The elements of set $Sm(F,X)$ - strictly efficient, $P(F,X)$ - Pareto-optimal, and $Sl(F,X)$ - weakly efficient solutions of a multicriteria combynatorial problem over polyrranqements of the form $Z(F,X)$ are located at the vertices of polyhedron of polyrranqements $\Pi_{qk}^{ns}(A,H)$.

Proof. Taking into account correlation (6) between the introduced sets of efficient solutions and also according to fact that the set of feasible solutions X is a subset of the set of vertices of the general polyhedron of polyrranqements, that $\Pi_{qk}^{ns}(A,H)$, and $x \subset \operatorname{vert} \Pi_{qk}^{ns}(A,H)$ we come to $x \subset \operatorname{vert} \Pi_{qk}^{ns}(A,H)$ the justice of inclusions $Sm(F,X) \subset P(F,X) \subset Sl(F,X) \subset \operatorname{vert} \Pi_{qk}^{ns}(A,H)$ takes place. The theorem is proved.

Let the functions of vectorial criterion $f_i(x), i \in N_l$, are linear, that is $F(x)$ Structural properties of feasible region and sets of different types of efficient decisions, marked in the theorem 5, and also linear of functions of

vectorial criterion allows to take the decisions of problem $f_i(x) = \langle c_i, x \rangle, i \in N_l$. X to the X decision of the problem $G = \Pi_{qk}^{ns}(A,H) \bigcap D$.

Statement 2. The following inclusions are true for sets of optimal solutions of the problem $Z(F,X)$: $Sm(F,X) \subset P(F,X) \subset Sl(F,X) \subset \operatorname{vert} \Pi_{qk}^{ns}(A,H)$.

Let us represent the polyhedron Π_{qk}^{ns} as $\Pi_{qk}^{ns}(A) = \{x \in R^n \mid \langle \pi_i, x \rangle \le \gamma_i, i \in N_p\}$.

Introduce a set $N(y) = \{i \in N_q \mid \pi_i, y = \gamma_i\}$, $0^+Q(y) = \{x \in R^n \mid \pi_i x \le 0, i \in N(y)\}$ is a cone that can be constructed for all points $y \in \operatorname{vert} \Pi_{qk}^{ns}(A,H)$. Obviously, if $N(y) = \varnothing$, then $X \subseteq y + 0^+\Pi(y)$.

The structural properties of the feasible domain X and of the sets of various types of efficient solutions mentioned in Statement 3 and the linearity of the functions of vector criterion allow reducing the problem $Z(F,X)$ to the problem $Z(F,G)$ defined on the continuous feasible set $G = \Pi_{qk}^{ns}(A) \bigcap D$.

Theorem 6. The following inclusions are true: $P(F,G) \bigcap \operatorname{vert} \Pi_{qk}^{ns}(A,H) \subset P(F,X)$, $Sl(F,G) \bigcap \operatorname{vert} \Pi_{qk}^{ns}(A,H) \subset Sl(F,X)$, and $Sm(F,G) \bigcap \operatorname{vert} \Pi_{qk}^{ns}(A,H) \subset Sm(F,X)$.

Proof. Since $\operatorname{vert} \Pi_{qk}^{ns}(A,H) \bigcap D \subset G$, we have

$$P(F,G) \bigcap \operatorname{vert} \Pi_{qk}^{ns}(A,H) \bigcap D \subset P(F, G \bigcap \operatorname{vert} \Pi_{qk}^{ns}(A,H) \bigcap D) = P(F,X).$$

Similarly, we can prove the relationships

$$Sm(F,X) = Sm(F, D \bigcap \operatorname{vert} \Pi_{qk}^{ns}(A,H)) \supset Sm(F,G) \bigcap \operatorname{vert} \Pi_{qk}^{ns}(A,H).$$

$$Sl(F,X) = Sl(F, \operatorname{vert} \Pi_{qk}^{ns}(A,H) \bigcap D) \supset Sl(F,G) \bigcap \operatorname{vert} \Pi_{qk}^{ns}(A,H).$$

Let the functions $f_i(x), i \in N_l$, of the vector criterion $F(x)$ be linear, i.e., $f_i(x) = \langle c_i, x \rangle, i \in N_l, C \in R^{n \times l}$ is a matrix and a linear mapping $C: R^n \to R^l$, and c_i is its row vector, $i \in N_l$. Denote by $K = \left\{x \in R^n \mid Cx \ge 0\right\}$ the cone of perspective directions [4] of the problem $Z(F,X)$, $K_0 = \left\{x \in R^n \mid Cx = 0\right\}$ is the kernel of the mapping C, and $\operatorname{int} K = \left\{x \in R^n \mid Cx > 0\right\}$ is the interior of the cone K.

As follows from (3)–(5), the statements below are true $\forall x \in X$:

$$x \in Sl(C,X) \Leftrightarrow (x + \operatorname{int} K) \cap X = \varnothing, \tag{7}$$

$$x \in P(C,X) \Leftrightarrow x + (K \setminus K_0) \cap X = \varnothing \tag{8}$$

$$x \in Sm(C,X) \Leftrightarrow (x + K) \cap X \setminus \{x\} = \varnothing. \tag{9}$$

Theorem 7. If the feasible set X of the problem $Z(F,X)$ contains no constraints that describe the convex polyhedral set D, or $\Pi_{qk}^{ns}(A,H) \subseteq D$, i.e., $X = \operatorname{vert} \Pi_{qk}^{ns}(A,H)$, then the following equalities are true $\forall x \in R^n$:

$$Sl(F, \Pi_{qk}^{ns}(A,H)) \bigcap \operatorname{vert} \Pi_{qk}^{ns}(A) = Sl(F,X), \quad P(F, \Pi_{qk}^{ns}(A,H)) \bigcap \operatorname{vert} \Pi_{qk}^{ns}(A,H) = P(F,X),$$

$$Sm(F, \Pi_{qk}^{ns}(A)) \bigcap \operatorname{vert} \Pi_{qk}^{ns}(A) = Sm(F, X)$$

Proof. As follows from Theorem 6 and conditions of this theorem, the statements below are true $\forall x \in R^n$:

$$x \in Sl(F, \Pi_{qk}^{ns}(A,H)) \bigcap \operatorname{vert} \Pi_{qk}^{ns}(A,H) \Rightarrow x \in Sl(F,X),$$

$$x \in P(F, \Pi_{qk}^{ns}(A,H)) \bigcap \operatorname{vert} \Pi_{qk}^{ns}(A,H) \Rightarrow x \in P(F,X),$$

$$x \in Sm(F, \Pi_{qk}^{ns}(A,H)) \bigcap \operatorname{vert} \Pi_{qk}^{ns}(A,H) \Rightarrow x \in Sm(F,X).$$

Let us prove inverse implications. Let $x \in Sl(F,X),$ whence $x \in \operatorname{vert} \Pi_{qk}^{ns}(A,H)$ by Statement 3. Suppose by contradiction that $x \notin Sl(F, \Pi_{qk}^{ns}(A,H))$. Since functions $f_i(x), i \in N_l,$ of the vector criterion $F(x)$ are linear, the condition $\operatorname{int} K \bigcap (\Pi(x) - x) \neq \varnothing$ is satisfied by Theorem 5 [6], i.e., the cone $(x + \operatorname{int} K)$ contains some points of the boundary of the polyhedron $\Pi_{qk}^{ns}(A,H)$. Therefore, there exists a vertex $\Pi_{qk}^{ns}(A)$ belonging to this cone. By virtue of (7), this means that $x \notin Sl(F,X)$, which leads to a contradiction with the condition of the theorem. Other statements of the theorem can be proved similarly.

Corollary 1. Under the conditions of Theorem 7, the following statements are true $\forall x \in X$:

$$x \in Sl(F,X) \Leftrightarrow x \in Sl(F, \Pi_{qk}^{ns}(A,H)) \bigcap \operatorname{vert} \Pi_{qk}^{ns}(A,H),$$

$$x \in P(F,X) \Leftrightarrow x \in P(F, \Pi_{qk}^{ns}(A,H)) \bigcap \operatorname{vert} \Pi_{qk}^{ns}(A,H),$$

$$x \in Sm(F,X) \Leftrightarrow x \in Sm(F, \Pi_{qk}^{ns}(A,H) \bigcap \operatorname{vert} \Pi_{qk}^{ns}(A,H).$$

If the feasible domain $X = \operatorname{vert} \Pi_{qk}^{ns}(A,H)$, then the necessary and sufficient optimality conditions obtained in [6] for all the above-mentioned types of efficient solutions are true for any point $x = \operatorname{vert} \Pi_{qk}^{ns}(A,H)$ of the problem $P(F,X)$. If $\Pi_{qk}^{ns}(A) \bigcap D \neq \Pi_{qk}^{ns}(A)$, then only sufficient optimality conditions are true.

Theorem 8. For an arbitrary $x = \operatorname{vert} \Pi_{qk}^{ns}(A)$,

$$x \in P(F, \Pi_{qk}^{ns}(A)) \bigcap D \Rightarrow x \in P(F,X),$$

$$x \in Sl(F, \Pi_{qk}^{ns}(A)) \bigcap D \Rightarrow x \in Sl(F,X),$$

$$x \in Sm(F, \Pi_{qk}^{ns}(A)) \bigcap D \Rightarrow x \in Sm(F,X).$$

Proof. Since $G = \Pi \bigcap D,$ the following implications are true:

$$\forall x \in \operatorname{vert} \Pi_{qk}^{ns}(A,H) : x \in P(F, \Pi_{qk}^{ns}(A,H)) \bigcap D \Rightarrow x \in P(F, \Pi_{qk}^{ns}(A,H) \bigcap D) = P(F,G) \Rightarrow x \in P(F,X),$$

$$x \in Sl(F, \Pi_{qk}^{ns}(A,H)) \bigcap D \Rightarrow x \in Sl(F,X),\ x \in Sm(F, \Pi_{qk}^{ns}(A,H)) \bigcap D \Rightarrow x \in Sm(F,X).$$

Thus, Theorems 5 – 8 establish an interrelation between the problem $Z(F,X)$ and the problem $Z(F,G)$ defined over a continuous feasible set. It enables to apply the classic methods of continuous optimization to the decision of vectorial combinatorial problems on polyrrangements and on this basis develop new original methods of decision, using properties of combinatorial sets and their protuberant shells.

If the problem $Z(F,X)$ does not contain linear limitations forming a convex polyhedral set $D \subset R^n$, or if we have $\Pi \subseteq D$, i.e. $X = \text{vert}\,\Pi$, then, taking into account necessary and sufficient optimality conditions (theorem 7), the process of its solution is reduced to the search for efficient solutions of the problem $Z(F,G)$ over the continuous feasible set $G = \Pi_{qk}^{ns}(A,H)$ with the subsequent choice of only the solutions thatf are vertices of the permutable polyhedron of polyrrangements $\Pi_{qk}^{ns}(A,H)$.

Analyzing theorems 6 and 8, we obtain the following relationships between the problems $Z(F,X)$ and $Z(F,G)$: if we have $x \in R(F,G) \bigcap vert\Pi_{qk}^{ns}(A,H)$, then $x \in R(F,X)$, and if we have $x \notin R(F,G) \bigcap vert\Pi_{qk}^{ns}(A,H)$, then this does not imply that $x \notin R(F,X)$, where $R(F,X)$ denotes the set set $P(F,X)$, $Sm(F,X)$ or $Sl(F,X)$.

4. Main approach of the task contains additional linear limitationds, the following approach its decision is offered.

If the problem $Z(F,X)$ contains additional linear constraints, then the following approach to its solve is proposed.

1. We find the efficient solutions of the problem $Z(F,\Pi_{qk}^{ns}(A,H))$.

2. We check their membership in the set D. If we have $x \in P(F,\Pi_{qk}^{ns}(A,H)) \bigcap D$, then $x \in P(F,X)$.

3. We will consider feasible solutions $x \in X$ to the problem $Z(F,X)$ that are inefficient in the problem $Z(F,\Pi)$, i.e. $x \in X \setminus P(F,\Pi_{qk}^{ns}(A,H)) \bigcap D$, and check them for efficiency. To this and, use the necessary and sufficient conditions formulated in [15].

Statement 3. A feasible solution x^0 is efficient if and only if it is an optimal solution of the following problem:

$$Z^1(F,X): \max\left\{\sum_{i=1}^{m} f_i(x) \mid x \in X,\ f_i(x) \geq f_i(x^0),\ i \in N_m\right\}.$$

If the solution x^0 is inefficient, then, as a result of solution of this problem, we find an efficient solution x^* that is more preferable than x^0, i.e., we have $F(x^*) \geq F(x^0)$.

Continuing investigations and developing the results of [1, 5, 6, 8-13], we propose an approach to the solution of the problem $Z(F,X)$ on the basis of linear convolution (aggregation) of its partial criteria and the further reduction of the search for solutions of the initial problem to the solution of a series of scalar (one-criterion) problems and the check of the obtained solutions for optimality. The method of solution of one-criterion problems is based on the ideas of decomposition, Kelly 's cutting-planes, and relaxation.

Next, we consider a method whose realisation takes into account the fact that the number of constraints is sufficiently large. Then it is expedient to use a relaxation procedure or temporary rejection of some constraints and the solution of a problem over a wider domain, i.e., under remained constraints.

At the initial stage of construction of the sought-for algorithm, we should determine the initial point. We will consider a one-criterion problem without constraints that describe a polyhedron *D* and call them additional constraints.

Statement 4. If, for the elements of the multiset *A* and coefficients c_j, $j \in N_n$, of the objective function of the problem a maximum functions $f(x)$, conditions $c_{i_1} \le c_{i_2} \le ... \le c_{i_n}$ and $a_1 \le a_2 \le \ldots \le a_n$, respectively, are satisfied, on the admissible set is attained at a point $x^* = (x^*_{i_1},...,x^*_{i_n}) \in \mathrm{vert}\, \Pi^{ns}_{qk}(A,H)$ that is specified as follows:

$$x^*_{i_j} = a_j \ \forall\ j \in N_n\,, \tag{10}$$

and its minimum is accordingly attained at a point $y = (y_{i_1}, y_{i_2},..., y_{i_n}),$ where

$$y_{i_{j+1}} = a_{n-j} \ \forall\ j \in N_{n-1} \bigcup \{0\}\,.$$

For description and basing of method of decision of the problem we will introduce next denotations. We write down the feasible region of the problem $Z(F,X)$ in the form $G = \{x \in R^n \mid Hx \le g\}$, $g = (g_1, g_2,..., g_q)$ is matrix, which is used for the matrix-vectorial form of record of limitations of the form (2) and linear inequalities describing a polyhedron D, where all limitations are taken to one $\le$ kind of inequalities. We will designate N_q the set, the elements of which determine the numbers of limitations of the system (2) and additional limitations describing the protuberant many-sided set $D := N_q$.

We define sets $G_i = \{x \in R^n \left| \langle h_i, x\rangle \le g_i\right\}, i \in N_q$ and, for an arbitrary $x^s \in R^n$, define sets $N^a(x^\nu) = \{i \in N_q \left| \left\langle h_i, x^\nu \right\rangle = g_i\right\}$ – accordingly active and nonactive limitations in the point; x^ν - accordingly $i \in N_q$ the vector-line of matrix and i component of vector H.

We will submit a problem in to consideration, the problem where $Z(F, G^\nu): \max\left\{F(x) \mid x \in G^\nu\right\}$ is set of indexes of limitations, describing the feasible region of problem $G^\nu = \left\{x \in R^n \left| \langle h_i, x\rangle \le g_i, i \in Q_\nu \subset N_q\right.\right\},$, which is solved on м step of algorithm, $Z(F, G^\nu)$, is set of numbers of limitations which were not included in this problem on м step.

Definition 5. We call the quantity $r_i(x) = \langle h_i, x\rangle - g_i,\ i \in N_q$, a deviation of point $x \in R^n$ from the boundary of a set G_i and the quantity $r(x) = \max\{r_i(x) \left| \ i \in N_q\right.\}$ a deviation of point $x \in R^n$ from the boundary of the set G. It is obvious that, for $i \in N_p$, we have

$$r_i(x) = \sum_{j=1}^{i} x_{\alpha_j} - \sum_{j=1}^{i} a^i_j\ , \tag{11}$$

and for $i \in N_q \setminus N_p$, we have

$$r_i(x) = \langle b_i, x\rangle - d_i\,, \tag{12}$$

where b_i is the i th row vector of the matrix $B, d_i \in R$.

Theorem 9. An efficient (Pareto-optimal, weakly efficient, and strictly efficient) solution x^0 of problem $Z(F,G^\nu)$ is an efficient (in the same sense) solution of the problem $Z(F,G)$ if and only if the condition $r(x)\le 0$ is true.

Proof. The necessity of this statement is obvious since the feasible solution x^0 of the problem $Z(f,G^\nu)$ is a feasible solution of the problem $Z(F,G)$ if and only if the condition $r(x)\le 0$ is satisfied. The sufficiency of this statement follows from the construction of the problem $Z(F,G)$ and the definition of $r(x)$.

The approach to solving the class of vector problems proposed here is to reduce the original multicriterion problem to an optimization problem with one criterion $f_r(x), r\in N_l$, which is declared principal or basic provided that the values of all other criteria should be no less than some prescribed (threshold) values $t_i, i\in N_l\setminus\{r\}$. Thus, we have the problem

$$Z(f_r,X(t_i)):\max\left\{f_r(x)\middle| f_i(x)\ge t_i, i\in N_l\setminus\{r\}, x\in X\right\}.$$

The optimal solution x^0 of this problem is always weakly efficient, and if it is unique (up to equivalence $\sim_f$), it is also efficient. If the solution x^0 is efficient, then it is a unique (up to equivalence $\sim_f$) solution of the problem $Z(f_r,X(t_i))$ for any fixed $r\in N_l$ and $t_i=f_i(x^0), i\in N_l\setminus\{r\}$. Choosing a criterion as the principal one does not limit the choice of the optimal solution. To determine the threshold values t_i, $i\in N_l\setminus\{r\}$, Statement 4 can be used, which makes it feasible to establish the upper and lower bounds for the criteria $f_i(x)$, $i\in N_l$, on a set of polyarrangements. We propose two approaches to solve the original problem $Z(F,X)$. The first approach assigns the minimum values of criteria $f_i(x)$, $i\in N_l$, on the set of polyarrangements to the thresholds $t_i\in N_l\setminus\{r\}$ followed by the reduction of the feasible set of the problem $Z(f_r,X)$ by choosing the values of thresholds $t_i\in N_l\setminus\{r\}$, arranged in increasing order, next to the minimum values of the criteria. The second approach searches for the optimal solution of the problem $Z(f_r,X)$ by assigning the greatest feasible values to the criteria $f_i(x)$, $i\in N_l$, and then expanding its feasible domain if the original problem appears inadmissible; if it is admissible, an efficient or weakly efficient solution is found.

The procedure of assigning constraints to a series of threshold values t_i is simple in both approaches. Using Statement 4 and ordering the coefficients of the criteria, we reduce this procedure to computing the scalar product of two vectors, i.e., to finding values of linear criteria. Taking into account the structural features of the set of polyarrangements, it is feasible to compute t_i more efficiently, using permutations of the elements of each i th, $i\in N_S$, subset of the multiset A.

The general idea of the method proposed to solve the problem $Z(F,X)$ consists in successive inclusion of constraints of the problem that describe the feasible region.

1. Reduce the multicriterion problem $Z(F,G)$ to a one-criterion problem $Z(f,G)$ using the principal-criterion method. Put $\nu=0$.

2. Select constraints of the original system of linear inequalities that describe the feasible set $G^\nu\subset G$ of the problem $Z(f,G^\nu)$ and use the simplex method to find its optimal solution x^ν.

3. If the optimal solution x^V is an element of the set of polyarrangements, then, at the point x^V, check the constraints that have not been taken into account. Obviously, they can only be constraints that describe the convex closed set D. If the solution x^V does not satisfy some constraints, then supplement constraints of the feasible region of the problem $Z(f, G^V)$, with the worst-satisfied constraint from the convex closed set *D*. If the solution x^V satisfies the above-mentioned constraints, it is an efficient solution of the problem $Z(F, G)$ and, hence, of the problem $Z(F, X)$.

4. If the solution x^V is not a point of the set of polyarrangements, make a cutoff through adjacent vertices that cuts off a vertex not being feasible (i.e., polyarrangement). Add this cutoff to constraints of the problem $Z(f, G^V)$.

5. Compare the value $f(x^V)$ of the objective function with its value found at the previous step. If it decreases, reject the constraints insignificant at the point x^V. If the value of $f(x^V)$ does not change, do not reject the constraints. Use the changed feasible region go to Step 2 to solve the problem $Z(f, G^V)$.

Obviously, the algorithm results in an efficient solution of the problem $f(x^V)$ or establishes its unsolvability by solving a finite number of subproblems of the form $Z(f, G^V)$.

Conclusions

A vector combinatorial problem has been analyzed using the information about the convex hull of the feasible region and the properties of polyhedra whose vertices determine a prescribed combinatorial set of polyrrangements. A method to solve complex multicriterion problems over this combinatorial set have been developed and substantiated. Using the structural properties of combinatorial polyhedra makes it possible to develop efficient algorithms to solve new classes of vector combinatoria optimization problems.

Acknowledgements

The paper is published with financial support by the project ITHEA XXI of the Institute of Information Theories and Applications FOI ITHEA Bulgaria www.ithea.org and the Association of Developers and Users of Intelligent Systems ADUIS Ukraine www.aduis.com.ua.

Bibliography

[1] I. V. Sergienko, Mathematical Models and Methods to Solve Discrete Optimization Problems [in Russian], Naukova Dumka, Kyiv (1988).

[2] I. V. Sergienko and M. F. Kaspshitskaya, Models and Methods of Computer Solution of Combinatorial Optimization Problems [in Russian], Naukova Dumka, Kyiv (1981).

[3] I. V. Sergienko and V. P. Shilo, Discrete Optimization Problems: Challenges, Solution Techniques, and Investigations [in Russian], Naukova Dumka, Kyiv (2003).

[4] I. V. Sergienko, L. N. Kozeratskaya, and T. T. Lebedeva, Stability and Parametric Analyses of Discrete Optimization Problems [in Russian], Naukova Dumka, Kyiv (1995).

[5] I. V. Sergienko, T.T. Lebedeva, and N. V. Semenova, “Existence of solutions in vector optimization problems,” Cybern. Syst. Analysis, 36, No. 6, 823–828 (2000).

[6] T. T. Lebedeva, N. V. Semenova, and T. I. Sergienko, “Optimality and solvability conditions in linear vector optimization problems with convex feasible region,” Dop. NANU, No. 10, 80–85 (2003).

[7] T. T. Lebedeva, N. V. Semenova, and T. I. Sergienko, “Stability of vector problems of integer optimization: Relationship with the stability of sets of optimal and nonoptimal solutions,” Cybern. Syst. Analysis, 41, No. 4, 551–558 (2005).

[8] N. V. Semenova, L. N. Kolechkina, and A. N. Nagirna, “An approach to solving discrete vector optimization problems over a combinatorial set of permutations,” Cybern. Syst. Analysis, 44, No. 3, 441–451 (2008).

[9] N. V. Semenova, L. M. Kolechkina, and A. M. Nagirna, “Vector combinatorial problems in a space of combinations with linear fractional functions of criteria,” Inform. Theor. Appl., 15, 240–245 (2008).

[10] V. V. Podinovskii and V. D. Nogin, Pareto-Optimal Solutions of Multicriterion Problems [in Russian], Nauka, Moscow (1982).

[11] Yu. G. Stoyan and S. V. Yakovlev, Mathematical Models and Optimization Methods of Geometric Design [in Russian], Naukova Dumka, Kyiv (1986).

[12] Yu. G. Stoyan, O. O. Yemets, and E. M. Yemets, Optimization over Polyarrangements: Theory and Methods [in Ukrainian], RVTs PUSKU, Poltava (2005).

[13] O. O. Yemets and L. M. Kolechkina, Combinatorial Optimization Problems with Linear Fractional Objective Functions [in Ukrainian], Naukova Dumka, Kyiv (2005).

[14] V. A. Emelichev, M. M. Kovalev, and M. K. Kravtsov, Polyhedra, Graphs, and Optimization [in Russian], Nauka, Moscow (1981).

[15] K. Aardal and S. Hoesel, “Polyhedral techniques in combinatorial optimization. II: Computations,” Statist. Neerlandica, 53, 131–177 (1999).

Authors' Information

Natalia V. Semenova *- V. M. Glushkov Institute of Cybernetics, National Academy of Science of Ukraine, Candidate of Physical & Mathematical Science, Senior Scientific Researcher, 40, Prospect Akademika Glushkova, 03187, Kyiv, Ukraine; e-mail: nvsemenova@meta.ua*

Lyudmyla N. Kolechkina *- V. M. Glushkov Institute of Cybernetics, National Academy of Science of Ukraine, Candidate of Physical & Mathematical Science, Senior Scientific Researcher, 40, Prospect Akademika Glushkova, 03187, Kyiv, Ukraine; e-mail: ludapl@ukr.net*

МЕТОДИКА ОПРЕДЕЛЕНИЯ КАЧЕСТВА ТЕСТОВЫХ ЗАДАНИЙ, ОЦЕНИВАЕМЫХ ПО НЕПРЕРЫВНОЙ ШКАЛЕ

Наталия Белоус, Ирина Куцевич, Ирина Белоус

***Аннотация**: В работе описывается методика определения качества тестовых заданий с помощью которой проводится выделение в тесте несостоятельных заданий и заданий плохого качества. В работе проведен сравнительный анализ применения дихотомической и непрерывной шкал для оценивания.*

***Ключевые слова**: качество тестовых заданий, коэффициент корреляции, валидность, несостоятельные тестовые задания, задания плохого качества, сложность тестового задания, субъект обучения.*

***ACM Classification Keywords**: K.3.1 Computer Uses in Education*

***Conference**: The paper is selected from XVth International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.*

Введение

Одним из сложных и противоречивых вопросов при проведении тестирования является оценивание знаний. В настоящее время в большинстве случаев используется дихотомическая шкала, по которой за каждое задание можно получить 0 или 1 балл. Данная шкала удобна при оценивании заданий т.н. закрытого типа, в которых выбирается один правильный ответ из многих. Существует многообразие типов тестовых заданий: закрытые (многоальтернативные и одноальтернативные), открытые, на установление соответствия между элементами, на установление правильной последовательности, ситуационные тестовые задания [Комплекс нормативних документів, 1998]. Для оценивания заданий разных типов часто применение дихотомической шкалы часто недостаточно, т.к. в случае, когда субъект обучения дает неполный или частично правильный ответ, он оценивается как неправильный. Кроме дихотомической шкалы в настоящее время используется политомическая шкала, в которой допускается несколько категорий ответа на задание, каждая из которых оценивается по-разному. Например, за полностью верный ответ назначается 2 балла, за частично верный – 1 балл и за неверный – 0 баллов. Недостатком этой шкалы является сложность вычисления общего результата на основе баллов, полученных за задания. Кроме того, в этом случае не учитываются неправильно выбранные варианты ответа. Простое суммирование баллов не соответствует истинному уровню знаний обучаемых. Чтобы избежать этих недостатков авторами предлагается введение непрерывной шкалы оценивания знаний на интервале от 0 до 1 и специализированные технологии определения оценок за выполнение каждого из типов тестовых заданий [Belous N., 2004].

Уровень знаний студентов варьирует от качества постановки учебного процесса, от количества выделяемых часов и от качества учебного материала, в том числе и тестового. Разнообразие причин, влияющих на качество знаний, подтверждает, что необходимо контролировать качество тестового материала с определенной периодичностью. Под качеством тестового материала принимают возможность различия субъектов обучения с высоким уровнем знаний и слабых [Аванесов В.С., 1989].

Для проведения качественного анализа тестового материала, оцениваемого по дихотомической шкале оценивания знаний, предлагается проведение статистической обработки результатов тестирования

[Аванесов В.С., 1989, Олейник Н.М., 1991, Комплекс нормативних документів, 1998]. Однако, для повышения точности оценивания знаний авторами предлагается применение непрерывной шкалы (в диапазоне от 0 до 1, где 1 –ответ полностью правильный, 0 – ответ полностью неправильый, промежуточные значения соответствуют неполным или частично правильным ответам). Для определения качества тестов, оцениваемых по непрерывной шкале, авторами предлагается методика, позволяющая по результатам предварительного тестирования выделять несостоятельные задания и задания плохого качества. К несостоятельным заданиям относятся те задания, которые не служат цели дифференцирования знаний и являются в этом случае бесполезными. К таким заданиям относятся слишком легкие (на которые ответили все) или слишком трудные (на которые никто не ответил), а также задания, не относящиеся к рассматриваемой в тесте предметной области. Задания, которые требуют корректировки, например, из-за неточности формулировок, относятся к заданиям плохого качества.

Целью работы является разработка методики определения качества тестовых заданий, оцениваемых по непрерывной шкале, частным случаем которой является дихотомическая шкала.

Методика Проведения Качественного Анализа Тестовых Заданий

В общем виде тест представляет собой систему, состоящую из набора тестовых заданий. Требование системности заключается в том, что между заданиями, включенными в тест, должны прослеживаться четкие связи, которые отражаются в результатах выполнения теста группой субъектов обучения. Для оценивания системных качеств теста применяется коэффициент корреляции, показывающий степень связи между случайными величинами, в данном случае, между тестовыми заданиями, на которые отвечала группа студентов.

Исходными данными для проведения определения качества тестов являются результаты тестирования выборки субъектов обучения, заданные с помощью неупорядоченной матрицы результатов, в которой столбцы соответствуют номерам тестовых зданий, строки – фамилиям субъектов обучения. Элементами неупорядоченной матрицы тестирования являются результаты res_{ij} i-го субъекта обучения за выполнение j-го задания, оцененного по непрерывной системе в диапазоне [0,1]. По исходной неупорядоченной матрице строится упорядоченная матрица, данные из которой являются исходными к проведению дальнейших вычислений. По неупорядоченной матрице результатов тестирования определяются несостоятельные задания (задания, с которыми не справился ни один субъект обучения и те, с которыми все справились). Эти задания в упорядоченную матрицу не включаются. Оставшиеся тестовые задания упорядочиваются следующим образом: строки матрицы упорядочиваются по суммарному баллу за выполнение всех заданий каждым субъектом обучения в порядке возрастания сверху вниз, столбцы матрицы – по суммарному баллу за выполнение каждого задания всеми субъектами обучения в порядке убывания слева направо. Упорядоченная матрица тестовых результатов приведена на рисунке 1.

После упорядочивания матрицы тестовых результатов вычисляются величины $\overline{R_j}$ - мера трудности задания (средний балл по всем заданиям) и $\overline{R_i}$ - средний балл по всем субъектам обучения.

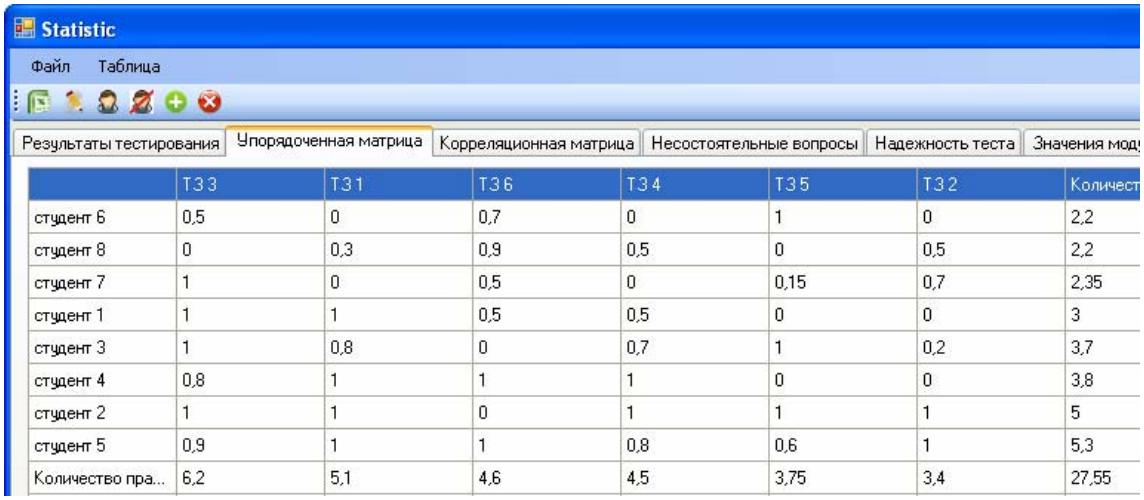
$$\overline{R_j} = \frac{1}{N}\sum_{i=1}^{N} res_{ij}\,, \tag{1}$$

$$\overline{R_i} = \frac{1}{n}\sum_{j=1}^{n} res_{ij}\,, \tag{2}$$

где n – количество заданий, включенных в тест;

N – количество студентов, прошедших предварительное тестирование.

По упорядоченной матрице строится корреляционная матрица тестовых заданий, которая отображает степень связи тестовых результатов субъектов обучения (рис. 2).



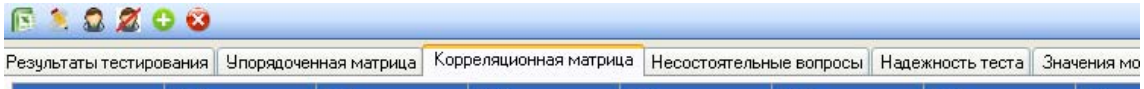
	T33	T31	T36	T34	T35	T32	Количест
студент 6	0,5	0	0,7	0	1	0	2,2
студент 8	0	0,3	0,9	0,5	0	0,5	2,2
студент 7	1	0	0,5	0	0,15	0,7	2,35
студент 1	1	1	0,5	0,5	0	0	3
студент 3	1	0,8	0	0,7	1	0,2	3,7
студент 4	0,8	1	1	1	0	0	3,8
студент 2	1	1	0	1	1	1	5
студент 5	0,9	1	1	0,8	0,6	1	5,3
Количество пра...	6,2	5,1	4,6	4,5	3,75	3,4	27,55

Рисунок 1 – Упорядоченная Матрица Результатов Тестирования

Элементами корреляционной матрицы являются коэффициенты корреляции C_{ij}, которые в случае применения непрерывной системы оценивания знаний авторами предлагается рассчитывать по формуле (3).

$$C_{ab} = \frac{\sum_{i=1}^{N}\left(res_{ia} \cdot res_{ib}\right) - N \cdot \overline{R_a} \cdot \overline{R_b}}{\sqrt{\left(\sum_{i=1}^{N} res_{ia}^{\,2} - N \cdot \left(\overline{R_a}\right)^2\right) \cdot \left(\sum_{i=1}^{N} res_{ib}^{\,2} - N \cdot \left(\overline{R_b}\right)^2\right)}}, \quad (3)$$

где a и b – тестовые задания, для которых рассчитывается коэффициент корреляции; $\overline{R_a}$ – мера трудности a-го тестового задания; $\overline{R_b}$ – мера трудности b-го тестового задания; C_{ab} – коэффициент корреляции тестовых заданий a и b.



	T33	T31	T36	T34	T35	T32	Rxy
T33	1	0,476	-0,507	0,234	0,235	0,151	0,524
T31	0,476	1	-0,124	0,895	0,00605	0,101	0,799
T36	-0,507	-0,124	1	-0,0777	-0,57	-0,125	-0,157
T34	0,234	0,895	-0,0777	1	0,0529	0,229	0,803
T35	0,235	0,00605	-0,57	0,0529	1	0,211	0,367
T32	0,151	0,101	-0,125	0,229	0,211	1	0,553
Сумма	1,589	2,35405	-0,4037	2,3332	0,93495	1,567	

Рисунок 2 – Корреляционная Матрица Результатов Тестирования

Корреляция заданий друг с другом не должна быть высокой, иначе задания начинают дублировать друг друга. По классификации коэффициентов корреляции Дворецкого связь тестовых заданий должна рассматриваться как слабая (коэффициент корреляции $C_{ab} < 0,3$). С другой стороны, при отрицательных значениях коэффициента корреляции ($C_{ab} < 0$) наблюдается обратная корреляция между тестовыми заданиями. Отрицательная корреляция между тестовыми заданиями нежелательна. Если задание отрицательно коррелирует с другими заданиями, то исход ответов на него противоположен результатам по другим заданиям. По всей вероятности у такого задания либо имеются грубые ошибки в содержании и

(или) оформлении (например, нет правильного ответа), либо проверяются знания из другой предметной области. Такие задания подлежат удалению. В случае, приведенном на рисунке 2, отрицательной корреляцией отличаются все тестовые задания. Следует обратить внимание на то, что отрицательная корреляция у заданий 1, 2, 3, 4 и 5 наблюдается именно с заданием 6. Это означает, что проблематичным является именно тестовое задание 6. У тестового задания 1 наблюдается сильная корреляция с заданиями 3 и 4, что свидетельствует также о проблематичности тестового задания. Разделим теперь тестовые задания на задания плохого качества и несостоятельные тестовые задания.

Важным параметром, применяемым при проведении качественного анализа тестовых заданий, является коэффициент валидности. Валидность – это мера соответствия того, насколько методика и результаты исследования соответствуют поставленным задачам. Для определения коэффициентов валидности тестовых заданий вычисляются коэффициенты корреляции заданий теста с суммой баллов субъектов обучения R_i. Следовательно, $V_j = C_{j\,R_i}$. Для вычисления коэффициентов валидности V_j авторами предлагается применение формулы (4).

$$V_j = \frac{\sum_{i=1}^{N}\left(res_{ij}\cdot R_i\right) - \overline{R_j}\cdot R}{\sqrt{\left(\sum_{i=1}^{N} res_{ij}^{\,2} - N\cdot\left(\overline{R_j}\right)^2\right)\cdot\left(\sum_{i=1}^{N} R_i^2 - \frac{R^2}{N}\right)}}, \qquad (4)$$

где R_i – суммарный балл за выполнение заданий i-тым субъектом обучения, $R_i = \sum_{j=1}^{n} res_{ij}$;

R – суммарный балл, набранный всеми испытуемыми за выполнение тестовых заданий, $R = \sum_{i=1}^{N}\sum_{j=1}^{n} res_{ij}$.

По коэффициенту валидности проводится выявление тестовых заданий плохого качества и несостоятельных тестовых заданий. Для практических целей Авонесов рекомендует использовать коэффициенты валидности $V_j \geq 0.5$ [Аванесов В.С., 1989]. В случае $V_j \geq 0.5$ тестовое задание разделяет субъектов обучения с высоким уровнем знаний и слабых. Все те задания, для которых коэффициент валидности меньше или равен нулю, несостоятельны и непригодны для контроля знаний, поэтому их надо удалять из создаваемого теста, а для включения данных заданий в другие тесты их необходимо существенно переделывать и улучшать. Задания же, для которых $0 < V_j < 0.5$, являются заданиями плохого качества и требуют коррекции. Для данных, приведенных на рисунке 2, заданием плохого качества является ТЗ 6, несостоятельным заданием является ТЗ 1. Таким образом, ТЗ 1 требует корректировки, а ТЗ 6 должно быть полностью удалено из теста (рис. 2).

Выполнение предложенных расчетов позволяет сделать чистку теста и первые выводы о его ожидаемых качественных характеристиках. Корреляционная матрица результатов тестирования после проведения чистки упорядоченной матрицы приведена на рисунке 3.

После проведения чистки теста до проведения текущего и итогового оценивания знаний субъектов обучения проиодится распределение тестовых заданий по уровням сложности с применением разработанной авторами технологии распределения тестовых заданий по уровням сложности [Белоус Н.В. и др., 2009].

С учетом приведенных в работе обозначений, вычисление начального уровня знаний субъектов обучения Θ_i^0, начального уровня сложности тестовых заданий β_j^0 и дифференцирующей способность тестовых

заданий a_j проводится по формулам (5).

Результаты тестирования | Упорядоченная матрица | Корреляционная матрица | Несостоятельные вопросы | Надеж

	Т3 3	Т3 4	Т3 5	Т3 2	Rxy
Т3 3	1	0,234	0,235	0,151	0,603
Т3 4	0,234	1	0,0529	0,229	0,58
Т3 5	0,235	0,0529	1	0,211	0,65
Т3 2	0,151	0,229	0,211	1	0,654

Рисунок 3 – Корреляционная Матрица Результатов Тестирования После Проведении Чистки Упорядоченной Матрицы

$$\Theta_i^0 = \ln\left(\frac{\overline{R_i}}{1-\overline{R_i}}\right), \qquad \beta_j^0 = \ln\left(\frac{\overline{R_j}}{1-\overline{R_j}}\right), \qquad a_j = \frac{V_j}{\sqrt{1-(V_j)^2}} \tag{5}$$

С помощью интегрированной функциональной модели вычисляется набор параметров β_j, которые соответствуют устойчивым оценкам уровня сложности тестовых заданий. После проведения описанных вычислений разработанный тест может применяться для проведения объективного оценивания знаний субъектов обучения [Белоус Н.В. и др., 2009].

Сравнительный анализ применения методики при дихотомической и непрерывной шкалах

Пусть для определения результатов тестирования применялась дихотомическая шкала оценивания знаний. Рассмотрим изложенный выше пример (рис. 1-3) с использованием дихотомической шкалы. Таким образом, если результат за выполнение тестового задания субъектом обучения меньше 1, то в матрицу результатов тестирования ставим 0, в противном случае – 1. Следовательно, упорядоченная матрица результатов тестирования для дихотомической шкалы примет вид:

Результаты тестирования | Упорядоченная матрица | Корреляционная матри

	Т3 1	Т3 3	Т3 5	Т3 2	Т3 4	Т3 6	Количе...
студент 8	0	0	0	0	0	0	0
студент 6	0	0	1	0	0	0	1
студент 7	0	1	0	0	0	0	1
студент 1	1	1	0	0	0	0	2
студент 3	0	1	1	0	0	0	2
студент 4	1	0	0	0	1	1	3
студент 5	1	0	0	1	0	1	3
студент 2	1	1	1	1	1	0	5
Количество пра...	4	4	3	2	2	2	17

Рисунок 4 – Упорядоченная Матрица Результатов Тестирования при Дихотомической Шкале Оценивания

Для полученной упорядоченной матрицы (рис. 4) построим корреляционную матрицу (рис. 5).

Из анализа результатов, внесенных в корреляционную матрицу, видно, что большая часть заданий демонстрируют отрицательную корреляцию между результатами их выполнения либо имеют сильную связь между заданиями, что свидетельствует об их плохом качестве. В свою очередь, плохое качество при оценивании заданий по непрерывной шкале, показало только Т3 1, а несостоятельным заданием является Т3 6 (см. рис. 2).

Результаты тестирования | Упорядоченная матрица | Корреляционная матрица | Несостоятельные вопросы | Надежность теста | Значения мод

	ТЗ 1	ТЗ 3	ТЗ 5	ТЗ 2	ТЗ 4	ТЗ 6	Rxy
ТЗ 1	1	0	-0,258	0,577	0,577	0,577	0,775
ТЗ 3	0	1	0,258	0	0	-0,577	0,258
ТЗ 5	-0,258	0,258	1	0,149	0,149	-0,447	0,289
ТЗ 2	0,577	0	0,149	1	0,333	0,333	0,745
ТЗ 4	0,577	0	0,149	0,333	1	0,333	0,745
ТЗ 6	0,577	-0,577	-0,447	0,333	0,333	1	0,348
Сумма	2,473	0,681	0,851	2,392	2,392	1,219	
-r	0,412	0,114	0,142	0,399	0,399	0,203	
R*R	0,17	0,013	0,0202	0,159	0,159	0,0412	

Рисунок 5 – Корреляционная Матрица Результатов Тестирования при Дихотомической Шкале Оценивания

При детальном анализе тестовых заданий, включенных в тест, было отмечено, что в ТЗ 6 как правильный был выбран не тот вариант ответа, а в ТЗ 1 сложность задания объясняется длинной формулировкой задания, что усложняет анализ его содержания субъектами обучения.

Таким образом, можно сделать вывод, что применение непрерывной шкалы при определении качества тестовых заданий дает данные с большим уровнем надежности, чем при применении дихотомической шкалы.

Выводы

Рассмотренная методика определения качества тестовых заданий, оцениваемых по непрерывной шкале, позволяет создавать качественные тесты для проведения объективного оценивания знаний субъектов обучения. Предлагаемая в работе методика программно реализована в виде системы проведения качественного анализа тестовых заданий. Разработанная система может применяться как при работе с авторским программным комплексом проведения контроля знаний [Belous N., 2005], так и с данными, полученными из внешних систем автоматизированного тестирования. Это делает разработанную систему универсальной.

Предлагаемая программная система прошла апробацию на достоверной выборке (более 100 субъектов обучения на тесте из 132 вопросов). При проведении анализа результатов предварительного тестирования были выявлены 12 заданий плохого качества и 5 несостоятельных заданий, что показало работоспособность предлагаемой в работе методики проведения качественного анализа тестовых заданий.

Внедрение предлагаемой системы в высших учебных заведениях позволит проводить качественный анализ тестовых заданий непосредственно перед проведением промежуточного либо итогового тестирования. Программная система применима как в учебных заведениях любого уровня аккредитации, так и в организациях и учреждениях, где проводится профессиональный отбор с помощью тестирования, а также на курсах повышения квалификации.

Важным моментом при оценивании результатов выполнения тестовых заданий субъектами обучения является возможность угадывания правильных ответов [Белоус Н.В., 2009]. Возможна ситуация, когда слабый студент справляется с более сложным заданием, а не может выполнить более простое. В данном случае речь идет не о плохом качестве простого вопроса, а о том, что на более сложный вопрос ответ был угадан. В дальнейшем планируется усовершенствовать методику путем проведения предварительного нормирования, т.е. перерасчета значений упорядоченной матрицы с учетом возможности угадывания правильного ответа и сложности тестовых заданий в зависимости от ответов субъектов обучения.

Благодарности

Работа опубликована при финансовой поддержке проекта **ITHEA XXI** Института информационных теорий и приложений FOI ITHEA Болгария www.ithea.org и Ассоциации создателей и пользователей интеллектуальных систем ADUIS Украина www.aduis.com.ua.

Библиография

[Belous N., 2004] Belous N., Voytovich I. Lifelong education conception using computer testing // Материалы VIII Международной конференции Украинской ассоциации дистанционного образования "Образование и виртуальность", 2004. – с. 307-313.

[Belous N., 2005] Белоус Н.В., Войтович И.В. Программный комплекс для проведения компьютерного и интерактивного обучения и тестирования знаний “КОДЭКС УМА” // Свидетельство про регистрацию авторского права №14030, Государственный департамент интеллектуальной собственности, 02.09.2005.

[Аванесов В.С., 1989] Аванесов В.С. Основы научной организации педагогического контроля в высшей школе. — М.: МИСиС, 1989. – 168 с.

[Белоус Н.В. и др., 2009] Белоус Н., Бондаренко М., Борисенко В., Семенец В., Куцевич И., Белоус И., Мележик О. Технология оценивания тестов в зависимости от типа и уровня сложности тестовых заданий на основе интегрированной модели // International Book Series "Information Science and Computing"№12. Methodologies and Tools of the Modern (e-) Learning // Supplement to International Journal "Information Technologies and Knowledge". ITHEA, SOFIA, 2009 с. 55-62

[Белоус Н.В., 2009] Белоус Н., Бондаренко М., Борисенко В., Семенец В., Куцевич И., Белоус И., Мележик О. Технология оценивания тестов в зависимости от типа и уровня сложности тестовых заданий на основе интегрированной модели // International Book Series "Information Science and Computing"№12. Methodologies and Tools of the Modern (e-) Learning // Supplement to International Journal "Information Technologies and Knowledge". ITHEA, SOFIA, 2009 с. 55-62

[Белоус Н.В., 2009] Белоус Н.В., Куцевич И.В. Дифференциальное оценивание знаний при дистанционном тестировании // Искусственный интеллект. – Донецк: ИПИИ. – 2009.– № 1. – С. 63–74

[Комплекс нормативних документів, 1998] Комплекс нормативних документів для розробки складових системи стандартів вищої освіти. Київ, 1998.

[Олейник Н.М., 1991] Олейник Н.М. Тест как инструмент измерения уровня знаний и трудности заданий в современной технологии обучения: Учеб. пособие по спец. курсу. – Донецк: ДонГУ, 1991. – 168 с.

Информация об авторах

Белоус Наталия *– заведующий лабораторией "Информационные технологии в системах обучения и машинного зрения (ITCO и M3)", к.т.н., профессор кафедры Программного обеспечения ЭВМ, Харьковский национальный университет радиоэлектроники, Харьков, Украина; e-mail: belous@kture.kharkov.ua*

Куцевич Ирина *– исследователь лаборатории "ITCO и M3"; Харьковский национальный университет радиоэлектроники, Харьков, Украина; e-mail: virishka@bk.ru*

Белоус Ирина *– исследователь лаборатории "ITCO и M3"; Харьковский национальный университет радиоэлектроники, Харьков, Украина; e-mail: belous@kture.kharkov.ua*

ПОСТРОЕНИЕ МАТЕМАТИЧЕСКОЙ МОДЕЛИ ОПТИМИЗАЦИИ ПРОИЗВОДСТВЕННОГО РАСПИСАНИЯ МЕТАЛЛУРГИЧЕСКОГО ПРОИЗВОДСТВА

Александр Турчин

Аннотация: *Рассматривается проблема оптимизации производственного портфеля металлургического предприятия в месячном разрезе. Предлагается двухуровневая математическая модель, в которой комбинируется как планирование по дням в течение месяца, так и суточное планирование производства. В результате возникает специальная задача оптимизации, для решения которой можно применять приближенные комбинаторные алгоритмы.*

Keywords: *combinatorial optimization, production portfolio metallurgical works.*

ACM Classification Keywords*: G.1.6 Nunerical Analysis; G.2.1 Discrete Mathematics.*

Conference*: The paper is selected from XVth International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.*

Постановка задачи оптимизации производственного расписания

Для постановки задачи оптимизации производственного расписания металлургического производства необходимо проанализировать основные технологические процессы сталеплавильного и листопрокатного производства, определить их значимые ограничения, сформулировать требования к функциональности системы автоматического формирования оптимального производственного расписания.

В производственных подразделениях металлургического комбината реализуются следующие основные процессы:

1. Литье слябов в машине непрерывного литья заготовки.
2. Разогрев слябов перед прокаткой в нагревательных печах.
3. Прокат слябов.
4. Порезка раската на отдельные листы и обработка листов.
5. Складирование готовой продукции и высылка заказов.

Соответственно, значимые ограничения, используемые при проектировании системы планирования, будут следующими:

1'. Суточная производительность машины непрерывного литья заготовки, объем одной плавки (вытопа).

2'. Суточная производительность, время выхода печи на рабочий режим.

3'. Суточная производительность, максимальный вес одного сляба.

4'. Суточный объем высылки, объем склада готовой продукции.

Следующим шагом является формирование месячного портфеля заказов, основными параметрами которого являются: вес готовой продукции – V (тонн), число отдельных заказов – n (единиц), общее количество листов проката – N (штук).

Каждый лист проката может быть охарактеризован определенным набором параметров (марка стали, метрические параметры, требования по отгрузке и др.), то есть каждому листу может быть приписан

уникальный набор параметров, позволяющий однозначно идентифицировать его на каждом этапе производства.

Требуется создать математическую модель проблемы, которую можно использовать в системах автоматического формирования производственного расписания, обеспечивающего выполнение принятых заказов наиболее оптимальным способом. Под производственным расписанием, подразумевается совокупность графиков выпуска продукции производственными подразделениями, в т.ч.:

- график литья слябов;
- график разогрева слябов;
- график проката слябов;
- график высылки заказов.

Основным критерием оценки эффективности создаваемого инструмента планирования является исполнимость генерируемого им производственного расписания, которая подразумевает соответствие создаваемых планов технологическим нормам и удовлетворение существующим ограничениям.

Также следует отметить, что на этапах формирования слябов возникают оптимизационные задачи, которые относятся к классу проблем оптимального раскроя (см., например, [1-3]). Отметим, что для решения одного из классов подобных задач успешно использовались *G*-алгоритмы [4].

Построение математической модели

Как видно из приведенной последовательности шагов 1-5, процесс выстраивания искомой последовательности листов является основополагающим по отношению ко всем другим процессам. Формирование последовательности листов однозначно определяет последовательность всех других технологических процессов – литья слябов, разогрева слябов и высылки сформированных заказов. При этом реализация указанных процессов не представляется сколь-нибудь сложной, т.к. перестановок каких либо элементов в сформированной последовательности не предполагается.

Упорядочение же последовательности листов из портфеля заказов является сложной оптимизационной задачей, учитывая размерности возникающих задач на перестановках, а также трудоемкость процесса вычисления значений целевой функции, поскольку для этого необходимо проимитировать выполняемые технологические процессы и операции по обработке листов, составляющих производственный портфель.

Предлагается такая двухуровневая математическая модель: на верхнем уровне происходит разделение слябов – и, как результат, листов – по дням планируемого периода с учетом ограничивающих условий, а на нижнем уровне – формирование и решение задач, которые относятся к суточному планированию.

Как по содержанию, так и с математической точки зрения центральной является задача распределения заказов между днями планового периода, поэтому сосредоточимся именно на этой задаче.

Введем обозначения:

$N = \{1, \ldots, n\}$ – множество номеров (идентификаторов) листов в портфеле заказов;

V_i – вес i-го листа, $i=1,\ldots, n$;

n – число всех листов;

N^k – листы, которые входят в k-й заказ, $k = 1, \ldots, K$;

K – число заказов;

S – число сформированных слябов, которые поступают на прокатное производство и для каждого из которых приписаны определенные листы заказа с учетом класса стали;

D^k – директивный срок исполнения k-го заказа (если для некоторого заказа $l \in \{1, \ldots, K.\}$ он отсутствует, то считаем $D^l = T$);

B – пропускная способность прокатного стана в сутки;

w^k – вес (значимость) k-го заказа;

m^s – класс (код) стали s-го сляба, $s = 1, \ldots, S$;

T – прогнозный период (для месячного планирования $T \in \{28, 29, 30, 31\}$),

Множество имеющихся в наличии слябов можно описать матрицей $Y=(y_{is})_{n\times S}$, где

$$y_{is} = \begin{cases} 1, \text{ если } i-\text{й лист входит в } s-\text{й сляб}, \\ 0-\text{в противном случае.} \end{cases}$$

Тут i=1,...,n;s=1,...,S . Понятно, что заказы формируют разбиение портфеля:

$$N = \bigcup_{k=1}^{K} N^k; \quad N^i \bigcap N^j = \varnothing \quad \forall\ i \neq j,\ i,j = 1, \ldots, K.$$

Для представления варианта решения введем матрицу $X = (x_{st})_{S\times T}$ размера $S \times T$, где

$$x_{st} = \begin{cases} 1, \text{ если } s-\text{й сляб изготовляется в } t-\text{й день}, \\ 0-\text{в противном случае.} \end{cases}$$

На основании имеющейся информации запишем матрицу $Z = (z_{it})_{n\times T}$, где

$$z_{it} = \begin{cases} 1, \text{ если лист } i \text{ изготовляется в день } t, \\ 0-\text{в противном случае.} \end{cases}$$

Для каждого заказа k, $k \in \{1, \ldots, K.\}$, введем множество, которое содержит все те дни планового периода, в которые исполняется этот заказ:

$$T^k(x) = \left\{ t \in \{1, \ldots, T\} : \sum_{i \in N^k} z_{it} \neq 0 \right\}.$$

Очевидно, что состав этих множеств зависит от варианта распределения слябов (листов) по дням, что и отображено в обозначении.

Пусть теперь

$$d_e^k = \max\{t : t \in T^k(X)\},$$

$$d_b^k = \min\{t : t \in T^k(X)\}.$$

Анализ современных металлургических производств показал, что наиболее важными критериями эффективности могут быть такие показатели.

Критерий 1. Равномерность распределения заказов между днями.

Оптимизация этого критерию преследует цель достижения ритмичности отгрузки готовых заказов, когда готовность к отгрузке определяется завершением изготовления последнего листа из данного заказа.

Критерий 2. Компактность срока (по дням) изготовления заказа отдельного заказчика.

Эта величина определяется как разница между последним и первым днем, когда исполняется данный заказ. Уменьшение числа дней, в которые изготавливается определенный заказа ("разброс дней заказа "), позволяет уменьшить потребность в складских помещениях.

Критерий 3. Максимизация ежедневной загрузки прокатного стана, то есть масса листов, которые катаются за сутки, должна в максимальной мере приближаться к показателю пропускной способности стана.

Указанный критерий учитывает величину отклонений плановых объемов проката от заданного ограничения на производительность прокатного стана.

Критерий 4. Минимизация штрафов за нарушение директивных сроков.

Критерий 5. Минимизация продолжительности исполнения плановых заказов.

Учитывается в том случае, когда загруженность (объем портфеля заказов) не является слишком большой по сравнению с плановым периодом.

Критерий 6. Число марок стали, которые прокатываются каждые сутки, должно быть минимальным.

Формализуем эти критерии минимизации следующим образом.

$$f_1(X) = \max_{i\in\{1,...,K-1\}}\{(d_e^{l_{i+1}} - d_e^{l_i})^2\} - \min_{i\in\{1,...,K-1\}}\{(d_e^{l_{i+1}} - d_e^{l_i})^2\}, \quad (1)$$

где предполагается, что числа $l_1,...,l_K$,$(l_i \in \{1,...,K\})$ такие, для которых $d_e^{l_1} \le d_e^{l_2} \le ... \le d_e^{l_K}$.

$$f_2(X) = \max_{1\le k\le K}\{(d_e^k - d_b^k)w^k\}. \quad (2)$$

$$f_3(X) = \max_{1\le t\le T}\{\psi_t\}, \quad (3)$$

где $\psi_t = \begin{cases} B - \sum_{s=1}^{S}(\sum_{i=1}^{n} y_{si}\cdot V_i)x_{st}, \text{если } B > \sum_{s=1}^{S}(\sum_{i=1}^{n} y_{si}\cdot V_i)x_{st}, \\ \beta(B - \sum_{s=1}^{S}(\sum_{i=1}^{n} y_{si}\cdot V_i)x_{st}) - \text{в противном случае.} \end{cases}$

$\beta > 0$ – штраф за перегрузку прокатного стана на единицу массы продукции.

$$f_4(X) = \sum_{k-1}^{K}\delta^k w^k, \quad (4)$$

где $\delta^k = \begin{cases} d_e^k - D^k, \text{если } d_e^k > D^k, \\ 0 - \text{в противном случае.} \end{cases}$.

$$f_5(X) = \max_{1\le k\le K} d_e^k. \quad (5)$$

$$f_6(X) = \sum_{t=1}^{T}\mu_t(X), \quad (6)$$

где $\mu_t(X)$ – число разных марок стали на множестве слябов $\{r\in\{1, \ldots, S\}: x_{rt}=1\}$.

Интегральную целевую функцию задачи представим в виде свертки указанных частичных критериев:

$$f(X) = \sum_{i=1}^{6}\alpha_i f_i(X), \quad (7)$$

где $\alpha_i > 0$ – весовые коэффициенты, устанавливаемые экспертным путем.

Ограничительные условия задачи:

$$\sum_{s=1}^{S} y_{si} = 1, \quad i = 1,...,n; \quad (8)$$

$$\sum_{i=1}^{n} y_{si} = 1, \quad s = 1,...,S; \quad (9)$$

$$\sum_{s=1}^{S} x_{st} = 1, \quad t = 1,...,T; \tag{10}$$

$$\sum_{t=1}^{T} x_{si} = 1, \quad s = 1,...,S; \tag{11}$$

$$\sum_{i=1}^{n} z_{it} = 1, \quad t = 1,...,T; \tag{12}$$

$$\sum_{t=1}^{T} z_{it} = 1, \quad i = 1,...,n; \tag{13}$$

$$d_e^k \geq d_b^k, \quad k = 1,...,K. \tag{14}$$

Задача состоит в поиске такой матрицы X_*, которая минимизирует функцию (7), т.е.

$$X_* = \arg\min_{X \in \Pi} f(X), \tag{15}$$

где область *П* определяется условиями (8)–(14).

Заключение

Для рассмотренной проблемы оптимизации производственного процесса металлургического производства предложена двухуровневая математическая модель комбинаторной оптимизации. Для решения сформулированной задачи был предложен комбинированный алгоритм, состоящий из алгоритма последовательного построения начального варианта решения и итерационного алгоритма ускоренного вероятностного моделирования [5]. Он был успешно применен на практике на одном из ведущих предприятий металлургического комплекса.

Благодарности

Работа опубликована при финансовой поддержке проекта **ITHEA XXI** Института информационных теорий и приложений FOI ITHEA Болгария www.ithea.org и Ассоциации создателей и пользователей интеллектуальных систем ADUIS Украина www.aduis.com.ua.

Список литературы

[1]. Сергиенко И.В. Математические модели и методы решения задач дискретной оптимизации. – К.: Наукова думка, 1985. – 384 с.

[2]. Пападимитриу Х., Стайглиц К. Комбинаторная оптимизация. Алгоритмы и сложность. – М.: Мир, 1985. – 512 с.

[3]. Гуляницкий Л.Ф. Решение задач комбинаторной оптимизации алгоритмами ускоренного вероятностного моделирования // Компьютерная математика. – Киев: Ин-т кибернетики им. В.М. Глушкова НАН Украины, 2004. – №1. – С. 64–72.

[4]. Гуляницкий Л.Ф., Гобов Д.А. О сравнении эффективности алгоритмов решения одного класса задач размещения прямоугольников на полубесконечной ленте //Компьютерная математика. – 2003. – №1. – С. 88-97.

[5]. Hulianytskyi L. F., Turchin A. Y. One class of stochastic local search algorithms // Int. J. "Information theories & applications". – 2008. – 15, N 3. – P. 245-252

Сведения об авторе

Александр Турчин *– Украинский центр оптимальных решений в информатике и связи, руководитель проекта. Украина, Киев, 03067, ул. Выборгская, 99, e-mail: turchin@ua.fm*

МНОГОКРИТЕРИАЛЬНЫЕ ЗАДАЧИ ЛЕКСИКОГРАФИЧЕСКОЙ ОПТИМИЗАЦИИ С ЛИНЕЙНЫМИ ФУНКЦИЯМИ КРИТЕРИЕВ НА НЕЧЕТКОМ МНОЖЕСТВЕ АЛЬТЕРНАТИВ

Наталия Семенова, Людмила Колечкина, Алла Нагорная

***Аннотация**: Рассматривается многокритериальная комбинаторная задача лексикографической оптимизации на нечетком множестве альтернатив с линейными критериями. Исследуются свойства области допустимых решений. Излагается один из возможных подходов к решению многокритериальной комбинаторной задачи с линейными целевыми функциями на нечетком множестве альтернатив.*

***Ключевые слова**: многокритериальная оптимизация, дискретная оптимизация, нечеткое комбинаторное множество, комбинаторное множество перестановок, Парето-оптимальные решения, слабо, строго эффективные решения, нечеткое множество альтернатив, нечеткое мультимножество, функция принадлежности.*

***ACM Classification Keywords**: G2.1 Combinatorics (F2.2), G1.6 Optimization*

***Conference**: The paper is selected from XVth International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.*

Введение

При решении многих прикладных задач довольно часто исходная информация для описания математических моделей задана нечетко. Такие ситуации отображают недостаток информации для решения задачи, так как при нечетких условиях и критериях становится проблемным принятие решения. При моделировании реальных задач нечеткость проявляется в форме описания функций и параметров, от которых они зависят.

Нечеткие множества широко используются в различных применениях искусственного интеллекта, теории распознавания образов, принятия решений и др. Во многих практических задачах исследования операций, проектирования сложных систем, возникает необходимость принятия решения с учетом нескольких критериев оптимальности. В то же время достаточно распространенными на практике являются задачи многокритериального выбора, в которых задано конечное множество альтернатив и альтернативы могут оцениваться как количественно, так и качественно. Особенность многокритериальных задач, как способа моделирования, состоит в том, что в условиях многокритериальности выбор наиболее целесообразного решения осуществляется из множества несравнимых альтернатив. Проблема нахождения этого множества имеет большое практическое и теоретическое значение. Кроме того, в реальных задачах экономики мощность множества альтернатив очень велика, что делает проблему принятия решения достаточно сложной. В работе [2] была дана математическая формулировка и приведено аксиоматическое обоснование известного еще с XIX в. принципа Эджворта–Парето для случая четкого отношения предпочтения лица, принимающего решение. Этот принцип является основополагающим при выборе наилучших решений в экономике и технике в тех случаях, когда приходится учитывать сразу несколько целевых функций (критериев). Выяснилось, что он не универсален, а справедлив лишь при

решении определенного (хотя и достаточно широкого) класса задач многокритериального выбора. За пределами этого класса его применение рискованно или же вообще недопустимо.

В настоящей работе принцип Эджворта-Парето распространяется на более широкий класс задач многокритериального выбора, в которых множество возможных решений является нечетким.

Различным аспектам теории задач векторной оптимизации посвящены работы многих ученых [1-6]. Как известно, задачи комбинаторной оптимизации, в том числе многокритериальные, могут быть сведены к задачам целочисленного программирования, однако это не всегда оправдано, потому что при этом теряется возможность учета комбинаторных свойств задачи [1], следовательно является целесообразной разработка новых подходов решения задач комбинаторного типа на основе исследования свойств области допустимых решений.

В настоящее время достаточно важны модели, учитывающие нечеткие параметры, как в целевых функциях, так и в ограничениях, заданных на комбинаторных множествах.

В данной статье рассматривается многокритериальная задача с линейными целевыми функциями, а также ее лексикографическая постановка, на нечетко заданной комбинаторной области допустимых решений. Исследованы свойства нечетко заданного комбинаторного множества перестановок, являющегося допустимой областью задачи. Предложен подход к решению изученной многокритериальной задачи.

1. Предварительные сведения. Основные понятия и определения

Нечеткие подмножества отличаются от обычных, или четких множеств тем, что в нечетком подмножестве степень принадлежности элемента множеству может быть любым числом единичного интервала $[0,1]$, а не только одним из двух значений элементов множества $\{0,1\}$, как в случае индикаторов обычных подмножеств. Это свойство нечетких подмножеств обеспечивает возможность теоретико–множественного представления реальных неточных понятий, в которых переход от непринадлежности к принадлежности происходит постепенно. Возможность формализации понятий такого типа оказалась очень полезной при разработке принципов искусственного интеллекта ЭВМ, моделирующих процессы мышления человека и др. Независимо от того, что использовать – нечеткие или четкие подмножества, определение степеней принадлежности опирается на некоторые субъективные критерии человека, принимающего решения. Однако если при определении четких подмножеств производится выбор одного из двух чисел – нуля или единицы, то для определения нечетких подмножеств возможности выбора степеней принадлежности намного разнообразнее. В ряде случаев определение подходящих значений степеней принадлежности элементов нечетких множеств приводит к значительным трудностям в работе с нечеткими понятиями.

В данной работе исследуется задача векторной оптимизации, заданная на комбинаторном множестве перестановок, следовательно, необходимо учесть то, что выпуклой оболочкой этого множества является общий перестановочный многогранник, множеством вершин которого есть рассматриваемое комбинаторное множество [6-8]. Изученные свойства комбинаторных множеств и их выпуклых оболочек дают возможность находить решение дискретной комбинаторной задачи как оптимизационной задачи на непрерывном допустимом множестве. На основании установленной взаимосвязи между многокритериальными задачами на комбинаторных множествах и оптимизационными задачами на непрерывном допустимом множестве, появляется возможность применять классические методы непрерывной оптимизации к решению векторных задач на различных комбинаторных множествах, а также развивать новые подходы их решения.

Определим, для дальнейшего изложения обобщение понятий n-выборки, мультимножества и комбинаторного множества перестановок на случай нечетко заданной информации.

Определение 1. Нечетким мультимножеством $\tilde{X}$, заданным на универсальном мультимножестве X, называется совокупность пар $\left(x,\mu_{\tilde{X}}(x)\right)$, где $x\in X$, а $\mu_{\tilde{X}}(x)$ – функция: $X\to[0,1]$, которая называется функцией принадлежности мультимножеству $\tilde{X}$.

Значение $\mu_{\tilde{X}}(x)$ для конкретного x называется степенью принадлежности этого элемента нечеткому мультимножеству $\tilde{X}$.

Согласно определению, обычные множества образуют подкласс нечетких множеств. Над нечеткими множествами, так же как и над обычными множествами, выполняется ряд операций, таких как объединение, пересечение, декартово произведение, разность и др. Эти же операции имеют место и для нечетких мультимножеств.

Пусть задано нечеткое мультимножество $\tilde{A}=\{a_1,\mu_{\tilde{A}}(a_1),a_2,\mu_{\tilde{A}}(a_2)...,a_q,\mu_{\tilde{A}}\left(a_q\right)\}$, его основание $S(\tilde{A})=\left\{\left(e_1,\mu_{\tilde{A}}(e_1)\right),\left(e_2,\mu_{\tilde{A}}(e_2)\right),...,\left(e_k,\mu_{\tilde{A}}(e_k)\right)\right\}$, где $e_j\in R^1\ \forall j\in N_k=\{1,...,k\}$, кратность элементов $k(e_j)=r_j, j\in N_k, r_1+r_2+...+r_k=q,$

$$\mu_{\tilde{A}}(e_i)=\min\left\{\mu_{\tilde{A}}\left(a_{i_j}\right)\middle| a_{i_j}=a_{i_t}, j\neq t,\forall i,j,t\in N_q\right\}.$$

Упорядоченной нечеткой n-выборкой из нечеткого мультимножества A называется набор

$$a=\left(\left(a_{i_1},\mu_{\tilde{A}}\left(a_{i_1}\right)\right),\left(a_{i_2},\mu_{\tilde{A}}\left(a_{i_2}\right)\right),\ldots,\left(a_{i_n},\mu_{\tilde{A}}\left(a_{i_n}\right)\right)\right), \qquad (1)$$

где $a_{i_j}\in\tilde{A}\ \ \forall i_j\in N_k,\ \ \forall j\in N_k,\ i_s\neq i_t$, если $s\neq t\ \ \forall s\in N_k,\ \ \forall t\in N_k$.

Определение 2. Нечеткое подмножество $P(\tilde{A})$, элементами которого являются нечеткие n-выборки вида (1) из нечеткого мультимножества $\tilde{A}$, называется нечетким евклидовым комбинаторным множеством, если для произвольной пары его элементов $a=\left(\left(a_1,\mu_{\tilde{A}}(a_1)\right),\left(a_2,\mu_{\tilde{A}}(a_2)\right),...,\left(a_n,\mu_{\tilde{A}}(a_n)\right)\right)$ и $b=(b_1,\mu_{\tilde{A}}(b_1),b_2,\mu_{\tilde{A}}(b_2),...,b_n,\mu_{\tilde{A}}(b_n))$ выполняется условие: $a\neq b\Leftrightarrow\left(\exists j\in N_n:\left(a_j,\mu_{\tilde{A}}(a_j)\right)\neq,\left(b_j,\mu_{\tilde{A}}(b_j)\right)\right)$, то есть множество $P(\tilde{A})$ имеет такое свойство: два элемента множества $P(\tilde{A})$ отличны один от другого, если они независимо от других отличий различаются порядком размещения символов, которые их образовывают и степенью принадлежности нечеткому множеству $P(\tilde{A})$.

Дадим определение нечеткого множества перестановок.

Определение 3. Нечеткое множество перестановок с повторениями из n действительных чисел, среди которых k различных, назовем общим нечетким множеством перестановок т.е. множеством упорядоченных n-выборок вида (1) из нечеткого мультимножества $\tilde{A}$, если $n=q>k$, и обозначим $P_{nk}(\tilde{A})$. При равенстве $n=q=k$ получаем нечеткое множество перестановок без повторений.

Будем рассматривать элементы множества перестановок как точки арифметического евклидового пространства R^n.

Известно, что каждый элемент множества $P_{nk}(A)$ является упорядоченным набором *n* действительных чисел, из которых *k* различны. Не теряя общности, упорядочим элементы мультимножества A следующим образом:

$$a_1 \le a_2 \le \ldots \le a_n. \tag{2}$$

Наряду с классическим перестановочным многогранником, введенным Радо [14], опишем общий перестановочный многогранник $\Pi_{nk}(A)$ [15, 16], являющийся выпуклой оболочкой общего множества перестановок $P_{nk}(A)$:

$$\sum_{j=1}^{n} x_j \le \sum_{j=1}^{n} a_j,\ \sum_{j=1}^{i} x_{\alpha_j} \ge \sum_{j=1}^{i} a_j\ , \tag{3}$$

а $\alpha_j \in N_n,\ \alpha_j \ne \alpha_t,\ \forall j \ne t,\ \forall j,\ t \in N_i,\ \forall i \in N_n$, а $P_{nk}(A) = \operatorname{vert} \Pi_{nk}(A)$.

Определение 4. Выпуклой комбинацией нечетких множеств $A_1, A_2, \ldots A_n$ в R^n называется нечеткое множество A с функцией принадлежности вида

$$\mu_A(x) = \sum_{i=1}^{n} \lambda_i \mu_i(x), \text{ где } \lambda_i \ge 0,\ i \in N_n,\ \sum_{i=1}^{n} \lambda_i = 1.$$

Нечеткий выпуклый многогранник $\Pi_{nk}(A)$ также можно представить, как выпуклую комбинацию его вершин.

2.Постановка задачи

Обычно под задачей математического программирования понимают задачу отыскания экстремума некоторой целевой функции на допустимом множестве альтернатив. С помощью целевой функции формально представляется одно из основных свойств альтернатив: ценность, полезность, стоимость и др. Нечеткость в постановке задачи математического программирования может содержаться как в описании множества альтернатив, так и в описании целевой функции. Различные формы описания исходной информации обусловливают существование различных формулировок задач нечеткого математического программирования: а) задача достижения нечетко поставленной цели при нечетких ограничениях; б) задача нечеткого математического программирования при нечетком множестве допустимых альтернатив; в) нечеткий вариант стандартной задачи математического программирования со «смягчением» целевой функции и / или ограничений, где вместо задачи оптимизации решается задача удовлетворения цели и соответствующие неравенства для целевой функции и ограничений могут нарушаться; г) задача оптимизации с нечеткими коэффициентами и др.

В данной статье исследуемая задача состоит в максимизации векторной функции F на нечетком евклидовом комбинаторном множестве $\tilde{X}$.

Рассматривается многокритериальная задача комбинаторной оптимизации следующего вида: $Z(F,X): \max\left\{F(x) \mid x \in \tilde{X} \subset R^n\right\},\ F(x) = (f_1(x), \ldots, f_l(x)),\ f_i : R^n \to R,\ i \in N_l,\ \tilde{X} \ne \varnothing,\ \tilde{X} -$ нечеткое подмножество множества $X = \operatorname{vert} \Pi_{nk}(A) \cap D, \quad \Pi_{nk}(A) = \operatorname{conv} P_{nk}(A), \quad P_{nk}(A)$ – комбинаторное множество перестановок, $D \subset R^n$ – выпуклый многогранник, нечеткое подмножество

$\tilde{X} = \{x, \mu_{\tilde{X}}(x)\}$, где $x \in X$, а $\mu_{\tilde{X}}(x): X \to [0,1]$ – функция принадлежности множеству X. $\tilde{X}$ будем называть нечетким множеством альтернатив.

Под максимизацией будем понимать выбор нечеткого подмножества R нечеткого множества $\tilde{X}$, которому отвечает наибольшее значение как векторной функции F, так и функции принадлежности $\mu_{\tilde{X}}(x)$ нечеткому множеству альтернатив. Эти альтернативы в задачах многокритериальной оптимизации в зависимости от способа их сравнения называются эффективными (оптимальными по Парето), слабо эффективными (по Слейтеру), строго эффективными (по Смейлу) и соответственно обозначаются: $P(F,X), Sl(F,X), Sm(F,X)$. Напомним [14], что альтернатива $x^* \in \tilde{X}$ называется эффективной, если не существует иной альтернативы $x \in \tilde{X}$ такой, что $F(x) \geq F(x^*)$, $\mu_D(x) \geq \mu_D(x^*)$ и хотя бы одно неравенство строгое; слабо эффективной, если $\exists x \in \tilde{X}$: $F(x) > F(x^*)$, и строго эффективной, если $\exists x \in \tilde{X}: x \neq x^*$, $F(x) \geq F(x^*)$.

Из определений следует, что $Sm(F,\tilde{X}) \subset P(F,\tilde{X}) \subset Sl(F,\tilde{X})$.

Исходную задачу $Z(F,X)$ представим в виде $(l+1)$ – критериальной задачи:

$$F(x) \to \max, \mu_{\tilde{X}}(x) \to \max, x \in X.$$

Под решением задачи с нечетким множеством альтернатив понимаем нечеткое множество с функцией принадлежности: $\mu(x) = \left\{\mu_{\tilde{X}}(x) \middle| x \in P(F,\tilde{X}); 0 \middle| x \notin P(F,\tilde{X})\right\}$.

Таким образом, нечеткое подмножество решений будет включать в себя те и только те альтернативы универсального множества X, которые дают значения векторной функции $F(x)$ и функции принадлежности $\mu_{\tilde{X}}(x), x \in X$, не улучшаемые одновременно.

Следует отметить, что нечеткий вариант этой задачи означает, что ограничения смягчаются, то есть допускается возможность их нарушения с той или иной степенью.

Пусть $P(\alpha)$ - множество всех эффективных альтернатив $(l+1)$- критериальной задачи:

$y_i \to \max, i \in N_l$, $\mu_D(x) \to \max$, $F(x,y) \geq \alpha$, $x \in X$, $y = (y_1, ..., y_l) \in Y$.

Тогда, решением векторной задачи нечеткой оптимизации с нечетким множеством альтернатив и со степенью недоминирующих альтернатив, не меньше α, называется нечеткое множество с функцией принадлежности вида:

$$\mu_\alpha(x) = \begin{cases} \mu_D(x), x \in P(\alpha), \\ 0, \qquad x \notin P(\alpha). \end{cases}$$

Таким образом, нечеткое множество решений исходной задачи будет включать в себя те и только те альтернативы со степенью недоминированности не меньше α, которые будут эффективными как по оценкам альтернатив $y_i, i \in N_l$, так и по функции принадлежности $\mu_D(x)$ нечеткому множеству альтернатив. Выбор некоторой конкретной из них альтернативы осуществляется с помощью методов многокритериальной оптимизации.

Более того, вместо задачи максимизации можно поставить задачу достижения некоторого наперед заданного значения функции цели, соответствующего удовлетворению исходной цели.

3.Подходы к решению задачи

Нередко на практике теория оптимизации применяется к неточным моделям, где нет никаких оснований задавать коэффициенты в виде точно определенных чисел. Такое искусственное сужение априорной информации может привести к искажению конечных результатов.

Разработка методов решения поставленной задачи $Z(F,X)$ в условиях нечеткой определенности предполагает знание и использование результатов операций нахождения суммы, произведения, минимума и максимума нечетких величин. При сравнении двух нечетких величин необходимо определение равенства этих величин.

Под нечетким числом здесь будем понимать нечеткое множество с областью определения в виде интервала действительной оси R^1. Множество всех нечетких чисел, определенных на R^1, обозначим через $\tilde{R}^1$. Пусть x и y - два нечетких числа с носителями $S_x=(a_1,a_2)$ и $S_y=(a_1,a_2)$ соответственно; $a_2>a_1, b_2>b_1$; $g:R^1\times R^1\to R^1$ - некоторая функция. Тогда согласно принципу обобщения нечеткое число $D=g(x,y)$ определяется функцией принадлежности

$$\mu_D(z)=\sup_{\substack{g(a,b)=z\\ a\in S_x,\ b\in S_y}} \min\{\mu_x(a),\mu_y(b)\}. \qquad (4)$$

Пусть $\otimes$ - одна из четырех арифметических операций: +, -, ·, /; $g(a,b)=a\otimes b$. Тогда формула (4) определяет результат арифметической операции $\otimes$ над нечеткими числами x и y. Если $g(\cdot)$ - функция не двух, а n аргументов, то принцип обобщения формулируется аналогично (4).

При сравнении двух нечетких величин необходимо определение равенства этих величин.

Определение 4. Две нечетких величины (два числа) $(x_1,\mu_1(x_1))$ и $(x_2,\mu_2(x_2))$ будем считать равными, если $x_1=x_2$ и $\mu_1(x_1)=\mu_2(x_2)$.

Определение 5. Если выполняется условие $x_1\geq x_2$, $\mu_1(x_1)\geq\mu_2(x_2)$ и одно из этих неравенств строгое, то нечеткая величина $(x_1,\mu_1(x_1))$ больше нечеткой величины $(x_2,\mu_2(x_2))$.

Разработан подход к решению задачи $Z(F,X)$, основанный на методе последовательных уступок. При решении многокритериальной задачи этим методом вначале производится качественный анализ относительной важности частных критериев. Особенностью данного метода есть то, что критерии задачи должны быть предварительно упорядочены по убыванию их важности, так что главным является критерий $f_1(x)$, менее важен $f_2(x)$, затем следуют остальные частные критерии $f_3(x),...,f_l(x)$. Максимизируется первый по важности критерий $f_1(x)$ и определяется его наибольшее значение f_1^*. Затем назначается величина допустимого снижения (уступки) $\Delta_1\geq 0$ критерия $f_1(x)$ и ищется наибольшее значение f_2^* второго критерия $f_2(x)$ при условии, что значение первого критерия должно быть не меньше, чем $f_1^*-\Delta_1$. Снова назначается величина уступки $\Delta_2\geq 0$, но уже по второму критерию, которая вместе с первой используется при нахождении условного максимума третьего критерия, т.д. Наконец, максимизируется последний по важности критерий $f_l(x)$ при условии, что значение каждого критерия $f_r(x)$ из $l-1$ предыдущих должно быть не меньше соответствующей величины $f_r^*-\Delta_r$, получаемые в итоге стратегии считаются оптимальными.

Таким образом, выбор решения задачи осуществляется путем выполнения многошаговой процедуры и состоит в последовательном включении ограничений задачи $Z(F,X)$ и учете структурных особенностей его допустимой области. Оптимальным считается решение, являющееся решением последней задачи из следующей последовательности задач:

$$f_1^* = \max\{f_1(x) | x \in X\},\ f_2^* = \max\left\{f_2(x) \middle| x \in X, f_1(x) \geq f_1^* - \Delta_1\right\},...,$$

$$f_l^* = \max\left\{f_l(x) \middle| x \in X, f_{r-1}(x) \geq f_{r-1}^* - \Delta_{r-1}, r \in N_l\right\}.$$

Следует отметить, что в случае, когда все Δ_r – нули, метод последовательных уступок выделяет только лексикографически оптимальные альтернативы; эти альтернативы доставляют наибольшее на множестве допустимых значений решение первому по важности критерию $f_1(x)$. Поэтому величины уступок, назначенные для многокритериальной задачи, можно рассматривать как своеобразную меру отклонения приоритета (степени относительной важности) частных критериев от жесткого, лексикографического.

Опишем применение этого метода для получения лексикографически оптимальных решений при нечетко заданном допустимом комбинаторном множестве перестановок.

Суть метода заключается в выделении сначала множества альтернатив с наилучшей оценкой по наиболее важному критерию. Если такая альтернатива одна, то она считается лучшей. Если их несколько, то из подмножества альтернатив, выделенного на предыдущем шаге, выделяются те альтернативы, которые имеют лучшую оценку по следующему критерию в ряду упорядочения их по важности и т.д.

Обозначим $f_{ij} = f_i(x_j)$ – нечеткие оценки альтернатив $x_j, j \in N_n$, по критериям $f_i,\ i \in N_l$.

Применение метода при нечеткой исходной информации сводится к следующим действиям:

1. Критерии упорядочить по важности: $f_1(x)$, $f_2(x)$,..., $f_l(x)$.

2. С согласия ЛПР назначить уровень $\alpha \in [0,1]$, для которого определяется множество лучших альтернатив в соответствии с пунктами 3 – 5:

3. Определить нижнюю (l) и верхнюю (u) границы $\alpha-$уровневых подмножеств для оценки альтернатив по рассматриваемому критерию: $l(f_{ij}) = \min\limits_{\mu_{f_{ij}}(x) \geq \alpha} x,\ u(f_{ij}) = \max\limits_{\mu_{f_{ij}}(x) \geq \alpha} x$.

4. Для каждой пары альтернатив $z, y \in \tilde{X}$ вычислить показатели взаимного превышения критериальных оценок $h_{zy}(z > y)$ и $h_{yz}(y > z)$:

а) если оценки таковы, что $f_{iy}^{\alpha} \subseteq f_{iz}^{\alpha}$, то

$$h_{zy} = \frac{u(f_{iz}) - u(f_{iy})}{u(f_{iz}) - l(f_{iz})},\ h_{yz} = \frac{l(f_{iy}) - l(f_{iz})}{u(f_{iz}) - l(f_{iz})}, \quad (5)$$

где $z, y \in \tilde{X}$;

б) если оценки не пересекаются и $\exists\, f_0^z \in S_{f_{iz}} : \forall f^y \in S_{f_{iy}} \quad f_0^z > f^y$, то

$$h_{zy} = 1 - \frac{u(f_{iy}) - l(f_{iz})}{\max\{r(z), r(y)\}},\ h_{yz} = 0, \quad (6)$$

где $r(z) = u(f_{iz}) - l(f_{iz})$, $r(y) = u(f_{iy}) - l(f_{iy})$;

в) если оценки не пересекаются и $\forall f^z \in S_{f_{iz}}, \forall f^y \in S_{f_{iy}} : f^z > f^y$, то

$$h_{zy} = 1, h_{yz} = 0. \quad (7)$$

5. Вычислить показатели μ_{ij}^D принадлежности j-й альтернативы множеству лучших (D-множеству) по i-му критерию альтернатив $\mu_{ij}^D = \sup\{0, (\max\limits_{\substack{y \in X \\ y \neq j}} h_{jy} - \max\limits_{\substack{y \in X \\ y \neq j}} h_{yj})\}$,

где h_{jy}, h_{yj} вычислены по формулам (5) – (6) для i-го критерия.

6. Если множество лучших по рассматриваемому критерию альтернатив содержит ровно одну альтернативу с $\mu_{ij}^D \geq \alpha$, то она считается лучшей. Если это множество содержит более чем одну альтернативу с $\mu_{ij}^D \geq \alpha$, то выбирается следующий по важности критерий и повторяются пункты 3 − 5. Если все критерии просмотрены и множество лучших по рассматриваемому критерию альтернатив содержит более одной альтернативы и $\alpha < 1$, то α можно увеличить и перейти к пункту 3. Если $\alpha = 1$, то окончательный выбор лучшей альтернативы предоставляется ЛПР.

Рассмотрим еще один простой метод решения задачи $Z(F, X)$, который является развитием метода идеальной точки на случай нечетко заданных альтернатив, при отсутствии информации о предпочтениях на множестве критериев.

Пусть заданы или вычислены нечеткие оценки $f_i(x_j)$ альтернатив $x_j, j \in N_n$, по критериям $f_i, i \in N_l$.

Рассмотрим метод выбора альтернатив по обобщенному критерию пессимизма (максимина). Он состоит в выполнении следующих шагов:

1. Для каждого критерия вычислить нечеткую максимальную критериальную оценку $f_{i\max} = \tilde{\max}\left\{f_i(x_j) \mid x_j \in X, j \in N_n\right\}, i \in N_l$, по каждому из критериев оценки.

2. Вычислить приведенные нормализованные оценки альтернатив по критериям $f_{ij}^* = f_i(x_j) / f_{i\max}, x_j \in X, j \in N_n$.

3. Вычислить минимальную критериальную оценку для каждой альтернативы $f_{j\min}$, определяемую как $f_{j\min} = \tilde{\min}\left\{f_{ij}^* \mid i \in N_l\right\}, j \in N_n$.

4. Определить обобщенный максимум найденных минимальных оценок $f_{0\max} = \tilde{\max}\left\{f_{j\min} \mid j \in N_n\right\}$.

5. Оценить степень сходства $f_{0\max}$ с каждой из оценок $f_{j\min}$. В качестве показателя сходства нечетких чисел может быть использована величина $\sigma_j = \sum\limits_{z \in [0,1]} \left|\mu_{f_{0\max}}(z) - \mu_{f_{j\min}}(z)\right|, j \in N_n$.

6. Выбираем альтернативу с максимальным индексом $\sigma_j, j \in N_n$.

Если альтернатива выбирается по максимаксному принципу, то на шаге 3 следует вместо $f_{j\min}$ вычислить $f_{j\max} = \max\left\{f_{ij}^{*} \mid i \in N_l\right\}, j \in N_n$, и использовать в остальных шагах алгоритма вместо $f_{j\min}$ значения $f_{j\max}$.

Выводы

В данной работе в результате проведенного исследования векторной комбинаторной задачи, основанного на использовании информации о выпуклой оболочке допустимой области, изучении свойств многогранников, вершины которых определяют заданное комбинаторное множество перестановок, разработан и обоснован метод решения сложных многокритериальных задач на нечетко заданном комбинаторном множестве. Использование структурных свойств комбинаторных многогранников дает возможность разрабатывать эффективные алгоритмы решения новых классов векторных задач комбинаторной оптимизации.

Благодарности

Работа опубликована при финансовой поддержке проекта **ITHEA XXI** Института информационных теорий и приложений FOI ITHEA Болгария www.ithea.org и Ассоциации создателей и пользователей интеллектуальных систем ADUIS Украина www.aduis.com.ua.

Библиография

[1] Сергиенко И.В. Математические модели и методы решения задач дискретной оптимизации. – Киев: Наук. думка, 1988. – 472 с.

[2] Сергиенко И.В., Каспшицкая М.Ф. Модели и методы решения на ЭВМ комбинаторных задач оптимизации. – Киев: – Наук. думка, 1981. – 287 с.

[3] Сергиенко И.В., Шило В.П. Задачи дискретной оптимизации: проблемы, методы решения, исследования. – К.: Наук. думка, 2003. – 264 с.

[4] Сергиенко И.В., Козерацкая Л.Н., Лебедева Т.Т. Исследование устойчивости и параметрический анализ дискретных оптимизационных задач. – Киев: Наук. думка, 1995. – 170 с.

[5] Сергиенко И.В., Лебедева Т.Т., Семенова Н.В. О существовании решений в задачах векторной оптимизации // Кибернетика и системный анализ. – 2000. - № 6. – С. 39 – 46.

[6] Лебєдєва Т.Т., Семенова Н.В., Сергієнко Т.І. Умови оптимальності та розв'язуваності в задачах лінійної векторної оптимізації з опуклою допустимою множиною // Доповіді НАНУ. – 2003. – № 10. – С. 80–85.

[7] Лебедева Т.Т., Семенова Н.В Сергиенко Т.И. Устойчивость векторных задач целочисленной оптимизации: взаимосвязь с устойчивостью множеств оптимальных и неоптимальных решений // Кибернетика и системный анализ. – 2005. – № 4. – С. 90–100.

[8] Семенова Н.В., Колечкина Л.Н., Нагорная А.Н. Подход к решению векторных задач дискретной оптимизации на комбинаторном множестве перестановок // Там же. – 2008. – № 3. – С. 158–172.

[9] Semenova N.V., Kolechkina L.M., Nagirna A.M. Vector combinatorial problems in a space of combinations with linear fractional functions of criteria // Inform. Theories and Appl. – 2008. – 15. − P. 240 – 245.

[10] Семенова Н.В., Колечкина Л.Н., Нагорная А.Н. Решение и исследование векторных задач комбинаторной оптимизации на множестве полиперестановок // Проблемы управления и информатики. – 2008. – № 6. – С. 26-41.

[11] Семенова Н. Векторные задачи на комбинаторном множестве полиразмещений: условия оптимальности и подход к решению // Inform. Theories and Knowledge.− 2008. − 2.– P. 187 – 195. (Intern. Book Series "Information science and computing", №7).

[12] Колечкина Л. Многокритериальные задачи на комбинаторном множестве полиразмещений: структурные свойства решений // Ibid. – 2008. − 2. − P. 180– 186.

[13] Подиновский В.В., Ногин В.Д. Парето–оптимальные решения многокритериальных задач. – М.: Наука, 1982. – 256 с.

[14] Rado R. An inequality. – J. London Math. Soc., 1952, 27.

[15] Емеличев В.А., Ковалев М.М., Кравцов М.К. Многогранники, графы, оптимизация. – М.: Наука, 1981. – 344 с.

[16] Стоян Ю.Г., Яковлев С.В. Математические модели и оптимизационные методы геометрического проектирования. – Киев: Наук. думка, 1986. – 265 с.

[17] Ємець О. О., Колєчкіна Л. М. Задачі комбінаторної оптимізації з дробово-лінійними цільовими функціями.– Київ: Наук. думка. – 2005. – 118 с.

[18] Семенова Н.В. Условия оптимальности в векторных задачах комбинаторной оптимизации // Теорія оптимальних рішень. – 2008. – № 7. – С. 153-160.

[19] Семенова Н.В., Колєчкіна Л.М., Нагірна А.М. Розв'язання багатокритеріальних задач оптимізації на множині поліперестановок // Доповіді НАНУ. – 2009. – № 2. – С. 41– 48.

[20] Колечкина Л.Н. Оптимальные решения многокритериальных комбинаторных задач на размещениях // Теорія оптимальних рішень. – 2007. – № 6. – С. 67–73.

[21] Обработка нечеткой информации в системах принятия решений / А.Н. Борисов, А.В. Алексеев, Г.В. Меркурьева и др. – М.: Радио и связь, 1989. – 304с.

[22] Нечеткие множества в моделях управления и искусственного интеллекта / А.Н. Аверкин, И.З. Батыршин, А.Ф. Блишун и др. / Под ред. Д.А. Поспелова. – М.: Наука, 1986. – 312с.

[23] Орловский С.А. Проблемы принятия решений при нечеткой исходной информации. – М.: Наука, 1981. – 208с.

[24] Bellman R., Zade L.A. Decision-making in fuzzy environment. – Management Science, 1970, v.17, P. 141-162.

Информация об авторах

Наталия Владимировна Семенова *– Институт кибернетики имени В.М. Глушкова НАН Украины, канд. физ.-мат. наук, старший научный сотрудник, 03680 МСП Киев 187, проспект академика Глушкова, 40, Украина; e-mail: nvsemenova@meta.ua*

Людмила Николаевна Колечкина *– Институт кибернетики имени В.М. Глушкова НАН Украины, канд. физ.-мат. наук, доцент, докторант, 03680 МСП Киев 187, проспект академика Глушкова, 40, Украина; e-mail: ludapl@ukr.net*

Алла Николаевна Нагирна *– Институт кибернетики имени В.М. Глушкова НАН Украины, аспирантка, 03680 МСП Киев 187, проспект академика Глушкова, 40, Украина; e-mail: vpn2006@rambler.ru*

Decision Making

PROCEDURES OF SEQUENTIAL ANALYSIS AND SIFTING OF VARIANTS FOR THE LINEAR ORDERING PROBLEM

Pavlo Antosiak, Oleksij Voloshyn

Abstract: *This paper investigates the procedures of sequential analysis and sifting of unpromising variants after restrictions and after the restriction on the goal function. On the basis of the modified procedure W and the general scheme of sequential analysis, the algorithm of solving the linear ordering problem is developed for the problems of discrete optimization.*

Keywords: *the linear ordering problem, the sequential analysis of variants.*

ACM Classification Keywords: *H.4.2 Information Systems Applications: Types of Systems: Decision Support.*

Conference*: The paper is selected from XV[th] International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.*

Introduction

The linear ordering problem (LOP) is NP-hard combinatorial optimization problem with wide application: in collective decision making, economics, archaeology and scheduling [Reinelt, 1985]. Many works are devoted to the development of efficient algorithms for solving this problem [Grotschel, 1984], [Chanas, 1996], [Laguna, 1999], [Campos, 2001].

Problem formulationa

LOP can be formulated as follows. Consider a set of alternatives $A = \{A_1,...,A_n\}$ and permutation $\pi : A \to A$. Each permutation $\pi = (\pi_1,...,\pi_n)$ uniquely determines some linear ordering of alternatives. Denote by e_{ij}, $i, j \in N = \{1,...,n\}$ the cost of disposition of alternative A_i to alternative A_j in linear order, and by E the n-square matrix of costs. Then LOP consists in finding such permutation π, in which the maximum total cost is achieved

$$E(\pi) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} e_{\pi_i \pi_j} \qquad (1)$$

Evidently, that (1) is the sum of elements above the main diagonal of matrix P, its' elements p_{ij} are the result of permutation of π lines and columns of matrix E, that is $P = XEX^T$, where X is the matrix, that corresponds to the permutation π [Reinelt, 1985].

As in most combinatorial optimization problems, LOP has many alternative formulations. In the paper [Grotschel, 1984] the linear ordering problem is considered in the equivalent raising of linear programming problem with Boolean variables.

$$E(x) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} d_{ij} x_{ij} \to \max \tag{2}$$

$$x_{ij} + x_{jk} - x_{ik} \le 1,\ 1 \le i < j < k \le n, \tag{3}$$

$$-x_{ij} - x_{jk} + x_{ik} \le 0,\ 1 \le i < j < k \le n, \tag{4}$$

$$x_{ij} \in X_{ij},\ 1 \le i < j \le n, \tag{5}$$

where $d_{ij} = e_{ij} - e_{ji}$, $X_{ij} = \{0,1\}$ are the set of possible variants of values of component x_{ij}, $X = \prod_{j>i} X_{ij}$ is the set of all possible variants.

Approach to solve the problem

The basis for developing an algorithm to solve the problem (2) – (5) is the method of sequential analysis, elimination and creation of variants [Mikhalevich, 1965], [Volkovich, Voloshin, 1978, 1984,1993].

Procedures of sequential analysis

For the discrete optimization problems the basis of schemes of sequential analysis and elimination of variants without step-by-step constructing solution is the procedure W of consistent elimination (exclusion from the consideration, generally speaking, on the current step) of values of variables. The procedure W, in its turn, consists of two procedures W_1 and W_2. Specify them for our case.

Let's choose any indexes $i, j, k \in N$, but such as $k > j > i$. Following the general scheme of sequential analysis and elimination of variants for the problem with linear restriction, we receive the following criteria of elimination after the restrictions (procedure W_1 on the s-step, $s \in \{1,..., s_0\}$):

for fixed $j > i$, if

$$\begin{cases} \bar{x}_{ij} > 1 - \min\limits_{x_{jk}^{(s)}} x_{jk} + \max\limits_{x_{ik}^{(s)}} x_{ik} \\ \bar{x}_{ij} < -\max\limits_{x_{jk}^{(s)}} x_{jk} + \min\limits_{x_{ik}^{(s)}} x_{ik} \end{cases} \tag{6}$$

at least for one $k : k > j > i$, then the component x_{ij} of permissible solution of problem (2) – (5) can't take the value equal to $\bar{x}_{ij} \in X_{ij}^{(s)}$, where $X_{ij}^{(s)}$ is the set of possible variants of component x_{ij} on the step s, $s \in \{1,..., s_0\}$;

for fixed $k > j$, if

$$\begin{cases} \bar{x}_{jk} > 1 - \min\limits_{x_{ij}^{(s)}} x_{ij} + \max\limits_{x_{ik}^{(s)}} x_{ik} \\ \bar{x}_{jk} < -\max\limits_{x_{ij}^{(s)}} x_{ij} + \min\limits_{x_{ik}^{(s)}} x_{ik} \end{cases} \tag{7}$$

at least for one $i : k > j > i$, then the component x_{jk} of permissible solution of problem (2) – (5) can't take the value equal to $\bar{x}_{jk} \in X_{jk}^{(s)}$;

for fixed $k > i$, if

$$\begin{cases} \bar{x}_{ik} < \min\limits_{X_{ij}^{(s)}} x_{ij} + \min\limits_{X_{jk}^{(s)}} x_{jk} - 1 \\ \bar{x}_{ik} > \max\limits_{X_{ij}^{(s)}} x_{ij} + \max\limits_{X_{jk}^{(s)}} x_{jk} \end{cases} \quad (8)$$

at least for one $j: k > j > i$, then the component x_{ik} of permissible solution of problem (2) – (5) can't take the value equal to $\bar{x}_{ik} \in X_{ik}^{(s)}$.

1) Let $\left|X_{ij}^{(s)}\right| \cdot \left|X_{jk}^{(s)}\right| \cdot \left|X_{ik}^{(s)}\right| \geq 4$. Realized the selection of all possible situations and analysed them by relevant criteria (6) – (8), it is easy to make sure, that the elimination won't be in this case.

2) Let $\left|X_{ij}^{(s)}\right| \cdot \left|X_{jk}^{(s)}\right| \cdot \left|X_{ik}^{(s)}\right| = 2$. Let $\left|X_{ij}^{(s)}\right| = 2$. It is easy to see, when $X_{jk}^{(s)} = \{0\}$ i $X_{ik}^{(s)} = \{1\}$, we'll have the elimination of value $\bar{x}_{ij} = 0$. Similarly if $X_{jk}^{(s)} = \{1\}$ and $X_{ik}^{(s)} = \{0\}$, we'll have the elimination of value $\bar{x}_{ij} = 1$. In case when $\left|X_{jk}^{(s)}\right| = 2$ and if $X_{ij}^{(s)} = \{0\}$, $X_{jk}^{(s)} = \{0,1\}$, $X_{ik}^{(s)} = \{1\}$, the value $\bar{x}_{jk} = 0$ will be eliminated, and in case $X_{ij}^{(s)} = \{1\}$, $X_{jk}^{(s)} = \{0,1\}$, $X_{ik}^{(s)} = \{0\}$ the value $\bar{x}_{jk} = 1$ will be eliminated. At $\left|X_{ik}^{(s)}\right| = 2$, when we have $X_{ij}^{(s)} = \{0\}$, $X_{jk}^{(s)} = \{0\}$, $X_{ik}^{(s)} = \{0,1\}$, the value $\bar{x}_{ik} = 1$ will be eliminated after the criterion (8), and if $X_{ij}^{(s)} = \{1\}$, $X_{jk}^{(s)} = \{1\}$, $X_{ik}^{(s)} = \{0,1\}$, then $\bar{x}_{ik} = 0$ will be eliminated. Like to the previous case, realized the selection of all other possible situations (related to the case examined by us), it is easy to make sure, that the elimination after the criteria (6) – (8) will never be.

3) Let $\left|X_{ij}^{(s)}\right| = \left|X_{jk}^{(s)}\right| = \left|X_{ik}^{(s)}\right| = 1$. If $X_{ij}^{(s)} = \{0\}$, $X_{jk}^{(s)} = \{0\}$ and $X_{ik}^{(s)} = \{1\}$, then, after the proper criterion of elimination, we receive $X_{ij}^{(s)} = \varnothing$. If $X_{ij}^{(s)} = \{1\}$, $X_{jk}^{(s)} = \{1\}$ and $X_{ik}^{(s)} = \{0\}$, then we'll also receive $X_{ij}^{(s)} = \varnothing$. Such variants make it impossible to build any allowed variant of problem (2) - (5) and describe the situation, when on the s-step we have an abnormal end of procedure W_1. In all other cases, evidently, it won't be the elimination.

From the total scheme follows that the criteria of elimination after the restriction on the objective function (**procedure** W_2 on the s-step ($s = 1,..., s_0$)) will be:

If

$$d_{ij}\bar{x}_{ij} < e_s^* - \max_{X^{(s)}/X_{ij}^{(s)}} \left\{ \sum_{\substack{t>k \\ (k\neq i)\vee(t\neq j)}} d_{kt}x_{kt} \right\} = e_s^* - \sum_{\substack{t>k \\ (k\neq i)\vee(t\neq j)}} \max_{X_{kt}^{(s)}} \{d_{kt}x_{kt}\}, (j > i) \quad (9)$$

then the component x_{ij} of permissible solution of problem (2)-(5) can't take the value equal to $\bar{x}_{ij}$, where e_s^* some value taken from the interval $[\min\limits_{X^{(s-1)}} E(x), \max\limits_{X^{(s-1)}} E(x)]$.

The follow cases are possible during the application of procedure W_2.

Case 1. It wasn't any elimination. Then the contraction of set of possible variants may occur, if reinforce the inequality (9), which, evidently, can be done only by increasing restrictions on the objective function, selecting, for example, the value e_{s+1}^* by the method of dichotomy from the interval $[e_s^*, e_{max}^{(s)}]$, where

$$e_{max}^{(s)} = \sum_{j>i} \max_{X_{ij}^{(s-1)}} \{d_{ij}x_{ij}\}$$

is the maximum possible value of objective function on the step s.

Case 2. It was the elimination, but the reduced set of possible variants $X^{(s)}$ is empty or $X^{(s)} \neq \varnothing$, but there is no allowable variant of problem (2) - (5). In this case, the abnormal end of procedure W_2 is carried out. It is also known [Volkovich, Voloshin, 1978] that, in this case, any allowable variant of problem (2) - (5) satisfies the conditions $E(x_D) < e_s^*$. Then the extension of set $X^{(s)}$ can be obtained by reducing the value e_s^*, thus weakening the further restriction on the objective function, by choosing $e_{s+1}^* < e_s^*$ after the rule of dichotomy from the interval $[e_{min}^{(s)}, e_s^*]$, where $e_{min}^{(s)} = \sum_{j>i} \min_{X_{ij}^{(s-1)}} \{d_{ij} x_{ij}\}$ is the minimum possible value of objective function on the step s.

Statement 1. The value $x_{ij}^{(max)} = \arg\max_{X_{ij}^{(s-1)}} \{d_{ij} x_{ij}\}$, $\forall\ i,j \in N$, $j > i$ and $\forall\ s \in \{1,...,s_0\}$ can't be eliminated for LOP as a result of procedure W_2 work

Proof. Suppose the opposite. Let $\overline{S} \subseteq \{1,...,s_0\}$ is the set of all steps for which our supposition is correct. And let $\tilde{s} = \arg\min_{s \in \overline{S}} s$ is the first of these steps (exactly at this step W_2 the elimination will be firstly done after the supposition), on which at least one pair of indexes $j > i$ will be found such that

$$d_{ij} x_{ij}^{(max)} < e_{\tilde{s}}^* - \sum_{\substack{t>k \\ (k \neq i) \vee (t \neq j)}} \max_{X_{kt}^{(\tilde{s})}} \{d_{kt} x_{kt}\} \tag{10}$$

Let $N_i^{(\tilde{s})}$, $N_j^{(\tilde{s})}$ are the sets that completely describe all pair of indexes $j > i$, $i \in N_i^{(\tilde{s})}$, $j \in N_j^{(\tilde{s})}$ for which on the step $\tilde{s}$ is executed (10). At first, let take the first of these pairs $j_{\tilde{s}} > i_{\tilde{s}}$ (exactly for this pair W_2 the elimination will be firstly done after the supposition), that is $i_{\tilde{s}} = \arg\min_{i \in N_i^{(\tilde{s})}} i$ and $j_{\tilde{s}} = \arg\min_{j \in N_j^{(\tilde{s})}} j$. In this case the rule of choice $\tilde{s}$ and pair $j_{\tilde{s}} > i_{\tilde{s}}$ gives an opportunity to rewrite the inequality (10) in the following equivalent form:

$$\max_{X_{i_{\tilde{s}} j_{\tilde{s}}}^{(\tilde{s}-1)}} \{d_{ij} x_{ij}\} + \sum_{\substack{t>k \\ (k \neq i_{\tilde{s}}) \vee (t \neq j_{\tilde{s}})}} \max_{X_{kt}^{(\tilde{s}-1)}} \{d_{kt} x_{kt}\} < e_{\tilde{s}}^*,$$

from where we get correlation

$$e_{max}^{(\tilde{s})} < e_{\tilde{s}}^*. \tag{11}$$

But (11) contradicts the rule of choice on any step s of value e_s^* from the interval $[e_{min}^{(s)}, e_{max}^{(s)}]$. By analogical reasoning, through the finite number of steps firstly we are convinced of the impossibility of our assumption for the step $\tilde{s}$, and then after a similar reasoning for the other steps we arrive at the correctness of statement. The statement is proved.

Corollary 1. For LOP the first situation described in case 2 can't arrise.

Corollary 2. For LOP the elimination of values of the component x_{ij} after the restriction on the objective function can occur only when $\left|X_{ij}^{(s-1)}\right| = 2$, thus the values $x_{ij}^{(min)} = \arg\min_{X_{ij}^{(s-1)}} \{d_{ij} x_{ij}\}$, $\forall\ i,j \in N$, $j > i$ and $\forall\ s \in \{1,...,s_0\}$ can only be eliminated.

Corollary 3. For LOP the condition of elimination through the objective function can be rewritten in the following equivalent form:

If

$$d_{ij}x_{ij}^{(min)} < e_s^* - e_{max}^{(s)} + \max_{X_{ij}^{(s-1)}}\{d_{ij}x_{ij}\} \quad (12)$$

then the component x_{ij} of permissible solution of problem (2) - (5) can't take value equal to $x_{ij}^{(min)}$.

From (12) we get:

$$e_{max}^{(s)} + \min_{X_{ij}^{(s-1)}}\{d_{ij}x_{ij}\} - \max_{X_{ij}^{(s-1)}}\{d_{ij}x_{ij}\} < e_s^* \quad (13)$$

Of the form (13) follows that for a given value e_s^* if the value $x_{i_1j_1}^{(min)}$ eliminates, we have

$$\min_{X_{i_1j_1}^{(s-1)}}\{d_{i_1j_1}x_{i_1j_1}\} - \max_{X_{i_1j_1}^{(s-1)}}\{d_{i_1j_1}x_{i_1j_1}\} \geq \min_{X_{i_2j_2}^{(s-1)}}\{d_{i_2j_2}x_{i_2j_2}\} - \max_{X_{i_2j_2}^{(s-1)}}\{d_{i_2j_2}x_{i_2j_2}\},$$

then the value $x_{i_2j_2}^{(min)}$ will be also eliminated. Then in order not to have the second situation of case 2, evidently, it is necessary to permit the elimination of all values $x_{i^*j^*}^{(min)}$,

where

$$(i^*, j^*) \in \mathrm{Arg}\min_{(i,j):i<j}\{\min_{X_{ij}^{(s-1)}}\{d_{ij}x_{ij}\} - \max_{X_{ij}^{(s-1)}}\{d_{ij}x_{ij}\}\}.$$

Remark 1. Note that the elimination of all such values can be obtained by using strict limit on the objective function $E(x) > e_s^*$ and by choosing

$$e_s^* = e_{max}^{(s)} + \min_{(i,j):i<j}\left(\min_{X_{ij}^{(s-1)}}\{d_{ij}x_{ij}\} - \max_{X_{ij}^{(s-1)}}\{d_{ij}x_{ij}\}\right).$$

Algorithm

Step 1. Calculation of values $e_{max}^{(s)}$ and e_s^*.

Go to the next step.

Step 2. Application of procedure W_2, that is the elimination of all values $x_{i^*j^*}^{(min)}$, where

$$(i^*, j^*) \in \mathrm{Arg}\max_{(i,j):\begin{cases} i<j \\ |X_{ij}^{(s-1)}|=2 \end{cases}} |e_{ij} - e_{ji}|.$$

Step 3. Application of procedure W_1.

At the abnormal end of procedure W_1 the restoring of all eliminated on a current and previous step values and the end of algorithm work are realized.

Otherwise, if $\prod_{j>i} |X_{ij}^{(s)}| = 1$,

then it is the end of algorithm, otherwise it is the passage to the step 1.

Acknowledgements

The paper is parttially financed by the projectITHEA XXI of the Institute of Information Theories and Applications FOI ITHEA and the Consortium FOI Bulgaria (www.itea.org, www.foibg.com).

Conclusion

From point of view of realization the proposed algorithm is simple. If the found variant of solution $\tilde{x}$ satisfies the condition $E(\tilde{x}) > e^*_{s_0}$, it is an optimum variant [Volkovich, Voloshin, 1978, 1984, 1993]. The variant $\tilde{x}$ can be taken as an approximate solution if $E(\tilde{x}) \le e^*_{s_0}$. In case of abnormal end of procedure W_1 we receive the reduction of set of possible variants. The algorithm work efficiency is investigated on the real test dataset containing 49 matrices «production cost» from the previous years of some European countries and is popular in Internet [LOLIB]. For 10 of 49 examples an optimum value was found.

Acknowledgements

The paper is published with financial support by the project ITHEA XXI of the Institute of Information Theories and Applications FOI ITHEA Bulgaria www.ithea.org and the Association of Developers and Users of Intelligent Systems ADUIS Ukraine www.aduis.com.ua.

Bibliography

[Campos, 2001] Campos V., Glover F., Laguna M., Marti R. An experimental evaluation of a scatter search for the linear ordering problem// Journal of Global Optimization. – 2001.– vol. 21.–№ 4.– P. 397-414.

[Chanas, 1996] Chanas S., Kobylanski P. A new heuristic algorithm solving the linear ordering problem// ComputationalOptimization and Applications.– 1996.– vol. 6.– P. 191-205.

[Grotschel, 1984] Grotschel M., Jünger M., Reinelt G. A cutting plane algorithm for the linear ordering problem// Operations Research. – 1984. – vol. 2.– №6.– P. 1195-1220.

[Laguna,1999] Laguna M., Marti R., Campos V. Intensification and diversification with elite tabu search solutions for the linear ordering problem// Computers \& Operations Research.–1999. –vol. 26.- P. 1217-1230.

[LOLIB] http://www.iwr.uni-heidelberg.de/groups/comopt/software/LOLIB.

[Mikhalevich, 1965] Mikhalevich V.S. Consecutive optimization algorithms and their application. I. II// Cybernetics. – 1965. − №1. – P. 45-55; − №2. – P. 85-88.

[Reinelt, 1985]. Reinelt G. The linear ordering problem: algorithms and applications. Research and Exposition in Mathematics. – Berlin, Germany: Heldermann Verlag, 1985.

[Volkovich, Voloshin, 1978] Волкович В.Л., Волошин А.Ф. Об одной схеме метода последовательного анализа и отсеивания вариантов//Кибернетика,1978,№4.-С.99-105.

[Volkovich, Voloshin, 1993] Волкович В.Л., Волошин А.Ф. и др.Модели и методы оптимизации надежности сложных систем.- Киев: Наукова думка,1993.-312с.

[Volkovich, Voloshin,, 1984] Волкович В.Л., Волошин А.Ф. и др. Методы и алгоритмы автоматизированного проектирования сложных систем управления. – Киев: Наукова думка, 1984. – 216 с.

Authors' Information

Antosiak Pavlo *- Assistant, Uzhgorod National University, Mathematics Department. Uzhgorod, Ukraine, e-mail: antosp@ukr.net*

Voloshyn Oleksij *- Professor, Kyiv National Taras Shevchenko University, Faculty of Cybernetics. Kyiv, Ukraine, e-mail: ovoloshin@unicyb.kiev.ua*

STATISTICAL MODELING OF SPREAD OF OPTICAL FIELDS IN FIRES

**Elena Visotskaja, Olga Kozina, Veronika Lachenova,
Olga Remaeva, Tatiana Remaeva**

***Abstract**: Application of Monte-Carlo method for determination of optical radiation flow density of forest fires in a cloudless atmosphere with dependence on optic-geometrical parameters of surface's relief for the case of multi and single scattering is offered in this work. The use of initial directions of photon trajectory, free path length of the photon, coordinates of interact points between the photon and environment, also probabilities of photon's getting to a receiver into the statistical modeling of direction of distribution of optical radiation of fires is proposed.*

***Keywords**: Monte-Carlo method, photon trajectory, radiation transfer.*

***ACM Classification Keywords**: I.6 Simulation and Modeling. I.6.8 Types of Simulation - Monte Carlo.*

***Conference**: The paper is selected from XVth International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.*

Introduction

The checking of anthropogenic and natural objects and phenomenon which because chemical (the aerosols, gases) and heat (thermal) soiling are source of potential or direct risk for biological systems and habitat, plays important role in applied ecology. To such objects first of all follows to attribute the fires and smokes, which origin can be due various reasons [Потапов, 2002]. Emphasizing unconditional importance of this problem, necessary also to take into attention the direct influence of optical radiation (its heat component first of all) on atmosphere and underlying surface. This influence has more energy power and duration during big fires that permits to include its in special aspect of applied ecology - heat soiling the ambience [Доррер, 1979]. Given problem is actual because directed checking optical (heat) of radiation overland objects and atmospheric phenomenon within the framework of separate regions, continents, water areas and planets as a whole have not sufficient volume. Measuring of the flow of heat radiation on fires is very complicated task that is why calculations parameters and possibility of their modeling are playing great practical importance.

The numerical method of calculating the characteristics of the light field

At present most acceptable for a calculation of optical fields is the Monte-Carlo method [Соболь, 1968] sense of which in modeling of paths of the photons and events, which with them can occur.

The strategy of calculation of density of flow of scattered radiation from point source with single power [Марчук, 1967] bases on using of three arrays of input data.

The parameters which characterize external habitat compose first array: meteorological visibility, S_M, km, aerosol indicatrix of scattering χ_a, molecular indicatrix of scattering χ_M, optical thickness of atmosphere τ_0, terrestrial surface albedo A, height of slopes of relief h, lay l, orientation of slopes β.

The second array of input data includes the spectral features of source - lengths of waves of radiation λ, μm.

The third array includes parameters which define geometry of source and receiver: height of a radiation source H_u, km, distance between the receiver and epicenter of the source R, km, height of receiver above the surface

H_n, km, visual angle of optical device 2ξ, grad, visual angles of device of observe ($2\xi_в$ – vertical, $2\xi_г$ – horizontal angular field of view) grad, orientation of the receiver to the source of radiation (θ,φ).

The parameters of first group $S_м$, τ_0, χ_a and $\chi_м$ are interdependent therefore at decision of problem they are represented like functions one variable – meteorological range of visibility $S_м$.

For calculating the parameters of the third group (H_u, R, H_n, ξ, θ,φ) proposes to use the method of dependent tests, according to which in the calculation of optical flow in several receivers can use the same path of photons. At fixed location of the receiver of radiation a calculation is conducted with changing arguments $S_м$, A, h, β.

The process of radiation transfer is composed from independent "paths" of photons. This possible consider as homogeneous Markov chain states of which are a positions of particles in space of coordinates and directions [Ермаков, 1975]. Casual transition of the photon from one state to other is considered as row of casual events: reflection, scattering or absorption. End result of calculating of free path length of the photon after a new collision is mean value of desired functional [Чандрасекар, 1993]. In this task such functional can be the stream of photons at the point of the receiver. The method of local estimation of the flow of photons is proposed [Зигель, 1975] to calculate such functional. Estimation of complete flow of photons in the receiver carry out by average of statistical weight of the photon – probabilities that the photon hit in receiver – on all collisions.

$$\Phi = \frac{e^{-L(r)}\chi_{(\mu)}T_{(r)}\cos\theta_1}{2\pi r^2}, \tag{1}$$

where r – the distance between the receiver and the point of the photon interaction with environment;

$L(r)$ – optical distance passable the photon along the path r, $L(r)=\int_0^r \sigma(r)dr$;

$\sigma(r)$ – scattering index of environment;

$\chi_{(\mu)}$ – indicatrix of scattering, defined by expressions (3.6-3.7);

$T(r)$ – function of radiation transmission by atmospheric gases along the path r;

θ_1 – angle between the normal to the plane of the receiver and the direction of motion of the photon from the point of collision to the receiver.

Since the calculation is conducted under restrictions imposed by the receiver parameters (2ξ), the receiver having a round hole, the condition of getting the photon in the receiver will be determined by the type of inequality

$$\cos\theta_1 \geq \cos\xi, \tag{2}$$

for receiver with square-wave input hole - system of inequalities

$$\begin{cases} \dfrac{z_1 - z_n}{\sqrt{(x_1-x_n)^2+(y_1-y_n)^2+(z_1-z_n)^2}} \leq \sin\xi_в \\ \dfrac{n_x(x_1-x_n)+n_y(y_1-y_n)}{\sqrt{(x_1-x_n)^2+(y_1-y_n)^2}} \geq cos\xi_г \end{cases}, \tag{3}$$

where, x_1, y_1, z_1 – coordinates of the photon interaction with environment;

x_n, y_n, z_n – coordinates of receiver П.

If values of *F* had found by expression (1), the expression for calculating the flow of radiation from source with single unit capacity of power at the point of the receiver has the following form:

$$\Phi^* = \frac{\sum^{\{R_r\}} \Phi}{N}, \tag{4}$$

where, $\{R_r\}$ – set of points of interaction within the solid angle Ω which depends on angle of view of optical device;

N – number of modeled paths of photons.

The finding of the photon in the solid angle carries out by system of inequalities

$$\begin{aligned} &-200 \text{ km} \le x \le 200 \text{ km}, \\ &-200 \text{ km} \le y \le 200 \text{ km}, \\ &0 \text{ km} \le z \le 80 \text{ km} \end{aligned} \tag{5}$$

The task of radiation transfer in the above formulation is solved by modeling the trajectories of particles (photons). Input data for modeling are parameters describing the environment (the first group of parameters), the spectral characteristics of the source (the second group of parameters), the geometry of the source and the receiver (the third group of parameters).

Values of scattered radiation of the first and second group of parameters, as well as the height of the source H_u and the receiver H_n are defined as constants. The distance from the fire epicenter R, angle of view of the receiver 2ξ and orientation of the receiver to the source (θ, φ), as well as the height h and orientation β of slope terrain appears to be an array of variable information.

Photons enumerator provides to go to the next photon, and to control the number of calculated cycles for a given number of photons. End result is formed when all photons are checked.

The first calculated cycle begins with the choice of initial trajectories for the photon and then calculated the length of free path of photons in a given direction. In a heterogeneous environment (clear atmosphere) the length of free path of photons is determined according

$$-\ln j_4 = \frac{a}{\alpha_0 m}\left[1+\left\{\frac{\lambda_0}{\lambda}\right\}^4\right](1-e^{-\alpha_0 m l_{np}})e^{-\alpha_0 h_\phi} + \frac{\varepsilon\lambda_0}{\beta_0 m\lambda}(1-e^{-\beta_0 m l_{np}})e^{-\beta_0 h_\phi}, \tag{6}$$

where $m = cos\zeta$ – cosine of the angle ζ between the initial direction of motion of the photon and the normal to the surface of the Earth;

h_ϕ – the height of the point of the photon scattering on the ground surface.

After that coordinates of the point of the photon interaction with the environment be calculated. But inside area of calculation the photon can interact with environment, for example, absorbed of atmospheric gases and the Earth, scattered on aerosol particles and the molecular environment, reflected from the Earth.

The probability that a photon is not absorbed by atmospheric gases (mainly water vapor), is characterized by function

$$T(r) = \frac{1}{1+K_\lambda G(r,\mu_1)}, \tag{7}$$

where K_λ – spectral absorption coefficient of water vapor in the visible light spectrum $K_\lambda \approx 0$, infrared K_λ = 0,076 cm^2 / g;

$G(r,\mu_1)$ – volume of «besieged» water along the path defined by the distance *r*, and $\mu_1 = \cos(90^0 - \theta_n)$, where θ_n – polar angle of vector between the interaction point and the receiver

$$G(r,\mu_1) = G_0 \int_0^r e^{-\sigma_0 z(r,\mu_1)} dr \,, \tag{8}$$

where G_0 – concentration ratio of water vapor in the surface layer, the value of which vary widely depending on weather conditions and geographical areas. The average value of this ratio for the summer is 10 g/m^3, for the winter – 2 g/m^3, σ_0 – approximation ratio equal to 0.6 km^{-1}.

Calculation of the integral of (8) along free path of photons gives us

$$G(l_{np}, m) = \frac{G_0}{\sigma_0 m} e^{-\sigma_0 Hu} (1 - e^{-\sigma_0 m l_{np}}) \,, \tag{9}$$

where $m = cos\zeta$.

Volume of «besieged» water along the path from the interaction point till the receiver

$$G(r,\mu_1) = \frac{G_0}{\sigma_0 \mu_1} e^{-\sigma_0 Hn} (1 - e^{-\sigma_0 r \mu_1}) \,, \tag{10}$$

Thus, by defining of function $G(r,\mu_1)$ it is possible to decrease statistical weight of the photon *Φ* by multiplying its on probability of "hit" *T*(*r*) during every part of interaction between the photon and environment.

Type of photon interaction with the environment is determined according to the condition

$$L(l_{np}) > -\ln j_\sigma \,, \tag{11}$$

where $L(l_{np})$ – optical distance passable along the length of the photon free path. If condition is not right, "albedo situation" exist. In other words, the photon can be absorbed with probability 1-*A* or be reflected with probability *A* which equivalent of terrestrial surface albedo.

In moment then the photon is absorbed *y* Earth, new photon are defined and modeled its trajectory.

After establishing the fact of crossing the path of the photon, the coordinates of the point of intersection are finding.

If a regular model of the terrain are using, coordinates (*x*, *y*, *z*) of the intersection point are searching from a set of three equations, two of which describe the motion of the photon after the last interaction point (x_0, y_0, z_0) (departure from the source) and third equation is represented by piecewise-linear approximation.

After this reflection of the photon with attention to orientation of terrain slope can be modeled: directional cosines normal $\overrightarrow{n_0}$ to the elementary surface area slope n_{0x}, n_{0y}, n_{0z} are calculating, the angle between the direction of the photon moving and the normal $\overrightarrow{n_0}$ determined, indicatrix of reflection is determined.

If the photon is in the direct line of sight from the point of reflection to the receiver, functional *Φ(R)* that determine the contribution of the act of reflection in the total flow of radiation *Φ* coming into the receiver are calculated according to the formula

$$\hat{O}(R) = \frac{\chi_{îòð}(\theta_1)\dot{A}}{2\pi r^2} \exp(-L(\vec{r}))\Delta(\vec{x}, \overrightarrow{x_n}) \,, \tag{12}$$

where $\chi_{omp}(\theta_1)$ – indicatrix of reflection from the surface with condition of normalization $\int_0^{\pi} \chi_{omp}(\theta_1)\sin\theta_1 d\theta_1 = 1$;

A – albedo into intersection point;

$\vec{r}$ – distance between intersection point and the receiver;

$\Delta(\vec{x},\overrightarrow{x_n})$ – pointer of condition of "hit" of the point into view area of the receiver. This pointer can be equal 1 if the condition is right, and 0 – in contrary.

$(\vec{x},\overrightarrow{x_n})$ – coordinates in phase space $R\times\Omega$ of points of intersections between trajectory of the photon and underlying surface or between trajectory of the photon and the receiver accordingly.

If the photon does not come into receiver after reflection, new direction of moving of the photon are determined. And new value of angle of scattering γ is selected and whole procedure of calculations of distances, orientations and angle fields are repeated.

Thus, result of one modeling cycle is the probability of getting a given number of photons $N = K$ in the receiver for the given matrixes of distances $\{R\}$, orientations $\{\theta,\varphi\}$, visual angles of the receiver $\{2\xi\}$ or $\{2\xi_в \times 2\xi_г\}$ is completed for condition $N \geq K$.

Qualitative estimation of the reliability of the developed method

Qualitative estimation of the reliability of the developed method was carried out by the way comparing received results to results by calculations based on the Monte Carlo method and based on finite-difference method of solving the transport equation.

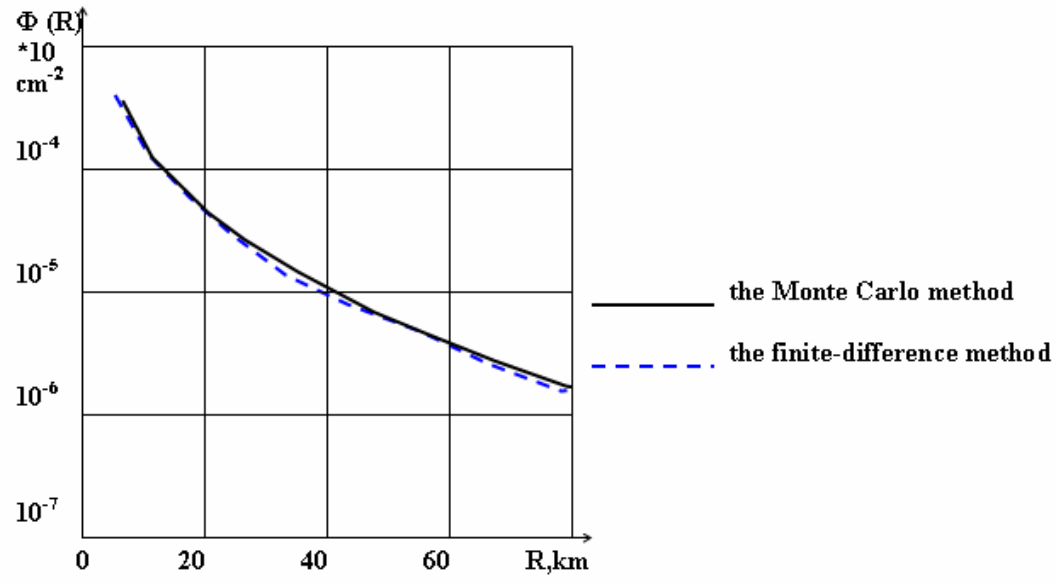


Fig. 1 Comparing received results to results by calculations based on the Monte Carlo method and based on finite-difference method.

Initial data for these calculations are: height of a radiation source H=1km, the Earth's albedo A = 0, lengths of wave of radiation λ = 0,45 μm, measuring range of distances – from 5 to 100 km, the radiation is taken from the whole upper hemisphere.

Due to fig.1 it is seen that the results in reference points at $h \to 0$ have satisfactory agreement with literature data. At the molecular atmosphere, the maximum divergence between the results is 15% for single scattering (include the two curves shown in fig. 1). Given that divergence between curves in fig. 1 obtained by different methods are varied in the same range, reference options can be considered like coordinated and responsive to the real picture of radiation transfer.

Conclusion

Proposed numerical method of calculating the characteristics of the light field, which is based on the Monte Carlo method, are well in line with probability sense of the problem. Equation of radiative transfer which solved by the Monte Carlo method, takes into account the many scattered and reflected components of radiation, allows the use of a stratified model of the atmosphere and any method of mathematical modeling of the underlying surface. Using the proposed technology for calculating the density of optical radiation as a component of geoinformation system allows simulating the spread of optical radiation fire that will enhance the effectiveness of measures to eliminate it.

Acknowledgements

The paper is published with financial support by the project ITHEA XXI of the Institute of Information Theories and Applications FOI ITHEA Bulgaria www.ithea.org and the Association of Developers and Users of Intelligent Systems ADUIS Ukraine www.aduis.com.ua.

Bibliography

[Доррер, 1979] Г.А. Доррер. Математические модели динамики лесных пожаров. М.: Лесн. пром-сть, 1979.

[Ермаков, 1975] С.М. Ермаков. Метод Монте-Карло и смежные вопросы. М.: Наука, 1975.

[Зигель, 1975] Р. Зигель, Дж. Хауэлл. Теплообмен излучением. М.: Мир, 1975.

[Марчук, 1967] Г.И. Марчук. Метод Монте-Карло в проблеме переноса излучения. М.: Высшая школа, 1967.

[Потапов, 2002] А.Д. Потапов. Экология. М.: Высшая школа, 2002.

[Соболь, 1968] И.М. Соболь. Метод Монте-Карло. М.: Наука, 1968.

[Чандрасекар, 1993] С. Чандрасекар. Перенос лучистой энергии. М.: ИЛ, 1993.

Authors' Information

Elena Visotskaja *– PhD, lecturer of Biomedical Electronic Devices and Systems Department of Kharkov National University of Radio Electronics, Lenina Av., 14, Kharkov, 61166, Ukraine; e-mail: diagnost@kture.kharkov.ua*

Olga Kozina *– PhD, lecturer of Computers and Programing Department of National Technical University 'KPI', Frunze street, 21, Kharkov, 61002, Ukraine; e-mail: okaraban@rambler.ru*

Veronica Lachenova *– engineer of Kharkov regional centre of medical statistics, Pravda Av., 8, Kharkov, 61100, Ukraine; e-mail: borman_d@mail.ru*

Olga Remaeva *– PhD, lecturer of Higher Mathematics Department of Kharkov National University of Radio Electronics, Lenina Av., 14, Kharkov, 61166, Ukraine.*

Tatiana Remaeva *– graduate student of Higher Mathematics Department of Kharkov National University of Radio Electronics, Lenina Av., 14, Kharkov, 61166, Ukraine.*

ПРИНЦИПЫ И ЗАДАЧИ ОПТИМИЗАЦИИ РАЗМЕЩЕНИЯ ДАТЧИКОВ ПОЖАРНОЙ СИГНАЛИЗАЦИИ В УСЛОВИЯХ НЕОПРЕДЕЛЕННОСТИ

Александр Землянский, Виталий Снитюк

Аннотация: *В статье рассмотрена проблема размещения датчиков пожарной сигнализации. Указаны особенности и недостатки современных методов размещения таких датчиков, связанные с ограниченностью учета факторов, которые влияют на процесс как их размещения, так и использования. Предложен комплексный подход к оптимизации размещения датчиков, учитывающий условия внешней среды и внутренние особенности помещений, а также экспертные заключения и их объективизацию.*

Ключевые слова*: Пожаротушение, нечеткая логика, принятие решений, оптимизация.*

ACM Classification Keywords*: I.2.1 Applications and Expert Systems*

Conference*: The paper is selected from XV[th] International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.*

Введение

Ежегодно на Земле возникает около семи миллионов пожаров. Исходя из прогнозов, сделанных на основе пожарной статистики, в мире в течение следующего года может погибнуть на пожаре около 225 тыс. человек, 2 миллиона 250 тыс. человек получить увечья, и 4,5 миллиона человек - тяжелые ожоговые травмы. Основными направлениями обеспечения пожарной безопасности являются устранение условий возникновения пожара и минимизация его последствий. Одним из способов решения таких задач есть установка автоматических средств предупреждения о возникновения пожара.

В Украине средствами пожарной автоматики оборудовано 89,1% объектов. Адресными установками сигнализации оборудовано около 78% объектов. Техническое обслуживание проводится на 72,6% объектов. 17,2% установок выработали свой ресурс и подлежат замене. На 11,8% объектов установки пожарной сигнализации подлежат замене. В 39,1% случаев пожарная автоматика не сработала, что привело к значительным материальным потерям. Таким образом, существует устойчивая тенденция к росту количества пожаров и аварий, уровню их последствий. Одной из причин этого является низкая эффективность систем пожарной автоматики и, в частности, систем пожарной сигнализации.

Настоящая статья продолжает цикл работ об оптимизации процессов принятия решений при пожаротушении [Snytyuk, 2007], [Снитюк, 2007] и является инициальной работой по оптимальному размещению датчиков с использованием методов теории нечетких систем.

Особенности размещения пожарной сигнализации, ее вероятностные характеристики

Рассмотрим один из подходов к оптимизации размещения датчиков пожарной сигнализации. На сегодня существуют традиционные способа их размещения. Согласно первому способу датчики размещаются по треугольной схеме (рис. 1), во втором случае датчики находятся в углах прямоугольника (рис. 2). Известны многочисленные исследования, результатом которых являются рекомендации по количеству датчиков того или иного типа, которые могут быть использованы в определенных помещениях, на самолетах, подводных лодках и т.п. В их основе лежат определение вероятностей:

– срабатывания датчика, если имеет место пожар;

– не срабатывания датчика, если пожара нет;

– срабатывания датчика, если пожара нет;

– несрабатывания датчика, если пожар есть.

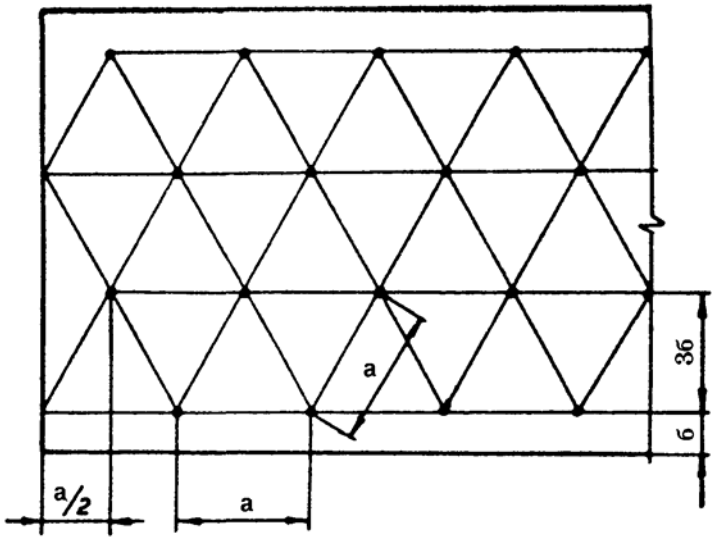


Рис. 1. Схема треугольного размещения датчиков

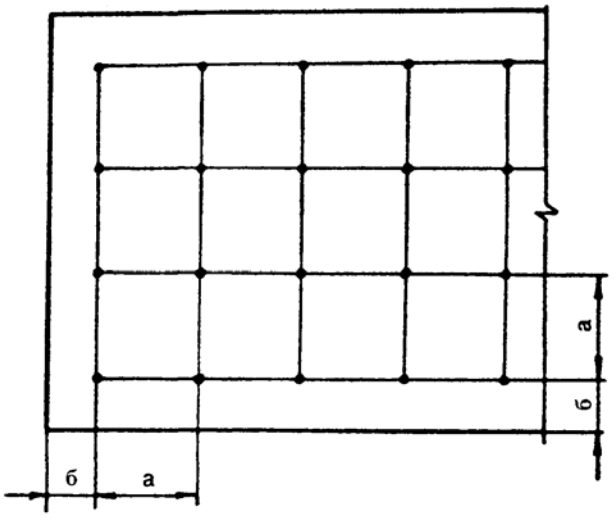


Рис. 2. Схема прямоугольного размещения датчиков

а – расстояние между датчиками; б – расстояние от стенки к датчику

Исходя из значений указанных вероятностей, осуществляется установка дополнительных датчиков и таким образом осуществляется резервирование и повышается надежность правильного срабатывания пожарной сигнализации. Заметим, что такая практика имеет место на высокотехнологичных объектах, уже перечисленных кораблях, самолетах, атомных электростанциях. Вместе с тем, остается немало предприятий, объектов жилого сектора, где сигнализация устанавливается, исходя из схем (рис. 1 и рис. 2). При этом учитывается высота размещения датчиков (как правило, она имеет интервальное представление), и как следствие, площадь зоны, контролируемой датчиком. Значительная часть

помещения остается неконтролируемой, или датчики срабатывают, когда пожар достигает своей максимальной точки.

Таким образом, учитывая особенности и недостатки размещения датчиков пожарной сигнализации, предлагаются следующие принципы:

- при установке датчиков пожарной сигнализации ориентироваться на возможные человеческие жертвы (X_1), величину материального ущерба (X_2), а также возможные последствия техногенных и экологических катастроф (X_3), вызванные несрабатыванием датчиков. Необходимо также учитывать убытки от их ложного срабатывания (X_4);
- для решения задачи оптимизации структуры системы пожарной сигнализации, необходимо разработать метод определения коэффициента, указывающего на величину опасности пожара на объекте, $k = F(X_1, X_2, X_3, X_4)$;
- при превышении значения коэффициента k некоторой величины $C > 0$ разработать технологию, позволяющую определить количество и структуру размещения датчиков пожарной сигнализации. Очевидно, что в таком случае, количество датчиков будет большим, чем рассмотрено в структуре на рис. 1 и рис. 2;
- при разработке структуры размещения датчиков базироваться на их технических характеристиках; результатах стендовых испытаний; особенностях помещений, где будут установлены датчики; экспертных заключениях.

Базируясь на предложенных принципах, для определения оптимальной структуры размещения датчиков пожарной сигнализации необходимо решить следующие задачи:

1. Осуществить идентификацию зависимости $k = F(X_1, X_2, X_3, X_4)$, исходя из экспертных заключений и системы нечетких продукционных правил:

$$Если\ X_1^i \in A_1^i, u\ X_2^i \in A_2^i, u\ X_3^i \in A_3^i, u\ X_4^i \in A_4^i,\ то\ k \in A^i,\ i = \overline{1,n},$$

где A_j^i – нечеткие множества со своими функциями принадлежности, n – количество экспертов.

2. Идентифицировать вероятность срабатывания датчика в зависимости от расстояния до очага возгорания и температуры горения:

$$Если\ d^i \in B_1^i\ u\ T^i \in B_2^i,\ то\ p \in P^i,\ i = \overline{1,n},$$

где d^i – расстояние от очага возгорания до датчика, B_1^i – соответствующее нечеткое множество, T^i – температура пламени, p – вероятность правильного срабатывания датчика.

3. Исходя из решений задач 1 и 2, определить оптимальную структуру размещения датчиков, используя нечеткие продукционные правила:

$$Если\ S_j^i \in C_j^i, u\ p \in B_j^i, u\ k \in A_j^i,\ то\ Q \in R_j^i,\ i = \overline{1,n},\ j = \overline{1,m},$$

где S_j^i – j-й вариант структуры, предложенный i-м экспертом, C_j^i – соответствующее нечеткое множество, Q – интегральный показатель, указывающий на оптимальность размещения датчиков, R_j^i – соответствующее нечеткое множество. Решение последней задачи и укажет на вариант структуры, являющийся оптимальным.

Необходимо заметить, что в приведенных задачах присутствует значительное количество нечетких продукционных правил, содержащих функции принадлежности, предложенные экспертами. Очевидно, что такие правила нередко имеют противоречивый характер. Их объективизация заключается в определении параметров функций принадлежности, исходя из значений, содержащихся в обучающих выборках, и решении задачи минимизации суммарной ошибки при равноправных заключениях, или минимизации взвешенной ошибки, в противном случае.

Для решения такой задачи возможно использование аппарата теории нечетких нейросетей или эволюционного моделирования. Последнее по ряду соображений представляется предпочтительным, хотя эксперименты остаются еще впереди.

Заключение

Рассмотренные аспекты решения задачи оптимизации системы пожарного мониторинга в условиях неопределенности вызваны низкой вероятностью срабатывания датчиков пожарной сигнализации и соответствующими последствиями. Зависимость размещения датчиков от факторов внешней среды и внутренних особенностей объектов должна учитываться наряду с экспертными заключениями. Соответствующие технологии оптимизации структуры датчиков пожаротушения необходимо базировать на технологиях «мягких» вычислений, поскольку жестко заданные правила их размещения оказываются достаточно часто неэффективными.

Благодарности

Работа опубликована при финансовой поддержке проекта **ITHEA XXI** Института информационных теорий и приложений FOI ITHEA Болгария www.ithea.org и Ассоциации создателей и пользователей интеллектуальных систем ADUIS Украина www.aduis.com.ua.

Библиография

[Snytyuk, 2007] Snytyuk V., Dghulay O. Evolutionary technique of shorter route determination of fire brigade following to fire place with the optimized space of search // Information Technologies and Knowledge. − 2007. − Vol. 1. − № 4. − Pp. 325-332.

[Гнатиенко, 2008] Гнатиенко Г.Н., Снитюк В.Е. Экспертные технологии принятия решений. – Киев: Маклаут, 2008. – 444 с.

[Снитюк, 2007] Снитюк В.Е., Быченко А.А. Эволюционное моделирование процесса распространения пожара // Bulgaria, Varna. Proc. XIII-th Int. Conf. Knowledge-dialogue-Solution", 2007 (June). – Pp. 247-254.

[Снитюк, 2008] Снитюк В.Е. Прогнозирование. Модели, методы, алгоритмы. – Киев: Маклаут, 2008. – 364 с.

[Снитюк, 2008] Снитюк В.Е., Быченко А.А., Джулай А.Н. Эволюционные технологии принятия решений в пожаротушении. – Черкассы: Маклаут, 2008. – 268 с.

Информация об авторе

Александр Землянский – *Адъюнкт, Академия пожарной безопасности имени Героев Чернобыля, 18000, Украина, Черкассы, Ул. Оноприенко, 8; e-mail: zemapb@gmail.com*

Виталий Снитюк *– Заведующий кафедрой информационных технологий проектирования, Черкасский государственный технологический университет, 18006, Украина, Черкассы, бул. Шевченко, 460; e-mail: snytyuk@gmail.com*

КООПЕРАТИВНЫЕ МОДЕЛЕ-ОРИЕНТИРОВАННЫЕ МЕТАЭВРИСТИКИ ДЛЯ ЗАДАЧ КОМБИНАТОРНОЙ ОПТИМИЗАЦИИ

Леонид Гуляницкий, Сергий Сиренко

***Аннотация**: Предлагается методология построения кооперативных метаэвристических методов решения задач комбинаторной оптимизации на основе моделе-ориентированных алгоритмов. Ее особенностью является решение задачи путем поиска (оптимизации) в пространстве моделей, который проводится на основе частных моделей, сформированных базовыми (составными) алгоритмами. Описана схема таких методов, разработана кооперативная метаэвристика на основе алгоритмов оптимизации муравьиными колониями, проведено исследование эффективности предлагаемой методологии на основе анализов результатов вычислительного эксперимента.*

***Ключевые слова**: комбинаторная оптимизация, моделе-ориентированные методы, кооперативные метаэвристики, оптимизация муравьиными колониями.*

***ACM Classification Keywords**: G.1.6 [Numerical Analysis] Optimization, I.2.8 [Artificial Intelligence]: Problem Solving, Control Methods, and Search – Heuristic methods, General Terms: Algorithms.*

***Conference**: The paper is selected from XVth International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.*

Введение

Среди прикладных методов решения задач комбинаторной оптимизации (ЗКО) в последнее время формируется класс моделе-ориентированных (model-based) алгоритмов – в отличие от большинства традиционных алгоритмов, которые относят к классу задаче-ориентированных (instance-based) [Zlochin et al., 2004]. Задаче-ориентированные алгоритмы генерируют новые варианты решений только на основе одного или нескольких текущих вариантов решений. К числу таких методов относятся генетические алгоритмы, повторяемый локальный поиск и другие [Hoos and Stützle, 2005]. В моделе-ориентированных методах варианты решений генерируются с использованием параметризированной вероятностной модели, которая обновляется на основе ранее рассмотренных вариантов решений так, чтобы поиск концентрировался в областях с вариантами решений „высокого качества" [Zlochin et al., 2004]. Как отмечено в [Dorigo and Blum, 2005], в моделе-ориентированном алгоритме происходит оптимизация параметров и/или структуры модели, т.е. вместо исходной ЗКО решается задача непрерывной оптимизации в пространстве параметров модели используемой этим алгоритмом. Широко известными моделе-ориентированными методами являются алгоритмы оптимизации муравьиными колониями (ОМК) [Dorigo and Stützle, 2004], метод кросс-энтропии [Rubinstein and Kroese, 2004] и метод вычисления оценки распределения [Larrañaga and Lozano, 2002].

Другой активно используемой технологией построения современных алгоритмов решения ЗКО является гибридизация высокоуровневых алгоритмов – метаэвристик [Raidl, 2006]. С точки зрения стратегии управления различают интегративные и кооперативные схемы [Cotta, 1998]. Наше внимание будет сосредоточено на втором типе схем, в которых составные алгоритмы обладают высокой степенью автономности, образовывая на нижнем уровне нечто вроде асинхронных групп [Taludar et al., 2003].

В работе предлагается и исследуется методология построения кооперативных метаэвристик на основе моделе-ориентированных алгоритмов [Гуляницький, 2009]. Особенностью предлагаемого подхода к построению кооперативных методов решения ЗКО является наличие управляющего алгоритма агрегирования (оптимизации) частных моделей, сформированных базовыми (составными) алгоритмами. В следующем разделе описана общая схема таких методов, далее излагаются детали разработанной

кооперативной метаэвристики на основе алгоритмов ОМК. Результаты и обсуждение проведенного исследования эффективности предлагаемой методологии на основе вычислительного эксперимента приведены в предпоследнем разделе. Заключение и описание направлений дальнейшего исследования завершает работу.

Под ЗКО мы будем понимать задачу минимизации заданной целевой функции на локально-конечном пространстве вариантов решения, состоящим из комбинаторных объектов [Гуляницький, 2008].

Построение кооперативных гибридных метаэвристик

Рассмотрим проблему построения гибридных кооперативных метаэвристик, которые базируются не на одном, а на нескольких моделе-ориентированных алгоритмах; пусть K – число таких алгоритмов, которые назовем базовыми. Будем считать, что каждый такой алгоритм оперирует со своей моделью $M_k, k = 1,...,K,$ а в результате их деятельности формируется один или несколько вариантов решения.

Рассмотрим одну из возможных реализаций. Предположим, что работа алгоритмов может осуществляться асинхронно, а их взаимодействие на итерации h происходит путем формирования вышестоящей процедурой агрегированной модели (метамодели) M^{h+1} как с учетом отдельных моделей $M_1^h,...,M_K^h$ и сформированной на предыдущей итерации модели M^h, так и, возможно, сгенерированных алгоритмами решений. Таким образом, кооперативная метаэвристика оперирует несколькими моделями и осуществляет оптимизацию в пространстве моделей, с целью предоставления возможности базовым алгоритмам (или некоторым из них) генерировать наилучшие решения ЗКО.

```
procedure БазовыйАлгоритм(k, M)
      Инициализация;
      while (не завершена работа) do
            ГенерированиеВариантовРешения (M_k, M);
            ОбновлениеМодели (M_k);
      end while
      return НаилучшееНайденноеРешение;
end procedure
```

```
procedure УправляющаяМетаэвристика()
      Инициализация;
      for k := 1 to K do
            Запустить(БазовыйАлгоритм(k, M^1));
      h := 1;
      while (не выполняется условие завершения) do
            if (выполняется условие обмена)     then
                  for k := 1 to K do
                        M_k^h := ПолучитьМодель(M_k);
                  end for
                  M^{h+1} := СформироватьМодель(M^h, M_1^h,..., M_K^h);
                  for k := 1 to K do
                        РазослатьМодель(k, M^{h+1});
                  end for
                  h := h + 1;
            end if
      end while
      ЗавершитьРаботуАлгоритмов();
      return НаилучшиеНайденныеРешения;
end procedure
```

Рис. 1 Схема гибридной кооперативной моделе-ориентированной метаэвристики

Предлагается методология создания метаэвристических алгоритмов решения ЗКО, общая схема которых состоит из этапов, представленных на Рис. 1. Здесь для простоты изложения при формировании агрегированной модели (метамодели) не отражен явно фактор учета решений, сгенерированных базовыми алгоритмами.

Опишем ключевые этапы кооперативного алгоритма. После этапа инициализации осуществляется запуск всех базовых алгоритмов, каждый из которых итерационно и независимо генерирует решения и обновляет свою собственную модель. При выполнении условий обмена, текущие модели базовых алгоритмов (и, возможно, найденные ими варианты решений) используются для формирования новой агрегированной модели (метамодели). При этом может также использоваться агрегированная модель с предыдущей итерации обмена, а сам процесс формирования может быть представлен в виде оптимизационной проблемы поиска наилучшего элемента в пространстве моделей. После этого агрегированная модель рассылается базовым алгоритмам, где может, как использоваться в сочетании с их собственной моделью, так и заместить ее полностью. При выполнении условий завершения работы кооперативная метаэвристика возвращает один или несколько наилучших из найденных базовыми алгоритмами вариантов решения.

В качестве базовых алгоритмов могут выступать отдельные экземпляры одного алгоритма, в этом случае схема называется гомогенной. Экземпляры алгоритмов могут иметь различные значения параметров. В случае гетерогенной схемы (когда базовые алгоритмы не основываются на одном алгоритме/методе), модели базовых алгоритмов должны быть согласованы, т.е. должен существовать способ преобразования значений параметров и/или структур моделей между собой.

Управляющая метаэвристика может формировать на одной итерации не одну, а сразу несколько разных агрегированных моделей (метамоделей). Это целесообразно как в случае гетерогенной схемы, когда базовые алгоритмы могут использовать модели разной структуры, так и когда требуется исследовать несколько различных моделей одним алгоритмом.

При использовании данной методологии необходимо определить:

1. тип(ы) базовых моделе-ориентированных алгоритмов;
2. способ использования базовыми алгоритмами агрегированной модели (метамодели);
3. условия обмена (схема коммуникации);
4. способ формирования агрегированной модели (метамодели);
5. условия завершения работы кооперативной метаэвристики.

Предполагается, что применение разработанной методологии позволит диверсифицировать работу базовых алгоритмов, что уменьшает вероятность завершения процесса поиска в областях, не содержащих глобально-оптимальное решение. Обмен информацией между базовыми алгоритмами позволяет надеется на повышение эффективности процесса поиска за счет проявления синергетических эффектов. Таким образом, выбор конкретной схемы обмена информации (способа кооперирования) определяет баланс между интенсификацией и диверсификацией поиска.

Разработанная схема может также служить основой для классификации гибридных алгоритмов КО, например, на основе параметризации таких показателей, как синхронность выполнения и схемы взаимодействия базовых алгоритмов, их однотипность, количество используемых моделей и методы формирования агрегированных моделей, степень использования предыстории.

Для иллюстрации и первичного исследования данной схемы была разработана кооперативная метаэвристика на основе алгоритмов ОМК. Она изложена после краткого описания метаэвристки ОМК.

Оптимизация муравьиными колониями

В последние годы активно развиваются методы так называемого роевого интеллекта, в которых совокупность сравнительно простых агентов конструирует стратегию своего поведения без наличия

глобального управления [Kennedy et al., 2001]. Одним из широко известных роевых методов является метод ОМК. По аналогии с биологической моделью, ОМК базируется на непрямом обмене информацией колонии агентов, называемых искусственными муравьями, использующих феромонные следы как коммуникационное средство [Dorigo and Stützle, 2004]. Феромонные следы в ОМК служат распределенной численной информацией, которая, наряду с эвристической информацией о задаче, используется муравьями для недетерминированного конструирования решений задачи и которую муравьи адаптивно изменяют для отображения опыта, накопленного в процессе поиска решения. Важно отметить, что хотя каждый муравей достаточно сложен, чтобы найти решение задачи, хорошие решения, обычно, появляются только в результате коллективного взаимодействия между муравьями, посредством записи/считывания значений феромонных следов. В известном смысле, это является распределенным процессом обучения, в котором отдельные агенты, муравьи, не адаптируются, а наоборот, адаптивно изменяют вид задачи и ее восприятие другими агентами.

Кроме построения муравьями решений, метаэвристика ОМК включает еще две процедуры (РИС. 2) [Dorigo and Stützle, 2004]: обновление феромонных значений и действия демона (включение процедуры действий демона необязательное). Главная процедура метаэвристики ОМК – ПланированиеДействий не определяет, каким образом распланированны и синхронизированны ПостроениеМуравьямиРешений, ОбновлениеФеромона и ДействияДемона, что оставляет разработчику свободу в определении способа взаимодействия этих трех процедур.

Обновление феромонных значений включает как увеличение значений – добавление муравьями феронона согласно построенному ими решению, так и уменьшение – испарение феромона, процесс с помощью которого интенсивность феромонного следа автоматически уменьшается со временем.

```
procedure ОМК()
      while (не выполняется условие завершения) do
            ПланированиеДействий
                  ПостроениеМуравьямиРешений();
                  ОбновлениеФеромона();
                  ДействияДемона(); {необязательно}
            end ПланированиеДействий
      end while
end procedure
```

Рис. 2 Метаэвристика оптимизации муравьиными колониями

Испарение осуществляет полезную форму «забывания», содействуя исследованию новых областей в пространстве поиска и избежанию очень быстрой сходимости алгоритма к субоптимальной области.

Действия демона могут использоваться для осуществления централизованных процедур, которые не могут быть выполнены отдельными муравьями. Например, активация процедуры локальной оптимизации или сбор глобальной информации, которая может быть использована для принятия решения при откладывании дополнительного феромона для отклонения процесса поиска от локальной перспективы.

Алгоритмы ОМК успешно применяются ко многим сложным ЗКО. Для отдельные их классов были получены теоретические результаты, свидетельствующие о сходимости по значению к глобально-оптимальному решению [Dorigo and Blum, 2005].

Детально метод описан, в частности, в работах [Dorigo and Stützle, 2004; Dorigo and Blum, 2005].

Кооперативная метаэвристика на основе алгоритмов ОМК

В рамках предложенной методологии была разработана кооперативная гибридная метаэвристика в виде последовательного (для одного процессора) алгоритма, который включает несколько алгоритмов ОМК (Рис. 3). Данная схема не предписывает, следует ли использовать экземпляры одного и того же алгоритма, либо она может быть гетерогенной, что оставляет разработчику выбор при реализации.

Алгоритмы работают синхронно и после завершения ими заданного количества IT итераций, совершается обмен информацией. В рамках данной схемы агрегированная модель формируется только на основе текущих моделей базовых алгоритмов (без учета предыстории) и после формирования замещает собою их модели.

В качестве процедуры агрегирования была реализована схема взвешенного суммирования значений параметров (феромонных значений) моделей базовых алгоритмов $\tau_1,...,\tau_K$ (для простоты изложения будем считать, что значения параметров нормированы). Веса вычисляются на основе значений целевой функции на текущих наилучших вариантах решений $f_1^{opt},...,f_K^{opt}$ и средних значений целевой функции $f_1^{avg},...,f_K^{avg}$ по найденным базовыми алгоритмами решениям на последней завершенной итерации по следующему соотношению:

$$\tau = \sum_{k=1,...,K} w_k \tau_k \ , \ w_k = 0.2/K + 0.6\delta_{f_k^{opt},f^{opt}} + 0.2\delta_{f_k^{avg},f^{avg}} \ , k = 1,...,K \ ,$$

где $f^{opt} = \min_{k=1,...,K}\left\{f_k^{opt}\right\}, f^{avg} = \min_{k=1,...,K}\left\{f_k^{avg}\right\}, \delta_{x,y} = \begin{cases} 0, & x \neq y; \\ 1, & x = y. \end{cases}$

```
procedure Кооперативная_метаэвристика_ОМК()
    Инициализация (τ^1);
    h := 1
    while (h · IT ≤ IT_max)
          for k := 1 to K do
                τ_k^h := τ^h;
          end for
          for i := 1 to IT do
                for k := 1 to K do
                      ПостороениеМуравьямиРешений(τ_k^h);
                      ОбновлениеФеромона(τ_k^h);
                      ДействияДемона();
                end for
          end for
          ВычислитьВеса(w_1,...,w_K);
          τ^h := Σ_{k=1}^{K} w_k τ_k^h ;
          h := h + 1;
    end while
    return НаилучшееНайденноеРешение;
end procedure
```

Рис. 3 Кооперативная моделе-ориентированная схема на основе алгоритмов ОМК

Поскольку исследовалось влияние наличия обмена информацией в кооперативной схеме на эффективность, то в качестве условия завершения было выбрано выполнение базовыми алгоритмами определенного количества итераций IT_{max}.

Вычислительный эксперимент

Поскольку теоретическое исследование метаэвристических алгоритмов решения ЗКО крайне редко позволяет получать практически применимые результаты, принято анализировать показатели эффективности путем проведения вычислительных экспериментов. С этой целью обычно используют

"классические" модели комбинаторной оптимизации – такие, например, как задача коммивояжера (ЗК) [Hoos and Stützle, 2005], которая состоит в поиске минимального гамильтонового цикла в полном взвешенном графе.

Проведенный эксперимент состоял в сравнении результатов, полученных как разработанной гибридной гетерогенной метаэвристикой, созданной на основе использования двух алгоритмов ОМК (следовательно, $K = 2$), так и алгоритмом, идентичным указанной метаэвристике, но с отключенным модулем обмена. Таким образом, в качестве последнего рассматривался комбинированный алгоритм, включающий два экземпляра алгоритма ОМК, работающих параллельно и независимо, который возвращал в конце работы лучшее из двух найденных решений.

Формирование управляющей метаэвристикой агрегированной модели осуществлялось после каждых 100 итераций, выполненных базовыми алгоритмами. Критерием завершения работы всего кооперативного метода являлось выполнение каждым из базовых алгоритмов 1000 итераций.

В качестве базовых алгоритмов при реализации были использованы экземпляры известного алгоритма Max-Min Ant System (MMAS) с реинициализацией феромонных значений [Stützle and Hoos, 2000] и экземпляры алгоритма Ant Colony System (ACS) [Dorigo and Gambardella, 1997], реализация которых взята из пакета ACOTSP [Stützle, 2004]. Феромонные значения реинициализируются, если решения сгенерированные на одной итерации достаточно близки друг к другу и на протяжении заданного количества итераций алгоритмом не было найдено улучшения. Модель в этих алгоритмах ОМК для ЗК представлена в виде квадратной матрицы, элементами которой являются феромонные значения. В алгоритме MMAS используются динамические ограничения τ_{max}, τ_{min} на значения элементов феромонной матрицы (параметров модели), поэтому перед агрегированием матриц осуществлялось масштабирование элементов с приведением их значений к отрезку [0,1]: $\tau_i^{norm} = \left(\tau_i - \tau_{min}\right) \Big/ \left(\tau_{max} - \tau_{min}\right)$. В алгоритме ACS такие ограничения отсутствуют и при масштабировании использовались минимальный и максимальный элемент матрицы, соответственно. Аналогичным образом осуществлялось масштабирование агрегированных матриц при их передаче базовым алгоритмам с учетом соответствующих ограничений.

Параметры базовых алгоритмов устанавливались равными стандартными значениям, рекомендуемыми в литературе для ЗК [Dorigo and Stützle, 2004], т.е. дополнительная их оптимизация для конкретного набора тестовых ЗК не проводилась. Количество муравьев составляло 25, коэффициент испарения феромона ρ = 0.5, в псевдослучайном пропорциональном правиле выбора, которое применяли муравьи при построении решений, были такие значения параметров: α = 1, β = 2. В качестве деятельности демона в алгоритмах ОМК ко всем построенным муравьями решениям применялся алгоритм простого локального поиска 3-opt [Hoos and Stützle, 2005], реализация которого также взята из пакета ACOTSP [Stützle, 2004].

В табл. 1 приведены результаты 20 вариантов решения каждой из представленных симметричных ЗК (таких, в которых граф является неориентированным) с известными оптимальными решениями, взятых из Интернет-библиотеки TSPLIB [TSPLIB, 2009]. Здесь число в названии задачи обозначает ее размерность, f_{opt} – известное значение целевой функции в точке глобального минимума, f_{min} – лучшее найденное соответствующим алгоритмом значение целевой функции, δ_{avg} – средняя относительная погрешность алгоритма (%), i_{avg} – среднее количество итераций на протяжении которых алгоритмом было найдено лучшее решение, t_{avg} – среднее время, на протяжении которого алгоритмом было найдено лучшее решение на ПЭВМ класса Pentium IV 2,66 ГГц (с). Разработанная гибридная кооперативная метаэвристика обозначена в таблице как **ACS_MMAS_кооп**, а алгоритм, объединяющий независимые экземпляры ACS и MMAS, – **ACS_MMAS_нез.**

Отметим, что разработанный метод сравнивался с комбинированным алгоритмом **ACS_MMAS_нез,** являющимся достаточно эффективным, поскольку для всех задач (кроме задачи d657), участвующих в

эксперименте, он уже находил глобально-оптимальные решения хотя бы в одной из попыток. Наличие обмена информации не ухудшило этот показатель в алгоритме **ACS_MMAS_кооп**, он также нашел глобально-оптимальные решения тестовых задач (за исключением задач d493 и d657). Но для большинства задач алгоритм **ACS_MMAS_кооп** показал лучшую среднюю относительную погрешность (исключение составляет задача ali535). Следует подчеркнуть, что именно улучшение показателя эффективности в среднем и является важным, поскольку в эксперименте исследовались алгоритмы, которые состояли в одном запуске алгоритма (без рестарта) **ACS_MMAS_нез** и **ACS_MMAS_кооп**, соответственно.

Таблица 1 Результаты решения ЗК

Задача	f_{opt}	**ACS_MMAS_нез**				**ACS_MMAS_кооп**			
		f_{min}	δ_{avg} ,%	i_{avg}	t_{avg} , сек	f_{min}	δ_{avg} ,%	i_{avg}	t_{avg} , сек
d493	35002	**35002**	0.021	620.05	61.30	35004	**0.013**	561.7	54.55
att532	27686	**27686**	0.051	302.90	34.38	**27686**	**0.032**	450.0	47.53
ali535	202339	**202339**	**0.005**	406.00	64.31	**202339**	0.007	268.5	40.92
u574	36905	**36905**	0.052	522.40	63.21	**36905**	**0.008**	412.7	49.91
p654	34643	**34643**	0.026	457.20	53.27	**34643**	**0.010**	635.1	73.69
d657	48912	**48913**	0.078	550.70	83.55	**48913**	**0.032**	574.3	80.94
u724	41910	**41910**	0.047	602.25	77.76	**41910**	**0.032**	350.8	45.47
rat783	8806	**8806**	0.065	522.50	77.73	**8806**	**0.026**	590.8	82.54
pr1002	259045	**259045**	0.160	721.90	153.80	**259045**	**0.077**	775.7	165.83
vm1084	239297	**239297**	0.025	580.90	212.40	**239297**	**0.012**	470.1	166.38

Время работы и количество итераций до нахождения лучшего варианта решения, у алгоритма **ACS_MMAS_кооп** были сравнимыми с соответствующими показателями алгоритма **ACS_MMAS_нез**.

Результаты вычислительного эксперимента свидетельствуют о том, что данная методология позволяет повышать эффективность моделе-ориентированных алгоритмов путем организации кооперативного решения задачи.

Заключение

В работе предложена методология разработки гибридных кооперативных моделе-ориентированных метаэвристик, предполагающая наличие управляющего уровня, на котором осуществляется построение агрегированной (мета) модели, а в качестве исходных данных для такого построения выступают частные модели, сформированные базовыми алгоритмами. Важнейшей особенностью предлагаемых кооперативных методов является то, что они оперируют с несколькими моделями, а оптимизацию осуществляют в пространстве моделей, а не в пространстве решений исходной ЗКО. В качестве базовых алгоритмов могут выступать как отдельные экземпляры одного моделе-ориентированного алгоритма, так и экземпляры алгоритмов разных методов.

Для оценки эффективности предлагаемой методологии в рамках описанного подхода был разработан и исследован экспериментально гетерогенный кооперативный метод на базе алгоритмов ОМК. Результаты вычислительного эксперимента по решению ЗК показали, что разработка методов согласно предложенной методологии может дать повышение эффективности по сравнению с параллельным и независимым выполнением составных алгоритмов с использованием лучшего результата.

Предметом дальнейшего исследования являются разработка и исследования схем агрегирования моделей (метаоптимизации), в частности, с использованием других базовых алгоритмов, формулирование условий на базовые алгоритмы, при которых эти схемы позволят достигнуть повышенных показателей эффективности, а также экспериментальное исследование при решении других ЗКО.

Предложенная методология может также служить основой для классификации прикладных алгоритмов решения ЗКО.

Благодарности

Работа опубликована при финансовой поддержке проекта **ITHEA XXI** Института информационных теорий и приложений FOI ITHEA Болгария www.ithea.org и Ассоциации создателей и пользователей интеллектуальных систем ADUIS Украина www.aduis.com.ua.

Список литературы

[Cotta, 1998] C. Cotta. A study of hybridisation techniques and their application to the design of evolutionary algorithms. In: AI Communications, 1998, Vol. 11, No. 3–4. pp. 223–224.

[Dorigo and Blum, 2005] M. Dorigo, C. Blum. Ant colony optimization theory: A survey. In: Theoretical computer science, 2005, Vol. 344, No. 2–3. pp. 243–278. http://code.ulb.ac.be/dbfiles/DorBlu2005tcs.pdf

[Dorigo and Gambardella, 1997] M. Dorigo, L.M. Gambardella. Ant Colony System: A cooperative learning approach to the traveling salesman problem. In: IEEE Transactions on Evolutionary Computation, 1997, Vol. 1, No. 1, pp. 53–66.

[Dorigo and Stützle, 2004] M. Dorigo, T. Stützle. Ant Colony Optimization. Cambridge: MIT Press, 2004. 348 p.

[Hoos and Stützle, 2005] H.H. Hoos, T. Stützle. Stochastic Local Search: Foundations and Applications. San Francisco: Morgan Kaufmann Publ, 2005. 658 p.

[Kennedy et al., 2001] J. Kennedy, R.C. Eberhart, Y. Shi. Swarm Intelligence. Morgan Kaufmann, San Francisco, CA, 2001. 512 p.

[Larrañaga and Lozano, 2002] P. Larrañaga, J.A. Lozano. Estimation of Distribution Algorithms. A New Tool for Evolutionary Computation. Boston, MA: Kluwer Academic, 2002. 382 p.

[Raidl, 2006] G.R. Raidl. A Unified View on Hybrid Metaheuristics. In: Lecture Notes in Computer Science. Springer-Verlag, 2006, Vol. 4030. pp. 1–12. http://www.ads.tuwien.ac.at/publications/bib/pdf/raidl-06.pdf

[Rubinstein and Kroese, 2004] R.Y. Rubinstein, D.P. Kroese. The cross-entropy method: a unified approach to combinatorial optimization, Monte-Carlo simulation, and machine learning. Springer, 2004. 300 p.

[Stützle and Hoos, 2000] T. Stützle, H. Hoos. Max – Min Ant System. In: Future Generation Computer Systems. 2000, No. 8. pp. 889–914.

[Stützle, 2004] T. Stützle. ACOTSP, Version 1.0. http://www.aco-metaheuristic.org/aco-code. 2004.

[Taludar et al., 2003] S. Talukdar, S. Murty, R. Akkiraju. Asynchronous teams. In: Handbook of Metaheuristics. Vol. 57. Kluwer Academic Publishers. 2003. pp. 537–556.

[TSPLIB, 2009] TSPLIB. http://comopt.ifi.uni-heidelberg.de/software/TSPLIB95/. Страница посещалась в июне 2009.

[Zlochin et al., 2004] M. Zlochin, M. Birattari, N. Meuleau, M. Dorigo. Model-Based Search for Combinatorial Optimization: A Critical Survey. In: Annals of Operations Research, 2004, Vol. 131. pp. 373–395. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.66.672&rep=rep1&type=pdf

[Гуляницький, 2008] Л.Ф. Гуляницький. До формалізації та класифікації задач комбінаторної оптимізації. В: Теорія оптимальних рішень. 2008, № 7. С. 45-49. http://www.nbuv.gov.ua/portal/natural/Tor/2008/TOP_Gulianickiy_2008.pdf

[Гуляницький, 2009] Л.Ф. Гуляницький. Розробка кооперативних метаевристик. In: Abstract of Int. Conf. "Problems of Decision Making under Uncertainties (PDMU–2009)" (April 27-30, 2009, Skhidnytsia, Ukraine). Kyiv, 2009. pp. 90–91.

Информация об авторах

Леонид Гуляницкий *(Hulianytskyi) – д.т.н., ведущий научный сотрудник Института кибернетики им. В.М. Глушкова НАН Украины, пр. Глушкова, 40, Киев, 03680, Украина. e-mail: lh_dar@hotmail.com*

Сергий Сиренко *(Sirenko) – аспирант, Институт кибернетики им. В.М. Глушкова НАН Украины, пр. Глушкова, 40, Киев, 03680, Украина. e-mail: mail@sirenko.com.ua*

ОБ ОСОБЕННОСТЯХ ПРИНЯТИЯ РЕШЕНИЙ ПРИ ОРГАНИЗАЦИИ ВЫЧИСЛЕНИЙ НА ОСНОВЕ МЕТОДА БАЗИСНЫХ МАТРИЦ

Алексей Волошин, Всеволод Богаенко, Владимир Кудин

Аннотация. *Использование разных типов данных при проведении вычислений (чисел с плавающей запятой одинарной, двойной, повышенной точности) существенно влияет на основные критерии оценки эффективности: быстродействие, точность и объемы вычислений. В работе исследовано влияние использования разных вариантов организации вычислений на эффективность алгоритмов метода базисных матриц. Предложен вариант построения системы поддержки принятия решения для организации вычислений на линейных моделях для достижения заданных значений параметров по основным критериям: точности и быстродействию.*

Ключевые слова*: линейная модель, типы данных, базисная матрица.*

ACM Classification Keywords: *H.4.2 Information Systems Applications: Types of Systems: Decision Support*

Conference*: The paper is selected from XVth International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.*

Введение

Известно, что большинство исследуемых физических процессов на определенном этапе моделирования описываются в классе линейных моделей, в частности, в виде системы плохо обусловленных линейных алгебраических уравнений(СЛАУ) с квадратной матрицей ограничений. Малые неточности представления подобных математических моделей могут существенно влиять на количественные и качественные характеристики получаемого решения при использовании конкретного метода (алгоритма) [Воеводин, 1979]. Такие неточности зачастую обусловлены ограниченностью длины мантиссы при представлении чисел с плавающей запятой. Важно отметить, что, несмотря на наличие ЭВМ, у которых операции округления реализованы самым лучшим образом, тем не менее, избежать ошибок округления и улучшить их известные теоретические оценки не удается [Воеводин, 1979]. Выбором конкретных типов данных с плавающей запятой, достигается различная эффективность по точности получаемого решения, быстродействию и объемам вычислений. Дополнительной возможностью для повышения эффективности вычислений является правильная организация использования регистровой, кэш-памяти, оперативной и внешней памяти [Деммель, 2001]. Известно, что скорость обменов между этими типами памяти зачастую намного выше скорости проведения вычислений и при плохой организации вычислений может "свести на нет" эффект от использования "быстрой" памяти. Все это обуславливает включение в контур принятия решения лица, принимающего решение (ЛПР), с целью правильной организации вычислительного процесса: указания механизма (процедуры) устранения неопределенностей при выборе приоритетов по критериям к решению и типам данных. Достигается это, например, использованием категории функций принадлежности [Орловский, 1981]. Такой подход усложняет исследования, но, в тоже время, открывает новые возможности для построения алгоритмов.

Постановка задачи

Предметом исследования является линейная модель – система линейных алгебраических уравнений вида:

$$Au = C, \quad (1)$$

где $A = \left\{a_{ij}\right\}_{\substack{i=\overline{1,m} \\ j=\overline{1,m}}}$ - квадратная матрица размерности $(m \times m)$, $a_j = (a_{j1}, a_{j2}, ..., a_{jm}),\ j \in J = I = \{1, 2, .., m\}$, - строки матрицы $\grave{A}$, $u = (u_1, u_2, ..., u_m)^T$ - вектор переменных, $C = (c_1, c_2, .., c_n)^T$ - вектор ограничений, $a_j u \le c_j,\ j \in J$, - полупространство, определенное гиперплоскостью $a_j u = c_j,\ j \in J$. Модель (1) исследуется в пространстве E^m.

Целью исследований является:

- экспериментальный анализ влияния использования различных типов данных, алгоритмов, уровня обусловленности системы при построении вычислительных схемы на основные параметры решения - точность решения и обращения матрицы, быстродействие;

- проверка эффективности вычислительных схем метода базисных матриц (МБМ) [Кудин, 2002] по указанным критериям на моделях заданной размерности;

- построение концепции принятия решения по достижению заданной эффективности вычислительного процесса;

Концепция анализа состоит из четырех стадий.

Первая стадия содержит анализ типовой модели заданной размерности на разрешимость (существование, единственность решений СЛАУ), исследование свойств модели на основе базисного метода и алгоритма.

Вторая – проведение расчетов на основе вычислительных алгоритмов (с выбором или без выбора главного элемента, без процедуры уточнения, с одно или двухстадийной процедурой уточнения в каждом из них), при разных сценариях объявления типов данных (с плавающей запятой размерностью 64, 128, 256 бит) и уровней плохой обусловленности (0-7). Фиксируются значения параметров быстродействия, точности решения и обращения матрицы, оценки количества используемой памяти.

Третья - построение функциональных зависимостей (интерполяционных многочленов) быстродействия, точности решения и обращения от типов объявлений переменных, уровня обусловленности системы.

Четвертая стадия построение на основе интерполяционных многочленов функций принадлежности с областью значений из интервала $[0,1]$. Формирование на основе функций принадлежности (при фиксированной размерности модели и уровня обусловленности) механизма выбора приемлемых значений параметров (алгоритма и типов данных) в вычислительной схеме для достижения желаемых значений основных критериев на найденном решении.

Основные положения метода базисных матриц (МБМ)

В предложенном МБМ введены в рассмотрение строчные базисные матрицы. Базисные матрицы последовательно изменяются замещением строк вспомогательной СЛАУ строками (нормалями ограничений) основной СЛАУ. В общем случае, в модели количество ограничений превышает количество

переменных вида (1), в данном случае $m = n$ (для анализа вводится в рассмотрение вспомогательная СЛАУ с известными свойствами соответствующей размерности) [Кудин, 2002].

Определение. Квадратную матрицу $A_б$, составленную из m линейно независимых нормалей ограничений, будем называть базисной, а решение соответствующей ей системы уравнений $A_б u = c^0$ базисным. Две базисные матрицы, отличающиеся одной строкой, будем называть смежными.

Установлены формулы связи базисного решения, коэффициентов разложения нормалей ограничений и целевой функции, коэффициентов обратной матрицы, невязок ограничений и значений целевой функции при переходе к базисной матрице $\overline{A}_б$ (смежной), которая образуется из матрицы $A_б$ заменой ее строки a_k на a_l, которая не входит в базисную матрицу $A_б$ [Волошин, 2009]. При нахождении формул и основных соотношений между элементами метода при переходе от одной базисной матрицы к следующей считаем $a_{i1}, a_{i2}, ..., a_{im}$ нормалями ограничений, $a_j u^T \le c_j, j \in J_б \quad J_б = \{i_1, i_2, ..., i_m\}$ - индексами ограничений, нормали которых образуют строки базисной матрицы $A_б$ a_l - нормалью ограничения $a_l u \le c_l$, $\alpha_l = (\alpha_{l1}, \alpha_{l2}, ..., \alpha_{lm})$ - коэффициентами разложения вектора a_l по строкам базисной матрицы $A_б$.

Теорема 1. Между элементами МБМ в смежных базисных матрицах имеют место соотношения:

$$\overline{\alpha}_{rk} = \frac{\alpha_{rk}}{\alpha_{lk}}, \ \overline{\alpha}_{ri} = \alpha_{ri} - \frac{\alpha_{rk}}{\alpha_{lk}}\alpha_{li}, \ r = \overline{0,n}; \ i = \overline{1,m}; \ i \ne k; \tag{2}$$

$$\overline{e}_{rk} = \frac{e_{rk}}{\alpha_{lk}}, \ \overline{e}_{ri} = e_{ri} - \frac{e_{rk}}{\alpha_{lk}}\alpha_{li}, \quad r = \overline{1,m}; \ i = \overline{1,m}; \ i \ne k; \tag{3}$$

$$\overline{u}_{0j} = u_{0j} - \frac{e_{jk}}{\alpha_{1k}}\Delta_l, \ j = \overline{1,m}, \tag{4}$$

$$\overline{\Delta}_k = -\frac{\Delta_l}{\alpha_{lk}}, \ \overline{\Delta}_r = \Delta_r - \frac{\alpha_{rk}}{\alpha_{lk}}\Delta_l, \ r = \overline{1,n}; \ r \ne k; \tag{5}$$

причем условием невырожденности является условие $\alpha_{lk} \ne 0$, допустимости опорного базисного решения - $\alpha_{lk} < 0$, роста значения целевой функции - $\alpha_{0k} < 0$. Здесь e_{ri} - элементы матрицы $A_б^{-1}$, обратной к $A_б$; $e_k = (A_б^{-1})_k$.- столбец обратной матрицы, $\Delta_r = a_r u_0^T - c_r$ - невязка r-го ограничения в вершине u_0.

Теорема 2. Если $\exists$ индекс k такой, что $\alpha_{0k} < 0$ и $\alpha_{rk} \ge 0$ для всех небазисных r, то целевая функция (4) задачи принимает неограниченные значения на множестве допустимых решений.

На основе (2)-(5) построена вычислительная схема метода базисных матриц.

Вычислительный эксперимент

Возможность применения различных типов данных на разных модификациях алгоритмов МБМ позволяет построить систему поддержки принятия решения, которая на основе заданных пользователем ограничений на такие параметры алгоритмов, как быстродействие и точность, позволит выбрать наилучший по этим критериям алгоритм. Выбор базируется на эвристических зависимостях параметров

алгоритма и значений критериев от параметров матрицы ограничений СЛАУ: размерности и оценки числа обусловленности. Быстродействие алгоритмов зависит только от размерности матрицы n и некоторого параметра быстродействия системы для заданного типа данных и может быть оценена как

$$T^0(n) = \sum_{i=1}^{n} \big(2ni + n\big) t_c = (n^3 + 2n^2) t_c, \tag{6}$$

для алгоритма без выбора ведущего элемента, и

$$T^2(n) = (3n^3 + n^2) t_c, \tag{7}$$

для алгоритма с выбором ведущего элемента, где значение параметра t_c должно оцениваться для каждой конкретной аппаратно-программной платформы экспериментальным путем.

Для построения эвристических зависимостей между точностью решения, числом обусловленности и размерностью матрицы ограничений, была проведена серия вычислительных экспериментов.

Проводилось решение разными алгоритмами МБМ СЛАУ размерностью 256x256 с матрицей ограничений:

$$A = \begin{cases} a_i = (rnd(1.0), ..., rnd(1.0)),\ i = 0 \\ a_i = a_{i-1} + \dfrac{\alpha}{n}(rnd(1.0), ..., rnd(1.0)),\ i > 0, \end{cases}$$

где $\| a_i - a_j \|^2 < \alpha$, α - число, которое коррелирует с числом обусловленности системы. В качестве критериев точности брались точность обращения матрицы $\varepsilon_1 = \left\| I - A^{-1} A \right\|$ и точность машинного решения u_0 (1) в сравнении с аналитическим (точным) $u = (1, 0, ..., 0)$, $\varepsilon_2 = \left\| u_0 - (1, 0, ..., 0) \right\|$. Полученные экспериментальные данные приведены в табл. 1,2. Из них следует, что каждый из алгоритмов дает решения, приближенные к точному, только в определённом диапазоне значений коэффициента α. Когда же α выходить за пределы этого диапазона, СЛАУ становится неразрешимой с помощью алгоритма. В пределах диапазона применения, зависимость погрешности решения от α является линейной для всех рассматриваемых алгоритмов. Эти зависимости, построенные на основе данных из табл. 1,2, приведены в табл. 3, в которой используются следующие обозначения: double –числа с плавающей запятой двойной точности (64битной размерности), dd –числа с плавающей запятой размерности 128 бит, qd – числа с плавающей запятой размерности 256 бит, +1, +2 - количество итераций уточнения, $\log_{10} \alpha$ - уровни плохой обусловленности системы (0-7).

Таблица 1. Порядок точности обращения матрицы ($\log_{10} \varepsilon_1$)

$\log_{10} \alpha$	0	1	2	3	4	5	6	7
Алгоритм без выбора ведущего элемента								
Double	-15.30	-12.48	-10.31	-9.15	-5.47			
double+1	-18.95	-17.62	-15.87	-13.96	-8.12	-9.63	-6.44	
double+2	-16.84	-15.80	-15.79	-14.08	-5.86	-9.30	-5.95	
Dd	-46.26	-43.88	-41.67	-39.77	-39.57	-37.05		
dd+1	-51.36	-48.99	-48.01	-44.47	-43.88	-41.87	-39.02	
dd+2	-51.22	-49.99	-47.43	-45.22	-43.99	-41.83	-39.68	-38.17
Qd	-108.82	-110.08	-108.03	-102.56	-102.79			

qd+1	-116.80	-112.97	-113.53	-110.23	-109.95	-108.10	-105.26	
qd+2	-117.52	-116.17	-113.75	-111.40	-108.72	-105.66	-106.13	-103.19
Алгоритм с выбором ведущего элемента								
Double	-17.70	-16.82	-14.54	-13.07	-9.87	-8.95	-7.10	
double+1	-19.02	-17.63	-15.83	-13.92	-8.27	-9.57	-6.33	-5.31
double+2	-16.83	-15.77	-15.76	-14.06	-5.89	-9.37	-5.93	-3.22
dd	-49.96	-47.71	-46.16	-43.81	-43.16	-40.95	-38.16	-35.42
dd+1	-51.31	-48.97	-47.99	-44.48	-43.85	-41.86	-39.33	-38.69
dd+2	-51.21	-49.90	-47.41	-45.21	-44.03	-41.83	-39.70	-38.16
qd	-115.70	-113.03	-111.33	-109.47	-106.46	-104.84	-103.51	-100.48
qd+1	-117.04	-114.58	-113.83	-110.06	-109.04	-106.80	-105.47	-102.29
qd+2	-112.25	-114.94	-113.79	-111.73	-108.24	-105.84	-104.93	-103.90

Таблица 2. Порядок отклонения машинного решения от аналитического, $\log_{10} \varepsilon_2$

$\log_{10} \alpha$	0	1	2	3	4	5	6	7
Алгоритм без выбора ведущего элемента								
double	-12.87	-7.95	-3.95					
double+1	-13.84	-13.07	-6.55	-4.04				
double+2	-9.08	-7.72	-7.11	-3.38				
dd	-44.92	-40.09	-36.43	-29.96	-29.35	-25.62		
dd+1	-45.81	-44.34	-38.23	-34.12	-30.16	-26.51	-22.26	
dd+2	-45.25	-42.67	-38.96	-32.50	-31.81	-26.97	-23.13	-20.36
qd	-106.98	-105.95	-103.00	-95.91	-93.20			
qd+1	-111.25	-104.62	-103.19	-98.63	-96.29	-94.11	-89.22	
qd+2	-111.67	-109.49	-107.34	-99.94	-96.08	-87.79	-89.67	-84.60
Алгоритм с выбором ведущего элемента								
double	-15.81	-12.29	-8.89	-2.75				
double+1	-12.40	-10.55	-5.37	-2.70				
double+2	-10.85	-8.55	-6.75	-2.27				
dd	-46.16	-42.45	-39.07	-34.66	-32.50	-27.71	-22.33	-19.33
dd+1	-45.92	-43.43	-36.86	-31.96	-29.12	-26.33	-22.43	-17.85
dd+2	-44.41	-42.15	-36.89	-32.16	-31.15	-26.44	-21.90	-20.09
qd	-113.35	-107.94	-104.76	-100.43	-96.25	-93.69	-86.65	-85.01
qd+1	-109.58	-108.57	-103.25	-95.93	-93.22	-91.00	-87.99	-80.48
qd+2	-101.21	-106.73	-107.70	-100.30	-93.89	-87.59	-86.52	-85.47

Таблица 3. Коэффициенты линейных зависимостей погрешностей обращения и решения от α

	Погрешность обращения $\mathrm{E}_i(256,\alpha) = a_1\alpha + b_1$		Погрешность решения $\mathrm{E}_i(256,\alpha) = a_2\alpha + b_2$	
	a_1	b_1	a_2	b_2
Алгоритм без выбора ведущего элемента				
double	2.298301	-15.13695	4.460865	-12.71689
double+1	2.187973	-19.50594	3.594783	-14.76752

double+2	1.984797	-17.89935	1.772247	-9.482547
dd	1.739039	-45.71417	3.862877	-44.0509
dd+1	1.977781	-51.30349	4.085718	-46.74769
dd+2	1.916184	-51.39725	3.674145	-45.56692
qd	1.959164	-110.3728	3.759974	-108.5303
qd+1	1.712103	-116.1141	3.357264	-109.6857
qd+2	2.113576	-117.7165	4.179782	-112.9506
Алгоритм с выбором ведущего элемента				
double	1.865153	-18.17445	4.258355	-16.31943
double+1	2.105514	-19.35422	3.425782	-12.8938
double+2	2.044694	-18.01112	2.752232	-11.2342
dd	1.974183	-50.0744	3.864485	-46.55131
dd+1	1.852254	-51.04376	3.999826	-45.73785
dd+2	1.908559	-51.36158	3.617025	-44.55755
qd	2.102657	-115.4602	4.073971	-112.7673
qd+1	2.034889	-117.0114	4.119735	-110.671
qd+2	1.61611	-115.1076	3.309921	-107.7606

Учитывая полученные зависимости, алгоритм работы системы поддержки принятия решений можно описать следующим образом.

- **Входные данные:** t_m - максимально допустимое время расчётов; ε_m - максимально допустимая погрешность решения; P - приоритетность критериев (времени и погрешности).
- **Алгоритм:**

1) Построение множества алгоритмов, которые удовлетворяют условию относительно времени расчётов:

   $A_t = \{A_i : T_i(n) < t_m\}$, где $T_i(n)$ - время работы алгоритма A_i, которое оценивается по формуле (1) или (2).

2) Построение множества алгоритмов, которые удовлетворяют условию относительно погрешности:

   $A_\varepsilon = \{A_i : \mathrm{E}_i(n,\alpha) < \varepsilon_m\}$, где $\mathrm{E}_i(n,\alpha)$ - оценка погрешности обращения матрицы или решения СЛАУ для алгоритма A_i (из табл.3).

3) Если $A_t \cap A_\varepsilon \neq \varnothing$, в зависимости от приоритетности выбирается самый быстрый или самый точный алгоритм из $A_t \cap A_\varepsilon$.

4) В случае, когда $A_t \cap A_\varepsilon = \varnothing,\ A_t \neq \varnothing,\ A_\varepsilon \neq \varnothing$, в зависимости от приоритетности выбирается самый быстрый алгоритм из A_ε или самый точный алгоритм из A_t

5) Если $A_t = \varnothing$ или $A_\varepsilon = \varnothing$, в зависимости от приоритетности, выбирается самый быстрый или самый точный алгоритм из $A_t \cup A_\varepsilon$.

Выводы

Концепция принятия решения на основе МБМ имеет следующие свойства:

- Возможность находить решение квадратной системы уравнений за фиксированное время с заданной точностью;
- Возможность использовать решение исходной модели при анализе возмущенной модели;
- Возможность контролировать или направлено изменять величину ранга системы;
- Возможность проводить анализ свойств системы при изменении значений отдельных элементов и ее компонент.

Благодарности

Работа опубликована при финансовой поддержке проекта **ITHEA XXI** Института информационных теорий и приложений FOI ITHEA Болгария www.ithea.org и Ассоциации создателей и пользователей интеллектуальных систем ADUIS Украина www.aduis.com.ua.

Список литературы

[Воеводин, 1979] Воеводин В.В. Вычислительные основы линейной алгебры. Макроэкономика.-К.:Вища школа,- 2004. - 207с.

[Волошин, 2009] Кудин. В., Богаенко В., Волошин А. Анализ малых возмущений линейных экономико-математических моделей // Information Science & Computing, International Book Series, Number 10, Volume 3, ITHEA, SOFIA, 2009.- P. 67-73.

[Деммель, 2001] Деммель Дж. Вычислительная линейная алгебра. Теория и приложение М.: Мир, -2001.-430с.

[Кудин, 2002] Кудин В.И. Застосування методу базисних матриць при дослідженні властивостей лінійної системи // Вісник Київського університету. Серія фіз.-мат. науки. - 2002. - 2. - С. 56-61.

[Орловский, 1981] Орловский С.А Принятие решений при нечёткой исходной информации.- М:, Наука, - 1981. - 206с.

Информация об авторах

Алексей Ф. Волошин *– Киевский национальный университет имени Тараса Шевченко, факультет кибернетики, Украина, д. т. н., профессор, E-mail: ovoloshin@unicyb.kiev.ua*

Всеволод А. Богаенко *– Киев,. Институт кибернетики имени В.Г. Глушкова НАН Украины, Украина,, к. т. н., н. с., E-mail: sevab@ukrnet.ua*

Владимир И. Кудин *– Киевский национальный университет имени Тараса Шевченко, факультет кибернетики, Украина,, д. т. н., с. н. с., E-mail: V_I_Kudin@mail.ru*

ПОИСК ИНДИВИДУАЛЬНО-ОПТИМАЛЬНЫХ РАВНОВЕСИЙ В УСЛОВИЯХ ЧАСТИЧНОЙ ИНФОРМИРОВАННОСТИ ИГРОКОВ

Сергей Мащенко

Аннотация: *Рассматривается понятие индивидуально- оптимального равновесия в некооперативных играх. Описаны методы параметризации множества этих равновесий. Предлагается процедура поиска индивидуально-оптимальных равновесий в условиях частичной информированности игроков.*

Ключевые слова*: некооперативные игры, равновесие по Нешу, оптимальность по Парето, индивидуально-оптимальное равновесие.*

ACM Classification Keywords*: H4.2 Decision support*

Conference*: The paper is selected from XVth International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.*

Введение

Наиболее привлекательными концепциями оптимальности в условиях полной информированности игроков являются принципы оптимальности по Парето и по Нешу [1].

Концепция оптимальности по Парето основана на идее кооперативного поведения игроков, когда они коллективно выбирают свои стратегии и совместно учитывают функции выигрыша. Поэтому не существует ситуаций, которые будут для всех игроков одновременно лучше любой Парето-оптимальной. В случае, когда игроки выбирают основой для соглашения между собой концепцию Парето-оптимальности, у некоторых из них может возникнуть соблазн при выборе конкретной Парето-оптимальной ситуации изменить свою стратегию на другую, которая будет лучше для них. В этом случае такая ситуация будет нестабильной и их договоренность может быть разрушенной.

Концепция равновесий по Нешу основывается на идее некооперативного поведения игроков, когда они индивидуально выбирают свои стратегии и каждый учитывает лишь свою функцию выигрыша. Ситуация игры называется равновесием Неша, если от нее невыгодно отклоняться любому одному игроку (все другие игроки свои стратегии не изменяют), поскольку значение его функции выигрыша не улучшится (будет для него оптимальным). Если игроки заключают соглашение о своем будущем поведении и его основой является равновесие Неша, то оно будет стабильным. "Ценой" привлекательности равновесий Неша являются серьезные проблемы, которые связаны с их существованием, сложностью нахождения, проблемой выбора единственного равновесия [1].

В определенном смысле, эти принципы оптимальности представляют собой крайности в поведении игроков между коллективным и индивидуальным выбором стратегий и учетом функций выигрыша игроков.

Принцип индивидуальной оптимальности [2], предоставляет возможность каждому игроку выбирать свои стратегии индивидуально (некооперативно), но учитывать при этом интересы всех других игроков (компромисс ради разрешения конфликта). Этот принцип обоснован в, так называемых, одноцелевых играх, где у всех игроков цель – одна, но она характеризуется для каждого игрока своей функцией выигрыша. В идеале, эта цель заключается в выборе игроками своих стратегий так, чтобы сложилась наиболее предпочтительная ситуация для всех игроков. Поскольку такие ситуации могут не существовать, то игроки могут согласиться на компромисс ради общей цели.

Индивидуально-оптимальные равновесия

Рассмотрим одноцелевую игру *G* в нормальной форме $(X_i, u_i; i \in N)$, где $N = \{1,2,...n\}$ - множество из *n* игроков; X_i - множество стратегий *i*-го игрока; $u_i(x)$ - его функция выигрыша, которая определена на множестве ситуаций $X = \prod_{i \in N} X_i$ игры и принимает действительные значения. Каждый из игроков стремится получить по возможности большее значения своей функции выигрыша. Поскольку игроки имеют общую цель, они могут соглашаться на компромисс.

Принцип индивидуальной оптимальности основывается на специальном отношении *NE*-доминирования. Будем говорить, что ситуация *y* игры *G* находится в отношении сильного *NE*-доминирования игрока $i \in N$ к ситуации *x* ($y \overset{NE(i)}{\succ\!\!\succ} x$), если $y = (y_i, x_{N\setminus i})$, где вектор $x_{N\setminus i} = (x_j)_{j \in N\setminus\{i\}}$, и $u_j(y_i, x_{N\setminus i}) > u_j(x), \forall j \in N$.

Ситуация *x** называется слабым индивидуально-оптимальным равновесием (множество этих равновесий обозначим через $WIOE$), если не существует такого игрока $i \in N$ и другой ситуации $y \in X$, которая бы сильно *NE*-доминировала *x**, т. е. $\nexists i \in N, \nexists y \in X : y \overset{NE(i)}{\succ\!\!\succ} x$.

Применение слабых индивидуально-оптимальных равновесий мотивируется следующим сценарием одноцелевой игры. Игроки договариваются о необязательном соглашении придерживаться ситуации $x^* = (x_i^*)_{i \in N}$. Далее они независимо один от другого принимают решение о выборе своих стратегий. В том, и только том случае, когда основой соглашения будет слабое индивидуально-оптимальное равновесие x^*, изменение любым игроком $i \in N$, согласованной с другими игроками, стратегии x_i^* на другую, всегда приведет к ситуации, которая не будет доминировать x^* хотя бы для одного игрока (в том числе и его самого). Поэтому цель игрока $i \in N$, которая состоит из его личных интересов и интересов других игроков, которые он учитывает, может быть не удовлетворенной и достигнутый в ходе предыдущих переговоров компромисс может быть разрушенным.

Параметризация множества индивидуально-оптимальных равновесий

Свойства и условия индивидуальной оптимальной рассмотрены в [2]. В этой работе мы рассмотрим один подход к поиску индивидуально-оптимальных равновесий.

Будем считать, что функции выигрыша игроков ограничены на множестве *X* ситуаций игры *G*. Обозначим через $S = \sum_{i \in N} \sup_{x \in X} u_i(x) < \infty$ - верхнюю границу суммы функций выигрыша игроков. Введем векторы параметров $\mu_i = (\mu_i^j)_{j \in N}$, $i \in N$. Обозначим множества этих параметров через $M_i^{\geq 0} = \left\{ \mu_i = (\mu_i^j)_{j \in N} \middle| \sum_{j \in N} \mu_i^j = 1; \mu_i^j \geq 0,\ j \in N \right\}, i \in N$. *Имеет место следующая теорема.*

Теорема 1. Если в игре *G* ситуация $x^* \in WIOE$, функции выигрыша игроков $u_i(x^*) > 0, i \in N$, то всегда существуют такие векторы параметров $\mu_i \in M_i^{\geq 0}$, $i \in N$, в частности с компонентами:

$$\overline{\mu}_i^j = u_j(x^*)/S + 1/n - \sum_{k \in N} u_k(x^*)/Sn,\ j \in N, i \in N, \quad (1)$$

что выполняются следующие неравенства:

$$\min_{j\in N}(u_j(x_i, x^*_{N\setminus i}) - S\mu_i^j) \le \min_{j\in N}(u_j(x^*) - S\mu_i^j),\ \forall x_i \in X_i, i \in N. \quad (2)$$

Любое решение x^* системы неравенств (2) при заданных $\mu_i \in M_i^{\ge 0}$, $i \in N$, является слабым индивидуально-оптимальным равновесием.

Доказательство. Пусть ситуация x^* удовлетворяет неравенствам (2). Отсюда следует, что для $\forall i \in N$ и $\forall x_i \in X_i$ имеют место неравенства: $\min_{j\in N}(u_j(x_i, x^*_{N\setminus i}) - S\mu_i^j) \le \min_{j\in N}(u_j(x^*) - S\mu_i^j) \le u_k(x^*) - S\mu_i^k$, $\forall k \in N$, поэтому для $\forall i \in N$, $\forall x_i \in X_i$ существует $j \in N$ такое, что $u_j(x_i, x^*_{N\setminus i}) \le u_j(x^*)$. Следовательно из определения *NE*-доминирования игрока $i \in N$ следует, что $\nexists x = (x_i, x^*_{N\setminus i}) \in X : x \overset{NE}{\succ\!\succ} x^*$. Отсюда получим, что $x^* \in WIOE$.

Возьмем векторы $\overline{\mu}_i = (\overline{\mu}_i^j)_{j\in N}$, $i \in N$, с компонентами (1). Несложно убедиться, что $\overline{\mu}_i \in M_i^{\ge 0}$ для $\forall i \in N$. Из того, что x^* - слабое индивидуально-оптимальное равновесие следует, что $\nexists x \in X : x \overset{NE}{\succ\!\succ} x^*$, т. е. для $\forall i \in N$, $\forall x_i \in X_i$, $\exists j \in N$, что $u_j(x_i, x^*_{N\setminus i}) \le u_j(x^*)$, а значит $u_j(x_i, x^*_{N\setminus i}) - S\overline{\mu}_i^j \le u_j(x^*) - S\overline{\mu}_i^j$. Поскольку для $\forall i \in N$ значения $u_j(x^*) - S\overline{\mu}_i^j = -\left(S - \sum_{k\in N} u_k(x^*)\right)\Big/ n$ для $\forall j \in N$, то для $\forall x_i \in X_i$ получим $\min_{j\in N}(u_j(x_i, x^*_{N\setminus i}) - S\overline{\mu}_i^j) \le -\left(S - \sum_{k\in N} u_k(x^*)\right)\Big/ n = \min_{j\in N}(u_j(x^*) - S\overline{\mu}_i^j)$. Поэтому ситуация x^* удовлетворяет неравенствам (2). Теорема доказана.

Следует отметить, что параметры $\mu_i \in M_i^{\ge 0}$ позволяют игроку $i \in N$ выразить свое предпочтение на множестве функций выигрыша $u_j(x), j \in N$, всех игроков. Так, например, если он считает, что для него более важны интересы игрока $j_1 \in N$, чем игрока $j_2 \in N$, то согласно теореме 1 ему следует выбрать $\mu_i^{j_1} > \mu_i^{j_2}$. Варьируя параметры $\mu_i \in M_i^{\ge 0}$, $i \in N$, можно находить те или другие индивидуально-оптимальные равновесия, решая неравенства (2). С другой стороны, каждое индивидуально-оптимальное равновесие характеризуется некоторым множеством параметров $\mu_i \in M_i^{\ge 0}$, $i \in N$, и, соответственно, предпочтением на множестве функций выигрыша всех игроков.

Следует также отметить, что параметрам μ_i^j, которые фигурируют в неравенствах (2), можно придать определенное игровое содержание. Пусть x^* - слабое индивидуально-оптимальное равновесие игры *G*. Тогда согласно теореме 1 оно будет решением системы неравенств (2), по крайней мере при $\overline{\mu}_i^j = u_j(x^*)/S + 1/n - \sum_{k\in N} u_k(x^*)/Sn$, $j \in N, i \in N$. Отсюда легко увидеть, что для каждого фиксированного игрока $i \in N$ значения параметра μ_i^j указывает каким, по его мнению, желательно, что бы был выигрыш игрока $j \in N$ в достигнутом компромиссе x^*. Тогда и только тогда игроку $i \in N$ будет невыгодно отклоняться от x^*. Поскольку величину $S = \sum_{i\in N} \sup_{x\in X} u_i(x)$ можно интерпретировать как суммарный идеальный выигрыш всех игроков, то можно также сказать, что вектор параметров μ_i, который отвечает слабому индивидуально-оптимальному равновесию x^*, является желаемым дележом суммарного идеального выигрыша *S* между игроками, с точки зрения игрока *i*. При таком дележе игроку *i* будет не

выгодно разрушать достигнутый компромисс изменением своей стратегии.

Неравенства (2) могут быть значительно упрощены, если функции выигрыша игроков будут вогнутыми, а множества стратегий игроков – выпуклыми.

Теорема 2. Пусть множества стратегий игроков $X_i, i \in N$, - выпуклы, а функции их выигрыша $u_i, i \in N$, - вогнуты по стратегиям каждого игрока в отдельности. Ситуация *x** будет слабым индивидуально-оптимальным равновесием тогда и только тогда, когда существуют векторы параметров $\mu_i \in M_i, i \in N$, такие, что ситуация *x** будет удовлетворять следующей системе неравенств:

$$\sum_{j \in N} \mu_i^j (u_j(x^*) - u_j(x_i, x^*_{N\setminus i})) \geq 0 \ \ \forall x_i \in X_i, i \in N. \qquad (3)$$

Доказательство. Пусть выполняются неравенства (3). Допустим от противного, что $x^* \notin WIOE$. Тогда $\exists i \in N$, $\exists x \in X : x \overset{NE(i)}{\succ\succ} x^*$. Отсюда следует, $\exists i \in N : u_j(x_i, x^*_{N\setminus i}) > u_j(x^*), \forall j \in N$. Взяв взвешенную сумму этих неравенств по векторам $\mu_i \in M_i$, получим противоречие, а именно $\sum_{j \in N} \mu_i^j (u_j(x^*) - u_j(x_i, x^*_{N\setminus i})) < 0$.

Пусть $x^* \in WIOE$. Тогда для $\forall i \in N$ система неравенств $u_j(x_i, x^*_{N\setminus i}) > u_j(x^*), j \in N$ будет несовместной на множестве X_i. Построим образы множества допустимых решений этой системы неравенств и множества стратегий X_i при отображении $y_i^j = u_j(x_i, x^*_{N\setminus i})$. Получим соответственно: $V_i = \left\{ (v_i^j)_{j \in N} \middle| v_i^j = u_j(x_i, x^*_{N\setminus i}) > u_j(x^*), j \in N \right\}$, $Y_i = \left\{ (y_i^j)_{j \in N} \middle| y_i^j = u_j(x_i, x^*_{N\setminus i}), x_i \in X_i, j \in N \right\}$.

Очевидно, $V_i \cap Y_i = \varnothing$. Обозначим через $\overline{V}_i = \left\{ (v_i^j)_{j \in N} \middle| v_i^j = u_j(x_i, x^*_{N\setminus i}) \geq u_j(x^*), j \in N \right\}$ - замыкание множества V_i. Очевидно, множество $\overline{V}_i$ - выпукло. Покажем выпуклость множества Y_i. Действительно, пусть $\bar{y}_i = (u_j(\bar{x}_i, x^*_{N\setminus i}))_{j \in N}$, $\bar{\bar{y}}_i = (u_j(\bar{\bar{x}}_i, x^*_{N\setminus i}))_{j \in N} \in Y_i$. Рассмотрим стратегию $\tilde{x}_i = \lambda \bar{x}_i + (1-\lambda)\bar{\bar{x}}_i, \lambda \in [0,1]$. Тогда, в силу вогнутости функций выигрыша по стратегиям игроков, получим: $u_j(\tilde{x}_i, x^*_{N\setminus i}) = u_j(\lambda \bar{x}_i + (1-\lambda)\bar{\bar{x}}_i, x^*_{N\setminus i}) \geq \lambda \bar{y}_i + (1-\lambda)\bar{\bar{y}}_i \geq u_j(x^*), \forall j \in N$. Поэтому $\lambda \bar{y}_i + (1-\lambda)\bar{\bar{y}}_i \in \overline{Y}_i$ и множество Y_i является выпуклым.

Таким образом, имеем выпуклые множества $\overline{V}_i$ и Y_i, и пустое пересечение множества Y_i с внутренностью V_i множества $\overline{V}_i$. Отсюда, согласно теореме про отделяющую гиперплоскость, всегда существует такой вектор $\mu_i = (\mu_i^j)_{j \in N} \neq 0$, что $\sum_{j \in N} \mu_i^j (v_i - y_i) \geq 0, \forall v_i \in V_i, \forall y_i \in Y_i$. Это неравенство будет верным, в частности, и для $v_i = (u_j(x^*))_{j \in N}$. Поэтому, поскольку $y_i = (u_j(x_i, x^*_{N\setminus i})_{j \in N}$, то получим (1). Теорема доказана.

Параметрам μ_i^j можно придать следующую интерпретацию. Величина μ_i^j характеризует относительный вес выигрыша игрока $j \in N$ в предпочтении игрока $i \in N$ на множестве функций выигрыша всех игроков необходимый и достаточный для того, чтобы игроку *i* было не выгодно отклоняться от достигнутого компромисса. Так, например, если он считает, что для него более важные интересы игрока $j_1 \in N$, чем игрока $j_2 \in N$, то согласно теореме 2 ему следует выбрать $\mu_i^{j1} > \mu_i^{j2}$. Выбирая различные параметры $\mu_i \in M_i^{\geq 0}, i \in N$, можно находить разные индивидуально-оптимальные равновесия как решения

неравенств (3). С другой стороны, каждое индивидуально-оптимальное равновесие характеризуется некоторым множеством векторов параметров $\mu_i \in M_i^{\geq 0}$, $i \in N$, и, соответственно, предпочтением на множестве функций выигрыша всех игроков.

Если подвести итог, то можно заметить, что для нахождения любого индивидуально-оптимального равновесия игрокам достаточно выбрать соответствующие векторы параметров $\mu_i \in M_i, i \in N$, и решить систему неравенств:

$$F_i(U(x_i, x^*_{N\setminus i}), \mu_i) \leq F_i(U(x^*), \mu_i) \ \forall x_i \in X_i, i \in N, \quad (4)$$

где функция $F_i(U(x), \mu_i) = \min\limits_{j \in N}(u_j(x) - S\mu_i^j)$ - в общем случае и $F_i(U(x), \mu_i) = \sum\limits_{j \in N} \mu_i^j u_j(x)$ - в случае выпуклой игры. Если игра происходит в условиях полной информированности (каждый игрок $i \in N$ знает функции выигрыша u_j, $j \in N \setminus \{i\}$, и предпочтения других игроков, которые характеризуются векторами параметров μ_j, $j \in N \setminus \{i\}$) для решения этой задачи можно применять те, или иные методы оптимизации для подобного класса задач.

Поиск индивидуально-оптимальных равновесий

Особенный интерес представляет случай частичной информированности игроков. Пусть каждый игрок $i \in N$ знает функции выигрыша u_j, $j \in N \setminus \{i\}$, всех других игроков, но векторы параметров μ_j, $j \in N \setminus \{i\}$, ему неизвестны. Такая информированность - естественна, поскольку предпочтение каждого игрока на множестве функций выигрыша других игроков, как правило, представляет собой конфиденциальную информацию. Кроме этого, это предпочтение может не всегда полностью осознаваться игроком и изменяться (уточняться) в процессе принятия решения.

Для реализации процедуры поиска индивидуально-оптимального равновесия игры в условиях частичной информированности игроков необходимо решать систему неравенств (4) так, чтобы каждый игрок оперировал лишь той информацией, которая ему известна.

Наиболее универсальной схемой, которую можно применить в условиях частичной информированности игроков есть, так называемая, процедура “нащупывания Курно” [1].

Пусть $x^0 = (x_i^0)_{i \in N}$ некоторая начальная ситуация игры. Для фиксированного вектора параметров $\mu = (\mu_i)_{i \in N}$ обозначим через $BR_i(\mu_i) = \left\{ x \in X \middle| F_i(U(y_i, x_{N\setminus i}), \mu_i) \leq F_i(U(x), \mu_i), \forall y_i \in X_i; x_{N\setminus i} \in X_{N\setminus i} \right\}$ - множество наилучших ответов игрока $i \in N$ на фиксированные наборы стратегий других игроков.

Процедура нащупывания Курно строит последовательность $x^1, \ldots, x^t, \ldots$ ситуаций игры G такую, что $x_i^t \in BR_i^{(t-1)}(\mu_i)$, t=1,2…, где $BR_i^{(t-1)}(\mu_i) = \left\{ x \in X \middle| F_i(U(x_i, x_{N\setminus i}^{t-1}), \mu_i) \leq F_i(U(x^{t-1}), \mu_i), \forall x_i \in X_i \right\}$ - множество наилучших ответов игрока $i \in N$ на набор $x_{N\setminus i}^{t-1} = (x_j^{t-1})_{j \in N\setminus\{i\}}$ фиксированных стратегий других игроков, которые получены на предыдущем шаге процедуры.

Таким образом, каждый игрок $i \in N$ оперирует лишь с известной ему информацией (функции выигрыша всех игроков и вектор параметров μ_i, который характеризует его собственное предпочтение на множестве функций выигрыша других игроков).

Если процедура нащупывания Курно сходится, то мы получим некоторое индивидуально-оптимальное равновесие x^*, которое отвечает набору предпочтений всех игроков. Такая процедура может и не сходиться. Недостатком процедуры нащупывания Курно является также низкая скорость сходимости, которая в общем случае не поддается оцениванию.

В этой работе мы рассмотрим другую процедуру поиска индивидуально-оптимальных равновесий, которая основана на распределенных методах решения оптимизационных задач [3].

Общая схема распределенного решения оптимизационных задач основана на определении, так называемой, функции рассогласования [3]. Эта функция может определяться разными способами, но должна быть количественной оценкой, характеризующей несогласованность решений, которые выбраны отдельными игроками, по их принадлежности к решению всей задачи в целом. Вообще, определение функции рассогласования близко к функции штрафа и характеризует величину агрегированной разницы значений связующих переменных, которые получены в локальных решениях взаимосвязанных подзадач. Общая идея построения распределенного решения систем взаимосвязанных задач заключается в пошаговом согласовании их решений с целью обеспечить получение следующего приближения к решению задачи, с меньшей величиной функции рассогласования, чем на предыдущем шаге. Сходимость таких процедур обеспечивается коррекцией моделей задач на каждом шаге.

Перепишем задачу (4) в эквивалентном виде распределив переменные по каждому игроку в отдельности: $x^i = \frac{1}{n}\sum_{j\in N} x^j, x^i \in BR_i(\mu_i), i \in N$. Здесь каждому игроку $i \in N$ отвечают векторы распределенных переменных $x^i = (x^i_j)_{j\in N}$. Вполне понятно, что $x^* = x^i, i \in N$ тогда и только тогда, когда x^* является решением задачи (4). Построим вспомогательную задачу с квадратичной целевой функцией рассогласования, решение которой будет также совпадать с решением (4):

$$g(x) = \sum_{i\in N}\left\| x^i - \frac{1}{n}\sum_{j\in N} x^j \right\|^2 \to \min,\ x^i \in BR_i(\mu_i), i \in N.$$

Для решения этой вспомогательной задачи используем итерационный метод спуска, в котором будем выбирать допустимые направления спуска, с использованием линейной аппроксимации целевой функции по векторам переменных $x^i, i \in N$.

Пусть начальное приближение $x^{i(0)} \in BR_i(\mu_i), i \in N$. Тогда для *k*-го приближения $x^{i(k)}$, $i \in N$, получим: $g^{(k)} = g(x^{(k)}) + 2\sum_{i\in N}\left\langle x^{i(k)} - \frac{1}{n}\sum_{j\in N} x^{j(k)}, x^i - x^{i(k)}\right\rangle$ $i = 1,2,\ldots$. Отбросив константы и постоянные множители, будем искать направление спуска на (*k*+1)-м шаге, путем решения следующих задач: $\sum_{i\in N}\left\langle x^{i(k)} - \frac{1}{n}\sum_{j\in N} x^{j(k)}, x^i\right\rangle \to \min,\ x^i \in BR_i(\mu_i), i \in N$. Эти задачи, в свою очередь, декомпозируются на *n* независимых подзадач:

$$\left\langle x^{i(k)} - \frac{1}{n}\sum_{j\in N} x^{j(k)}, x^i\right\rangle \to \min \quad x^i \in BR_i(\mu_i) \tag{5}$$

(обозначим векторы их решений через $\bar{x}^{i(k+1)}$, $i \in N$).

Следующее (*k*+1)-е приближение определяется из условий уменьшения функции вдоль допустимого направления следующим образом:

$$x^{i(k+1)} = x^{i(k)} + \lambda^{(k)}(\bar{x}^{i(k+1)} - x^{i(k)}),\ i \in N, \tag{6}$$

где $\lambda^{(k)}$ -находиться из условий наибольшего уменьшения значения функции рассогласования:

$$\lambda^{(k)} = \arg\min_{\lambda\in[0,1]} g(x^{(k)} - \lambda(x^{(k)} - \bar{x}^{(k+1)})) = \sum_{i\in N}\left\langle x^{i(k)} - \frac{1}{n}\sum_{j\in N} x^{j(k)}, x^{i(k)} - \bar{x}^{i(k+1)}\right\rangle \Big/ \sum_{i\in N}\left\|(x^{i(k)} - \bar{x}^{i(k+1)})\right\|^2 \tag{7}$$

Таким образом, в соответствии с заданными предпочтениями на множестве функций выигрыша, игроки предлагают один другому к рассмотрению ситуации игры. Эти ситуации определяются ими по формулам (7) благодаря решению каждым игроком отдельно оптимизационной задачи (5). Эти задачи описываются на основе только той информации, которой владеет соответствующий игрок, т.е. информацией о функции его выигрыша и ситуации игры, которая наблюдается всеми игроками вместе.

Обоснуем сходимость рассмотренной выше процедуры и оценим скорость ее сходимости.

Для доказательства теоремы будем использовать общеизвестное свойство выпуклых дифференцированных функций на выпуклых множествах, приведенное в следующей лемме.

Лемма. Пусть $\varphi(x)$ - дифференцированная функция, которая определена на некотором множестве X эвклидового пространства E^n, градиент которой удовлетворяет условию Липшица, а именно $\exists L > 0$, для которого $\|\nabla\varphi(x) - \nabla\varphi(y)\| \le L\|x - y\|$ для $\forall x, y \in X$. Тогда для $\forall x, y \in X$ имеет место неравенство $\varphi(x) - \varphi(y) \ge \langle\nabla\varphi(x), x - y\rangle - \frac{L}{2}\|x - y\|^2$.

Докажем следующую теорему.

Теорема 3. Пусть для фиксированного вектора параметров $\mu = (\mu_i)_{i\in N}$ множества $BR_i(\mu_i)$ наилучших ответов игроков $i \in N$ на фиксированные наборы стратегий других игроков - выпуклы и замкнуты, а их пересечение $X^*(\mu) = \bigcap_{i\in N} BR_i(\mu_i)$ - непусто. Тогда процедура (5)-(9) генерирует последовательности приближений $x^{i(k)} \in BR_i(\mu_i), i \in N;\ k = 1,2,\ldots$, которые сходятся к некоторому индивидуально-оптимальному равновесию $x^* \in X^*(\mu)$ со скоростью геометрической прогрессии с оценкой $O\left(\frac{n}{k}\right)$.

Доказательство. Сначала определим константу Липшица для функции $\varphi(x_1, x_2) = \|x_1 - x_2\|^2$. Поскольку $\nabla\varphi = (2(x_1 - x_2), -2(x_1 - x_2))$, то $\|\nabla\varphi(x) - \nabla\varphi(y)\| =$
$= \|(2(x_1 - x_2), -2(x_1 - x_2)) - (2(y_1 - y_2), -2(y_1 - y_2))\| = \|2(x_1 - x_2) - 2(y_1 - y_2), -2(x_1 - x_2) + 2(y_1 - y_2))\| =$
$= 2\|(x_1 - x_2) - (y_1 - y_2)\| = 2\|\varphi(x) - \varphi(y)\|$. Таким образом, константа Липшица $L = 2$.

Для некоторого $k = 0,1,\ldots$ рассмотрим разность

$$\Delta^{(k)} = g^{(k)} - g^{(k+1)}, \tag{8}$$

которая согласно лемме при $L = 2$ может быть оценена следующим образом:

$$\Delta^{(k)} \ge \left\langle\nabla g(x), x^{(k)} - x^{(k+1)}\right\rangle - \left\|x^{(k)} - x^{(k+1)}\right\|^2 = \sum_{i\in N}[2\left\langle x^{i(k)} - \frac{1}{n}\sum_{j\in N} x^{j(k)}, x^{i(k)} - x^{i(k+1)}\right\rangle - \left\|(x^{i(k)} - x^{i(k+1)})\right\|^2].$$

Пусть $x^{i(k+1)} = x^{i(k)} - \lambda(x^{i(k)} - \bar{x}^{i(k+1)}),\ i \in N$. Тогда для $\forall\lambda \in [0,1]$ получим:

$$\Delta^k \geq 2\lambda \sum_{i\in N}\left\langle x^{i(k)} - \frac{1}{n}\sum_{j\in N} x^{j(k)}, x^{i(k)} - \bar{x}^{i(k+1)}\right\rangle - \lambda^2 \sum_{i\in N}\left\|(x^{i(k)} - \bar{x}^{i(k+1)})\right\|^2 . \quad (9)$$

Выберем $\lambda = \lambda^{(k)}$ по формуле (7). Тогда в случае, когда $\sum_{i\in N}\left\langle x^{i(k)} - \frac{1}{n}\sum_{j\in N} x^{j(k)}, x^{i(k)} - \bar{x}^{i(k+1)}\right\rangle \geq \sum_{i\in N}\left\|(x^{i(k)} - \bar{x}^{i(k+1)})\right\|^2$. Значение $\lambda^{(k)} = 1$ и с (9) мы получим:

$$\Delta^{(k)} \geq 2\sum_{i\in N}\left\langle x^{i(k)} - \frac{1}{n}\sum_{j\in N} x^{j(k)}, x^{i(k)} - \bar{x}^{i(k+1)}\right\rangle - \sum_{i\in N}\left\|(x^{i(k)} - \bar{x}^{i(k+1)})\right\|^2 \geq \sum_{i\in N}\left\langle x^{i(k)} - \frac{1}{n}\sum_{j\in N} x^{j(k)}, x^{i(k)} - \bar{x}^{i(k+1)}\right\rangle .$$

Поскольку по предположению теоремы $\|x - y\| \leq R < \infty$ для $\forall x, y \in BR_i(\mu_i), i \in N$, то величина $\left\langle x^{i(k)} - \frac{1}{n}\sum_{j\in N} x^{j(k)}, x^{i(k)} - \bar{x}^{i(k+1)}\right\rangle \leq \left\|x^{i(k)} - \frac{1}{n}\sum_{j\in N} x^{j(k)}\right\| \cdot \left\|x^{i(k)} - \bar{x}^{i(k+1)}\right\| \leq R^2$. Не трудно убедиться, что отсюда мы получим неравенство $\Delta^{(k)} \geq \frac{1}{R^2}\sum_{i\in N}\left\langle x^{i(k)} - \frac{1}{n}\sum_{j\in N} x^{j(k)}, x^{i(k)} - \bar{x}^{i(k+1)}\right\rangle^2$. В случае, когда $\sum_{i\in N}\left\langle x^{i(k)} - \frac{1}{n}\sum_{j\in N} x^{j(k)}, x^{i(k)} - \bar{x}^{i(k+1)}\right\rangle < \sum_{i\in N}\left\|(x^{i(k)} - \bar{x}^{i(k+1)})\right\|^2$, значение $\lambda^{(k)} < 1$ и из (9) мы получим неравенство $\Delta^{(k)} \geq \sum_{i\in N}\left\langle x^{i(k)} - \frac{1}{n}\sum_{j\in N} x^{j(k)}, x^{i(k)} - \bar{x}^{i(k+1)}\right\rangle \Big/ \sum_{i\in N}\left\|(x^{i(k)} - \bar{x}^{i(k+1)})\right\|^2$. Отсюда, поскольку по предположению теоремы $\|x - y\| \leq R < \infty$ для $\forall x, y \in BR_i(\mu_i), i \in N$, получим неравенство

$$\Delta^{(k)} \geq \frac{1}{R^2}\sum_{i\in N}\left\langle x^{i(k)} - \frac{1}{n}\sum_{j\in N} x^{j(k)}, x^{i(k)} - \bar{x}^{i(k+1)}\right\rangle^2 . \quad (10)$$

Таким образом, в обоих случаях, когда $\lambda^{(k)} = 1$ или $\lambda^{(k)} < 1$, для того, чтобы показать, что $\Delta^{(k)} > 0$ достаточно доказать справедливость неравенства

$$\delta^{(k)} = \sum_{i\in N}\left\langle x^{i(k)} - \frac{1}{n}\sum_{j\in N} x^{j(k)}, x^{i(k)} - \bar{x}^{i(k+1)}\right\rangle > 0 . \quad (11)$$

Действительно, из (5) получим: $\delta^{(k)} = \sum_{i\in N} \max\left\{\left\langle x^{i(k)} - \frac{1}{n}\sum_{j\in N} x^{j(k)}, x^{i(k)} - x^i\right\rangle \middle| x^i \in BR_i(\mu_i)\right\} \geq$

$$\geq \sum_{i\in N}\left\langle x^{i(k)} - \frac{1}{n}\sum_{j\in N} x^{j(k)}, x^{i(k)} - x^i\right\rangle, \forall x^i \in BR_i(\mu_i) . \quad (12)$$

В частности, неравенство (12) будет выполняться и для $x^* = x^i, i \in N$, где x^* - решение задачи (4). Тогда

$$\delta^{(k)} \geq \sum_{i\in N}\left\langle x^{i(k)} - \frac{1}{n}\sum_{j\in N} x^{j(k)}, x^{i(k)} - x^*\right\rangle = \sum_{i\in N}\left\langle x^{i(k)} - \frac{1}{n}\sum_{j\in N} x^{j(k)}, x^{i(k)}\right\rangle =$$

$$= \sum_{i\in N}\left\langle x^{i(k)} - \frac{1}{n}\sum_{j\in N} x^{j(k)}, x^{i(k)} - \frac{1}{n}\sum_{j\in N} x^{j(k)}\right\rangle = \sum_{i\in N}\left\|x^{i(k)} - \frac{1}{n}\sum_{j\in N} x^{j(k)}\right\|^2 \geq 0 . \quad (13)$$

Вполне понятно, что $\delta^{(k)} = 0$ тогда и только тогда, когда $x^{i(k)} = x^*, i \in N$, где x^* - решение задачи (4), т.е. $x^* \in X^*(\mu) = \bigcap_{j\in N} BR_j(\mu_j)$. Таким образом, если $x^{i(k)} \neq x^*$ хотя бы для одного $i \in N$, то неравенство (11) будет строгим. Тогда $\Delta^{(k)} > 0$, $k = 1,2,...$ и мы получим строго монотонно убывающую последовательность

$\{\Delta^{(k)}\}_{k=0,1,...}$, которая ограничена снизу. Поэтому $\Delta^* = \lim_{k\to\infty} \Delta^{(k)} = 0$, а $x^* = \lim_{k\to\infty} x^{i(k)}$, $i \in N$.

Построим оценку скорости сходимости процедуры. Из (10) посредством неравенства Минковского $(\sum_{i\in N} a_i)^2 \le n\sum_{i\in N} a_i^2$ справедливо неравенство $\Delta^{(k)} \ge \frac{1}{nR^2}(\sum_{i\in N}\left\langle x^{i(k)} - \frac{1}{n}\sum_{j\in N} x^{j(k)}, x^{i(k)} - \bar{x}^{i(k+1)}\right\rangle)^2$. Отсюда и из (11), (13) получим неравенство $\Delta^{(k)} \ge \frac{1}{R^2 n}\left(\sum_{i\in N}\left\| x^{i(k)} - \frac{1}{n}\sum_{j\in N} x^{j(k)}\right\|^2\right)^2$. Тогда согласно обозначению (8) $g^{(k)} - g^{(k+1)} \ge \frac{1}{nR^2}[g^{(k)}]^2$, $k = 0,1,...$. Поскольку $g^{(k)} > 0$, то $\frac{g^{(k)}}{g^{(k+1)}} \ge 1 + \frac{1}{nR^2}\frac{[g^{(k)}]^2}{g^{(k+1)}} \ge 1$. Тогда $\frac{1}{g^{(k+1)}} - \frac{1}{g^{(k)}} = \frac{g^{(k)} - g^{(k+1)}}{g^{(k)} g^{(k+1)}} \ge \frac{1}{nR^2}\frac{g^{(k)}}{g^{(k+1)}} \ge \frac{1}{nR^2}$. Если рассмотреть любую конечную сумму последних неравенств, получим: $\sum_{k=0}^{k_0}\left(\frac{1}{g^{(k+1)}} - \frac{1}{g^{(k)}}\right) = \left(\frac{1}{g^{(k_0)}} - \frac{1}{g^{(0)}}\right) \ge \sum_{k}^{k_0}\frac{1}{nR^2} = \frac{k_0}{nR^2}$. Тогда $g^{(k_0)} = \frac{nR^2 g^{(0)}}{nR^2 + g^{(0)}k_0} = O\left(\frac{n}{k_0}\right)$. Теорема доказана.

Заключение

Принцип индивидуальной оптимальности обобщает классические принципы оптимальности в некооперативных играх и расширяет класс конфликтно разрешимых игр. Применение этого принципа обосновано, если конфликт между игроками не может быть разрешен согласно классическим принципам оптимальности и игроки соглашаются идти на компромисс ради его достижения.

Благодарности

Работа опубликована при финансовой поддержке проекта **ITHEA XXI** Института информационных теорий и приложений FOI ITHEA Болгария www.ithea.org и Ассоциации создателей и пользователей интеллектуальных систем ADUIS Украина www.aduis.com.ua.

Литература

[1] Мулен Э. Теория игр с примерами из математической экономики. – М.: Мир, 1985. -200 с.

[2] Мащенко С.О. Исследование стабильности равновесий на основе принципа индивидуальной оптимальности// Кибернетика и системный анализ. -2007.-4.–С.162-169.

[3] Волкович В.Л., Коленов Г.В., Мащенко С.О. Алгоритмы поиска допустимого решения в линейных распределенных системах// Автоматика. -1988.-4.–С. 70-77.

Информация об авторе

Мащенко Сергей Олегович *– Киевский национальный университет имени Тараса Шевченко, доцент, кандидат физ.-мат. наук; Проспект академика Глушкова, 6, Киев – 207, Украина;*
e-mail: msomail@yandex.ru

НЕЧЕТКИЙ АЛГОРИТМ ПОСЛЕДОВАТЕЛЬНОГО АНАЛИЗА ВАРИАНТОВ

Алексей Волошин, Николай Маляр, Оксана Швалагин

Аннотация: *Предлагается модификация процедур отсева декомпозиционного метода последовательного анализа и отсеивания вариантов при нечетко заданных допустимых множествах альтернатив дискретных моделей принятия решений.*

Ключевые слова: *дискретное программирование, последовательный анализ вариантов, нечеткость.*

ACM Classification Keywords: *H.4.2 Information Systems Applications: Types of Systems: Decision Support.*

Conference*: The paper is selected from XV[th] International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.*

Введение

Одним из наиболее эффективных методов решения оптимизационных задач различных классов, возникающих при моделировании процессов принятия решений, является метод последовательного анализа вариантов (ПАВ), предложенный в 60-х г. XXст. в Институте кибернетики АН Украины [Михалевич, Шор, 1961]. С точки зрения формальной логики схема ПАВ является повторением таких действий:

- разбиение множества вариантов решений задачи на подмножества с дополнительными свойствами;
- использование этих свойств при поиске логических противоречий в описании отдельных множеств;
- исключение из рассмотрения подмножеств, в описании которых имеются противоречия.

В [Михалевич, Шор, 1977] общая схема ПАВ конкретизируется для различных классов многовариантных задач, описываются алгоритмы "последовательного конструирования, анализа и отсеивания вариантов", в частности, широко известный алгоритм "Киевский веник". Основным правилом отсева бесперспективных вариантов в этих алгоритмах являлся принцип монотонной рекурсивности [Михалевич, Шор, 1961], родственной критерию оптимальности динамического программирования [Беллман, 1960]. Наряду с известными достоинствами алгоритмы пошагового конструирования обладают и определенными недостатками [Волкович, Волошин, 1984]. В развитие общих концепций ПАВ в работах [Волкович, Волошин, 1978-2009] предложены процедуры параллельного отсева подвариантов, в частности, единичной длины (алгоритм W [Волкович-Волошин,1978]). При этом возникает проблема конструирования полного варианта, которая решается путем последовательного введения ограничений на значение функционала [Волошин, 1987]. Предложенную схему можно интерпретировать как декомпозиционный метод решения задачи, в которой анализ подвариантов проводится независимо (параллельно) друг от друга, конструирование полных вариантов осуществляется на верхнем уровне при анализе агрегирующей (координирующей) задачи [Лэсдон, 1975]. Эффективность алгоритмов, базирующихся на предложенных принципах, подтверждена вычислительными экспериментами, теоретическими исследованиями и решением практических задач [Волкович, Волошин, 1987, 1993, 2004, 2009]. При решении прикладных задач исходные данные задаются, как правило, приближенно, неточно, нечетко. В [Волошин, 1980] предложен алгоритм ПАВ для нахождения ''субоптимальных" решений (ε – приближенных по функционалу и δ – приближенных по ограничениям). При этом любые ''ε, δ – отклонения'' считались

''равноприемлемыми''. В данной работе процедуры отсева в приближенных алгоритмах ПАВ обобщаются на нечеткий случай [Орловский,1981]. Предполагается, что в оптимизационной задаче математического программирования при четко заданных множествах возможных решений и значений функционалов нечетко задаются допустимых множества решений и множества уровней целевой функции.

Постановка задачи

Пусть задано некоторое конечное множество $X = \{\mathbf{x} = (x_1,...,x_j,...,x_n)\}$ возможных вариантов задачи. Множество возможных вариантов построения j-ой компонента x_j вектора $\mathbf{x}$ обозначим X_j, $j = \overline{1,n}$. Множества X_j могут задаваться множеством значений функции дискретного аргумента; задаваться таблично; быть множеством перестановок некоторого базисного множества параметров. Вектор $p = (x_{j1}, x_{j2},..., x_{jk})$, где $x_{j1} \in X_{j1}$, $x_{j2} \in X_{j2}$,..., $x_{jk} \in X_{jk}$ $(1 \le j_1 < \cdots < j_k \le n)$, назовется подвариантом длины k. Множество всех подвариантов обозначим через $P(X)$, в котором выделим множество допустимых подвариантов $D(X) \subseteq P(X)$. В множестве $D(X)$ выделяется подмножество полных допустимых вариантов (вариантов длинны n) $D \subseteq D(X)$, для которого справедливо соотношение $D \subseteq X$.

Пусть $F(p)$ – функционал, определенный на множестве $P(X)$.

Рассматривается следующая задача нахождения варианта $\mathbf{x}$:

$$\mathbf{x} \in \underset{x \in D}{\operatorname{Argmax}} F(x);$$

$$\underset{x \in D}{\operatorname{Argmax}} F(x) = \{\mathbf{x} \in D \mid F(x) = \max_{x' \in D} F(x')\}.$$

Вариант $\mathbf{x}^* \in D$, который максимизирует функционал $F(x)$, назовем оптимальным. Вариант $\mathbf{x}$, при котором достигается максимум функционала $F(\mathbf{x})$ на некотором множестве $X^{(s_0)} \subseteq X$, назовем максимальным. Полученные максимальные варианты исследуются с помощью критериев оптимальности. Множество допустимых подвариантов $D(X)$ опишем следующим образом:

$$D(X) = \left\{ p \mid G_i(p) \overset{\alpha}{\prec} G_i,\ p \in P(X),\ 0 \le \alpha \le 1 \right\},$$

где $G_i(p) = \{(g_i(p), \mu_{g_i}(g_i(p))),\ i = \overline{1,m}\}$, – функционал подварианта p, определенный на множестве $P(X)$, $G_i = \{(g_i, \mu_i(g_i)),\ g_i \in [\underline{g}_i, \overline{g}_i]\ i = \overline{1,m}\}$ – заданные нечеткие множества, $\overset{\alpha}{\succ}$ – нечеткое отношение предпочтения, α – степень предпочтения [Орловский, 1981].

Очевидно, что множество полных допустимых вариантов определяется как:

$$D = \left\{ \mathbf{x} \mid G_i(\mathbf{x}) \overset{\alpha}{\prec} G_i, \mathbf{x} \in X,\ 0 \le \alpha \le 1 \right\}.$$

Родовым множеством подварианта $p \in D(X)$ называется множество $R(p)$, состоящее из тех вариантов $\mathbf{x}$, которые содержат подвариант p. В свою очередь, допустимым родовым множеством $\overline{R}(p)$ подварианта p называется такое подмножество родового множества $R(p)$, которое состоит из допустимых вариантов $\mathbf{x} \in D$. Родовым множеством $R(p)$ некоторого множества подвариантов

$P \subseteq P(X)$ называется множество, которое является объединением родовых множеств подварианта $p \in P$, то есть $R(p) = \bigcup_{p \in P} R(p)$.

Величина $\Delta_p^i = \left\{ \left(\Delta g_p^i, \mu\left(\Delta g_p^i \right) \right), \Delta g_p^i \in \left[\underline{\Delta g}_p^i, \overline{\Delta g}_p^i \right] \right\}$ называется допуском для подварианта p относительно множества D по i-ому ограничению, если из того, что $G_i(p) \overset{\alpha}{\succ} \Delta_p^i$, следует, что допустимое родовое подмножество $\overline{R}(p)$ подварианта p пусто.

В соответствии с методологией ПАВ необходимо построить процедуру анализа родовых множеств подвариантов и исключения тех из них, допустимые родовые множества которых пусты или не содержат оптимальных вариантов. Эта задача сводится к задаче вычисления допусков и проверки множества подвариантов на удовлетворение этим допускам.

Обозначим через $Q(P) = \left\{ \Delta_p^i \right\}$, $i = \overline{1,m}$, множество допусков относительно множества D для множества подвариантов P по всем ограничениям.

На множестве P зададим оператор $E = E(P,Q)$, отсеивающий те подварианты, которые не удовлетворяют множеству допусков Q. Применив оператор E на множестве P, получим множество подвариантов $P' \subseteq P$, такое что

$$\overline{R}(P) = \overline{R}(P'). \tag{1}$$

Тем самым, из рассмотрения выключаются подварианты $p \in \widetilde{P} = P \setminus P'$, для которых $\overline{R}(\widetilde{P}) = \varnothing$.

Обозначим $\wp = \cup P$ конечное семейство множеств подвариантов $P \subseteq P(X)$, для которого справедливо $R(\wp) = X$, $\overline{R}(\wp) = D$. Пусть $Q(\wp) = \cup Q(P)$ - система допусков семейства $\wp$.

Определим на множестве $\wp$ оператор Ω аппроксимации множества $D \subseteq X$, который вычисляет систему допусков $Q(\wp)$ и сужает множество X путем применения оператора E на множествах $R \subseteq \wp$.

Описание алгоритма

Опишем схему работы оператора Ω.

Шаг 0. Вычисляем начальную систему допусков $Q\left(\wp^{(0)}\right)$ на множестве $\wp^0 = \wp$.

Шаг s. На подмножествах $P^{(s-1)} \subset \wp^{(s-1)}$ для заданной системы допусков $Q\left(\wp^{(s-1)}\right)$ применяем оператор отсева E. В результате получаем новое семейство множеств подвариантов $\wp^{(s)} = \cup P^{(s)}$, такое, что

$$\wp^{(s)} \subseteq \wp^{(s-1)} \tag{2}$$

Учитывая, что $R\left(\wp^{(s-1)}\right) = X^{(s-1)}$ и справедливо (2), получим, что справедливо следующее условие:

$$X^{(s)} \subseteq X^{(s-1)}..$$

Поэтому, учитывая (1), получим

$$\overline{R}\left(\wp^{(s-1)}\right) = \overline{R}\left(\wp^{(s)}\right) = D.$$

Далее пересчитываем систему допуском для множества $\wp^{(s)}$.

На шаге s оператора Ω возможны следующие случаи:

a. Справедливо условие предпочтения $Q\left(\wp^{(s)}\right) \succ Q\left(\wp^{(s-1)}\right)$, то есть дальнейшее применение оператора E на множествах подвариантов $P^{(s)}$ при допусках $Q\left(\wp^{(s)}\right)$ может привести к сужению множества $P^{(s)}$. Переходим к шагу s+1.

b. Справедливо условие равности $Q\left(\wp^{(s)}\right) = Q\left(\wp^{(s-1)}\right)$, то есть $P^{(s)} = P^{(s-1)}$. Отсюда следует, что $X^{(s)} = X^{(s-1)}$. Эти условия свидетельствуют об окончании работы оператора Ω.

Аналогично [Волкович, Волошин, 1984] можно показать, что оператор Ω оканчивает работу за конечное число шагов.

В случае, когда оператор Ω недостаточно сужает множества $X^{(s_0)}$, применяются процедуры анализа по функционалу, которые заключаются в выделении из множества $X^{(s_0)}$ подмножеств перспективных по функционалу вариантов $X_\gamma^{(s_0)}$ и выборе в этом множестве максимального варианта.

На функционал вводится дополнительное ограничение:

$$F(\mathbf{x}) \overset{\beta}{\succ} F^{*(\gamma)}, \quad 0 \le \beta \le 1, \tag{3}$$

где $F^{*(\gamma)} \in \left[\min\limits_{x \in X^{(s_0)}} F(\mathbf{x}), \max\limits_{x \in X^{(s_0)}} F(\mathbf{x})\right]$ - нечеткий параметр с функцией принадлежности $\nu\left(F^{*(\gamma)}\right)$.

Обозначим множество вариантов, которые удовлетворяют (3) при выбранном значении $F^{*(\gamma)}$, через

$$D_F^\gamma = \left\{\mathbf{x} \mid F(\mathbf{x}) \overset{\beta}{\succ} F^{*(\gamma)}\right\},$$

и множество допустимых вариантов во множестве D_F^γ через

$$D_\gamma = D \cap D_F^\gamma.$$

Обозначим через $P\left(X^{(s_0)}\right)$ множество всех подвариантов. Сформируем семейство $\wp_\gamma = \cup P^\gamma$ множеств подвариантов $P^\gamma \subseteq P\left(X^{(s_0)}\right)$ такое, что $R(\wp_\gamma) = X^{(s_0)}$ и $\overline{R}(\wp_\gamma) = D_F^\gamma$.

Величина ΔF_p^γ называется допуском для множества подвариантов $P^\gamma \subseteq \wp_\gamma$ относительно множества D_F^γ по функционалу, если из того, что $F(p) \overset{\beta}{\prec} \Delta F_p^\gamma$ следует, что допустимое родовое множество $\overline{R}^\gamma(p) = R(p) \cap D_\gamma$ подварианта p пусто.

Исключение вариантов из множества $X^{(s_0)}$ по ограничению (3) осуществляется путем отсева с помощью допусков тех подвариантов $p \in P^\gamma$, которые не входят в максимальные варианты.

Аналогично [Волкович, Волошин, 1984] можно показать, что если множество D_γ не пусто и подвариант p не удовлетворяет допуску ΔF_p^γ по функционалу, то допустимое родовое множество $\overline{R}^\gamma(p)$ подварианта p не содержит оптимального варианта.

Обозначим систему допусков для семейства $\wp_\gamma$ через $Q(\wp_\gamma) = \cup\left\{\Delta F_p^\gamma\right\}$. Система допусков вычисляется с помощью оператора аппроксимации Ω, определенного на семействах $\wp_\gamma$ и $\wp$. В результате получим полную систему допусков $Q(\wp, \wp_\gamma) = Q(\wp) \cup Q(\wp_\gamma)$.

Тогда схема работы «нечеткого» оператора Ω будет следующей:

Шаг 0. Для заданного значения $F^{*(\gamma)}$ вычисляем для семейств $\wp^{(0)} = \wp$ и $\wp_\gamma^{(0)} = \wp_\gamma$ полную систему допусков $Q\left(\wp^{(0)}, \wp_\gamma^{(0)}\right)$, где $\wp = \cup P$, $\wp_\gamma = \cup P^\gamma$.

Шаг s. На множествах $P^{(s-1)}$ и $P^{\gamma(s-1)}$, выбранных из семейств $\wp^{(s-1)}$ и $\wp^{(s)}$, при заданной полной системе допусков $Q\left(\wp^{(s-1)}, \wp_\gamma^{(s-1)}\right)$, применяем оператор отсева E. В результате получим новые семейства $\wp^{(s)} = \cup P^{(s)}$, $\wp^{(s)} \subseteq \wp^{(s-1)}$ и $\wp_\gamma^{(s)} = \cup P^{\gamma(s)}$, $\wp_\gamma^{(s)} \subseteq \wp_\gamma^{(s-1)}$ и множество $X_\gamma^{(s)}$ такое, что $X_\gamma^{(s)} \subseteq X_\gamma^{(s-1)} \subseteq X$. Для семейств $\wp^{(s)}$ и $\wp_\gamma^{(s)}$ вычисляем полную систему допусков $Q\left(\wp^{(s)}, \wp_\gamma^{(s)}\right)$. При вычислении допусков предлагается использовать нечеткие множества α-уровней [Орловский, 1981] и смещение вправо нижней границы интервала допуска. В случае, если $Q\left(\wp^{(s-1)}, \wp_\gamma^{(s-1)}\right) = Q\left(\wp^{(s)}, \wp_\gamma^{(s)}\right)$, то вычисление оканчивается, при этом $X_\gamma^{(s_0)} = X_\gamma^{(s)}$, иначе переходим к шагу *s*+1.

Множество вариантов $X_\gamma^{(s_0)}$, которые образовывается в результате применения оператора Ω, является аппроксимацией множеств допустимых вариантов $D_\gamma \subseteq X_\gamma^{(s_0)}$, среди которых ищется максимальный вариант. Рассмотрим следующие случаи.

1. $X_\gamma^{(s_0)} = \varnothing$, то есть множество D_γ не содержит допустимых вариантов. Аналогично [Волкович, Волошин, 1984] можно показать, что в этом случае любой вариант $\mathbf{x} \in D$ удовлетворяет условию $F(\mathbf{x}) \overset{\beta}{\prec} F^{*(\gamma)}$. Таким образом, $F^{*(\gamma)}$ является оценкой функционала F на множестве X. В этом случае уменьшаем значение $F^{*(\gamma)}$, выбираем величину $F^{*(\gamma+1)} < F^{*(\gamma)}$ и опять применяем оператор Ω.

2. Множество $X_\gamma^{(s_0)}$ объемно, то есть $\left|X_\gamma^{(s_0)}\right| > N$. Тогда в нем находим максимальный вариант, который проверяем на оптимальность с помощью критериев оптимальности. Если максимальный вариант $\mathbf{x}^*$ не удовлетворяет условию (3), то он может быть выбран в качестве приближенного варианта.

3. $\left|X_\gamma^{(s_0)}\right| \le N$. В множестве $X_\gamma^{(s_0)}$ выбираем максимальный вариант $\mathbf{x}^*$, который проверяем на оптимальность. В случае, если он допустим, но не оптимален, то выбираем его в качестве приближенного. Если он недопустим, то на семействах $\wp^{(s_0)}$ и $\wp_\gamma^{(s_0)}$ применяем оператор конструирования κ, который строит множество $\widetilde{\wp} = \wp \cup \wp_\gamma$ и формирует агрегированную задачу A_γ: максимизировать

$$F\left(p_1, \ldots, p_j, \ldots p_r\right)$$

при $g_i\left(p_1, \ldots, p_j, \ldots, p_r\right) \le g_i^*$, $i = \overline{1, m}$, $\left(p_1, \ldots, p_j, \ldots, p_r\right) \in \widetilde{\wp}$, $r \le n$, $\cup p_j = \mathbf{x} \in X_\gamma^{(s_0)}$.

Если для максимального варианта $\mathbf{x} = \cup p_j$ не выполняется критерий оптимальности, то решается задача A_γ и полученное решение проверяется на оптимальность.

Заключение

При решении практических задач принятия решений необходимо учитывать приближенность, неточность, нечеткость исходных данных. Во многих случаях ''основной компонентой нечеткости'' в ситуациях принятия решений, которые описываются моделями математического программирования, является описание допустимого множества решений (правых частей ограничений). В докладе предлагается модификация процедур метода последовательного анализа вариантов, одного из наиболее эффективных методов решения многовариантных задач выбора, на случай нечетких данных.

Благодарности

Работа опубликована при финансовой поддержке проекта **ITHEA XXI** Института информационных теорий и приложений FOI ITHEA Болгария www.ithea.org и Ассоциации создателей и пользователей интеллектуальных систем ADUIS Украина www.aduis.com.ua.

Список литературы

[Беллман, 1960] Беллман Р. Динамическое программирование.- Москва: Иностранная литераткура.,1960.-400с.

[Волкович, Волошин, 1978] Волкович В.Л., Волошин А.Ф. Об одной схеме метода последовательного анализа и отсеивания вариантов//Кибернетика,1978,№4.-С.99-105.

[Волкович, Волошин, 1984] Волкович В.Л., Волошин А.Ф. и др. Методы и алгоритмы автоматизированного проектирования сложных систем управления. – Киев: Наукова.думка, 1984. – 216 с.

[Волкович, Волошин, 1993] Волкович В.Л., Волошин А.Ф. и др.Модели и методы оптимизации надежности сложных систем.- Киев: Наукова думка,1993.-312с.

[Волошин, 1980] Волошин А.Ф. Нахождение субоптимальных решений в дискретных оптимизационных задачах методом ПАВ//Вычислительные аспекты в пакетах прикладных программ.- Киев, ИК АН УССР,1980.-С.25-35.

[Волошин, 1987] Волошин А.Ф. Метод локализации области оптимума в задачах математического программирования//Доклады АН СССР,1987,том 293,№3.-С.549-553.

[Волошин, 2004] Волошин А.Ф. Методы анализа статических балансовых эколого-экономических моделей большой размерности//Издательство ”Педагогика”, Научные записки, том 7, Киев, 2004.-С.43-55. (укр. яз.).

[Волошин, 2009] Волошин А., Кудин В., Богаенко В. Анализ малых возмущений линейных экономико-математических моделей//International Book Series ”Information Science&Computing”,N10,2009.-P.67-73.

[Гаращенко, Волошин, 2009] Гаращенко Ф.Г., Волошин А.Ф. и др. Развитие методов и технологий моделирования и оптимизации сложных систем:Монография.-Киев:Издательство ”Сталь”, 2009.-668с.

[Михалевич, Шор, 1961] Михалевич В.С., Шор Н.З. Метод последовательного анализа вариантов при решении вариационных задач управления, планирования и проектирования//Труды IV Всесоюз.мат.съезда,М.,1961.-С.91.

[Михалевич, Шор, 1977] Михалевич В.С., Шор Н.З. и др. Вычислительные методы выбора оптимальных решений.- Киев: Наукова думка,1977.-178с.

[Орловский, 1981] Орловский С.А. Проблемы принятия решений при нечеткой исходной информации. – Москва: Наука. Главная редакция физико-математической литературы, 1981. – 208 с.

Сведения об авторах

Волошин Алексей Федорович *– профессор, Киевский национальный университет имени Тараса Шевченко, факультет кибернетики. Киев, Украина. E-mail: ovoloshin@unicyb.kiev.ua*

Маляр Николай Николаевич *– декан математического факультета, заведующий кафедрой кибернетики и прикладной математики Ужгородского национального университета, кандидат технических наук, доцент. Ужгород, Украина. E-mail: cyber@mail.uzhgorod.ua*

Швалагин Оксана Юрьевна *– аспирант, Ужгородский национальный университет, математический факультет. Ужгород, Украина.*

КАЧЕСТВЕННЫЙ АНАЛИЗ МНОГОМЕРНЫХ НЕЧЕТКИХ РАЗНОСТНЫХ СИСТЕМ

Евгений Ивохин

Аннотация: *Предлагается способ исследования устойчивости нечетких разностных систем большой размерности на основе применения метода вектор-функций Ляпунова и специальной процедуры декомпозиции системы на подсистемы.*

Ключевые слова: *разностные системы, функции Ляпунова, нечеткость*

ACM Classification Keywords: *H.4.2 Information Systems Applications: Types of Systems: Decision Support.*

Conference*: The paper is selected from XV[th] International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.*

Введение

Использование нечетких множеств для описания ситуаций с неопределенностью находит все большее применение в разных прикладных отраслях [Чуличков, 2000]. Однако, следует заметить, что в основном рассматриваются системы, размерность которых невелика [Меренков, 2000; Пушков, 2001; Ивохин, Волчков, 2006]. Исследования многомерных систем усложняются не только проблемами, связанными с размерностью, но и с отсутствием специфичных операций декомпозиции и агрегирования нечетких множеств.

В данной работе рассматривается одна из возможных реализаций процедуры декомпозиции нечетких систем на подсистемы и ее использование для анализа устойчивости многомерных разностных нечетких систем методом векторних функций Ляпунова.

Постановка задачи

Рассмотрим произвольное конечномерное пространство над полем действительных чисел $X = R^n$. Будем называть X универсальным множеством.

Определение. Нечетким множеством $\widetilde{A}$ универсального множества $X = \{x\}$, называется совокупность пар $A = \{(\mu_{\widetilde{A}}(x), x)\}$, где $\mu_{\widetilde{A}}(x) : X \to [0,1]$ – отображение множества X в единичный отрезок [0,1], которое называется функцией принадлежности нечеткого множества A.

Рассматривается нечеткая разностная система:

$$\widetilde{X}_{k+1} = R_k \circ \widetilde{X}_k, \quad k = 0,1,2,... \tag{1}$$

где $\widetilde{X}(t_0) = \widetilde{X}_0 \in E^n$ – компактное множество начальных состояний, $\widetilde{X}(t_k) = \widetilde{X}_k$ – нечеткие множества в $X = R^n$ возможных состояний системы в моменты времени $t_k \in T$, которые определяют решения системы, $\mathrm{R}(t_k) = \mathrm{R}_k$ – некоторые нечеткие отображения из X в X, определяющие переходы системы, $\circ$ - операция композиции, $k = 0,1,2,...$.

Допустим, що существует траектория $\{\bar{x}_k\}_{k=0,1,2,...}$, $\bar{x}_k \in \operatorname{supp} \widetilde{X}_k$, $k = 0,1,2,...$, которая является регулярной траекторией (РТС) системы (1).

Определение. РТС (1) $\{\bar{x}_k\}_{k=0,1,2,...}$ называется устойчивой по Ляпунову, если $\forall \bar{T} > 0, \forall \varepsilon > 0, \forall \eta > 0 \ \exists \ \delta(\varepsilon, \eta, \bar{T}) > 0, \gamma(\varepsilon, \eta, \bar{T}) > 0$ такие, что для любой траектории $\{x_k\}_{k=0,1,2,...}$ системы (1), начальные условия которой удовлетворяют неравенствам:

$$|| \bar{x}_0 - x_0 || < \delta ,$$

$$| \mu(x_0) - 1 | < \gamma ,$$

для $\forall k : t_k \geq \bar{T}$ справедливы неравенства:

$$|| \bar{x}_k - x_k || < \varepsilon , \ k = 0,1,2,... ,$$

$$| \mu(x_k) - 1 | < \eta , \ k = 0,1,2,.... .$$

Пример разностной нечеткой системы.

Рассмотрим динамику системы

$$\widetilde{X}_{k+1} = \mathrm{R}_k \circ \widetilde{X}_k , \ \ k = 0,1,2,...$$

$$x_k \in \operatorname{supp} \widetilde{X}_k = [a_k, b_k] \subseteq X = [-1,1], \ k = 0,1,2,...$$

$$a_{k+1} = sin(a_k), b_{k+1} = sin(b_k), a_0 = -1, b_0 = 1,$$

$$\mu_k = \begin{cases} 1 - cos(x_k), x_k \in [a_k, b_k] \\ 0, x_k \notin [a_k, b_k] \end{cases}$$

Траектории данной системы имеют вид, показанный на рис.1.

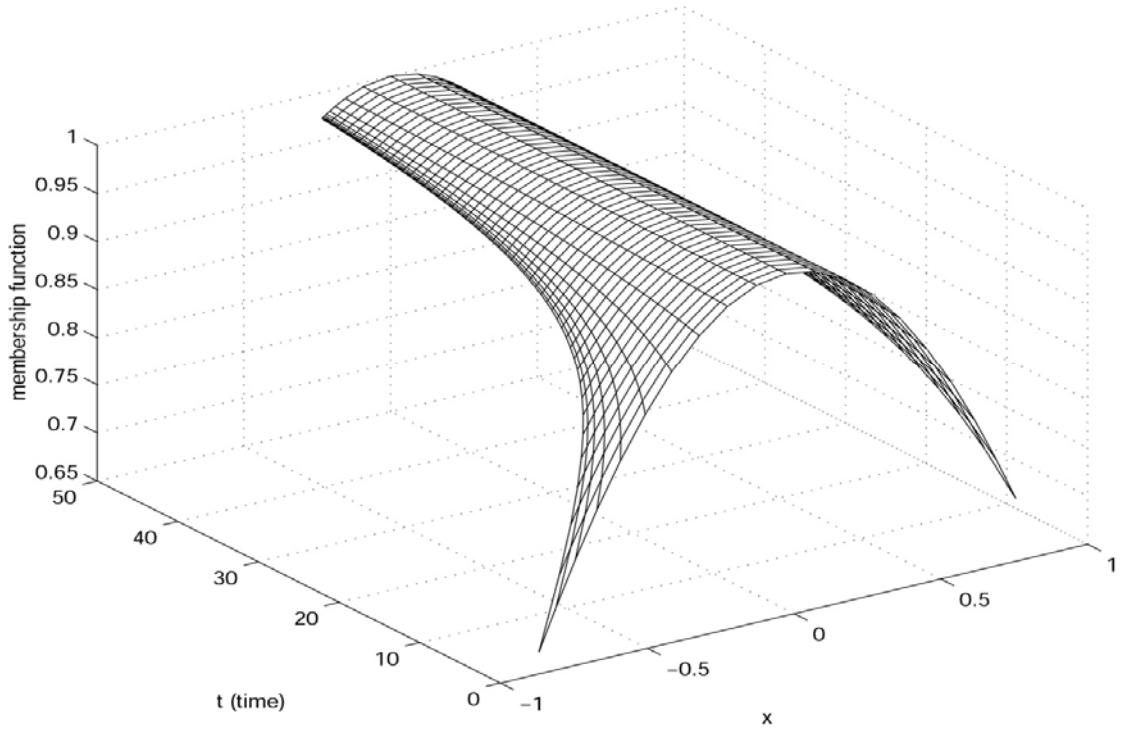


Рис.1. Траектории разностной нечеткой системы

Утверждение. Если система (1) имеет регулярную траекторию $\{\bar{x}_k\}_{k=0,1,2,\ldots}$ и существуют функции $V: X \to [0,1]$ та $G: [0,1] \to [0,1]$ такие, что для любой траектории системы (1) $\{x_k\}_{k=0,1,2,\ldots}$ выполняются неравенства:

$$\begin{aligned} &V_k = V(x_k) \geq 0 \\ &V(\bar{x}_k) = 0 \\ &\Delta V_k = V_{k+1} - V_k \leq 0 \\ &G_k = G(\mu_k) \geq 0 \\ &G(\bar{\mu}_k) = G(1) = 0 \\ &\Delta G_k = G_{k+1} - G_k \leq 0 \end{aligned},$$

то регулярная траектория системы (1) является устойчивой по Ляпуновим.

Результаты исследования многомерных систем

I. Случай многомерной четкой разностной системы (1).

Пусть система (1) имеет вид

$$x_i(k+1) = f_i(x_i(k),k) + \sum_{j=1, j\neq i}^{m} C_{ij} x_j(k), \quad i = \overline{1,m}, \ k = 0,1,2,\ldots \quad . \tag{2}$$

Рассмотрим каждую из подсистем

$$x_i(k+1) = f_i(x_i(k),k), \quad i = \overline{1,m} \tag{3}$$

отдельно. Пусть $f_i(0,k) \equiv 0$, то есть подсистемы имеют нулевые положения равновесия, которые асимптотически устойчивы. Существуют функции Ляпунова $\upsilon_i(x_i,k)$, которые удовлетворяют вдоль решений $x_i(k)$ каждой из подсистем (3) условиям:

1. $c_{i1}\|x_i(k)\|^2 \leq \upsilon_i(x_i(k),k) \leq c_{i2}\|x_i(k)\|^2$, (4)
2. $\Delta\upsilon_i(x_i(k),k) = \upsilon_i(x_i(k+1),k+1) - \upsilon_i(x_i(k),k) \leq -c_{i3}\|x_i(k)\|^2$,
3. $\left|grad\,\upsilon_i(x_i(k),k)\right| = c_{i4}\|x_i(k)\|^2$, $c_{ij} > 0$, $i = \overline{1,m}$, $j = \overline{1,4}$.

При этом нулевое решение $x_i(k) \equiv 0$ каждой из подсистем (3) будет экспоненциально устойчивым, то есть выполняется неравенство

$$\|x_i(k)\| \leq \sqrt{c_{i2}/c_{i1}}\,(1 - c_{i3}/c_{i2})^{k/2}\|x_i(0)\|.$$

Рассмотрим случай линейных подсистем

$$x_i(k+1) = A_i x_i(k), \ k = 0,1,2,\ldots \ i = \overline{1,m}. \tag{5}$$

Замкнутая система имеет вид

$$x_i(k+1) = A_i x_i(k) + \sum_{j=1, j\neq i}^{m} C_{ij} x_j(k), \ k = 0,1,2,\ldots \ i = \overline{1,m}. \tag{6}$$

Обозначим

$$\varphi\ (H_0) = \lambda_{max}(H_0) / \lambda_{min}(H_0),\ \ \gamma\ (H_0) = \lambda_{min}(Q_0) / \lambda_{max}(H_0),$$

где $H_0,\ Q_0$ - симметрические положительно определенные матрицы, связанные матричным разностным уравнением Ляпунова

$$H_0 - R^T H_0 R = Q_0. \tag{7}$$

Здесь $R = \{r_{ij}\},\ i, j = \overline{1,m}$, - квадратная матрица, элементы которой определяются соотношениями

$$r_{ij} = \begin{cases} 1 + \left[-\lambda_{\ min}(Q_i) + \sum\limits_{s=1, s\neq i}^{m} \left| A_i^T H_i C_{is} \right| \right] / \lambda_{\ max}(H_i),\ i = j \\ \left[\left| A_i^T H_i C_{ij} \right| + \sum\limits_{s=1, s\neq i}^{m} \left| C_{ij}^T H_i C_{is} \right| \right] / \lambda_{\ min}(H_j),\ i \neq j \end{cases} \tag{8}$$

$$K = \max_{i=1,m} \{\lambda_{max}(H_i)\} / \min_{i=1,m} \{\lambda_{min}(H_i)\}. \tag{9}$$

а $H_i,\ Q_i,\ i = \overline{1,m}$ симметрические положительно определенные матрицы, которые входят в матричные разностные уравнения Ляпунова

$$H_i - A_i^T H_i A_i = Q_i,\ i = \overline{1,m}. \tag{10}$$

Теорема. *Пусть для каждой из подсистем (5) существуют положительно определенные матрицы* $H_i,\ i = \overline{1,m}$, *такие, что матрица* $R = \{r_{ij}\},\ i, j = \overline{1,m}$ *вида (8) асимптотически устойчива (имеет собственные числа* $|\lambda_{\ i}(R)| < 1,\ i = \overline{1,m}$*). Тогда исходная система (6) также асимптотически устойчива и для ее решения* $x(k)$ *справедливо*

$$\|x(k)\| < K\sqrt{\varphi\ (H_0)}\left(1 - \gamma\ (H_0)\right)^{k/2} \|x(0)\|,\ k = 1,2,\ldots. \tag{11}$$

II. Подход к качественному анализу многомерных нечетких разностных систем.

Пусть задана динамическая система *S*, множество состояний которой описывается нечеткими множествами $\widetilde{A}_S = \{(x, \mu_S(x)), x \in X\}$, где *X* - некоторое универсальное множество. Предположим, что в системе выделено две подсистемы S_1, S_2, состояния которых описываются нечеткими множествами $\widetilde{A}_{S_1} = \{(x_1, \mu_{S_1}(x_1)), x_1 \in X_1\}$ и $\widetilde{A}_{S_2} = \{(x_2, \mu_{S_2}(x_2)), x_2 \in X_2\}$ универсальных множеств X_1, X_2 соответственно, $X_1 \cup X_2 = X$. Отождествляя $\mu_{S_1}(x_1)$ та $\mu_{S_2}(x_2)$ с полезностью подсистем S_1, S_2, будем представлять величины меры принадлежности $\mu_S(x)$ в виде

$$\mu_S(x) = [\mu_{S_1}(x_1)]^{\alpha} [\mu_{S_2}(x_2)]^{\beta}, \tag{12}$$

где $\alpha, \beta \geq 0$.

Нахождение α, β можно провести на основе методики принятия решения в многокритериальной ситуации с учетом важности критериев.

Пусть элементы квадратной матрицы $P=\{p_{ij}\}_{i,j=\overline{1,2}}$, $p_{ij}=1,\ p_{ij}\geq 0,\ i,j=\overline{1,2},\ i\neq j$ определяют важности подсистем S_1, S_2. Матрица с неотрицательными элементами называется *M*-матрицей (позитивной матрицей). По теореме Перрона-Фробениуса она имеет максимальное собственное число $\lambda(P)\geq 0$ и неотрицательный собственный вектор $v(P)=(v_1,v_2)$, соответствующий собственному числу $\lambda(P)$. Выбираем $\alpha=v_1, \beta=v_2$, получаем показатели степеней в (12). Кроме этого, $\mu_{S_1}(x_1)+\mu_{S_2}(x_2)=1$. В результате нахождение функций $\mu_{S_1}(x_1)$ и $\mu_{S_2}(x_2)$ сводится к решению нелинейной системы уравнений

$$\begin{aligned}&[\mu_{S_1}(x_1)]^{\alpha}[\mu_{S_2}(x_2)]^{\beta}=\mu_S(x)\\&\mu_{S_1}(x_1)+\mu_{S_2}(x_2)=1\end{aligned}. \qquad (13)$$

Заключение

Предложенные процедуры декомпозиции и агрегирования нечетких систем на основе введенной операции с нечеткими множествами позволяют рассмотреть задачи качественного анализа динамики многомерных нечетких систем вида (1) с использованием метода вектор-функций Ляпунова.

Благодарности

Работа опубликована при финансовой поддержке проекта **ITHEA XXI** Института информационных теорий и приложений FOI ITHEA Болгария www.ithea.org и Ассоциации создателей и пользователей интеллектуальных систем ADUIS Украина www.aduis.com.ua.

Список литературы

[Ивохин, Волчков, 2006] Ивохін Е.В., Волчков С.А. Качественное исследование динамики нечетких дискретных систем// Информационно-управляющие комплексы и системы. —2006. — №1. — С. 114-122. (укр. яз.).

[Меренков, 2000] Меренков Ю.Н. Устойчивоподобные свойства дифференциальных включений, нечетких и стохастических дифференциальных уравнений М.:Ун-т Др.Нар., 2000. 123с.

[Пушков, 2001] Пушков С.Т. Об общей теории нечетких систем: глобальное состояние и нечеткая глобальная реакция нечеткой системы// Изв.РАН. Теория и системы управления, 2001, №5, стр. 105-109.

[Чуличков, 2000] Чуличков А.И. Математические модели нелинейной динамики. М: Физматлит, 2000. 294с.

Сведения об авторе

Ивохин Евгений Викторович *– доцент, Киевский национальный университет имени Тараса Шевченко, факультет кибернетики. Киев, Украина. E-mail: ivohin@univ.kiev.ua*

Economics Decision Support Systems

COLLECTIVE PRODUCT COST SHARING IN CONDITIONS OF MANAGED ECONOMY

Olexij Voloshin, Vasyl Laver

Abstract: *The collective product cost-sharing problems in the conditions of the managed economy are being observed. The fuzzy generalization and "dual" approach to the solution of this problem is offered.*

Keywords: *decision making, cost-sharing, regulated monopoly.*

ACM Classification Keywords: *H.4.2 Information Systems Applications: Types of Systems: Decision Support.*

Conference*: The paper is selected from XV[th] International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.*

Introduction

The distribution of costs, profits and resources (financial, energy e.t.c) is one of the most important problems existing today. Thus, in 2007 the Nobel Prize in economy was granted to U.S. scientists Leonid Hurwicz, Eric Maskin and Roger Myerson for research of cost sharing processes in the conditions of managed economy [Nobelprize].

Liberal economists proclaim that state shall not interfere into economic relations. However, the classic liberal economic model with its cult of liberation of economic relations from the state's interference proved to be unable to cope effectively with multiple relevant challenges. Unregulated economic relations are abstract. Maximizing their profits, major economic players (monopolists and monopsonists [Ponomarenko, Perestyuk, Burym 2004]) do not pay much attention to solving social problems (including ecological). At the same time, existence of monopolies creates certain advantages. For example, in the situation when the production technology has increasing returns to scale, monopoly is more efficient at the production. Even under decreasing returns to scale, the case for joint production frequently arises (the workers of the cooperative jointly own its machines, and the partners of the law firm share its patrons [Moulin, 1991]). Considering the above, state regulation of economic activities is becoming a very important issue. The symbiosis of market and state methods of regulation is very popular, especially nowadays during the world economy crisis.

Collective product manufacturing problem

The model of production of collective product is examined [Voloshin, Mashchenko, 2006]:

$$u_i(M_i - x_i, y) \to \max, i \in N = \overline{\{1, n\}}, n \geq 2,$$
$$\sum_{i=1}^{n} x_i = c(y), \tag{1}$$

M_i - initial amount of money of the agent *i*; x_i - his contribution to the production of *y* units of the collective product; *c(y)* – the cost of producing y units of collective product; u_i - his utility function. Let us assume that

cooperation is efficient: $c(y) \le \sum_{i=1}^{n} M_i$. We impose on functions c and u_i terms that are ordinary for the models of the microeconomics [Ponomarenko, Perestyuk, Burym 2004]: $c(0) = 0$, c is non-decreasing and convex; u_i are cuasi-convex functions, increasing on every variable ($m_i = M_i - x_i$ and y) for all agents; c and u_i are differentiable. The economy has two goods, one pure public good and one private good (money).

Using additive aggregation $u = \sum_{i=1}^{n} u_i$ (which refers to the utilitarian approach [Voloshin, Mashchenko, 2006]), we can find Pareto optimal allocations (in assumptions of the (1) problem) from next necessary and sufficient condition (Samuelson's condition [Moulin, 1991]):

$$\sum_{i=1}^{n} u_{iy} / u_{im_i} = c'(y), \tag{2}$$

where u_{iy} (u_{im_i}) – is the partial derivative of each agent's utility function w.r.t the public (private) good. $c'(y)$ - is the derivative of the production function w.r.t. the collective product.

Assume that all preferences are additively separable in the input and output goods and linear on the output: $u_i(M_i - x_i, y) = b_i(y) + (M_i - x_i)$, where $b_i(y)$ is the monetary equivalent of *y* units of public good.

Under quasi-linearity of utilities, Samuelson's condition (2) simplifies to $\sum_{i=1}^{n} b_i'(y) = c'(y)$, namely, the sum of marginal benefits equals the marginal cost of the public good. Note that the input levels x_i disappeared from the formula. One of the approaches to the solving of the (1) problem is modeling it as the following transferable utility (TU) cooperative game: for all $S \subseteq \{1,...,n\}$: $v(S) = \max_{y \ge 0} \left\{ \sum_{i \in S} M_i + \sum_{i \in S} b_i(y) - c(y), 0 \right\}$ (the reflection v puts in accordance to every coalition its surplus)

A feasible allocation $(x_1, x_2, ..., x_n, y)$ is then in the core of our TU public good economy if and only if $\sum_{i \in S} (b_i(y) + (M_i - x_i)) \ge v(S)$ for all of the coalitions. Note that this inequality for $S = N$ means precisely that the level *y* is efficient (i.e., maximizes total surplus).

In [Moulin, 1991] a numerical example is examined. The technology has constant returns to scale $c(y) = \frac{3}{2} y$.

There are two agents with quasi-linear preferences: $b_1(y) = \ln(1+y)$, $b_2(y) = 2\sqrt{y}$. Initial money values are ignored (accepted equal to zero). Unique maximum y^* is computed by solving the Samuelson's condition (2): $y^* = 1$. The corresponding surplus is $v(12) = (b_1 + b_2 - c)(y^*) = 1{,}19$. Costs for production of y^* are equal $c(y^*) = 1{,}5$. The surplus that single agent coalitions can achieve are $v(1) = 0$ and $v(2) = 0{,}67$.

The core allocations divide arbitrarily the cooperative surplus $\Delta v(12) = v(12) - v(1) - v(2) = 0{,}53$ between the two agents. Particularly, in [Moulin, 1991] two allocations are proposed. At one extreme, agent 1 keeps all the cooperative surplus, implying the following cost shares: $x_1 = 0{,}17$ and $x_2 = 1{,}33$. At the other extreme, agent 2 keeps the cooperative surplus, and hence the cost shares: $x_1' = 0{,}69$ and $x_2' = 0{,}81$.

"Dual" approach to the solution of problem (1)

The "dual" approach to the solution of problem (1) with quasi-linear preferences is proposed. Three cost-sharing principles are a priori set between agents [Voloshin, Mashchenko, 2006]:

1. Costs equality ($x_i = \frac{c(y)}{n}$ for all *i*);
2. "Free money" equality ($x_i = b_i - \frac{M - c(y)}{n}$);
3. Cost proportionality (to initial money amounts $x_i = \left(M_i / M\right) / c(y)$).

Note that in this model, agents can freely transfer the input (money), so these cost-sharing principles are not ideal. Thus, when using the cost equality principle there can be a situation, when some agent's cost share will exceed his initial wealth. Also, when using "free money" equality principle there can be a situation when some agents will receive a cash payment for using the public good. In such cases the will to co-operation decreases. The cost proportionality principle always leads to the allocations from the core of the game. But there can be a situation, when some of the agents will not be satisfies by his part of the collective surplus – in this case he might give up the co-operation.

Let's illustrate these cost-sharing principles on the above-mentioned example. Initially the agents are endowed with following amounts of money: $M_1 = 0{,}19$, $M_2 = 1{,}81$. Application of the cost equality principle leads us to the following allocation: $x_1 = x_2 = 0{,}75$. As we can see, cost share of the agent 1 exceeds his initial wealth, so he won't co-operate. Maximal coalition surplus in this case equals $v(12) = \left(M + b_1 + b_2 - c\right)(y^*) = 3{,}19.$

Using "free money" equality principle gives us the next allocation: $x_1 = -0{,}06$ and $x_2 = 1{,}56$. In this case the agent 1 receives a cash payment of 0,06. It's natural, that the agent 2 will refuse the co-operation.

The cost proportionality principle gives us a following cost sharing: $x_1 = 0{,}1425$, $x_2 = 1{,}3575$. Agents' surpluses are $b_1(y^*) + \left(M_1 - x_1\right) = 0{,}7405$ and $b_2(y^*) + \left(M_2 - x_2\right) = 2{,}4525$. If agent 1 will not be satisfied with his surplus, he can give up co-operation and coalition will disintegrate.

Let us consider the case, when all preferences are additive in the input and output goods: $u_i(m_i, y) = a_i \ln(m_i + M) + b_i \ln y$ ($m_i + M > 0$, because we suppose that the expenses of each agent can not exceed the sum of money in the economy), where a_i, b_i are positive constants, $i = \overline{1,n}$.

In this case there are difficulties with the search of the optimal cost share allocation, because in general we have two equations with *n+1* variables. The above-mentioned cost-sharing principles allow avoiding these difficulties. Beforehand setting some cost-sharing principle and putting the proper cost share expressions in the Samuelson's condition we can get an equation with one variable - *y*.

Let us consider an economy with constant returns to scale: $c(y) = \alpha y$. In this case the optimal amount of the public good y^* can be find analytically (in general it can be found numerically [Bandy, 1988]).

We find the expressions for the partial derivatives of each agent's utility function:

$$u_{iy}(m_i, y) = \frac{b_i}{y}, \qquad u_{im_i}(m_i, y) = \frac{a_i}{m_i + M} \text{ (где } m_i = M_i - x_i, \ \forall i \in N \text{).}$$

As the technology has constant returns to scale, $c'(y) = \alpha$.

Putting these expressions in equation (2), we get:

$$\sum_{i=1}^{n}\frac{b_i}{y}\frac{M_i - x_i + M}{a_i} = \alpha,$$

$$\sum_{i=1}^{n}\frac{b_i}{a_i}\left[M_i - x_i + M\right] = \alpha y.$$

Let's examine the case of equal cost-sharing ($x_i = \alpha y / n$, $\forall i \in N$). Then

$$\sum_{i=1}^{n}\frac{b_i}{a_i}\left[M_i + M\right] - \frac{\alpha y}{n}\sum_{i=1}^{n}\frac{b_i}{a_i} = \alpha y, \qquad \sum_{i=1}^{n}\frac{b_i}{a_i}\left[M_i + M\right] = y\left(\frac{\alpha}{n}\sum_{i=1}^{n}\frac{b_i}{a_i} + \alpha\right)$$

From the last expression we can get the formula for y^*: $$y^* = \frac{\sum_{i=1}^{n} b_i / a_i \left[M_i + M\right]}{\alpha\left[1 + n^{-1}\sum_{i=1}^{n}\left(b_i / a_i\right)\right]}$$

In a similar manner we can get formulas for finding the optimal value of the public good in case of other cost-sharing principles:

– $y^* = \dfrac{\sum_{i=1}^{n} b_i / a_i \left[M(1 + n^{-1})\right]}{\alpha\left[1 + n^{-1}\sum_{i=1}^{n}\left(b_i / a_i\right)\right]}$ (in case of "free money" equality);

– $y^* = \dfrac{\sum_{i=1}^{n} b_i / a_i \left[M_i + M\right]}{\alpha\left[1 + \sum_{i=1}^{n}\left(b_i M_i / a_i M\right)\right]}$ (in case of cost proportionality to initial money amounts).

Let's consider a numerical example. There are two agents with initial wealth $M_1 = 0{,}19$, $M_2 = 1{,}81$ and additive preferences:

$u_1(m_1, y) = 2\ln(m_1 + M) + \ln y$, $u_2(m_2, y) = 3\ln(m_2 + M) + \ln y$.

Technology has constant returns to scale: $c(y) = \frac{3}{2}y$.

For different cost-sharing principles we will get different optimal values for public good production. Thus, using cost equality principle gives us $y^* = 1{,}113$ with corresponding cost share $x_1 = x_2 = 0{,}835$ and surplus $v(12) = 4{,}093$. Using "free money" equality principle we get the next values: $y^* = 1{,}176$, $x_1 = 0{,}072$, $x_2 = 1{,}692$, $v(12) = 4{,}077$. In case of cost proportionality $y^* = 1{,}169$. Corresponding allocation is $x_1 = 0{,}167$ and $x_2 = 1{,}586$. Cooperative surplus in this case equals $v(12) = 4{,}119$. As we see, cost proportionality principle gives the biggest cooperative surplus. In the case of equal cost sharing agent 1 will refuse the co-operation, because his cost share exceeds his initial wealth.

Fuzzy generalization of problem (1)

Let us consider a fuzzy generalization of problem (1) [Voloshyn, Laver, 2008]. Assume the agents are normalized in order by their initial wealth M_i. Then *N* can be presented as a union of sets: $N = N_1 \cup N_2 \cup N_3$, where

$N_1 = \overline{\{1, n_1\}}$ - "poor" agents; $N_2 = \overline{\{n_1 + 1, n_2\}}$ - "middle class"; $N_3 = \overline{\{n_2 + 1, n\}}$ - "wealthy" agents, $n_1 \geq 1,\ n_2 \geq 2,\ n_1 < n_2 < n.$

Assume that there is some optimal value of the public good production y^*, found from the distinct problem using some cost-sharing principle. There can be a situation that "poor" agents will not have enough money to cover their cost share. On the other hand, "wealthy" agents can consent to cover the cost deficit, to guarantee the issue of optimum amount of public good.

Agents' cost shares satisfy the following terms: $0 \leq x_i \leq M_i,\ \forall i \in N$. Each agent sets a membership function $\mu_i : [0, M_i] \to [0,1]$, which characterizes the agent's satisfaction with his cost share.

Then the optimal allocation $(x_1, x_2, \ldots, x_n)$ can be found as a solution of the following problem:

$$\begin{cases} \mu_A(x) \to \max; \\ \sum_{i=1}^{n} x_i = c, \end{cases} \quad \text{where } \mu_A(x) = \min\{\mu_1(x_1), \mu_2(x_2), \ldots, \mu_n(x_n)\},\ c = c(y^*). \tag{3}$$

(3) can be equivalently offered as:
$$\begin{cases} \alpha \to \max; \\ \mu_i(x_i) \geq \alpha \ \ (\forall i \in N); \\ \sum_{i=1}^{n} x_i = c. \end{cases} \tag{4}$$

Let us examine a numerical example. There are three agents with the following initial amounts of money: $M_1 = 2,\ \ M_2 = 5,\ M_3 = 9$. Collective product worth $c=9$ has to be produced. Using equal cost-chare we will get the following allocation: $x_1 = x_2 = x_3 = 3$. Agent 1 will refuse the co-operation, as his cost share exceeds his initial wealth. We can avoid this problem if the agents' cost-share satisfy the following terms $0 \leq x_i \leq M_i,\ \forall i \in N$.

One of the possible cost-sharing methods in this case is the head tax [Voloshin, Mashchenko, 2006] with corresponding allocation $x_1 = 2, x_2 = 3{,}5, x_3 = 3{,}5.$ Assume that the agents agree to deviate from their cost shares and their priorities are described with following membership functions:

$$\mu_1 = \begin{cases} 1, x_1 \in [0;0{,}8]; \\ -0{,}6x_1 + 1{,}48, x_1 \in [0{,}8;1{,}3]; \\ -x_1 + 2,\ x_1 \in [1{,}3;2]; \end{cases} \qquad \mu_2 = \begin{cases} 1, x_2 \in [0;3]; \\ -0{,}2x_2 + 1{,}6, x_2 \in [3;4]; \\ -0{,}5x_1 + 2{,}8, x_2 \in [1{,}3;2]; \end{cases}$$

$$\mu_3 = \begin{cases} 1, x_3 \in [0;4{,}5]; \\ -0{,}05x_3 + 1{,}225, x_3 \in [4{,}5;6{,}5]; \\ -0{,}6x_3 + 4{,}8, x_3 \in [6{,}5;8]; \\ 0, x_3 \in [8;9]. \end{cases}$$

Optimal allocations can be found as the solution of the next problem:

From next system we find the high boundaries of the agents' maximum permissible cost share changes:

$$\begin{cases} \alpha \to \max; \\ -0{,}6x_1 + 1{,}48 \ge \alpha; \\ -0{,}2x_2 + 1{,}6 \ge \alpha; \\ -0{,}05x_3 + 1{,}225 \ge \alpha; \\ \sum_{i=1}^{n} x_i = c. \end{cases} \qquad \begin{array}{l} x_1 \le \dfrac{1{,}48-\alpha}{0{,}6}; \\ x_2 \le \dfrac{1{,}6-\alpha}{0{,}2}; \\ x_3 \le \dfrac{1{,}225-\alpha}{0{,}05}. \end{array}$$

Summing up the right parts of these inequalities and comparing them to *c*, we get the equation for finding the α :

$$\frac{1{,}48-\alpha}{0{,}6} + \frac{1{,}6-\alpha}{0{,}2} + \frac{1{,}225-\alpha}{0{,}05} = 9, \text{ so } \alpha \approx 0{,}97$$

Optimal (with 0,97 degree) allocation we find from the following terms:

$$\begin{cases} x_1 \le 0{,}85; \\ x_2 \le 3{,}15; \\ x_3 \le 5{,}1; \\ x_1 + x_2 + x_3 = 9. \end{cases}$$

One of the possible allocations is $x_1 = 0{,}85; x_2 = 3{,}15; x_3 = 5.$

Conclusion

The considered models enable to adequately describe the process of the collective good production and cost sharing processes. Unevenness of money allocation in society and individual preferences are taken in account.

Acknowledgements

The paper is published with financial support by the project ITHEA XXI of the Institute of Information Theories and Applications FOI ITHEA Bulgaria www.ithea.org and the Association of Developers and Users of Intelligent Systems ADUIS Ukraine www.aduis.com.ua.

Bibliography

[Bandy, 1988] Bandy B. Optimization Methods. – M.:. Radio I Svjaz', 1988. -128pg. (In Russian)

[Moulin, 1991] Herve Moulin. Axioms of Cooperative Decision Making. -M: Mir, 1991.-464p. (In Russian)

[Nobelprize] http://nobelprize.org/nobel_prizes/economics/laureates/2007/press.html

[Ponomarenko, Perestyuk, Burym 2004] Ponomarenko O.I., Perestyuk M.O., Burym V.M. Microeconomics. – K: Vyshcha Shkola, 2004.-262p. (In Ukrainian)

[Voloshin, Mashchenko, 2006] Voloshin O.F., Mashchenko S.O. Decision Making Theory - K.: VTC "Kyivj University", 2006.-304p. (In Ukrainian)

[Voloshyn, Laver, 2008] Voloshyn O.F., Laver V.O. Fuzzy Models of Regulated Monopoly in Production of the Collective Product – Works of the IV international school-seminar "Decision Making Theory". – Uzhgorod, UzhNU, 2008. - 175 p. (In Ukrainian)

Authors' Information

Olexij Voloshin *– Professor, Taras Shevchenko National University, faculty of cybernetics. Kyiv, Ukraine. E-mail: ovoloshin@unicyb.kiev.ua*

Vasyl' Laver *– Aspirant, Uzhgorod National University, faculty of mathematics. Uzhgorod, Ukraine. E-mail: v.laver@gmail.com*

THE FUZZY GROUP METHOD OF DATA HANDLING WITH FUZZY INPUTS

Yuriy Zaychenko

Abstract: *The problem of forecasting models constructing using experimental data in terms of fuzziness, when input variables are not known exactly and determined as intervals of uncertainty is considered in this paper. The fuzzy group method of data handling is proposed to solve this problem. The mathematical model of the problem mentioned above is built and fuzzy GMDH with fuzzy inputs is elaborated in the paper. The corresponding program which uses the suggested algorithm was developed. The experimental investigations and comparison of FGMDH with neural nets in problems of stock prices forecasting were carried out and presented in this paper.*

Keywords *Group Method of Data Handling, fuzzy, economic indexes, stock prices, forecasting*

ACM Classification Keywords: *H.4.2 [Information Systems Applications] Types of Systems - Decision support*

Conference: *The paper is selected from XVth International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.*

Introduction

The problem of forecast in financial sphere is of great importance. Financial processes have some specific features which prevent the application classical statistic methods and stipulate the development of novel approaches and methods based on artificial intelligence. One of such novel methods is *fuzzy group method of data handling (FGMDH)* developed in [Зайченко, 2003, Zaychenko, 2006]. As is well known, fuzzy GMDH allows to construct fuzzy models and has the following advantages:

1. The problem of optimal model finding is transformed to the problem of linear programming, which is always solvable;
2. There is interval regression model built as the result of method work out. Fuzzy GMDH in its original form has one limitation: the input variables must be crisp, non-fuzzy while the model constructed is fuzzy.

At the same time situations in financial sphere often occur when the input data are uncertain, given in the form of uncertainty intervals.

The goal of this paper is the development and investigation of extended version of FGMDH working in fuzzy environment when input variables are fuzzy and its comparison with fuzzy neural networks.

Mathematical Model of Group Method of data Handling with Fuzzy Inputs

The initial fuzzy GMDH considers a linear interval regression model:

$$Y=A_0Z_0+A_1Z_1+\ldots+A_nZ_n, \quad (1.1)$$

- where A_i are fuzzy numbers, which are described by threes of parameters $A_i = (\underline{A_i}, \breve{A}_i, \overline{A_i})$, where $\breve{A}_i$ – interval center, $\overline{A_i}$ – upper border of the interval, $\underline{A_i}$ - lower border of the interval,
- and Z_i are input variables which are also fuzzy numbers determined by parameters $(\underline{Z_i}, \breve{Z}_i, \overline{Z_i})$, $\underline{Z_i}$ - lower border, $\breve{Z}_i$ - center, $\overline{Z_i}$ - upper border of fuzzy number.

Then *Y* is output fuzzy number, which parameters are defined as follows (in accordance with L-R numbers multiplying formulas):

Center of interval:

$$\breve{y} = \sum \breve{A}_i * \breve{Z}_i .$$

Deviation in the left part of the membership function:

$$\breve{y} - \underline{y} = \sum (\left|\breve{A}_i\right| * (\breve{Z}_i - \underline{Z}_i) + (\breve{A}_i - \underline{A}_i) * \left|\breve{Z}_i\right|) .$$

Thus the upper border of the interval

$$\overline{y} = \sum (\left|\breve{A}_i\right| * (\overline{Z}_i - \breve{Z}_i) + \left|\breve{Z}_i\right| * (\overline{A}_i - \breve{A}_i) + \breve{A}_i * \breve{Z}_i) .$$

For the interval model to be correct, the real value of input variable Y should lay in the interval got by the method workflow.

So, the general requirements to estimation linear interval model are following:

to find such values of parameters $(\underline{A_i}, \breve{A}_i, \overline{A_i})$ of fuzzy coefficients, which enable:

a) Observed values y_k lay in estimation interval for Y_k;

b) Total width of estimation interval be minimal.

Input data for this task is Z_k=[Z_{ki}]$_i$ - input training sample, and also y_k – known output values, $k = \overline{1,M}$, *M* is a sample size (number of observation points).

There are two cases of fuzzy values membership function investigated in this work:

- Triangular membership functions
- Gaussian membership functions.

FGMDH with Fuzzy Inputs for Triangular Membership Functions

Let's consider the special case of the linear interval regression model:

$$Y = A_0Z_0+A_1Z_1+\ldots+A_nZ_n, \quad (1.2)$$

where A_i – fuzzy number of triangular shape, which is described by threes of parameters $A_i = (\underline{A_i}, a_i, \overline{A_i})$, where a_i – center of the interval, $\overline{A_i}$ – its upper border, $\underline{A_i}$ - its lower border.

Current task contains the case of symmetrical membership function for parameters A_i, so they can be described via pair of parameters (a_i , c_i).

$\underline{A_i} = a_i - c_i$, $\overline{A_i} = a_i + c_i$, c_i – interval width, $c_i \geq 0$,

Z_i – also fuzzy numbers of triangular shape, which are defined by parameters $(\underline{Z_i}, \breve{Z}_i, \overline{Z_i})$, $\underline{Z_i}$ - lower border, $\breve{Z}_i$ - center, $\overline{Z_i}$ - upper border of fuzzy number.

Then *Y* is a fuzzy number, whose parameters are defined as follows:

Center of the interval:

$$\breve{y} = \sum a_i * \breve{Z}_i .$$

Deviation in the left part of the membership function:

$$\breve{y} - \underline{y} = \sum (a_i * (\breve{Z}_i - \underline{Z}_i) + c_i \left| \breve{Z}_i \right|).$$

The lower border of the interval:

$$\underline{y} = \sum (a_i * \underline{Z}_i - c_i \left| \breve{Z}_i \right|).$$

The upper border of the interval:

$$\overline{y} = \sum (a_i * \overline{Z}_i + c_i \left| \breve{Z}_i \right|).$$

For the interval model to be correct, the real value of input variable Y should lay in the interval got by the method It can be described in such a way:

$$\begin{cases} \sum (a_i * \underline{Z}_{ik} - c_i \left| \breve{Z}_{ik} \right|) \le y_k, \\ \sum (a_i * \overline{Z}_{ki} + c_i \left| \breve{Z}_{ik} \right|) \ge y_k, k = \overline{1,M}. \end{cases}$$

Where Z_k=[Z_k]$_i$ is input training sample, y_k –known output values, $k = \overline{1,M}$, M – number of observation points.

So, the general requirements to estimation linear interval model are to find such values of parameters (a_i, c_i) of fuzzy coefficients, which enable:

a) Observed values y_k lay in estimation interval for Y_k ;

b) Total width of estimation interval is minimal.

These requirements can be redefined as the following task of linear programming:

$$\min_{a_i, c_i} \sum_{k=1}^{M} (\sum (a_i * \overline{Z}_i + c_i \left| \breve{Z}_i \right|) - \sum (a_i * \underline{Z}_i - c_i \left| \breve{Z}_i \right|)) \tag{1.3}$$

under constraints:

$$\begin{cases} \sum (a_i * \underline{Z}_{ik} - c_i \left| \breve{Z}_{ik} \right|) \le y_k, \\ \sum (a_i * \overline{Z}_{ki} + c_i \left| \breve{Z}_{ik} \right|) \ge y_k, k = \overline{1,M}. \end{cases} \tag{1.4}$$

Formalized problem formulation in case of triangular membership functions

Let's consider partial description

$$f(x_i, x_j) = A_0 + A_1 x_i + A_2 x_j + A_3 x_i x_j + A_4 x_i^2 + A_5 x_j^2. \tag{1.5}$$

Rewriting it in accordance with the model (1.1) needs such substitution $z_0 = 1$, $z_1 = x_i$,

$z_2 = x_j$, $z_3 = x_i x_j$, $z_4 = x_i^2$, $z_5 = x_j^2$.

Then math model (1.3)-(1.4) will take the form

$$\min_{a_i,c_i} (2Mc_0 + a_1\sum_{k=1}^{M}(\bar{x}_{ik} - \underline{x}_{ik}) + 2c_1\sum_{k=1}^{M}|\breve{x}_{ik}| + a_2\sum_{k=1}^{M}(\bar{x}_{jk} - \underline{x}_{jk}) + 2c_2\sum_{k=1}^{M}|\breve{x}_{jk}| +$$
$$+a_3\sum_{k=1}^{M}(|\breve{x}_{ik}|(\bar{x}_{jk} - \underline{x}_{jk}) + |\breve{x}_{jk}|(\bar{x}_{ik} - \underline{x}_{ik})) + 2c_3\sum_{k=1}^{M}|\breve{x}_{ik}\breve{x}_{jk}| + 2a_4\sum_{k=1}^{M}|\breve{x}_{ik}|(\bar{x}_{ik} - \underline{x}_{ik}) + \quad (1.6)$$
$$+2c_4\sum_{k=1}^{M}\breve{x}_{ik}^2 + 2a_5\sum_{k=1}^{M}|\breve{x}_{jk}|(\bar{x}_{jk} - \underline{x}_{jk}) + 2c_5\sum_{k=1}^{M}\breve{x}_{jk}^2)$$

with the following conditions:

$$a_0 + a_1\underline{x}_{ik} + a_2\underline{x}_{jk} + a_3(-|\breve{x}_{ik}|(\breve{x}_{jk} - \underline{x}_{jk}) - |\breve{x}_{jk}|(\breve{x}_{ik} - \underline{x}_{ik}) + \breve{x}_{ik}\breve{x}_{jk}) +$$
$$+a_4(-2|\breve{x}_{ik}|(\breve{x}_{ik} - \underline{x}_{ik}) + \breve{x}_{ik}^{\,2}) + a_5(2|\breve{x}_{jk}|(\breve{x}_{jk} - \underline{x}_{jk}) + \breve{x}_{jk}^{\,2}) - c_0 - c_1|\breve{x}_{ik}| -$$
$$-c_2|\breve{x}_{jk}| - c_3|\breve{x}_{ik}\breve{x}_{jk}| - c_4\breve{x}_{ik}^2 - c_5\breve{x}_{jk}^2 \le y_k, \quad (1.7)$$
$$a_0 + a_1\bar{x}_{ik} + a_2\bar{x}_{jk} + a_3(|\breve{x}_{ik}|(\bar{x}_{jk} - \breve{x}_{jk}) + |\breve{x}_{jk}|(\bar{x}_{ik} - \breve{x}_{ik}) - \breve{x}_{ik}\breve{x}_{jk}) + a_4(2|\breve{x}_{ik}|(\bar{x}_{ik} -$$
$$-\breve{x}_{ik}) - \breve{x}_{ik}^{\,2}) + a_5(2|\breve{x}_{jk}|(\bar{x}_{jk} - \breve{x}_{jk}) - \breve{x}_{jk}^{\,2}) + c_0 + c_1|\breve{x}_{ik}| + c_2|\breve{x}_{jk}| + c_3|\breve{x}_{ik}\breve{x}_{jk}| +$$
$$c_4\breve{x}_{ik}^2 + c_5\breve{x}_{jk}^2 \ge y_k.$$

$$c_l \ge 0,\ l = \overline{0,5}. \quad (1.8)$$

As we can see, this is the linear programming problem, but there are still no limitations for non-negativity of variables a_i, so we may pass to a dual problem, introducing dual variables $\{\delta_k\}$ and $\{\delta_{k+M}\}$.

Write down dual problem:

$$\max(\sum_{k=1}^{M} y_k \cdot \delta_{k+M} - \sum_{k=1}^{M} y_k \cdot \delta_k). \quad (1.9)$$

Under constraints:

$$\sum_{k=1}^{M}\delta_{k+M} - \sum_{k=1}^{M}\delta_k = 0.$$
$$\sum_{k=1}^{M}\bar{x}_{ik} \cdot \delta_{k+M} - \sum_{k=1}^{M}\underline{x}_{ik} \cdot \delta_k = \sum_{k=1}^{M}(\bar{x}_{ik} - \underline{x}_{ik}), \quad (2.0)$$
$$\sum_{k=1}^{M}\bar{x}_{jk} \cdot \delta_{k+M} - \sum_{k=1}^{M}\underline{x}_{jk} \cdot \delta_k = \sum_{k=1}^{M}(\bar{x}_{jk} - \underline{x}_{jk}).$$

And several constraints inequalities corresponding to variables C_i are not presented here for brevity.

It was proved that the dual LP problem is always solvable as well as the primary LP problem (1.6)-(1.8).

Experimental Investigations of FGMDH and Comparison with Fuzzy Neural Networks

The numerous experimental investigations of the suggested method-FGMDH with fuzzy inputs were carried out at the problem of forecasting stock prices of shares at Russian stock market RTS and comparison of efficiency of FGMDH and fuzzy neural networks. Some of the obtained results are presented below.

Experiment 1. "LUKOIL" stock price forecasting.

1. Stock price forecasting using fuzzy group method of data handling with triangular MF and Gaussian MF:

a) Forecasting based on previous data about stock prices at the period: 01.12./2005 - 20.12.2005.

The following results were obtained which are presented at the table 1 and Fig. 1

Table 1. Experiment 1 results using FGMDH with fuzzy data and triangular MF

Date	Real Values	Lower Border	Center	Upper Border	Deviation
01.12.2005	58.1	57.5984	58.30609	59.05755	0.2060885
02.12.2005	58.7	58.22393	58.70375	59.29865	0.00375055
05.12.2005	59.4	58.35602	59.10925	59.85003	0.290755
06.12.2005	59	57.94641	58.70476	59.55968	0.2952399
07.12.2005	59.85	58.78311	59.60075	60.4198	0.2492475
08.12.2005	59.6	58.6168	59.42533	60.16382	0.17467435
09.12.2005	59.9	59.43436	60.13189	60.94283	0.2318917
12.12.2005	60.65	59.70792	60.46069	61.22393	0.1893122
13.12.2005	60.65	60.10314	60.77584	61.39809	0.1258443
14.12.2005	61.15	60.63159	61.44362	62.25126	0.2936235
15.12.2005	60.25	59.7932	60.44087	61.0663	0.1908689
16.12.2005	61	60.15238	60.88398	61.60982	0.116021
19.12.2005	61.01	60.58471	61.24406	61.85153	0.234062
20.12.2005	60.7	60.02635	60.71058	61.35923	0.01058345

MSE = 0.215421

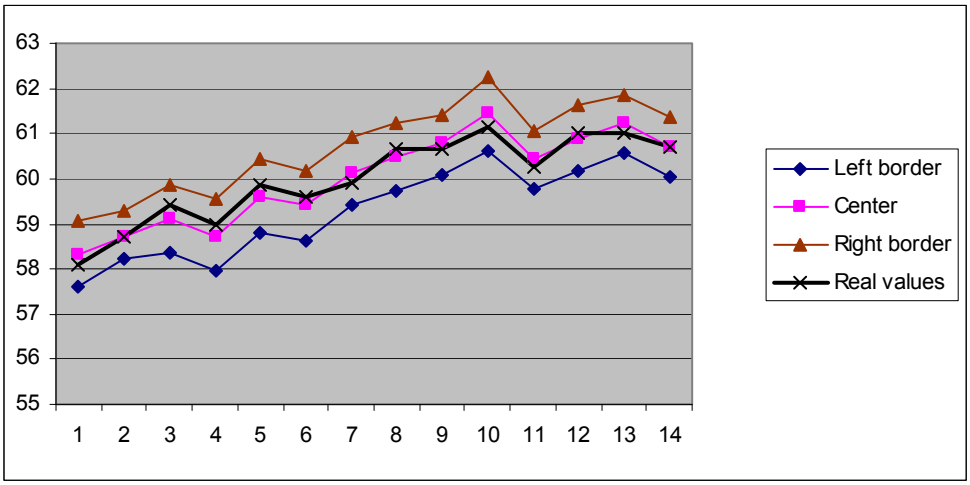


Fig.1. Experiment 3 results using FGMDH with fuzzy input data and triangular MF

Experiment 2. **"LUKOIL" stock price forecasting** based on previous data about stock prices of leading Russian energetic companies for the same period:

EESR – shares of "РАО ЕЭС России" joint-stock company, YUKO – shares of "ЮКОС" joint-stock company,

SNGSP – privileged shares of "Сургутнефтегаз" joint-stock company, SNGS – common shares of "Сургутнефтегаз" joint-stock company.

The following results were obtained:

MSE = 0.115072

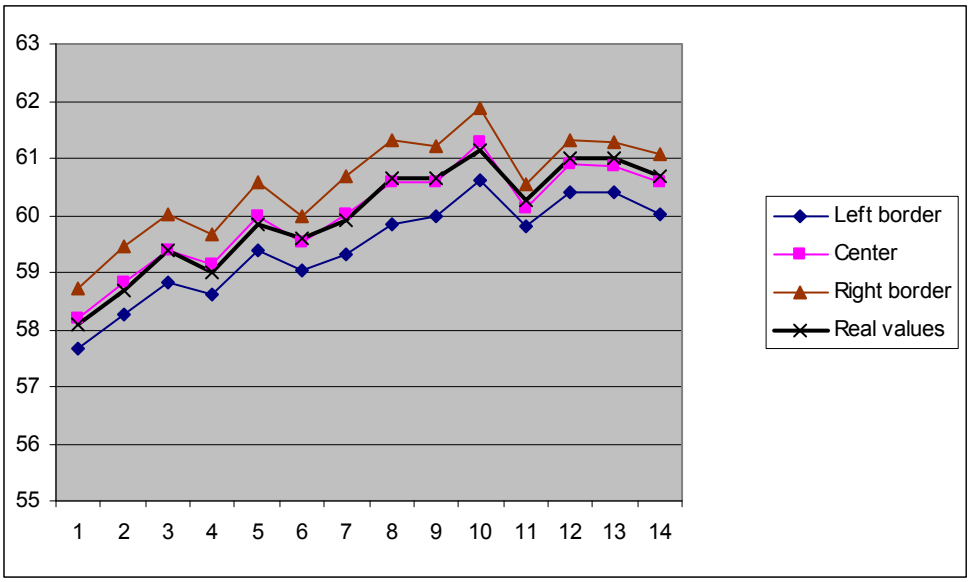


Fig. 2. Experiment 2 results using FGMDH with fuzzy data and Gaussian MF

Experiment 3. Stock price forecasting using fuzzy neural nets with Mamdani and Tsukamoto algorithm.

267 everyday indexes of stock prices during period from 1.04.2005 to 30.12.2005 were used for neural net training. The following results were obtained:

Table 2. Experiment 3 results using FNN

Date	Real Value	Mamdani with Gaussian MF	Mamdani with Triangular MF	Tsukamoto with Gaussian MF	Tsukamoto with Triangular MF
01.12.2005	58.1	58.23	58.41	58.37	58.48
02.12.2005	58.7	58.54	58.46	58.47	58.37
05.12.2005	59.4	59.14	59.11	59.1	59.05
06.12.2005	59	59.11	59.15	59.25	59.27
07.12.2005	59.85	59.97	60.1	60.19	60.23
08.12.2005	59.6	59.416	59.313	59.37	59.23
09.12.2005	59.9	60.12	60.22	60.27	60.32
12.12.2005	60.65	60.5	60.45	60.48	60.4
13.12.2005	60.65	60.54	60.31	60.42	60.28
14.12.2005	61.15	61.32	61.39	61.4	61.42
15.12.2005	60.25	60.1	59.99	60.06	59.97
16.12.2005	61	61.2	61.25	61.22	61.27
19.12.2005	61.01	61.24	61.34	61.28	61.34
20.12.2005	60.7	60.54	60.44	60.48	60.38
MSE		***0.18046***	***0.28112***	***0.268013***	***0. 10093***

As experiment 3 results show, forecasting using FNN with Mamdani controller with Gaussian MF was the best place, Tsukamoto controller with Gaussian MF takes the second place.

Forecasting using FGMDH with fuzzy input data gives best results when using Gaussian MF.

Table 3. Summarizing forecasting results for FGMDH and FNN with Gaussian MF:

	Mamdani controller	Tsukamoto Controller	FGMDH With fuzzy inputs (4 input variables)	FGMDH with fuzzy inputs (previous values on input)
MSD	***0.18046***	***0.268013***	***0.115072***	***0.094002***

For triangular MF:

Table 4. Summarizing forecasting results for FGMDH and FNN with triangular MF

	Mamdani controller	Tsukamoto Controller	FGMDH with fuzzy inputs (4 input variables)	FGMDH with fuzzy inputs (previous values on input)
MSD	***0.28112***	***0.34443***	***0.210865***	***0.215421***

As we can see from tables 2 and 3, the best MSD are given by FGMDH with fuzzy inputs, and this method also allows to build interval estimation of forecasted value. Gaussian MF is more accurate than triangular MF in FGMDH with fuzzy inputs, and also in FNN.

Experiment 4. RTS index forecasting (opening price)

Forecasting of RTS index (opening price) using fuzzy neural nets with Mamdani and Tsukamoto algorithm.

267 everyday indexes of stock prices during period from 1.04.2005 to 30.12.2005 were used for neural net training.

The following final results were obtained: which are presented at the table 5 (for Gaussian MF) and table 6 (triangular MF)

As experiment 4 results show, forecasting using FNN with Mamdani controller with Gaussian MF was the best, Mamdani controller with triangular MF is on the second place.

Forecasting using FGMDH with fuzzy input data gives best results when using Gaussian MF and using 5 input variables:

Table 5. Forecasting results of experiment 4 for FGMDH and FNN with Gaussian MF

	Mamdani Controller	Tsukamoto Controller	FGMDH with fuzzy inputs (5 input variables)	FGMDH with fuzzy inputs (previous values on the input)
MSD	***3.692981***	***7.002467***	***2.1151183***	***2.886697***
MAPE. %	***0.256091***	***0.318056***	***0.179447***	***0.256547***

Table 6. Forecasting results of experiment 4 for FGMDH and FNN with triangular MF

	Mamdani Controller	Tsukamoto Controller	FGMDH with fuzzy inputs (5 input variables)	FGMDH with fuzzy inputs (previous values of the input variable)
MSD	***3.34179***	***5.119318***	***4.717268***	***4.977901***
MAPE. %	***0.318056***	***0.419659***	***0.40437***	***0.415434***

Experiment 5. RTS 2 index forecasting (opening price)

Table 7. Forecasting results of experiment 5 for FGMDH and FNN with Gaussian MF

	Mamdani controller	Tsukamoto Controller	FGMDH with fuzzy inputs (4 input variables)	FGMDH with fuzzy inputs (previous input values)
MSE	***0.18046***	***0.268013***	***0.115072***	***0.094002***

Table 8. Forecasting results of experiment 5 for FGMDH and FNN with triangular MF

	Mamdani controller	Tsukamoto Controller	FGMDH with fuzzy inputs (4 input variables)	FGMDH with fuzzy inputs (previous values on input)
MSE	***0.28112***	***0.34443***	***0.210865***	***0.215421***

As current experiment results show, forecasting using FGMDH with fuzzy input data using Gaussian membership function was the best, fuzzy Mamdani controller with triangular and Gaussian MF takes the second place.

Best MSD are given by FGMDH with fuzzy inputs, and this method also allows to build interval estimation of forecasted value. Gaussian MF is more accurate than triangular MF in FGMDH with fuzzy inputs, and also in FNN.

Conclusion

In this paper new method of inductive modeling FGMDH with fuzzy inputs is described. This method represents the development of fuzzy GMDH when information is fuzzy and given in the form of uncertainty intervals. The mathematical model was constructed and corresponding algorithm was elaborated. The experimental results of application of the suggested method in the forecasting of market index and stock prices and their comparison with FNN are presented and discussed. The main advantages of the suggested method are following:

- It operates with fuzzy and uncertain input information and constructs the fuzzy model;

- The constructed model has minimal possible total width and in this sense it is optimal;

- For finding optimal model we solve corresponding linear programming problem which is always solvable.

Experiments have shown that FNN and FGMDH with fuzzy inputs are effective methods for forecasting stock prices and market indexes.

The best results were obtained in cases of Gaussian MF both for fuzzy neural nets and FGMDH.

Besides accurate results, which are in many cases better than the results of FNN, FGMDH also has the following advantages: possibility to work with arbitrary number of fuzzy input variables.

Acknowledgements

The paper is published with financial support by the project ITHEA XXI of the Institute of Information Theories and Applications FOI ITHEA Bulgaria www.ithea.org and the Association of Developers and Users of Intelligent Systems ADUIS Ukraine www.aduis.com.ua.

Bibliography

[Zaychenko Yu, 2006] Zaychenko Yu. "The Fuzzy Group Method of Data Handling and Its Application for Economical Processes Forecasting" - *Scientific Inquiry,* - Vol. 7, No.1, June, 2006 - p.83-96.

[Зайченко, 2001]. Ю.П. Зайченко, І.О. Заєць. Синтез та адаптація нечітких прогнозуючих моделей на основі методу самоорганізації //Наукові вісті НТУУ КПІ, №3, 2001.-с. 34-41.

[Зайченко, 2003] Ю.П. Зайченко. Нечеткий метод индуктивного моделирования в задачах прогнозирования макроэкономических показателей. // Системні дослідження та інформаційні технології, № 3, 2003.-с.25-45.

Authors' Information

Yuriy Zaychenko – P*rofessor NTUU "Kiev Polytechnic Institute", Institute of Applied System Analysis, Politehnicheskaya 14, Kiev, Ukraine, phone: 8-044-2418693, e-mail: baskervil@voliacable.com*

ЭКОНОМИЧЕСКИЕ ГИПОТЕЗЫ И ДВОЙСТВЕННАЯ ДИНАМИЧЕСКАЯ МОДЕЛЬ ЛЕОНТЬЕВА “ЗАТРАТЫ-ВЫПУСК”

Игорь Ляшенко, Елена Ляшенко, Андрей Онищенко

Аннотация: *В статье авторами предложены новые двойственные модели цен для динамической модели Леонтьева «затраты-выпуск», которые используют разные гипотезы относительно финансовых балансов. Параллельно к классической гипотезе о отсутствии денежных запасов используется одна из трех альтернативных: гипотеза о неизменности во времени общей стоимости капитала, гипотеза о расширении производства только за счет инфляции, а также гипотеза о коррупции при расширении производства. В зависимости от принятой гипотезы получены разные магистральные траектории цен.*

Ключевые слова*: динамическая модель Леонтьева «затраты-выпуск», двойственная динамическая модель, экономические гипотезы относительно финансовых балансов, динамика равновесных цен.*

ACM Classification Keywords: *I. Computing Methodologies, H.4.2 Information Systems Applications: Types of Systems: Decision Support*

Conference*: The paper is selected from XV[th] International Conference “Knowledge-Dialogue-Solution” KDS-2 2009, Kyiv, Ukraine, October, 2009.*

Введение

Изучение отраслевой структуры экономики в натуральных показателях приводит к необходимости перехода к соответствующей стоимостной структуре – ценовым показателям. Определение динамики цен с учетом разнообразных экономических гипотез позволяет проанализировать существующую систему цен, сравнить ее с расчетными значениями, установить взаимосвязь основных показателей динамики натуральных и стоимостных моделей.

Постановка задачи

В развитии экономико-математического моделирования, как в теоретическом так и практическом аспектах, особую роль сыграла модель Леонтьева [1, 2], ставшая основой линейной экономики. Математическая формализация модели имеет вид:

$$x = Ax + y, \quad x \geq 0, \tag{1}$$

где $x = (x_1, x_2, \ldots, x_n)^T$ – вектор-столбик валового выпуска продукции, $y = (y_1, y_2, \ldots, y_n)^T$ – вектор-столбик конечного продукта, $A = (a_{ij})_1^n \geq 0$ – неотрицательная матрица коэффициентов прямых производственных затрат.

Данной модели соответствует двойственная модель цен [2]:

$$p = pA + r, \quad p \geq 0, \tag{2}$$

где $p = (p_1, p_2, \ldots, p_n)$ – вектор-строка цен на продукцию, $r = (r_1, r_2, \ldots, r_n)$ – вектор-строка коэффициентов прибавочной стоимости.

Умножив соотношение (1) слева на вектор p, а соотношение (2) – справа на вектор x, приходим к двум соотношениям:

$$px = pAx + py, \quad x \geq 0, \tag{3}$$

$$px = pAx + rx, \quad p \geq 0.$$

Сравнивая полученные соотношения, приходим к условию:

$$py = rx,\ x \geq 0,\ p \geq 0\,, \tag{4}$$

левая часть которого отражает стоимость конечного продукта, а правая – суммарную прибавочную стоимость. Таким образом, соотношение (4) является гипотезой отсутствия денежных запасов, т.е. созданная стоимость равна использованной стоимости. Что является классической экономической гипотезой всех балансовых экономико-математических моделей.

Следующим шагом развития модели (1) стал переход к динамической модели "затраты-выпуск" [2]:

$$x = Ax + B\dot{x} + y, \quad x \geq 0, \tag{5}$$

где $\dot{x}(t) = (\dot{x}_1(t), \dot{x}_2(t), \ldots, \dot{x}_n(t))^T$ – вектор-столбик приростов производства продукции, $B = (b_{ij})_1^n \geq 0$ – неотрицательная матрица коэффициентов фондоемкости приростов производства.

Умножая соотношение (5) слева на вектор p и учитывая гипотезы (4) получаем:

$$px = pAx + pB\dot{x} + rx, \quad x \geq 0, \quad p \geq 0. \tag{6}$$

Экономический смысл слагаемого $pB\dot{x}$ заключается в отражении им стоимости капитала необходимого для обеспечения прироста производства $\dot{x}$.

В математических исследованиях переход к двойственной дифференциальной модели осуществляют с помощью постановки задачи оптимального управления и построения гамильтониана, относительно которого записываются необходимые условия оптимальности. В механике, как правило, гамильтониан отражает суммарную энергию: потенциальную и кинетическую. В случае модели (5) гамильтониан

$$H = pBx$$

отражает суммарную стоимость капитала. При этом классическая гипотеза о неизменности во времени этой стоимости формализируется следующим образом:

$$\frac{d}{dt}(pBx) = \dot{p}Bx + pB\dot{x} = 0\,. \tag{7}$$

Учитывая данный результат, соотношение (6) можно записать в виде:

$$(p - pA + \dot{p}B - r)x = 0, \quad x \geq 0, \quad p \geq 0. \tag{8}$$

Для выполнения условия (8) при любом $x \geq 0$ необходимо и достаточно, чтобы выполнялось соотношение

$$p = pA - \dot{p}B + r, \quad p \geq 0. \tag{9}$$

Полученная модель является двойственной по отношению к модели (5) при условии, что выполняются две классические гипотезы: статическая (4) и динамическая (7).

Возможно рассмотрение иных альтернативных к введенным гипотез. В частности,

$$pB\dot{x} = \dot{p}Bx, \tag{10}$$

которая отражает условие увеличения общей стоимости прироста капитала за счет инфляции стоимости общего капитала, а также

$$(1+\varepsilon)pB\dot{x} = \dot{p}Bx, \quad \varepsilon > 0, \tag{11}$$

которая вводит понятие коррупции ("отката") в объеме $\varepsilon pB\dot{x}$ при осуществлении прироста производства.

Задача данной работы состоит в обосновании новых гипотез (10) и (11), а также в исследовании магистральных траекторий валового выпуска продукции и соответствующих траекторий цен, характерных для введенных трех случаев.

Результаты

1. Магистральные траектории классических прямой и двойственной моделей Леонтьева "затраты-выпуск".

Проводя исследование прямой модели, будем искать общее решение системы линейных дифференциальных уравнений (5) как общее решение системы однородных уравнений (при $y \equiv 0$) и частное решение системы неоднородных уравнений (5) при общем значении $y(t)$.

Рассмотрим неотрицательное общее решение системы дифференциальных уравнений (5) при условии $y \equiv 0$ в виде

$$x(t) = e^{\mu t}x(0), \tag{12}$$

где μ – неотрицательный числовой параметр, $x(0) \geq 0$ – произвольный неотрицательный вектор.

Подставляя условие (12) в (5), приходим к соотношению

$$x(0) = \mu(I - A)^{-1}Bx(0). \tag{13}$$

Вводя предположение о продуктивности матрицы прямых затрат $A \geq 0$ [2], приходим к выводу, что ее корень Фробениуса $\lambda_A \in (0, 1)$ и существует обратная неотрицательная матрица $(I - A)^{-1} \geq 0$.

Перепишем соотношение (13) в виде

$$\left((I - A)^{-1}B - \lambda I\right)x(0) = 0, \quad x(0) \geq 0, \quad \mu = \lambda^{-1}. \tag{14}$$

Условие (14) имеет место лишь в случае равенства λ корню Фробениуса, а $x(0)$ правому вектору Фробениуса неотрицательной матрицы $(I - A)^{-1}B$, то есть

$$\lambda = \lambda_{(I-A)^{-1}B} > 0, \quad x(0) = x_{(I-A)^{-1}B} \geq 0. \tag{15}$$

Тогда темп роста валового выпуска продукции удовлетворяет неравенству:

$$\mu = \lambda^{-1}_{(I-A)^{-1}B} > 0. \tag{16}$$

Следующее неотрицательное частное решение системы неоднородных уравнений (5) будем искать при условии $y = y_0 e^{st}$, где s – числовой параметр, а $y_0 \geq 0$ – заданный неотрицательный вектор. Согласно общей теории дифференциальных уравнений оно также будет иметь вид

$$x_0(t) = e^{st} x_0, \quad x_0 \geq 0,$$

что позволяет перейти к соотношению

$$(I - A - sB)x_0 = y_0 \geq 0,$$

откуда следует

$$\left(I - s(I - A)^{-1} B\right)x_0 = (I - A)^{-1} y_0 \geq 0. \tag{17}$$

Соотношение (17) при условии, что матрица $s(I - A)^{-1} B \geq 0$ продуктивна, имеет неотрицательное решение

$$x_0 = \left(I - s(I - A)^{-1} B\right)^{-1} (I - A)^{-1} y_0 \geq 0. \tag{18}$$

Из условия продуктивности матрицы $s(I - A)^{-1} B \geq 0$ следует, что

$$s\lambda_{(I-A)^{-1}B} < 1,$$

то есть

$$s < \mu = \lambda_{(I-A)^{-1}B}. \tag{19}$$

Таким образом, общее неотрицательное решение системы дифференциальных уравнений (5) при $y(t) = e^{st} y_0$, $s < \mu = \lambda_{(I-A)^{-1}B}$ задается соотношением

$$x(t) = Ce^{\mu t} x_{(I-A)^{-1}B} + e^{st} (I - A - sB)^{-1} y_0, \tag{20}$$

где $C > 0$ произвольное положительное число, а μ и s удовлетворяет условию (19).

Анализ полученного решения при $t \to \infty$ указывает на то, что доминирующим является первое слагаемое. Таким образом, магистральной траекторией валового выпуска продукции является правый вектор Фробениуса матрицы $(I - A)^{-1} B$ с темпом роста μ, который равен обратному значению корня Фробениуса этой же матрицы (16).

Придерживаясь проведенной выше процедуры, проведем исследование двойственной модели цен (9), а именно определим общее неотрицательное решение соответствующей системы дифференциальных уравнений и магистральную траекторию развития. Для этого на первом этапе будем искать решение при условии $r \equiv 0$ в виде:

$$p(t) = e^{\nu t} p(0), \tag{21}$$

где ν неизвестный числовой параметр, $p(0) \geq 0$ – произвольный неотрицательный вектор.

Подставив соотношение (21) в (9) приходим к равенству

$$p(0)\left(I + \nu B(I - A)^{-1}\right) = 0, \quad p(0) \geq 0. \tag{22}$$

Последнее перепишем в виде

$$p(0)\left(B(I - A)^{-1} - \lambda I\right) = 0, \quad p(0) \geq 0, \quad \nu = -\lambda^{-1}. \tag{23}$$

Из условия (22) следует, что λ является корнем Фробениуса, а $p(0)$ – левым вектором Фробениуса неотрицательной матрицы $B(I-A)^{-1} \geq 0$, то есть

$$\lambda = \lambda_{B(I-A)^{-1}} > 0, \quad p(0) = p_{B(I-A)^{-1}} \geq 0. \tag{24}$$

Поскольку $\lambda_{B(I-A)^{-1}} = \lambda_{(I-A)^{-1}B}$, то

$$\nu = -\mu = -\lambda^{-1}_{(I-A)^{-1}B} < 0. \tag{25}$$

На следующем этапе ищем решение частное неотрицательное решение системы неоднородных уравнений (9) при условии $r(t) = e^{kt} r_0$, где k – некоторый числовой параметр, а $r_0 \geq 0$ – заданный неотрицательный вектор. Частное неотрицательное решение также будем искать в виде

$$p_0(t) = e^{kt} p_0, \quad p_0 \geq 0,$$

что приводит к соотношению

$$p_0(I - A + kB) = r_0 \geq 0,$$

откуда получаем

$$p_0\left(I + kB(I-A)^{-1}\right) = r_0(I-A)^{-1} \geq 0. \tag{26}$$

Последнее соотношение при условии, что матрица $-kB(I-A)^{-1} \geq 0$ продуктивна, имеет неотрицательное решение

$$p_0 = r_0(I-A)^{-1}\left(I + kB(I-A)^{-1}\right)^{-1} \geq 0.$$

Из продуктивности матрицы $-kB(I-A)^{-1} \geq 0$ следует, что

$$-k\lambda_{B(I-A)^{-1}} < 1,$$

то есть

$$k > \nu = -\mu = -\lambda^{-1}_{(I-A)^{-1}B}. \tag{27}$$

Таким образом, общее неотрицательное решение системы дифференциальных уравнений (9) при условии $r(t) = e^{kt} r_0 \geq 0$, $k > \nu = -\lambda^{-1}_{(I-A)^{-1}B}$ задается в виде:

$$p(t) = C_1 e^{-\mu t} p_{B(I-A)^{-1}} + e^{kt}(I - A + kB)^{-1} r_0, \tag{28}$$

где $C_1 > 0$ – произвольное положительное число, а $k > -\mu$.

Анализ полученного решения при $t \to \infty$ позволяет утверждать, что в (28) доминирующим является второе слагаемое. Отсюда приходим к выводу о существовании магистральной траектории цен – вектор r_0, который изменяется с темпом $k > -\mu$ во времени.

2. Гипотеза о расширении производства за счет инфляции.

Введенная гипотеза описывается соотношением (10). С учетом этого, модель (6) принимает вид:

$$\left(p - pA - \dot{p}B - r\right)x = 0, \quad x \geq 0, \quad p \geq 0. \tag{29}$$

Полученное равенство имеет место при любых $x \geq 0$ тогда и только тогда, когда выполняется соотношение

$$p = pA + \dot{p}B + r, \quad p \geq 0. \tag{30}$$

Полученная модель является двойственной моделью к (5) при условии, что выполняются две гипотезы: классическая гипотеза (4) и новая динамическая – (10).

Рассмотрим задачу построения магистральной траектории для двойственной модели (30). Для этого необходимо определить общее неотрицательное решение соответствующей системы дифференциальных уравнений. Общее неотрицательное решение системы (30) при $r \equiv 0$ будем искать в виде

$$p(t) = e^{\nu t} p(0), \tag{31}$$

где ν – неизвестный числовой параметр, $p(0) \geq 0$ – произвольный неотрицательный вектор.

Подставив соотношение (31) в (30) получаем

$$p(0)\left(I - \nu B(I - A)^{-1}\right) = 0, \quad p(0) \geq 0. \tag{32}$$

Уравнение (32) перепишем в виде

$$p(0)\left(B(I - A)^{-1} - \lambda I\right) = 0, \quad p(0) \geq 0, \quad \nu = \lambda^{-1}. \tag{33}$$

Из полученного результата следует, что λ является корнем Фробениуса, а $p(0)$ – левым вектором Фробениуса неотрицательной матрицы $B(I - A)^{-1} \geq 0$, то есть

$$\lambda = \lambda_{B(I-A)^{-1}} > 0, \quad p(0) = p_{B(I-A)^{-1}} \geq 0. \tag{34}$$

Таким образом, приходим к условию

$$\nu = \mu = \lambda^{-1}_{(I-A)^{-1}B} > 0. \tag{35}$$

Далее будем искать частное неотрицательное решение системы (30) при условии $r = e^{kt} r_0$, где k – некоторый числовой параметр, а $r_0 \geq 0$ – заданный неотрицательный вектор в виде

$$p_0(t) = e^{kt} p_0, \quad p_0 \geq 0,$$

что позволяет прийти к соотношению

$$p_0(0)\left(I - A - kB\right) = r_0 \geq 0,$$

из которого следует

$$p_0\left(I - kB(I - A)^{-1}\right) = r_0(I - A)^{-1} \geq 0. \tag{36}$$

Полученное соотношение при условии, что матрица $kB(I-A)^{-1} \geq 0$ продуктивна, имеет неотрицательное решение

$$p_0 = r_0(I-A)^{-1}\left(I - kB(I-A)^{-1}\right)^{-1} \geq 0.$$

Из условия продуктивности матрицы $kB(I-A)^{-1} \geq 0$ следует, что

$$k\lambda_{B(I-A)^{-1}} < 1,$$

то есть

$$k < \nu = \mu = \lambda^{-1}_{(I-A)^{-1}B}. \tag{37}$$

Таким образом, общее неотрицательное решение системы дифференциальных уравнений (30) при $r = e^{kt}r_0 \geq 0$, $k < \nu = \mu = \lambda^{-1}_{(I-A)^{-1}B}$ задается соотношением

$$p(t) = C_2 e^{\mu t} p_{B(I-A)^{-1}} + e^{kt}_{r_0}(I - A - kB)^{-1}, \tag{38}$$

где $C_2 > 0$ – произвольное положительное число, а μ и k удовлетворяют условию (37).

Поскольку при $t \to \infty$ в решении (38) доминирующим является первое слагаемое, то можно утверждать, что левый вектор Фробениуса матрицы $B(I-A)^{-1}$ является магистральной траекторией, которая возрастает с темпом μ, равному обратному значению корня Фробениуса этой же матрицы.

Проведенное исследование позволяет сделать следующий вывод: в условиях выполнения гипотезы о увеличении производства только за счет инфляции цен возможно прийти к сбалансированному состоянию экономики, когда и производство продукции, и цены возрастают с одним и тем же темпом.

3. Гипотеза о коррупции ("откат") при расширении производства.

Введенная гипотеза отражена соотношением (11). Включение ее в модель (6) приводит к следующему равенству

$$\left(p - pA - (1+\varepsilon)^{-1}\dot{p}B - r\right)x = 0, \quad x \geq 0, \quad p \geq 0. \tag{39}$$

Для выполнения полученного соотношения при любом $x \geq 0$, необходимо и достаточно, чтобы выполнялось равенство

$$p = pA + (1+\varepsilon)^{-1}\dot{p}B + r, \quad p \geq 0. \tag{40}$$

Полученная модель является двойственной по отношению к (5) при условии, что выполняются две гипотезы: классическая (4) и динамическая (11).

Рассмотрим задачу построения магистральной траектории двойственной модели (40). Учитывая идентичность алгоритма построения выше рассмотренному кратко остановимся на основных результатах.

Найдем общее неотрицательное решение системы (40) при $r \equiv 0$ в виде

$$p(t) = e^{\nu t}p(0), \quad p(0) \geq 0, \tag{41}$$

от которого переходим к соотношению

$$p(0)\left(B(I-A)^{-1}-\lambda I\right)=0,\quad p(0)\geq 0,\quad \nu=(1+\varepsilon)\lambda^{-1}. \tag{42}$$

Из последнего следует, что λ является корнем Фробениуса, а $p(0)$ – левым вектором Фробениуса неотрицательной матрицы $B(I-A)^{-1}\geq 0$, то есть

$$\nu=(1+\varepsilon)\lambda_{(I-A)^{-1}B}=(1+\varepsilon)\mu>0. \tag{43}$$

Соответствующее частное неотрицательное решение системы (40) при $r(t)=e^{kt}r_0$ в виде $p_0(t)=e^{kt}p_0,\quad p_0\geq 0,$ где k – некоторый числовой параметр, а $r_0\geq 0$ – заданный неотрицательный вектор, приводит к соотношению

$$p_0\left(I-k(1+\varepsilon)^{-1}B(I-A)^{-1}\right)=r_0(I-A)^{-1}\geq 0. \tag{44}$$

Полученное уравнение при условии продуктивности матрицы $k(1+\varepsilon)^{-1}B(I-A)^{-1}\geq 0$ имеет неотрицательное решение

$$p_0=r_0(I-A)^{-1}\left(I-k(1+\varepsilon)^{-1}B(I-A)^{-1}\right)^{-1}\geq 0.$$

Из условия продуктивности матрицы $k(1+\varepsilon)^{-1}B(I-A)^{-1}$ следует неравенство

$$k(1+\varepsilon)^{-1}\lambda_{B(I-A)^{-1}}<1,$$

то есть

$$k<\nu=(1+\varepsilon)\mu=(1+\varepsilon)\lambda^{-1}_{(I-A)^{-1}B}. \tag{45}$$

Таким образом, общее неотрицательное решение системы дифференциальных уравнений (40) при $r(t)=e^{kt}r_0\geq 0$, $k<(1+\varepsilon)\mu=(1+\varepsilon)\lambda^{-1}_{(I-A)^{-1}B}$ задается соотношением

$$p(t)=C_3e^{(1+\varepsilon)\mu t}p_{B(I-A)^{-1}}+e^{kt}_{r_0}\left(I-A-k(1+\varepsilon)^{-1}B\right)^{-1}, \tag{46}$$

где $C_3>0$ – произвольное положительное число, а μ и k удовлетворяют (45).

Анализ полученного решения при $t\to\infty$ указывает на доминирующее первое слагаемое, что позволяет рассматривать левый вектор Фробениуса матрицы $B(I-A)^{-1}$ магистральной траекторией цен, которая возрастает с темпом $(1+\varepsilon)\mu$, связанным условием (16) с соответствующим корнем Фробениуса.

Таким образом, при выполнении гипотезы о коррупции при расширении производства достигается состояние сбалансированности экономики, в которой цены возрастают более быстрым темпом, чем производство продукции.

Замечание. Соотношение темпов магистральных траекторий выпуска продукции и цен

$$\nu=(1+\varepsilon)\mu$$

дает возможность оценить уровень коррупции

$$\varepsilon = \frac{\nu}{\mu} - 1 \tag{47}$$

В качестве примера приведем следующие данные: при $\mu = 0{,}08$, $\nu = 0{,}12$ получаем $\varepsilon = 0{,}5$, а при $\mu = 0{,}08$, $\nu = 0{,}16$ – $\varepsilon = 1$.

Выводы

Исследование магистральных траекторий в случае классической гипотезы о постоянстве во времени суммарного объема капитала показало, что темп выпуска продукции определяется только матрицами прямых затрат A и фондоемкости прироста продукции B, а темп роста цен – вектором прибавочной стоимости r. Введение новой гипотезы о расширении производства только за счет инфляции цен приводит к результату сбалансированности темпов валового выпуска продукции и роста цен (выполняется строгое равенство). Вторая рассмотренная гипотеза о коррупции при расширении производства приводит к закономерным результатам: темп роста цен опережает темп роста объемов валового выпуска продукции. Наличие информации о данных характеристиках позволяет оценивать уровень коррупции.

Благодарности

Работа опубликована при финансовой поддержке проекта **ITHEA XXI** Института информационных теорий и приложений FOI ITHEA Болгария www.ithea.org и Ассоциации создателей и пользователей интеллектуальных систем ADUIS Украина www.aduis.com.ua.

Литература

[Леонтьев, 1997] Леонтьев В.В. Межотраслевая экономика. – М.: ОАО «Издательство «Экономика», 1997. – 479 с.

[Ляшенко, 2009] Ляшенко І.М., Онищенко А.М. Прямі та двоїсті балансові моделі “витрати-випуск” // Економічна кібернетика. – Донецьк, 2009. – №1-2.

[Пономаренко, 1995] Пономаренко О.І., Перестюк М.О., Бурим В.М. Основи математичної економіки. – К.: Інформтехніка. 1995. – 320 с.

Информация об авторах

И.Н. Ляшенко *– д-р физ.-мат. наук, заслуженный профессор Киевского национального университета имени Тараса Шевченко, 01601, ул. Владимирская 64, Киев, Украина*

Е.И. Ляшенко *– д-р екон.наук, профессор кафедры экономической кибернетики Киевского национального университета имени Тараса Шевченко, 01601, ул. Владимирская 64, Киев, Украина*

А.М. Онищенко *– канд. екон. наук, доцент, докторант Киевского национального университета имени Тараса Шевченко, 01601, ул. Владимирская 64, Киев, Украина, e-mail: onyshchenko@yandex.ru*

МЕТОДОЛОГИЧЕСКИЕ ОСНОВЫ ПРОГНОЗИРОВАНИЯ ИНФЛЯЦИИ

Ирина Горицына, Виктория Сатыр

***Аннотация**: Рассмотрена методология прогнозирования, основные принципы, подходы и методы расчетов динамики инфляционных процессов.*

***Ключевые слова**: инфляционный индекс, прогнозные модели, методы прогнозирования*

***ACM Classification Keywords:** I. Computing Methodologies, H.4.2 Information Systems Applications: Types of Systems: Decision Support.*

***Conference**: The paper is selected from XVth International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.*

Введение

Инфляция является одним из индикаторов макроэкономической стабильности и фактором, от которого в значительной степени зависит социально-экономическое развитие страны. Поэтому сдерживание роста инфляции и ее поддержание на благоприятном для экономики уровне является ключевой проблемой государственной экономической политики. Высокая инфляция разрушает денежную систему, провоцирует вывоз национального капитала за пределы страны, ослабляет национальную валюту, способствует ее вытеснению во внутреннем обращении иностранной валютой, подрывает возможности финансирования государственного бюджета. Достоверный прогноз инфляции является одной из предпосылок успешного проведения макроэкономической политики. Однако сделать его довольно сложно, по крайней мере, в Украине. Причиной этому - особенности экономических преобразований, происходивших в нашей стране на протяжении последних лет. За это время, в частности, произошли фундаментальные изменения в поведении населения и предприятий относительно способов сохранения денег; возникли каналы унификации наличных и безналичных средств; у людей появилась возможность выбора валюты в процессе накопления. При высоких непрогнозируемых темпах роста инфляции трудно предугадать поведение населения и субъектов хозяйствования. Эти и другие факторы затрудняют прогнозирование макроэкономических показателей, в частности, инфляции.

Применения прогнозных моделей

Прогнозируя инфляцию, следует определять сроки, на которые делается прогноз. В долгосрочном периоде основным фактором инфляции является рост количества денег в экономике. В короткосрочном прогнозе существенными могут быть немонетарные факторы или факторы косвенного монетарного влияния. Такими можно считать административные изменения цен, внешние шоки, инфляционные ожидания и т. п. Ни одна математическая модель не в состоянии дать исчерпывающего ответа на вопрос о том, какой именно будет инфляция в будущем. Однако, применение метода экспертных оценок может дать более обоснованный прогноз инфляции, как в краткосрочном, так и в долгосрочном периодах. Интерес к проблеме инфляции усиливался или ослабевал в зависимости от роста или уменьшение ее темпов, хотя она никогда не оставалась без внимания экономистов. В мировой экономической мысли

сформировалось несколько основных теоретических концепций инфляции, которые отличаются между собой по форме данного явления. По этому признаку можно выделить два направления:

1. Отождествление инфляции с любым ростом цен на товары и услуги;

2. Отождествление инфляции с обесцениванием неразменных на золото бумажных денег независимо от формы, в которой оно происходит (рост цен и заработной платы, падение курса национальной валюты, углубления товарного дефицита и др.) [Дербенцев, 2001].

Стабильное развитие любой экономики невозможно без эффективного управления инфляционным процессом и определения перспектив его развития. Результаты прогнозирования ныне являются одним из решающих факторов формирования макроэкономической политики государства. В связи с этим возникает объективная необходимость в прогнозировании будущих темпов инфляции, выявлении факторов, которые будут влиять на инфляционный рост, разработке действенных антиинфляционных мер, то есть в формировании методологии прогнозирования, которая должна определять основные принципы, подходы и методы осуществления расчетов динамики инфляционного процесса. В практике прогнозирования макроэкономических параметров чаще всего используются следующие:

1. Статистические методы (метод факторного анализа; экстраполяции тренда и регрессионного анализа).

2. Сценарный прогноз (дерево решений, экспертный метод).

Статистические методы

Как правило, методика факторного анализа используется при краткосрочном прогнозировании (предвидение сроком до 1 года). На долгосрочный период модель должна корректироваться с помощью дополнительных факторов и даже моделей, которые связывают между собой ВВП, денежную массу, скорость ее вращения и инфляцию (например, использование взаимосвязи, которые описаны уравнением обмена).

Эта модель базируется на уравнении индекса инфляции как средней взвешенной величины трех других индексов. При этом принимается во внимание общая структура потребления населения и изменение весовых соотношений между отдельными группами товаров и услуг. Основным принципом прогнозирования является принцип компромисса между максимальным ростом цен и эффектом от действия сдерживающих факторов [Министерство экономики и по вопросам европейской интеграции Украины, 1998]. Такой компромисс должен быть найден по каждой позиции, которая изучается Госкомстатом в структуре товаров и услуг. Это можно выразить следующей формулой:

$$I(t) = \sum \frac{S_j}{T_j} W_j \,, \qquad (1)$$

где W_j- весовые коэффициенты, отражающие структуру потребительской корзины, $\sum W_j = 1$; отношение коэффициента влияния стимулирующих факторов (S_j) к коэффициенту влияния сдерживающих (T_j) -- ожидаемый коэффициент роста цен по каждой группе товаров.

Данная модель основывается на предположениях относительно поведения определенных факторов инфляции в прогнозном периоде, оценке конъюнктуры внутреннего и внешнего рынков, а также предвидения изменений в макроэкономической политике (Рис.1.). Учет подобных предположений возможен лишь при условии использования метода экспертных оценок и свидетельствует о большой зависимости предложенного метода прогнозирования от субъективного фактора.

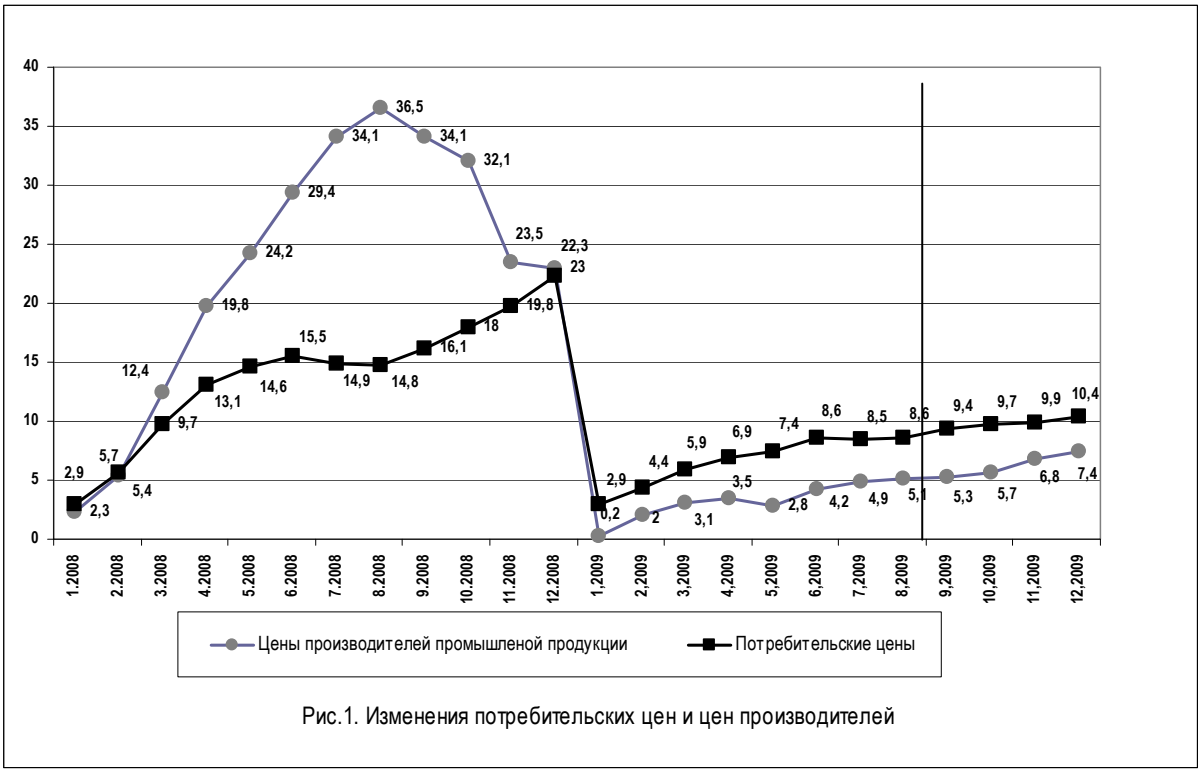


Рис.1. Изменения потребительских цен и цен производителей

Когда информация об уровне цен дана за ряд периодов, прогнозное значение факторов-аргументов может быть определено на основе экстраполяции трендов. Такую устойчивую тенденцию можно изобразить или аналитически - в виде уравнения (модели) тренда (например, в виде функции или графически).

Сущность методов прогнозной экстраполяции заключается в изучении динамики изменения экономического явления в предпрогнозном периоде и перенесения найденной закономерности на некоторый период будущего.

Однако степень реальности такого рода прогнозов и соответственно степень доверия к ним в значительной мере зависят от аргументации выбора пределов экстраполяции и стабильности «измерителей» по отношению к сущности рассматриваемого явления. Следует обратить внимание на то, что сложные объекты, как правило, не могут быть охарактеризованы одним параметром.

С помощью регрессионного анализа устанавливается степень влияния факторов на изучаемый показатель и на основе этого (в зависимости от их значения в будущем) делается расчет прогнозируемого значения функции. Такие модели, как правило, строятся без дробления индекса потребительских цен и в качестве факторов принимаются значения макроэкономических величин. Для построения регрессионной модели используют информацию, которая характеризует уровень цен по отдельным видам экономической деятельности за определенный промежуток времени (квартал, год), а также данные прогнозирования осуществляются путем подстановки в построенную модель значений независимых переменных в будущем периоде. Устанавливается корреляционная связь показателей, при которой каждому значению факторного значения могут соответствовать несколько значений результативного. Влияние того или иного фактора изменяет среднее значение уровня цен. При этом допускается, что значения коэффициентов регрессии будут сохраняться неизменным в течение всего прогнозного периода. Таким образом, получение прогнозных оценок базируется на определении прогнозных показателей факторов-аргументов.

Многофакторные модели прогнозирования индекса инфляции могут строиться на изучении влияния таких факторов, как темпы изменения реального ВВП, объемов наличных денег в обращении (M_0), объема агрегата M_2, кредитной задолженности между предприятиями, девальвации гривны, объемов кредитов НБУ, доходов населения, цен на энергоносители, объемов начисленной и невыплаченной заработной платы и т.д.Такие модели имеют следующий вид:

$$P_t = a_0 + a_1 Y_t + a_2 M_t + \ldots + a_n K_t + u_n, \quad (2)$$

где P_t - темп инфляции за период t, Y_t- объем реального ВВП за период t, M_t- объем денежной массы за период t, K_t - фактор K, осуществляющий существенное влияние на темпы инфляции в период t, u_n- погрешность, $a_0, \ldots a_n$ - коэффициенты регрессии.

На практике прогнозирование динамики инфляционных процессов означает расчет одного или нескольких индексов цен, с помощью которых измеряется инфляция. Главными из них является индекс потребительских цен и дефлятор ВВП. Если основной причиной инфляции является рост расходов, то моделирование инфляции учитывает изменение мировых цен, номинального валютного курса, уровня заработной платы и изменение косвенных налогов. В данном случае целесообразно рассчитывать индекс оптовых цен. Если же мы имеем дело с инфляцией спроса, то ее основными причинами выступают динамика денежной массы, дефицит государственного бюджета, ожидания людей относительно роста цен и существующий режим цен и валютного курса (как правило, эти показатели отражаются при расчете индекса потребительских цен и дефлятора ВВП).

С помощью данных, которые предоставляет Государственный комитет статистики Украины, можно проследить за ростом потребительских цен (с помощью индекса потребительских цен), а также проанализировать возможные причины стремительного изменения показателя инфляции в текущем году и построить прогноз на будущий год. Прежде всего, построим график изменения индекса потребительских цен по годам (прогноз на конец 2007 года на основе данных Госкомстата за январь-сентябрь 2009 года и прогноза Европейского банка реконструкции и развития):

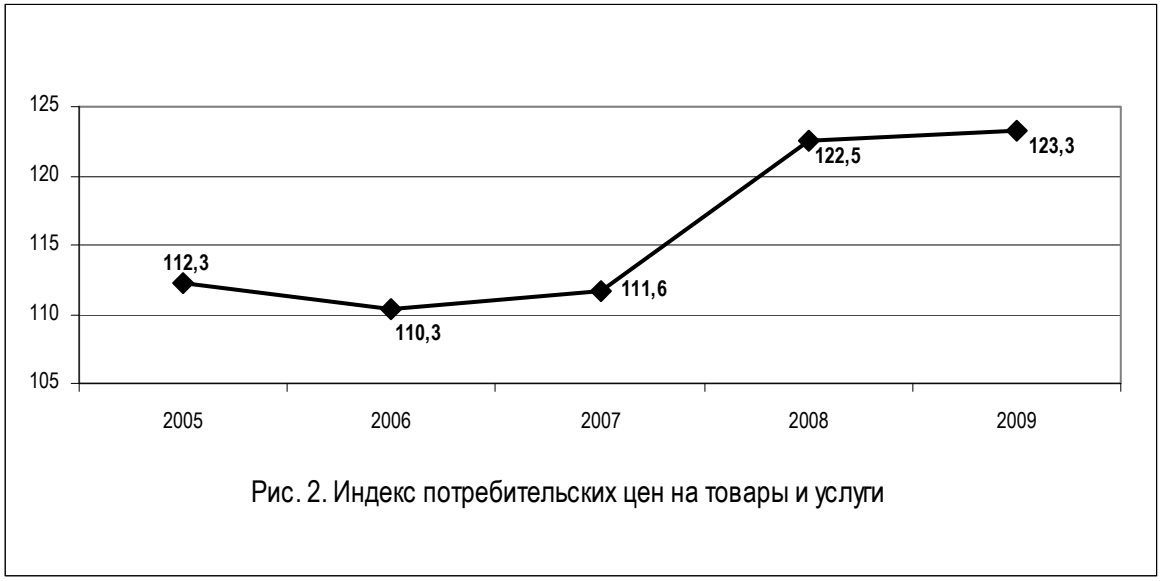


Рис. 2. Индекс потребительских цен на товары и услуги

Видно, что после незначительного снижения темпов роста индекса потребительских цен в 2005 году (он практически приблизился к «ползущему» уровню), остальные годы наблюдался значительный рост (Рис.2.). В чем же причина такого роста? Сравним показатель индекса потребительских цен и темпы роста денежной массы. Анализируя график, мы видим, что прямой зависимости между этими показателями нет, хотя некоторое влияние чрезмерного темпа роста денежной массы на изменение индекса потребительских цен оказывает (с небольшой временной задержкой, поскольку экономика обладает некоторой инерционностью).

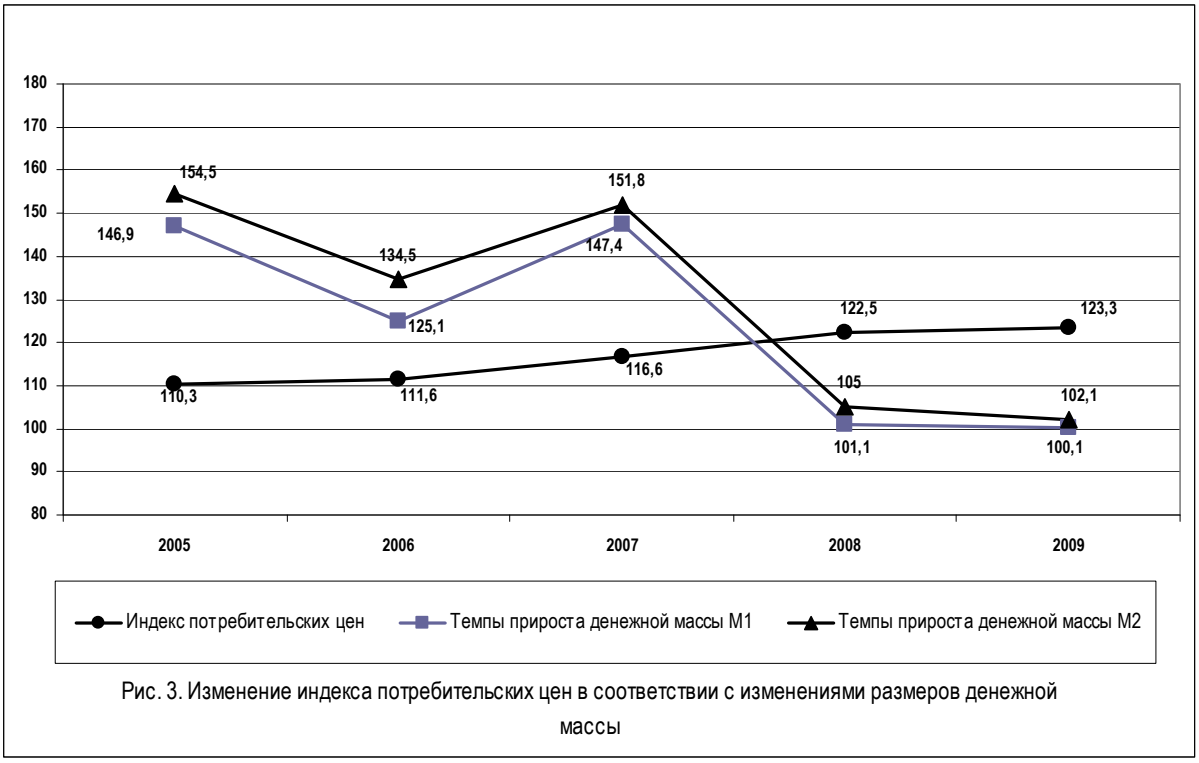


Рис. 3. Изменение индекса потребительских цен в соответствии с изменениями размеров денежной массы

Делаем вывод, что высокая инфляция в Украине вызвана не только приростом денежной массы («инфляцией спроса»). Поскольку антиинфляционная политика государства в данный момент направлена на ограничение роста денежной массы, а индекс потребительских цен имеет тенденцию к повышению с каждым месяцем, значит такая денежно-кредитная политика, направленная против классической «инфляции спроса», не имеет существенного эффекта (Рис.3.). Некоторое снижение темпов прироста индекса в летние месяцы объясняется обычными сезонными тенденциями украинской экономики.

Сценарный подход

Специфическим методом прогнозирования является Сценарный прогноз. Это - своего рода метод описания логически последовательного процесса или события, исходя из сложившейся ситуации. Описание сценариев ведется с учетом временных оценок. Основное назначение сценария - определение генеральной цели развития прогнозируемого объекта, явления и формулирование критериев для оценки верхних уровней «дерева целей». Сценарии обычно разрабатываются на основе данных предварительного прогноза и исходных материалов по развитию прогнозного объекта. К исходным

материалам следует отнести технико-экономические характеристики и показатели основных процессов производственной и научной базы для решения поставленной цели.

Сценарий - это картина, отображающая последовательное детальное решение задачи, выявление возможных препятствий, обнаружение серьезных недостатков, с тем чтобы предрешить вопрос о возможном прекращении начатых или завершении проводимых работ по прогнозируемому объекту. Сценарий, по которому должен составляться прогноз развития объекта или процессов, должен содержать в себе вопросы развития не только науки и техники, но и экономики, внешней и внутренней политики. Поэтому сценарии должны разрабатываться высококвалифицированными специалистами с учетом соответствующего профиля прогнозируемого объекта. Сценарий по своей описательности является аккумулятором исходной информации, на основе которой должна строиться вся работа по развитию прогнозируемого объекта. Поэтому сценарий в готовом виде должен быть подвергнут тщательному анализу.

В работах [Волошин, Пихотник, 1999], [Voloshin, Panchenko, 2001], [Voloshin, Panchenko, 2003], [Волошин, 2005], [Волошин, Головня, 2005], большинство которых представлялось на конференциях KDS, развивается концепция «качественного прогнозирования на основе многопараметрических зависимостей, представляемых деревом решений». Считается, что следствие определяется множеством причин, степень влияния которых на следствие определяется «субъективно» (экспертным измерением). Чем больше параметров, которые «формируют» причину, тем лучше (для адекватности модели). Однако это приводит к сложностям в анализе модели (возникает «проклятие размерности», с которым необходимо бороться). В [Волошин, Панченко, 2002] предлагаются алгоритмы последовательного анализа вариантов, позволяющие обрабатывать деревья с сотнями вершин.

Инструментарий создания прикладных систем прогнозирования

Построение прикладной системы поддержки принятия решений (СППР) сводится к выделению экспертами проблем и подпроблем (вершин дерева) и связей между ними (дуг дерева). Эксперты определяют веса (вероятности) переходов между вершинами. Допускаются нечеткие оценки экспертов с помощью логических переменных, описываемых значениями функции принадлежности (векторами действительных чисел от 0 до 1). Каждый эксперт задает три оценки – оптимистическую, реалистическую и пессимистическую, скаляризация которых осуществляется с учетом психологического типа эксперта. Тип определяется на основании психологических тестов, заложенных в систему. На основе психологических тестов определяются также коэффициенты «правдивости», «независимости», «осторожности» и т.д.

Дерево строится на основе коллективных оценок экспертов с применением метода парных сравнений. Для построения результирующего дерева применяют алгебраические методы обработки экспертной информации, в качестве расстояния между ранжировками применяют метрику Хемминга и меру несовпадений рангов объектов. Результирующее дерево определяется как медиана Кемени-Снелла или как компромисс [Волошин, 2005]. В случае задания приоритетов в нечеткой форме элементы матрицы задаются через функции принадлежности.

Для определения оптимальных путей в дереве предлагаются алгоритмы последовательного анализа вариантов [Волошин, Панченко, 2002], позволяющие обрабатывать деревья с сотнями вершин.

Дерево решений задается таблицами. Каждая таблица – это отдельный уровень дерева, каждая строка таблицы – отдельная вершина на этом уровне. Каждый элемент строки – это вероятность, с которой возможен переход из данной вершины в вершину нижнего уровня. Эти вероятности задаются функциями принадлежности, представляющие собой вектора действительных чисел от 0 до 1 любой длины. Таблица заполняется путем опроса экспертов. Существующие функции позволяют добавлять столбцы, строки, задавать словарь (который позволяет вербальным оценкам эксперта ставить в соответствие вероятности, путем задания определенных уровней), сохранять таблицы в файле, считывать таблицы из файла.

Экспертным путем задаются матрицы – результат сравнения вариантов вершин, которые могут быть включены в дерево. На основе анализа матриц определяются вершины, которые включаются в дерево и вероятности, с которыми возможен переход в них из вершин верхнего уровня. Если дерево решения можно декомпозировать на несколько поддеревьев, которые имеют одинаковые листья, вначале вычисляются вероятности этих листьев в каждом из них, а затем находятся вероятности для всего дерева в целом.

Прогнозирование индекса инфляции

Напомним прогнозные данные наших исследований: индекс инфляции в Украине за 2005-2006 гг составил 3-5% [Voloshin, Satyr, 2007]. Прогнозное значение индекса инфляции за 2007 г. в I квартале 2007г. [Волошин, 2007], принадлежало интервала 16-18% (с уровнем достоверности 0.9, поскольку индекс инфляции рассматривался как нечеткий параметр, то максимум функции принадлежности - "наиболее вероятно" значение - равнялось 17.3%). Национальный банк Украины прогнозировал инфляцию в Украине в 2007 г. на уровне 7%, правительство - 8%, официальный уровень инфляции за 2007 г. составил 16.6%. На 2008 г. нами получен прогноз в 28-30% [Волошин, 2008] (уровень достоверности 0.95, "наиболее вероятно" значение - 29.3%), в бюджете было заложено 9.6%, официальные прогнозы давали значение 11-13%. Министерство экономики прогнозирует, что по итогам текущего года инфляция составит 21,9%. Наш прогноз- 20%-24%, с уровнем достоверности 0.9, 21%-23% с уровнем достоверности 0.85 с максимумом функции принадлежности на этих интервалах 22,4%.

Выводы

Методология прогнозирования включает основные принципы, подходы и методы осуществления прогнозных расчетов динамики инфляционного процесса. В связи с этим, сформулированы основные принципы прогнозирования инфляции (целенаправленности, системности, научной обоснованности, адекватности и альтернативности) и определены основные подходы и методы, которые можно применять при разработке инфляционных прогнозов (факторный анализ, регрессионный анализ, метод экстраполяции тренда и конечно метод экспертных оценок).

Благодарности

Авторы благодарны проф. Волошину А.Ф. за ценные консультации при написании статьи.

Работа опубликована при финансовой поддержке проекта **ITHEA XXI** Института информационных теорий и приложений FOI ITHEA Болгария www.ithea.org и Ассоциации создателей и пользователей интеллектуальных систем ADUIS Украина www.aduis.com.ua.

Список литературы

[Voloshin, 2001] Voloshin O.F., Panchenko M.V. The Forecasting of Stable Processes by a Tree Solution Method using a Pairwise Comparison Method for Analysis of Expert Information. //Труды международной конференции «KDS-2001», Том 1, Санкт-Петербург, 2001- С.50-53 (англ.).

[Voloshin, 2003] Voloshin O.F., Panchenko M.V. The System of Quality Prediction on the Basis of a Fuzzy Data and Psychografphy of the Experts // Int. Journal “Information Theories & Applications", 2003, Vol.10, №3- P.261-265.

[Voloshyn, 2007] Voloshyn A, Satyr V.The inflation index prognosis based on the method of decision-making “tree”// Int. Journal “Informations Theories & Applications”, 2007, vol.14, N1.-P.63-67.

[Волошин, 1999] Волошин А., Пихотник Е. Экспертная система прогнозирования курса гривны. //"Искусственный интеллект", 1999, №2. С.354-359 (укр.).

[Волошин, 2002] Волошин А.Ф., Панченко М.В. Экспертная система качественного оценивания на основе многопараметрических зависимостей. //Проблемы математических машин и систем, 2002, №2-. С.83-89 (укр.).

[Волошин, 2005] Волошин А.Ф. О проблемах принятия решений в социально-экономических системах //Труды конференции «KDS-2005», Том 1, София, 2005- С.205-212.

[Волошин, 2006] Волошин А.Ф. Системы поддержки принятия решений как персональный интеллектуальный инструментарий лица, принимающего решение //Труды конференции «KDS-2006», София, 2006- С.149-153.

[Волошин, 2007] Волошин А., Сатыр В. Проблемы прогнозирования экономических макропараметров// Труды конференции «KDS-2007», София, 2007, том 1.-С.264-269.

[Волошин, Головня, 2005] Волошин А.Ф., Головня В.М. Система качественного прогнозирования на основе нечетких данных и психографии экспертов //Труды конференции «KDS-2005», Том 1, София, 2005- С.237-243.

[Дербенцев, 2001] Дербенцев В. Зависимость инфляционных процессов от технологической структуры экономики / / Банковское дело.- 2001.- № 2.- С. 57-60.

[Министерство экономики и по вопросам европейской интеграции Украины, 1998] Методика прогнозирования макроэкономических показателей / Министерство экономики и по вопросам европейской интеграции Украины - К. - 1998.- 115с.

Authors' Information

Горицына Ирина*–старший научный сотрудник, Киевский национальный университет имени Тараса Шевченко, кандидат экономических наук. Киев, Украина; e-mail: goritsyna@mail.ru*

Сатыр Виктория*- аспирант, Киевский национальный университет имени Тараса Шевченко. Киев Украина; e-mail: vsatyr@mail.ru*

ОБ РАСПРЕДЕЛЕНИИ ИНВЕСТИЦИЙ В ЭКОНОМИКО-СОЦИАЛЬНОЙ СФЕРЕ

Марина Коробова

Аннотация: *Рассматривается проблема оптимального распределения инвестиций между экономическим и социальным направлениями, которая сводиться к задаче оптимального управления. Находятся оптимальные траектории динамики социально-экономической системы с учетом соответствующих оптимальных капиталовложений. Таким образом, предлагается концепция для принятия оптимальных управленческих решений по инвестированию на макро- и микроуровнях.*

Ключевые слова: *социально-экономическая система, инвестиции, возобновимый ресурс, управление, принятие решений.*

ACM Classification Keywords: *I. Computing Methodologies, H.4.2 Information Systems Applications: Types of Systems: Decision Support.*

Conference*: The paper is selected from XVth International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.*

Введение

Особенностью современного этапа хозяйственного развития является формирование концепции о тесной взаимосвязи между экономическим и экологическим благополучием.

В наше время эколого-экономическая проблематика определяет не только эффективность функционирования всех видов и форм хозяйственной деятельности, но и принципиальные условия нормального функционирования каждого человека. Это включает, во-первых, эффективное использование экономикой природных ресурсов, а во-вторых, отыскание и обоснование методов предотвращения и ликвидации ущерба от загрязнения окружающей среды. Эти проблемы должны решаться на основе закономерностей естественноисторического характера, а также с учетом постоянно меняющихся потребностей общества.

Развитие экологической экономики тесно связано с приоритетами внутренней и внешней политики государства. Они должны обеспечить природо- и ресурсосберегающие национальные интересы. Очевидно, что исчерпание природных ресурсов является побочным результатом экономической деятельности. Поэтому о восстановлении ресурсов можно говорить как об отдельном производственном секторе, объемы «производства» в котором целесообразно было бы максимизировать. В этой связи процесс инвестиционных вложений на развитие экономики обязательно должен сопровождаться выделением части инвестиций на восстановление ресурсов. В частности, природный ресурс можно рассматривать как материальное обеспечение трудового ресурса (природный и трудовой ресурс являются взаимодополняющими). В таком случае есть основания говорить не о эколого-экономическом взаимодействии, а о взаимодействии экономики с социальной сферой. Ниже предложено модель оптимального распределения инвестиций между сектором материального производства и возобновлением трудовых ресурсов, причем последнее будет трактоваться как социальные капиталовложения.

Постановка задачи

На макроуровне социально-экономическую систему можно рассматривать как двухсекторную экономику (каждый сектор производит свой единственный продукт), где основным сектором является сектор материального производства, а вспомогательным сектором – состояние природного (трудового) ресурса.

Воспользуемся описанной в [Григорків, 1998], [Коробова, 2001] методикой построения и исследования такой системы.

Пусть K – основные производственные фонды (капитал); $\mu \in (0, 1)$ коэффициент износа фондов; R – объем восстановленных ресурсов (трудовых); R^* – значение показателя состояния невозмущенных ресурсов; c, $c > 0$ – коэффициент естественного восстановления ресурсов; I – общая величина инвестиций; I_K – инвестиции, вложенные в развитие экономики; I_R – инвестиции, вложенные в восстановление ресурсов (социальную сферу); ξ, $\xi > 0$ – количество единиц ресурса, необходимое для создания единицы инвестиций I_K; u_1, u_2 – весовые коэффициенты соответственно инвестиций I_K и I_R; v_1, v_2 – весовые коэффициенты компонент K и R состояния системы.

Поставим перед инвестором следующую задачу: распределить I единиц инвестиций в течение заданного промежутка времени $[0, T]$ так, чтобы в конце данного периода достичь максимально возможной величины капитала K и минимально возможной величины отклонения ресурса от невозмущенного состояния R^*, и при этом минимизировать объем инвестиций, вложенных в развитие экономики и на воспроизводство ресурсов (трудовых). Для сведения этой двухкритериальной задачи к однокритериальной, применяется метод аддитивной свертки [Волошин, 2006].

Математическую модель такой задачи можно записать в виде:

$$\begin{cases} \int\limits_0^T (-u_1 I_K - u_2 I_R)\,dt + v_1 K(T) - v_2 (R^* - R(T))^2 \to \max\limits_{I_K, I_R}, \\ \dot{K} = I_K - \mu K, \\ \dot{R} = c(R^* - R) + I_R - \xi I_K, \\ K(0) = K^{(0)}, \quad R(0) = R^{(0)}, \\ I_K + I_R \le I, \quad I_K \ge 0, \quad I_R \ge 0. \end{cases} \tag{1}$$

Итак, имеем систему дифференциальных уравнений, первое из которых отображает в стоимостной форме рост капитала, а второе – динамику ресурсов. Считаются известными исходное состояние $(K^{(0)}, R^{(0)})$ системы и ограничения на инвестиции I_K и I_R.

Сведение исходной задачи оптимального управления к задаче линейного программирования

Поскольку модель (1) является задачей оптимального управления, то для ее исследования используем принцип максимума Понтрягина [Кротов, 1990].

Построим функцию Гамильтона:

$$H(t, K, R, I_K, I_R, \psi_1, \psi_2) = \psi_1(I_K - \mu K) + \psi_2(c(R^* - R) + I_R - \xi I_K) + (-u_1 I_K - u_2 I_R), \tag{2}$$

и решим следующую задачу:

$$H(t, K, R, I_K, I_R, \psi_1, \psi_2) \to \max\limits_{I_K, I_R}.$$

Очевидно, эта задача эквивалентна такой задаче:

$$L(I_K, I_R) = (\psi_1 - \xi\psi_2 - u_1) I_K + (\psi_2 - u_2) I_R \to \max\limits_{I_K, I_R},$$

которая, в свою очередь, является задачей линейного программирования вида:

$$\begin{cases} c_1 I_K + c_2 I_R \to \max, \\ I_K + I_R \le I, \\ I_K \ge 0, \ I_R \ge 0, \end{cases} \tag{3}$$

где $c_1 = \psi_1 - \xi\psi_2 - u_1$, $c_2 = \psi_2 - u_2$.

Теперь перейдем к исследованию вариантов решения задачи (3), которое обозначим $(I_K^*, I_R^*))$.

Анализ значений коэффициентов целевой функции в задаче линейного программирования

Отметим, что задача (3) имеет реальный смысл лишь при $(c_1^2 + c_2^2) \neq 0$. В случае, когда $c_1 = c_2 = 0$ в качестве решения (3) можно выбрать произвольную допустимую точку, в частности (0, 0). Анализ поведения градиента $c = (c_1, c_2)$ целевой функции приводит нас к следующим ситуациям:

1) $c_1 = 0,\ c_2 > 0$; 2) $c_1 = 0,\ c_2 < 0$; 3) $c_1 > 0,\ c_2 = 0$; 4) $c_1 < 0,\ c_2 = 0$;

5) $c_1 < 0,\ c_2 < 0$; 6) $c_1 < 0,\ c_2 > 0$; 7) $c_1 > 0,\ c_2 > 0$; 8) $c_1 > 0,\ c_2 < 0$.

В зависимости от того, каковы значения коэффициентов c_1, c_2, имеем следующие варианты выбора оптимального распределения капиталовложений:

$$\begin{cases} 1)\ I_K^* = 0,\ I_R^* = I; & 2)\ I_K^* \in [0, I],\ I_R^* = 0; & 3)\ I_K^* = I,\ I_R^* = 0; \\ 4)\ I_K^* = 0,\ I_R^* \in [0, I]; & 5)\ I_K^* = 0,\ I_R^* = 0; & \\ 6)\ I_K^* = 0,\ I_R^* = I; & 7)\ I_K^* = I,\ I_R^* = 0. & \end{cases}$$

Ситуацию 8) необходимо исследовать отдельно.

8.1) Рассмотрим линию уровня $c_1 I_K + c_2 I_R = 0$. Если $c_1 \setminus c_2 = 1$ (или $(c_1 - c_2) = 0$), то линия уровня параллельна прямой $I_K + I_R = I$, то есть $I_K^* \in [0, I]$, $I_R^* = I - I_K^*$.

8.2) Пусть α – это угол наклона линии уровня к прямой $I_R = 0$ ($\frac{\pi}{2} < \alpha < \pi$), тогда $\frac{c_1}{c_2} = \mathrm{tg}\alpha$. Поскольку на промежутке $\left(\frac{\pi}{2}, \pi\right)$ тангенс возрастает, то из условия $(c_1 - c_2) > 0$ следует, что $\alpha < \frac{3\pi}{4}$, а значит $I_K^* = I$, $I_R^* = 0$.

8.3) Аналогичные соображения приводят к выводу, что при $(c_1 - c_2) < 0$ имеем: $I_K^* = 0$, $I_R^* = I$.

Дадим каждой из этих ситуаций соответствующее социально-экономическое толкование. Ситуация 1) может возникнуть, когда количество трудовых ресурсов уже очень мало, а капитала достаточно для того, чтобы не инвестировать сектор материального производства. Повышение уровня трудовых ресурсов происходит благодаря уменьшению производства (а тем самым и уменьшения использования ресурсов), а также инвестированию сектора по восстановлению ресурсов (социальной сферы).

В ситуации 2) трудовых ресурсов достаточно для того, чтобы их можно было бы в ближайшее время не восстанавливать, капитал представлен в относительно достаточном объеме, поэтому в этой ситуации можно частично проинвестировать сектор материального производства.

Ситуация 3) означает, что состояние трудовых ресурсов находятся в пределах допустимых значений, и поэтому все инвестиции было бы целесообразно вложить в сектор материального производства.

В ситуации 4) состояние трудовых ресурсов также находится в пределах нормы, а сектор материального производства не инвестируется, поэтому увеличения уровня трудовых ресурсов можно достичь естественным восстановлением ресурсов, а также возможно вложение любой суммы инвестиций для достижения большего восстановления трудового ресурса, и, тем самым, получить запас на будущее.

Ситуация 5) означает, что увеличения социального уровня трудовых ресурсов можно достичь за счет уменьшения объемов производства и естественного восстановления ресурсов. Эта ситуация приводит к уменьшению капитала, который можно уменьшать до тех пор, пока это не начнет сказываться на жизненном уровне людей или экономическом состоянии страны.

Ситуация 6) полностью аналогична ситуации 1); ситуация 7) аналогична ситуации 3).

Определение оптимальных инвестиций

Для нахождения решения исходной задачи (1), учитывая (2), решим сопряженную систему:

$$\begin{cases} \dot{\psi}_1 = -\dfrac{\partial H}{\partial K} = \mu\psi_1, \\ \dot{\psi}_2 = -\dfrac{\partial H}{\partial R} = c\psi_2. \end{cases} \tag{4}$$

Решениями системы (4) будут функции:

$$\psi_1(t) = d_1 e^{\mu t}, \qquad \psi_2(t) = d_2 e^{ct}, \tag{5}$$

где d_1 и d_2 – произвольные постоянные.

Так как (1) является задачей со свободным правым концом, то справедливы условия трансверсальности:

$$\begin{aligned} \psi_1(T) &= -v_1, \\ \psi_2(T) &= -2v_2(R^* - R(T)). \end{aligned} \tag{6}$$

Учитывая (5), (6), можем найти решение системы (4), которое будет иметь вид:

$$\psi_1(t) = -v_1 e^{\mu(t-T)}, \qquad \psi_2(t) = -2v_2(R^* - R(T))e^{c(t-T)}.$$

Очевидно, что $\psi_1(t) < 0$ для произвольного значения t.

Изучим поведение функций $c_i(t)$, $i = \overline{1, 2}$. В дальнейшем будем считать, что весовые коэффициенты u_i, v_i, $i = \overline{1, 2}$, являются положительными.

Рассмотрим лишь тот случай, когда $R^* \geq R(T)$. Тогда $\psi_2(t) \leq 0$. Следовательно, $c_2(t) < 0$ при $t \in [0, T]$. Теперь необходимо только исследовать знак коэффициента c_1. Найдем $c_1(t)$:

$$c_1(t) = \psi_1 - \xi\psi_2 - u_1 = -v_1 e^{\mu(t-T)} + 2\xi v_2(R^* - R(T))e^{c(t-T)} - u_1. \tag{7}$$

В качестве упрощения предположим, что $\mu = c$, тогда (7) будет иметь следующий вид:

$$c_1(t) = (2\xi v_2(R^* - R(T)) - v_1)e^{c(t-T)} - u_1.$$

Далее исследуем знак коэффициента $c_1(t)$:

1. Если $2\xi v_2(R^* - R(T)) - v_1 \leq 0$, то всегда $c_1(t) < 0$ при любом t.

2. Если $2\xi v_2(R^* - R(T)) - v_1 > 0$, то $c_1(t) < 0$ при $t < t^*$ и $c_1(t) > 0$ при $t > t^*$, где t^* имеет вид:

$$t^* = T + \frac{\ln\left(\dfrac{u_1}{2\xi v_2(R^* - R(T)) - v_1}\right)}{c},$$

причем $c_1(t^*) = 0$. Поскольку мы предположили, что $c_2 < 0$, то для данной ситуации ($R^* \geq R(T)$) существуют три варианта оптимального распределения инвестиций:

$$(I_K^*, I_R^*) = \begin{cases} c_1 < 0,\ c_2 < 0,\ I_K^* = 0,\ I_R^* = 0, \\ c_1 = 0,\ c_2 < 0,\ I_K^* \in [0, I],\ I_R^* = 0, \\ c_1 > 0,\ c_2 < 0,\ I_K^* = I,\ I_R^* = 0. \end{cases}$$

1. В случае $c_1 < 0,\ c_2 < 0$ мы имеем случай 5) из рассмотренных выше ситуаций. С точки зрения экономики в этом случае сектор материального производства не инвестируется, следовательно, не получает должного развития, и поэтому из (1) следует, что и природный (трудовой) ресурс способен

самовоспроизводиться без дополнительных капиталовложений. Данную ситуацию целесообразно использовать, если уменьшение капитала не слишком скажется на жизненном уровне людей или экономическом состоянии страны.

2. В случае $c_1 = 0,\ c_2 < 0$ мы имеем случай 2) с оптимальным распределением инвестиций $I_K^* \in [0, I],\ I_R^* = 0$. Реально такая ситуация может означать, что состояние имеющихся трудовых ресурсов находится в пределах нормы или даже есть избыток ресурсов, а состояние капитала приближается к норме, поэтому можно осуществить некоторые капиталовложения в сектор материального производства с тем, чтобы получить дополнительную прибыль и довести величину капитала до нормы, или даже получить запас производственного капитала.

3. Эта ситуация совпадает со случаем 7) и означает, что уровень капитала стал критически малым, а состояние имеющихся трудовых ресурсов находятся в пределах нормы, поэтому целесообразно было бы инвестировать сектор материального производства и наращивать объемы производственного капитала.

Как видим, ситуация, когда $R^* \geq R(T)$ (то есть, когда запас ресурса в конце планового периода не превышает значение показателя состояния невозмущенных ресурсов), свидетельствует о том, что природный ресурс способен самовоспроизводиться на данном промежутке времени без дополнительных капиталовложений, что и подтверждается результатами исследования.

С целью нахождения соответствующих оптимальных траекторий достаточно решить систему дифференциальных уравнений в задаче (1), подставляя в нее найденные оптимальные значения инвестиций. Итак, имеем следующие три случая:

1. Для всех $t \in [0, T]$ $c_1 < 0$, $c_2 < 0$. Оптимальные значения капиталовложений: $I_K^* = 0$, $I_R^* = 0$. Подставляя эти значения в систему дифференциальных уравнений задачи (1) и решая ее, получим такие оптимальные траектории:

$$\tilde{\tilde{K}}(t) = K^{(0)} e^{-\mu t}, \qquad \tilde{\tilde{R}}(t) = (R^{(0)} - R^*) e^{-ct} + R^*.$$

2. Для всех $t \in [0, T]$ $c_1 = 0$, а $c_2 < 0$. Оптимальные значения соответствующих инвестиций: $I_K^* \in [0, I]$, $I_R^* = 0$, а соответствующие им оптимальные траектории:

$$\tilde{\tilde{K}}(t) = (K^{(0)} - \frac{I_K^*}{\mu}) e^{-\mu t} + \frac{I_K^*}{\mu}, \qquad \tilde{\tilde{R}}(t) = (R^{(0)} - R^* + \frac{\xi}{c} I_K^*) e^{-ct} + R^* - \frac{\xi}{c} I_K^*. \qquad (8)$$

3. Для всех $t \in [0, T]$ $c_1 > 0$, а $c_2 < 0$. Решением будут значения $I_K^* = I$, $I_R^* = 0$, а оптимальные траектории для этого случая имеют вид, аналогичный (8) при $I_K^* = I$.

Выводы

В данном исследовании была рассмотрена динамика экономической системы, согласованную с динамикой самовосстанавливающихся природных ресурсов. В частности, если в роли последних рассматривать трудовые ресурсы, то модель приобретает признаки социально-экономического характера. В рамках такой социально-экономической системы рассмотрена противоречивая задача об оптимальном распределении определенного ограниченного объема капиталовложений (инвестиций) в течение заданного промежутка времени таким образом, чтобы в конце данного периода достичь как максимально возможного результата в экономической сфере, так и наилучшего эффекта относительно природного (трудового) ресурса, и при этом достичь этой цели необходимо минимальными инвестиционными расходами.

В работе предложен один из вариантов решения такой проблемы. Исходная задача оптимального управления сведена к задаче линейного программирования. Последняя была решена для одной из

возможных ситуаций, касающихся состояния социальной составляющей (трудового ресурса). Для указанной ситуации были найдены оптимальные инвестиций и для каждого из оптимальных значений капиталовложений были построены оптимальные траектории, характеризующие состояния экономической и социальной составляющих системы.

Заключение

Экономическое развитие в целом определяется тремя факторами экономического роста: трудовыми ресурсами, искусственно созданными средствами производства, природными ресурсами. Экономическая наука мало обращала внимание на экологические и социальные проблемы, что и послужило одной из причин формирования техногенного типа экономического развития. Этот тип можно охарактеризовать как природоразрушающий, основанный на использовании искусственных средств производства, созданных без учета экологических ограничений.

С целью предотвращения глобального и локального экологических кризисов необходимо заменить техногенный тип развития на устойчивый. Последний дает возможность удовлетворить потребности современных поколений, однако не ставит под угрозу существование последующих. Этот тезис, очевидно, касается и социального контекста. Развитие экологической экономики тесно связано с приоритетами внутренней и внешней политики любого государства. Они должны обеспечить природо- и ресурсосберегающие национальные интересы в рамках концепции устойчивого экономического развития.

Результаты исследования могут служить методологической базой для принятия оптимальных управленческих решений, касающихся согласованного финансирования вышеуказанных составляющих устойчивого экономического развития.

Благодарности

Автор благодарен проф. Волошину А.Ф. за ценные консультации при написании статьи.

Работа опубликована при финансовой поддержке проекта **ITHEA XXI** Института информационных теорий и приложений FOI ITHEA Болгария www.ithea.org и Ассоциации создателей и пользователей интеллектуальных систем ADUIS Украина www.aduis.com.ua.

Библиография

[Волошин, 2006] Волошин О.Ф., Мащенко С.О. Теорія прийняття рішень: Навчальний посібник. – К.: Видавничо-поліграфічний центр «Київський університет», 2006. – 304 с.

[Григорків, 1998] Григорків В.С. Оптимальне інвестування еколого-економічної системи // Вісник Київського Університету, серія фізико-математичні науки. Випуск 3, 1998. С. 169-174.

[Коробова, 2001] Коробова М.В. Агрегированная оптимизационная модель эколого-экономического взаимодействия // Проблемы управления и информатики, 2001, № 4. С. 144-154.

[Кротов, 1990] Основы теории оптимального управления / Под ред. В.Ф. Кротова. – М.: Высшая школа, 1990. – 430 с.

Сведения об авторе

Коробова Марина Витальевна *– Доцент, Киевский национальный университет имени Тараса Шевченко, факультет кибернетики. Киев, Украина. E-mail: mkorobova@rambler.ru*

МОДЕРНИЗАЦИЯ СИСТЕМЫ МАРТИНГЕЙЛ И АНАЛИЗ ЕЕ РАБОТЫ НА ВАЛЮТНЫХ РЫНКАХ

Владислав Плаксин

***Аннотация**: Предлагается изменение и адаптация известной торговой системы "Мартингейл" для прогнозирования кросс-курсов иностранных валют. Представленные экспериментальные результаты использования модернизированной версии МБАС "Мартингейл" доказывают эффективность предлагаемого подхода.*

***Ключевые слова**: рынок валюты, прогнозирования кросс-курсов, торговая система*

***ACM Classification Keywords**: H.4.2 [Information Systems Applications] Types of Systems - Decision support*

***Conference**: The paper is selected from XVth International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.*

Введение

У любой компании вне зависимости от вида деятельности в определенный момент времени бывают временно свободные денежные средства. И часто возникает вопрос, как их использовать с наибольшей полезностью для организации и, конечно же, с минимальными рисками от потерь. Существуют множество вариантов размещения своих средств для получения прибылей, к примеру:

1. Разместить денежные средства на депозите в коммерческом банке.
2. Приобрести различные векселя как промышленных предприятий, так и коммерческих банков.
3. Купить государственные ценные бумаги.
4. Приобрести акции корпоративных предприятий.
5. Работа на западных финансовых рынках
6. Спекуляция на валютных курсах

Данный набор возможностей по размещению денежных средств можно разделить на два важных вида:

Безрисковые операции — операции, имеющие наименьший риск (приобретение государственных ценных бумаг, депозитов и векселей Сбербанка, векселей и облигаций крупных экспортных компаний, кредитование со 100% обеспечением). Следует особо отметить, что по определению государственные ценные бумаги в национальной валюте являются самым надежным финансовым инструментом, однако чиновники в нашей стране заставили нас пересмотреть эту аксиому. Говоря об отсутствии риска данного финансового инструмента, мы, к сожалению, должны не забывать, что в странах СНГ безрисковость данных операций существует до определенного момента. Например, этот момент может наступить при сильном снижении цены на нефть, высоких внутренних политических колебаниях и многих других внешнеэкономических и политических факторах. Рисковые операции — операции, в основе которых предусмотрен риск не только в получении дополнительного дохода, но и возможной потери основного капитала (покупка корпоративных акций вне зависимости от срока инвестирования, работа на западных финансовых рынках, приобретение облигаций корпоративных компаний и др.).

Основным отличием первых операции от вторых является возможность понимания и реальной оценки текущей ситуации. Поэтому в условиях жизни на территории СНГ, а в частности Украины, более безопасно заниматься рисковыми операциями и адекватно оценивать риски этих операций, чем постоянно пытаться вытащить «кота из мешка».

Целью настоящего доклада является развитие известной системы Мартингейла, ее адаптация для валютных финансовых рынков а также ее тестирование и применение на валютном рынке и сравнительный анализ с базовой системой.

Управление рисками на валютных рынках

Рассмотрим кратко вопрос рисков на валютных рынках и управления денежными средствами. Анализируя основные ошибки многих трейдеров (компаний или частных лиц), которые торгуют на рынке Форекс, можно выделить следующие типы ошибок:

1. Неправильное соотношение реальных средств к заемным средствам.
2. Отсутствие психологического контроля при неудачных сделках.
3. Азарт.
4. Неумение правильно рассчитывать временные рамки существования сделки.

Все данные проблемы являются первичными и без их устранения никаких положительных результатов добиться, к сожалению, нельзя. По этой причине каждый из трейдеров, на определенном этапе своей эволюции начинает использовать конкретную систему по управлению денежными средствами, для минимизации потерь и максимизации прибылей. В данном докладе представляется адаптированная и модернизированная нами система управления денежными средствами «Мартингейл».

«Система Мартингейла» названа в честь удачливого картёжника XIX века, который был завсегдатаем казино Французской Ривьеры, но можно с уверенностью сказать, что этот термин ввёл в начале XX века математик П. Леви (Levy), изучавший парадоксы азартных игр. Концепция Леви получила широкое развитие в трудах другого математика — Дуба. Система Мартингейла (так сокращенно трейдеры называют данную систему в честь картежника и афериста Мартингейла), иллюстрирует всю простоту и естественность удвоения вероятности. Проще говоря, основной принцип работы системы заключается в том, что в случае неудачной сделки следует увеличивать следующую ставку в два раза от потери, чтобы в итоге компенсировать все потери одним выигрышем.

Именно такой принцип очень часто придумывают новички, впервые попавшие на финансовый рынок или в казино. Обратим наше внимание на основные плюсы и минусы данного подхода:

1. Самым очевидным минусом является постоянно растущий капитал следующей ставки при большом количестве проигрышей.
2. С другой стороны, одним из сильных плюсов ого принципа является высокая вероятность выигрыша при не сильных колебаниях финансовых рынках.

Модификация базовой системы Мартингейла, ее тестирование и анализ

В начале опишем условия, в которых собираемся использовать Мартингейла и адаптируем под нее алгоритм системы.

Начнем с математической части на примере расчета вероятности зависимости вероятности проигрыша от возможной прибыли при игре в монетку с помощью мартингейла. Так как каждый раз в результате проигрышного броска мы удваиваем ставку, то эти величины можно связать следующей последовательностью:

1v, 2v, 4v, 8v, 16v, 32v, 64v …. 1024v …. Степень двойки.

Теперь опишем базовые условия торговли на рынке форекс, для любой пары валют:

- Кредитное плечо реальных денег к залоговым – 1 к 100
- Стоимость одного 1 лота – 1000 долларов
- Стоимость движения рынка на один пункт при кредитном плече 1 к 100 равно 10 долларам прибыли или убытка.

Используя эти данные, попытаемся смоделировать ситуацию с использованием стандартных условий алгоритма Мартингейла

№	Объем ставки	Прибыль	Выигрыш
1	1	1	101000
2	2	-1	101000
3	4	-2	100000
4	8	-4	98000
5	16	-16	94000
6	32	-32	78000
7	64	-64	46000

У Вас, как и у всех начинающих трейдеров, сразу возникает вопрос: почему рабочая система в азартных играх при тех же условиях не работает на рынке Форекс. Ситуация до предела проста, в рынке Форекс Вам приходится или фиксировать убыток, или держать открытой сделку с этим убытком, ожидая ситуации, когда рынок вернется в исходное положение для сделки. Поэтому по системе Мартингейла для валютных рынков надо учитывать еще и предыдущие убытки.

Поэтому модифицированная базовая адаптированная система Мартингейла (МБА С) для финансовых рынков выглядит следующим образом:

$$Vn=sum(v1+v2+..Vn-1)*2$$

Смоделируем аналогичную ситуацию на валютном рынке, но уже для МАБС Мартингейл, с условием не полного возвращения на 100 пунктов в обратную сторону, а движения на половину, то есть 50 пунктов:

№	Объем ставки	Прибыль	Выигрыш
1	1	1	101000
2	2	-1	101000
3	6	-2	101000
4	18	-6	101000
5	54	-18	101000
6	162	-54	101000
7	486	-162	101000

Как видим, при использовании этой формулы для любой валютной пары, результат будет положительный, но большим минусом является быстро растущий объем ставки, который мгновенно может уничтожит весь депозит, но при этом длина отката уменьшена в 2 раза, что повышает сильно повышает вероятность положительной сделки.

Для детального тестирования была специально создана программа и проведены исторические тесты на длинных промежутках с использованием разных объемов депозита, для получения максимальных прибылей с минимальным риском. В итоге мы получили следующие результаты, при веденные на рис. 1:

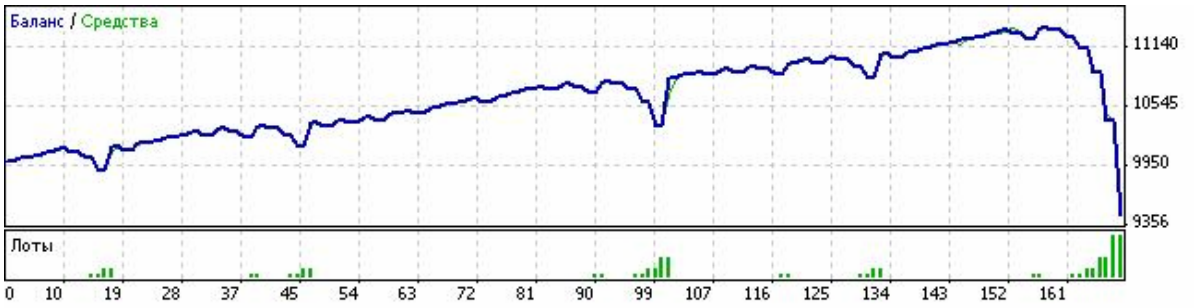


Рис. 1

Модификация базовой системы Мартингейла для работы на валютном рынке Форекс

В результате проведенных исследований с МБАС Мартингейл был сделан вывод: о нецелесообразности использования системы в том виде, в котором она есть на данный момент. Причиной такого вывода стал анализ исторических данных относительно откатов курса котировок валют во время длинных движений в одну сторону.

В результате исследований был разработан следующий алгоритм работы программы, который позволил модернизировать МБАС Мартингейл и вывести ее на совершенно новый уровень работы, что привело к положительным результатам.

Было предложено сделать следующее:

1. Открывать сразу две встречных сделки для того чтобы максимизировать прибыль во время любого движения рынка (вверх или вниз).

2. Уменьшить входящие лоты, для получения более стабильного и постоянного дохода.

3. Оптимизировать входные данные для открытия стартовых сделок.

4. Использовать плавающие «тейк профиты» для получения максимальной прибыли.

В результате выполнения данной работы была написана новая программа для торговли на рынке Форекс и проведены эксперименты на предистории, результаты которых демонстрируются ниже:

Символ EURUSD (Euro vs US Dollar)

Период 1 Минута (M1) 2008.01.02 09:01 - 2008.09.16 23:56 (2008.01.01 - 2008.09.17)

Начальный депозит 30000.00

Общая прибыль 22938.70

Общий убыток -14882.19

Чистая прибыль 8056.51

График изменения дохода представлен на рис. 2.

Рис. 2

Анализ работы «советника трейдера» на графике Евро- Доллара в период с 1 февраля 2008 года по 16 сентября 2008 года показывает, что «советник трейдера» дал чистую прибыль в размере 29%, что является высоким показателем, однако в первой четверти периода работы, «советник» показал сильную просадку средств, что чуть не привело к полной потери депозита и данная ситуация является неприемлемой. Просадка составила 8860.19 (29.11%). (см. Рис.2)

Выводы

В результате данного и множества других экспериментов сделаны следующие выводы:

1. Использование модернизированной версии МБАС Мартингейл является более прибыльной системой управления денежными средствами, чем базовая система Мартингейл

2. Для безопасной торговли с минимальными рисками приведенные системы не подходят, так как они являются системами управления денежными средствами, а не системами торговли.

3. Использование модернизированной версии МБАС Мартингейл в техническом анализе приведет к положительным результатам, и этим следует в дальнейшем заняться.

В качестве перспективных направлений дальнейших исследований в данной проблематике можно выделить следующие:

1. Интеграция модернизированной МБАС Мартингейл с техническими инструментами для финансовых рынков. Разработка автоматизированных систем и проведение анализа их работы с последующими выводами

2. Использования нейронных сетей для определения спокойных и возбужденных состояний на рынке Форекс с целью правильного использования модернизированной МБАС Мартингейл.

Благодарности

Работа опубликована при финансовой поддержке проекта **ITHEA XXI** Института информационных теорий и приложений FOI ITHEA Болгария www.ithea.org и Ассоциации создателей и пользователей интеллектуальных систем ADUIS Украина www.aduis.com.ua.

Литература

Abu-Mostafa, Y.S. (1995). "Financial market applications of learning from hints". In *Neural Networks in Capital Markets.* Apostolos-Paul Refenes (Ed.), Wiley, 221-232.

Beltratti, A., Margarita, S., and Terna, P. (1995). *Neural Networks for Economic and Financial Modeling*. ITCP.

Chorafas, D.N. (1994). *Chaos Theory in the Financial Markets*. Probus Publishing.

Colby, R.W., Meyers, T.A. (1988). *The Encyclopedia of Technical Market Indicators*. IRWIN Professional Publishing.

Ehlers, J.F. (1992). *MESA and Trading Market Cycles*. Wiley.

Kaiser, G. (1995). *A Friendly Guide to Wavelets*. Birk.

Шарп, У.Ф., Александер Г.Дж., Бэйли, Дж. В. (1997). *Инвестиции*. Инфра - М.

Информация об авторах

Плаксин Владислав *– Управляющий диллинг центра FOREX4you, Киев, Урицкого 36, тел:+380678722222, E-mail: vlad@forex4you.com.ua*

СИСТЕМА МОДЕЛИРОВАНИЯ И АНАЛИЗА ВАЛЮТНЫХ РИСКОВ

Виктор Бондаренко

***Аннотация**: В статье описывается система построенная на основе реализации метода Монте-Карло для анализа рисков и управления валютным портфелем.*

***Ключевые слова**: Валютные риски, валютный портфель, VaR метод, метод Монте-Карло.*

***ACM Classification Keywords**: J.4. Social and Behavioral Sciences*

***Conference**: The paper is selected from XVth International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.*

Введение

Валютный риск - это риск потерь обусловленный неблагоприятным изменением курсов иностранных валют в ходе осуществления сделок по купле-продаже этих валют. Валютный риск, или риск курсовых потерь, связан с интернационализацией рынка банковских операций, созданием транснациональных предприятий и банковских учреждений и представляет собой возможность денежных потерь в результате колебаний валютных курсов.

При этом, изменение курсов валют происходит в силу многочисленных факторов, например, в связи с изменением внутренней стоимости валют, постоянным перетеканием денежных потоков из страны в страну, спекуляцией и т.д. Ключевым фактором, характеризующим любую валюту является степень доверия к валюте.

Доверие к валюте сложный многофакторный критерий состоящий из нескольких показателей, например: показатель доверия к политическому режиму, степени открытости страны, либерализации экономики и режима обменного курса, экспортно-импортного баланса страны, базовых макроэкономических показателей и веры инвесторов в стабильность развития страны в будущем.

Вместе с тем, из всех факторов, влияющих на курс валют в долгосрочной перспективе, можно выделить два основных.

1. Первый из них это темп инфляции, наблюдаемой закономерностью которого является то, что в стране с более высокими темпами инфляции понижается курс национальной валюты по отношению к валютам стран с более низким темпом инфляции.
2. Резкие колебания курсов валют могут быть связаны причинами, как экономическими и политическими, так и чисто спекулятивными. Рынок чутко реагирует на все изменения экономических показателей, прогнозы экспертов, политические кризисы и политические слухи.

Необходимо отметить, что риску обусловленному трудно прогнозируемыми колебаниями валют подвержены как страны, где происходят эти колебания, так и страны, соседствующие с кризисными странами или имеющие с ними значительные экономические или политические связи.

Цель анализа валютных рисков заключается в их изучении и оценке с целью минимизации валютного риска и исключении убытков, вызываемых колебаниями курсов валют.

Одним из наиболее популярных средств снижения валютных рисков является диверсификация т.е. тактика формирования активов, при которой активы учреждения состоят из набора разных валют, благодаря чему, потери обусловленные снижением курса одной валюты могут быть компенсированы

прибылью, полученной от повышения курса другой валюты. Такой набор валют принято называть валютным портфелем.

Учитывая вышеизложенное, большой интерес представляют методики и технологии повышения качества анализа валютного портфеля и прогноза потерь, обусловленных валютными рисками. Эти методики позволяют численно оценить риск потерь валютного портфеля.

Для численной оценки валютного риска необходимо использовать меру риска. Мера риска - это способ, с помощью которого, можно численно оценить величину риска. Для нужд хеджирования необходимо, чтобы с помощью меры риска можно было оценивать риск потерь как отдельной валюты входящей в портфель, так и всего портфеля валют в целом.

Существует множество технологий оценки рисков. Среди них можно выделить как наиболее известные Value-at-Risk (VaR) и ее модификации Marginal VaR, Incremental VaR, EaR, Cash Flow-at-Risk (C-FaR). Широко известны и другие методики - бета-анализ, теории CAPM, APT, Short Fall, Capital-at-Risk, Maximum Loss. Некоторые из этих технологий известны давно, а другие только начинают завоевывать популярность в банках, инвестиционных и страховых компаниях, пенсионных фондах.

Остановимся на технологии риск-менеджмента VaR [Лобанов,Чугунов], которая в последнее время находит все большее распространение в среде инвесторов. Например, как было отмечено в исследовании New York University Stern School of Business, около 60% пенсионных фондов США используют в своей работе VaR метод.

VaR — это статистическая оценка выраженная в денежных единицах базовой валюты V (к примеру, в гривнах), которую не превысят, с заданной вероятностью α (обычно α=0,95), ожидаемые максимальные убытки X валютного портфеля за заданный интервал времени t (например, за 10 дней).

Таким образом, VaR определяет значение V из соотношения

$$P(X \leq V)=\alpha,$$

где P(A) – вероятность события A (в нашем случае, событие A заключается в том, что $X \leq V$).

Расчетом VaR занимается довольно много специализированных компаний, а зачастую и собственные подразделения финансовых структур. Как правило, расчеты проводятся вручную на невысоком уровне автоматизации, что требует использования ряда плохо обоснованных допущений.

Например, классическая техника расчета VaR основана на предположении о нормальном распределении курсов валют входящих в портфель. Однако в силу нестабильности рынка значения финансовых величин далеко не всегда подчиняются закону нормального распределения, что требует **восстановления плотности вероятности** для точной оценки VaR.

Проведенный анализ показал, что один из наиболее эффективных методов решения указанных задач, является метод моделирования основанный на использовании методов статистических испытаний (методы Монте-Карло).

Поэтому, целью работы является разработка технологичной системы анализа валютного портфеля, которая базируется на методе Монте-Карло. Система, должна позволять финансовому менеджеру:

1. На основе исторического анализа курсов валют, делать эффективные прогнозы прибыли (потерь) по каждой валюте портфеля так и по всему валютному портфелю в целом.
2. Оптимизировать структуру портфеля таким образом, чтобы прогнозируемые потери были минимальными.
3. Эффективно определять различные характеристики портфеля (вероятность безубыточности портфеля, вероятность получения минимального убытка и т.п.).

Описываемая технология внедряется на банковском факультете Киевского национального экономического университета.

Алгоритм метода Монте-Карло, реализованный в системе моделирования и анализа валютных рисков

Реализованный алгоритм моделирования и анализа рисков валютного портфеля основанный на использовании метода Монте-Карло имеет ряд достоинств среди которых наиболее важные:

- высокая точность расчетов;
- возможность моделирования любых исторических и гипотетических распределений;

Однако методы Монте-Карло не свободны от недостатков, среди которых можно отметить:

- высокая сложность моделей и соответственно большой риск ее неадекватности;
- высокие требования к вычислительной мощности и значительные затраты времени на проведение расчетов;

Реализованный алгоритм метода Монте-Карло для решения задач анализа валютного портфеля имеет следующий вид:

1. По историческим данным курсов валют портфеля рассчитываются оценки математического ожидания $\overline{x}$ и среднеквадратического отклонения (волатильности) σ по каждой валюте.

2. С помощью генератора псевдослучайных чисел генерируются случайные значения курсов валют распределенные по необходимому закону распределения с математическим ожиданием равным $\overline{x}$ и среднеквадратическим отклонением σ. Как показали исследования, распределение случайных величин курсов валют близко к нормальному распределению поэтому, в качестве новых гипотетических значений курсов в будущем можно использовать генератор псевдослучайных чисел распределенных по нормальному закону. Однако, расчеты будут более точными если в качестве распределения прогнозируемых курсов валют выбирать эмпирические распределения, которым подчинялись курсы валют в ретроспективной выборке. Именно такие распределения курсов валют используется в проектируемой системе.

3. Полученные в пункте 2 значения курсов валют используются для построения таблицы. Столбцы таблицы задают временные траектории моделируемых курсов валют (сценарии вариантов расчетов), а строки определяют период времени моделирования. Такая таблица строится для каждой валюты. Размерность таблицы произвольная и зависит от имеющихся вычислительных мощностей, но чтобы метод обеспечивал приемлемую точность, таблица должна быть достаточно большой. В наших экспериментах была использована таблица из 11 столбцов (вариантов временных траекторий) и 20 строк (количество торговых дней моделирования). Однако, в случае необходимости размерность таблицы может быть легко увеличена.

4. На основании таблиц построенных в пункте 3, вычисляются временные траектории моделируемых курсов валют. Для этого, необходимо выполнить такие действия:

Сначала выбирается период глубины времени *T* (например 20 торговых дней, за который отслеживаются исторические изменения курсов P_{it} всех *N* входящих в портфель валют. Где P_{it} курс i-ой валюты портфеля в t-ый торговый день. Здесь i= 1,2,3,….*N*, а t=1,2,3…..*T*.

Для каждого сценария вычисляется приращения курсов валют входящих в портфель. Эти приращения можно определить таким соотношением:

$$\Delta P_{it} = P_{it} - P_{it-1}, \quad (1)$$

где i= 1,2,3,….N, t=1,2,3…..T, P_{i0} - текущее значение курса i-ой валюты.

Для каждого сценария моделируется гипотетический курс каждой из валют портфеля такое моделирование выполняется по формуле:

$$P^*_{it} = P_{i0} + \Delta P_{it} \quad (2)$$

То есть гипотетический курс каждой из валют в будущем моделируется как текущее значение курса плюс прирост курса, соответствующей выполняемому сценарию.

Выполняется переоценка стоимости всего текущего портфеля по смоделированным курсам валют. Таким образом, для каждого сценария вычисляется насколько изменилась бы стоимость сегодняшнего валютного портфеля. Такая переоценка выполняется на основе соотношений:

$$\Delta V_{it} = V_{it} - V_{i0} = Q_i (P^*_{it} - P_{i0}) \quad (3)$$

где ΔV_{it} - приращение стоимости валютного портфеля по i-ой валюте для t-ого момента времени, V_{it} - стоимость i- валюты портфеля в t-ый момент времени, Q_i - количество i- ой валюты.

5. Определяем прогнозируемое значение стоимости i-ой валюты портфеля в базовой валюте (гривне) используемой для расчета результата для t-ого момента времени V^*_{it}:

$$V^*_{it} = (V_{i0} + \Delta V_{it}) K_i \quad (4)$$

где K_i - стоимость i-ой валюты портфеля в базовой валюте (гривне) в текущий момент времени.

6. Определяем новую прогнозируемую стоимость валютного портфеля S_t в базовой валюте для t-ого момента времени:

$$S_t = \sum_{i=1}^{N} V^*_{it}. \quad (5)$$

Для получения эмпирического закона распределения курсов валют на основе ретроспективной выборки использовался метод табличного получения значений курсов валют. Такой подход относительно легко формализуется и может быть запрограммирован, что дает возможность по заданным статистическим значениям курсов валют автоматически строить эмпирическую функцию распределения и определять их прогнозируемые значения в краткосрочной перспективе.

Программная реализация системы моделирования и анализа валютных рисков

Для реализации системы был выбран мощный язык C++ и его реализация в среде программирования Borland C++ Builder. Общий вид системы моделирования и анализа валютных рисков наведен на рис.1.

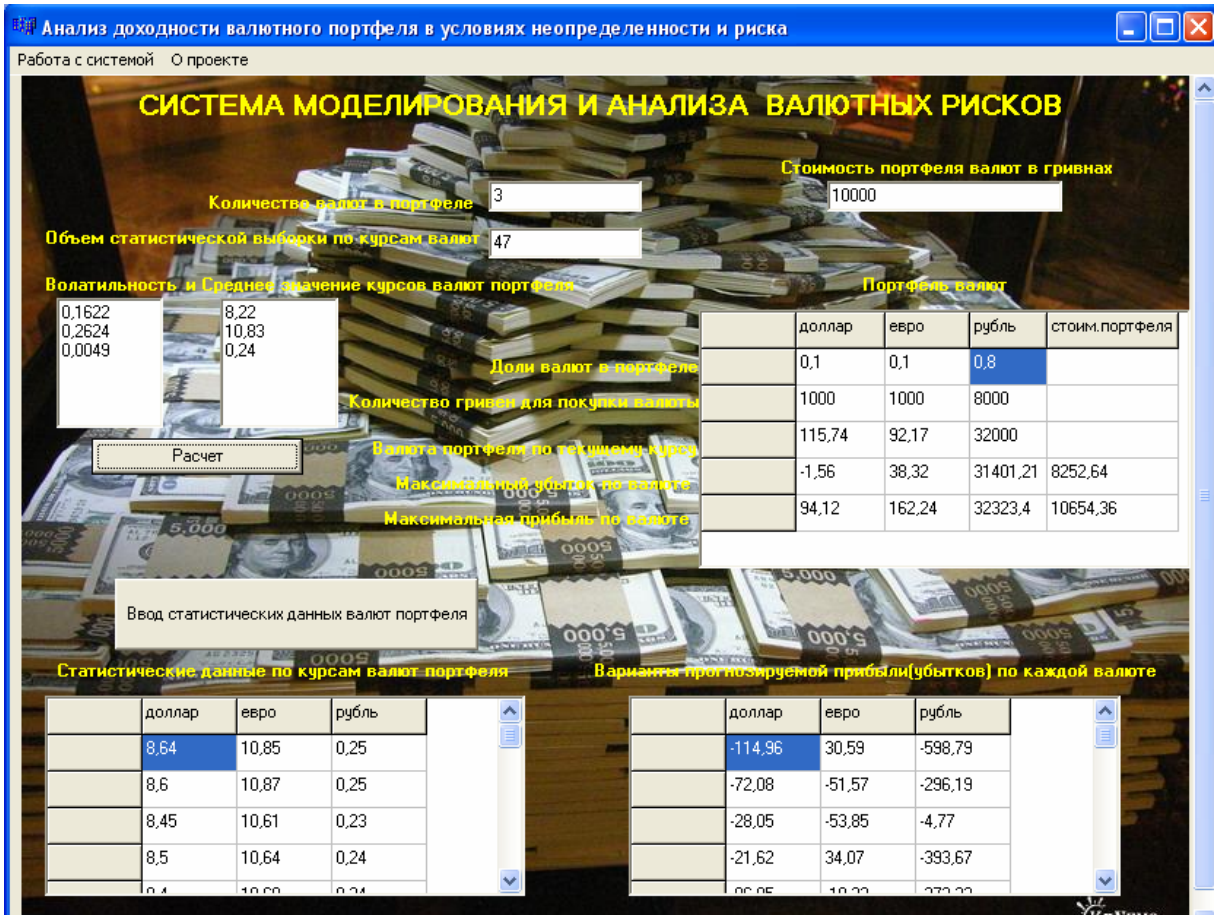
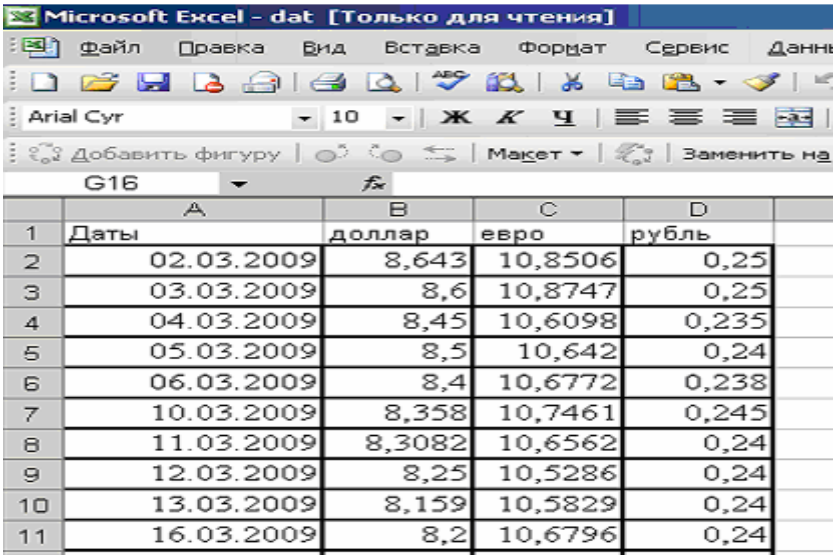


Рис.1. Общий вид системы моделирования и анализа валютных рисков.

Для работы с системой моделирования и анализа валютных рисков, необходимы статистические данные о курсах валют входящих в портфель. Такие данные удобно формировать в системе Excel. Лист с фрагментом значений ретроспективной статистики курсов валют наведен на рис.2.



	A	B	C	D
1	Даты	доллар	евро	рубль
2	02.03.2009	8,643	10,8506	0,25
3	03.03.2009	8,6	10,8747	0,25
4	04.03.2009	8,45	10,6098	0,235
5	05.03.2009	8,5	10,642	0,24
6	06.03.2009	8,4	10,6772	0,238
7	10.03.2009	8,358	10,7461	0,245
8	11.03.2009	8,3082	10,6562	0,24
9	12.03.2009	8,25	10,5286	0,24
10	13.03.2009	8,159	10,5829	0,24
11	16.03.2009	8,2	10,6796	0,24

Рис.2. Фрагмент рабочего листа системы Excel со статистикой курсов валют входящих в портфель.

Как видно из рис. 2. первая строка статистических данных по курсам валют содержит наименования валют. В нашем случае доллар, евро и российский рубль.

Для работы с системой необходимо в системе Excel подготовит статистические данные по курсам валют, которые необходимо включить в портфель. Фрагмент такого рабочего листа приведен на рис. 2. Как видно из рис.2., в столбце А начиная с клетки А2 располагаются даты на которые фиксировались курсы валют. В столбце В располагается первый вид валюты которая включена в портфель, значения валюты начинаются с клетки В2. В клетке В1 располагается наименование валюты. Следующая валюта включается в следующий столбец листа Excel (столбец С). Расположение значений аналогично расположению курсов валют записанных в столбец В. То есть, значения располагаются начиная с клетки С2, а в клетке С1 записывается наименование валюты. Аналогично курсы последующих валют портфеля располагаются в следующих столбцах рабочего листа D,E,F и т.д. После формирования исходных данных в системе Excel указанных выше образом, запускается система анализа валютного портфеля (Запускающий файл Project1.exe). В открывшемся окне необходимо занести количество валют входящих в портфель, объем статистических данных по каждой валюте, общая стоимость портфеля валют в гривнах, доли каждой валюты в портфеле.

Понятно, что сумма всех долей входящих в портфель валют должна равняться единицы. После нажатия кнопки «Статистические данные по валютам портфеля» на экране в окне высветится диалоговое окно общий вид которого приведен на рис.3. Это диалоговое окно позволяет выбрать Excel файл содержащий статистическую информацию по курсам валют портфеля.

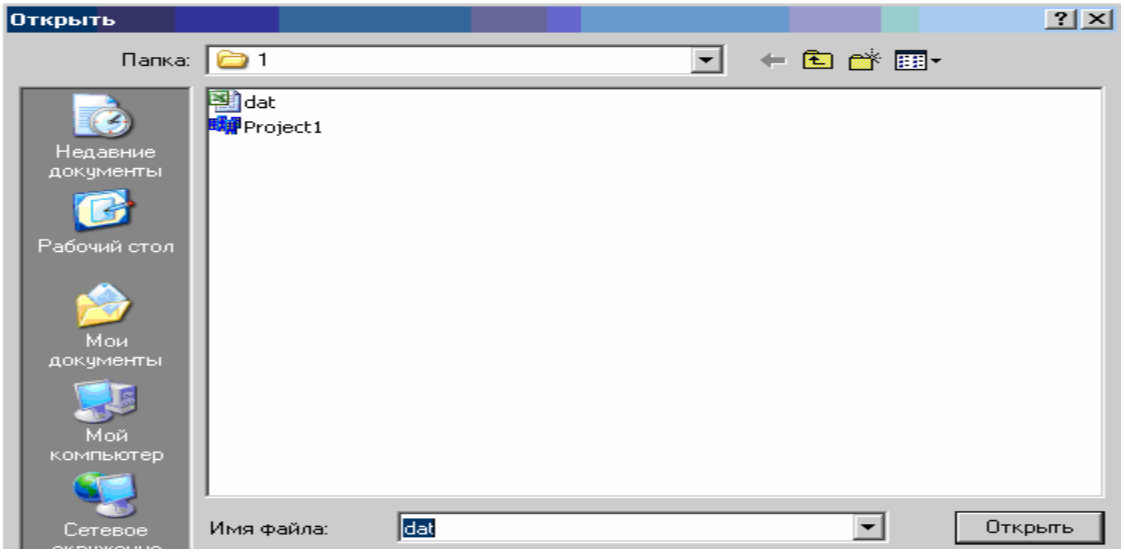


Рис.3. Вид диалогового окна для выбора Excel-файла, содержащего статистическую информацию по курсам валют портфеля.

Далее выбираем необходимый файл и нажимаем кнопку «Открыть». После чего, статистическая информация по курсам валют считывается из Excel-файла (к примеру, dat.xls) и заносится в систему, где отображается в окне «Статистические данные по курсам валют портфеля».

После этого, нажимаем кнопку «Расчет» и система выполняет расчеты по рассмотренному выше алгоритму. Результаты расчета будут высвечены в разных окнах на форме «Анализ рисков валютного портфеля». В окнах «Волатильность» и «Среднее значение курсов валют портфеля» будут высвечены волатильность и среднее значение по каждому виду валюты портфеля. В окне «Портфель валют» в первой строке будут находится введенные ранее менеджером доли валют составляющих портфель. Во

второй строке исходя из введенной менеджером стоимости портфеля, вычисляется сумма в гривнах выделенная для покупки разного вида валют портфеля. В третей строке «Валюта портфеля» вычисляется сумма валют разного вида составляющих портфель, исходя из текущего курса этих валют. В четвертой строке будет определен максимальный убыток по каждой валюте и суммарный максимальный убыток по всему портфелю (столбец «Стоим. портфеля»), а в пятой строке «Максимальная прибыль по валюте» будет вычислена максимальная прибыль по каждой валюте и суммарная прибыль по всему портфелю (столбец «Стоим. портфеля»).

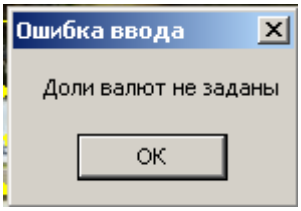
Результаты последних двух строк рассчитываются на основе данных вычисленных в окне «Прибыль (убытки) по каждой валюте». Указанная прибыль (убытки) рассчитываются по каждой временной траектории моделируемых курсов валют (сценарию вариантов расчетов), которых в нашей модели используется одиннадцать. Отсюда, помимо максимального убытка по валютному портфелю и максимальной прибыли по валютному портфелю вычисленного в окне «Портфель валют», можно рассчитывать различные характеристики портфеля, которые интересны для менеджера на основе данных рассчитанных в окне «Варианты прогнозируемой прибыли (убытков) по каждой валюте».

Поскольку система рассчитана на работу людей не являющихся профессиональными специалистами в области компьютерной техники, то в систему необходимо включить блокировки не позволяющие пользователю работать с ошибочными данными.

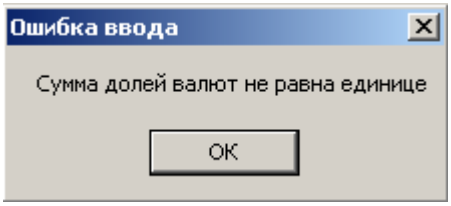
С этой целью, в систему включено три блокировки:

1. Эта блокировка запрещает выполнение расчетов если исходная информация о статистике курсов валют не введена в систему. В этом случае кнопка «Расчет» будет заблокирована.

2. Возможно менеджер нажмет кнопку «Расчет» забыв предварительно ввести доли валют в портфеле. В этом случае расчеты произведены быть не могут и система уведомляет менеджера о ошибке таким сообщением.



3. Могут иметь место ошибки при наборе долей валют портфеля, заключающиеся в том, что сумма долей всех валют портфеля не равна единице. В этом случае система уведомляет пользователя таким сообщением.



Следует отметить, что для работы с системой можно использовать и элемент меню «Работа с системой». Для просмотра сведений о авторских правах и версиях системы нужно выбрать из меню пункт «О проекте».

Работа с системой моделирования и анализа валютных рисков

Имея прогнозируемые значения стоимости i-ой валюты портфеля мы можем определять и прогнозировать различные характеристики портфеля.

Например:

1. Вероятность того что портфель будет безубыточным.
2. Максимально возможные убытки портфеля.
3. Максимально возможную прибыль портфеля и тому подобное.

Рассмотрим конкретные примеры такого анализа. Пусть прибыль (убытки) по валютным составляющим портфеля для каждого из одиннадцати сценариев вариантов расчетов имеют вид:

Доллар	Евро	Рубль
39,05969	38,05965	347,2579
38,47365	24,97965	231,6622
30,81099	15,36359	175,00443
26,371	7,779714	120,4757
16,6762	1,370843	40,05551
10,07833	-0,24964	26,81973
1,12364	-1,032232	23,08937
-18,2897	-31,1727	-43,7303
-23,019	-33,7671	-106,588
-23,9159	-36,2926	-233,458
-59,4184	-39,0728	-442,936

Исходя из указанных данных легко видеть, что наибольший убыток по доллару -59,4184, по евро -39,07, по рублю -442,92. Таким образом при максимальных потерях стоимость портфеля может составить 8950,61 грн, а при максимальной прибыли которая составляет по доллару 39,05, по евро 38,06, по рублю 347,26 стоимость портфеля составит 10290,94 грн. Таким же образом, можно определить различные вероятности характеризующие портфель. Например, вероятность того что все валюты портфеля будут прибыльными. По доллару имеем из одиннадцати статистических экспериментов долларовые составляющая была прибыльна 7 раз, таким образом, вероятность того, что долларовая составляющая будет прибыльной:

$$P_{дол.}=7/11 =0,64.$$

Для составляющей евро из одиннадцати статистических экспериментов прибыльность евро была 5 раз.

$$P_{евро} =5/11=0,45.$$

По рублю из одиннадцати статистических экспериментов рублевая составляющая была прибыльна 7 раз.

$$P_{рубль} =7/11=0,64.$$

Таким образом, вероятность события заключающегося в том, что каждая составляющая портфеля будет прибыльной выражается соотношением произведений вероятностей.

$$P_{портфеля} =P_{дол.}*P_{евро}*P_{рубль}=7/11*5/11*7/11=245/1331=0,18.$$

Аналогично, можно определить вероятности минимального убытка по каждой составляющей портфеля. Минимальные прибыльности по каждой составляющей портфеля и другие характеристики интересующие финансового менеджера.

Выбирая различные варианты структуры валютного портфеля, можно подбирать ее таким образом, чтобы минимизировать ожидаемые потери.

В последующих версиях предполагается включить в систему модель [Бондаренко], позволяющую автоматически подбирать структуру валютного портфеля минимизирующую риск ожидаемых потерь и максимизирующую ожидаемую прибыль.

Система может использоваться не только риск-менеджерами для принятия решений о риске валютного портфеля, но и в учебных целях как тренажер для отработки у студентов навыков работы по управлению валютными рисками.

Заключение

В работе рассмотрено использования метода Монте-Карло для моделирования и анализа рисков валютного портфеля. На основании проведенных исследований была разработана система моделирования и анализа рисков валютного портфеля. Эта система была разработана на языке C++ и реализована в интегрированной среде проектирование Borland C++ Builder.

На основании приведенных в литературе данных для прогнозирования курсов валют можно использовать нормальное распределение, однако результаты будут более точными если распределение курсов валют будут строится как эмпирические распределения полученные на основе использования ретроспективных статистических данных о курсах валют. Поэтому, в работе была отработана методика построения такого эмпирического распределения курса валют.

С целью удобства работы, в систему было включено ряд блокировок затрудняющих пользователю выполнение ошибочные действий в процессе работы системы. Интерфейс системы, проектировался максимально удобным и наглядным для пользователя, что позволяет освоить работу с системой за минимальное время.

Благодарности

Работа опубликована при финансовой поддержке проекта **ITHEA XXI** Института информационных теорий и приложений FOI ITHEA Болгария www.ithea.org и Ассоциации создателей и пользователей интеллектуальных систем ADUIS Украина www.aduis.com.ua.

Библиография

[Бондаренко] Бондаренко В.Є. Модель оптимального формування структури кредитного портфеля. Актуальные проблемы финансово-денежной политики и трансформация экономики Украины. Приложение №9(14) к журналу “Персонал” №4(58), 2000, с.101-104. с.101-104.

[Лобанов,Чугунов] Энциклопедия финансового риск-менеджмента. /Под ред А.А.Лобанова и А.В.Чугунова. - М:, Альпина Паблишер, 2003. – 786 с.

Сведения об авторе

Виктор Бондаренко – *Киевский национальный экономический университет; просп. Победы, 54, Киев-047, Украина, 03047; e-mail: victorbondarenko@rambler.ru*

Philosophy and Methodology of Informatics

ЗАКОН БОЛЬШИХ ЧИСЕЛ ДЛЯ ГИПЕРСЛУЧАЙНОЙ ВЫБОРКИ

Игорь Горбань

***Аннотация**: Доказаны две теоремы, определяющие условия сходимости выборочного среднего гиперслучайной величины и алгоритм расчета состоятельных оценок границ начальных моментов гиперслучайных величин при быстрых изменениях статистических условий. Доказана для гиперслучайных событий теорема, определяющая предельные границы функции распределения частоты гиперслучайных событий при устремлении объема выборки к бесконечности. Доказано, что нижняя граница математического ожидания гиперслучайной величины не меньше математического ожидания верхней границы распределения, а верхняя граница математического ожидания гиперслучайной величины не больше математического ожидания нижней границы распределения.*

***Ключевые слова**: теория гиперслучайных явлений, выборка, сходимость.*

***ACM Classification Keywords**: G3 Probability and Statistics*

***Conference**: The paper is selected from XV[th] International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.*

Введение

Недавно была опубликована монография [Горбань, 2007], посвященная новой теории гиперслучайных явлений, под которыми подразумевается множество условных случайных явлений – множество условных случайных событий, величин и функций. Гиперслучайную величину X наиболее полно характеризуют условные функции распределения $F_{x/g}(x)$, соответствующие множеству условий $g \in G$, менее полно – верхняя и нижняя границы функции распределения $F_{Sx}(x)$, $F_{Ix}(x)$. Представление о гиперслучайной величине дают начальные и центральные моменты границ распределения, в частности, математические ожидания границ m_{Sx}, m_{Ix}, дисперсии границ D_{Sx}, D_{Ix} и пр., а также границы моментов, к числу которых относятся границы математического ожидания m_{sx}, m_{ix}, границы дисперсии D_{sx}, D_{ix} и др.

В книге особое внимание уделено методам формирования соответствующих состоятельных оценок: оценок условных функций распределения $F^*_{x/g}(x)$, оценок границ функции распределения $F^*_{Sx}(x)$, $F^*_{Ix}(x)$, оценок математических ожиданий m^*_{Sx}, m^*_{Ix} этих границ, оценок дисперсий границ D^*_{Sx}, D^*_{Ix}, оценок границ математического ожидания m^*_{sx}, m^*_{ix}, оценок границ дисперсии D^*_{sx}, D^*_{ix} и пр. Исследована сходимость этих оценок к соответствующим точным характеристикам и параметрам. Доказан ряд теорем, определяющих условия сходимости.

Описанная в книге процедура формирования оценок построена по следующей схеме. Для всего множества G статистических условий g формируется выборка

$$\vec{x} = \{x_1, ..., x_N / g \in G\},$$

по ней для каждого условия g в отдельности рассчитывается оценка условных характеристик и условных параметров: оценка условной функции распределения $F^*_{x/g}(x)$, оценка условного математического ожидания $m^*_{x/g}$, оценка условной дисперсии $D^*_{x/g}$ или др. По оценкам условной функции распределения $F^*_{x/g}(x)$ вычисляются оценки границ функции распределения

$$F^*_{Sx}(x) = \sup_{g \in G} F^*_{x/g}(x), \quad F^*_{Ix}(x) = \inf_{g \in G} F^*_{x/g}(x)$$

и оценки характеристик, характеризующие эти границы: оценки математических ожиданий границ m^*_{Sx}, m^*_{Ix}, оценки дисперсий границ D^*_{Sx}, D^*_{Ix} и пр. По оценкам условных параметров определяются оценки границ соответствующих величин, например, по оценкам условных математических ожиданий $m^*_{x/g}$ – оценки границ математического ожидания $m^*_{sx} = \sup_{g \in G} m^*_{x/g}$, $m^*_{ix} = \inf_{g \in G} m^*_{x/g}$, по оценкам условных дисперсий $D^*_{x/g}$ – оценки границ дисперсии $D^*_{sx} = \sup_{g \in G} D^*_{x/g}$, $D^*_{ix} = \inf_{g \in G} D^*_{x/g}$ и т.д.

В работах [Горбань, 2007, Горбань, 2006] отмечалось, что определенные трудности можно ожидать при формировании выборки $\vec{x} = \{x_1, ..., x_N / g \in G\}$ из-за сложности обеспечения, контроля и поддержания условий $g \in G$. Однако, в ряде случаев вопрос облегчается тем, что для расчета искомых характеристик не требуются знания того, в каких именно условиях получены условные характеристики или параметры. Главное, чтобы были представлены все возможные условия g множества G и в массив данных, используемый для расчета условных характеристик и параметров, не попадали данные, соответствующие другим условиям.

Иногда можно предположить, что статистические условия меняются непрерывно и относительно медленно. Кроме того, может быть известна максимально возможная скорость изменения условий. На основе этих данных нетрудно рассчитать максимальное число последовательных элементов выборки $N_{\max}$, для которых условия остаются практически неизменными. Если объем выборки $N_{\max}$ достаточен для получения условных оценок приемлемого качества, то алгоритм нахождения искомых граничных оценок сводится к сбору данных на большом интервале наблюдения, разбиению полученного массива на фрагменты по $N_{\max}$ последовательных элементов, расчету по ним промежуточных условных оценок, а затем вычислению граничных оценок.

Описанный алгоритм достаточно прост и прозрачен. Его недостатком является то, что требуется вспомогательная априорная информации о скорости изменении статистических условий и, то, что он применим лишь при достаточно медленной смене условий.

До настоящего времени остаются не ясными целый ряд вопросов, касающихся различных оценок характеристик и параметров гиперслучайных явлений.

Целью настоящей статьи является прояснение некоторых из них на основе доказываемых ниже теорем для гиперслучайных событий и величин. Первая теорема, аналогичная известной теореме Чебышева,

определяет условия сходимости выборочного среднего гиперслучайной величины, вторая теорема, базирующаяся на первой, обосновывает возможность расчета состоятельных оценок начальных моментов гиперслучайных величин при быстрых изменениях статистических условий. Третья теорема, аналогичная теореме Бернулли [Королюк, 1985, Гнеденко, 1961], определяет условия сходимости границ частоты гиперслучайного события, четвертая теорема устанавливает связь между математическими ожиданиям границ распределения и границами математического ожидания.

Теорема о сходимости границ выборочного среднего

Теорема 1. Пусть гиперслучайная величина $X = \{X / g \in G\}$ представляет собой совокупность случайных величин X / g для различных статистических условий $g \in G$ с условными математическими ожиданиями $m_{x/g}$ и ограниченными условными дисперсиями $D_{x/g}$. Нижняя и верхняя границы математического ожидания гиперслучайной величины X равны соответственно m_{ix}, m_{sx}. Из генеральной совокупности гиперслучайной величины X, полученной в неконтролируемо меняющихся статистических условиях, формируется гиперслучайная выборка $\{X_1,\ldots,X_N\}$ объемом N. По этой выборке рассчитывается гиперслучайное выборочное среднее $m_x^* = \frac{1}{N}\sum_{n=1}^{N} X_n$.

Тогда при устремлении объема выборки к бесконечности ($N \to \infty$) гиперслучайное выборочное среднее m_x^* сходится по вероятности к множеству чисел $m_x = \{m_{x/\vec{g}},\ \vec{g} = (g_1,...,g_N),\ g_n \in G,\ n = \overline{1,N}\}$, представляющих собой множество средних $m_{x/\vec{g}} = \frac{1}{N}\sum_{n=1}^{N} m_{x_n/g_n}$ условных математических ожиданий $m_{x_1/g_1},\ldots,m_{x_N/g_N}$ случайных величин $X_1/g_1,\ldots,X_N/g_N$, рассчитанных для всевозможных $g_n \in G,\ n = \overline{1,N}$, а нижняя и верхняя границы этого выборочного среднего сходятся соответственно к нижней и верхней границам математического ожидания гиперслучайной величины X:

$$\begin{aligned} &\lim_{N\to\infty} P(\left|\inf m_x^* - m_{ix}\right| > \varepsilon) = 0, \\ &\lim_{N\to\infty} P(\left|\sup m_x^* - m_{sx}\right| > \varepsilon) = 0, \end{aligned} \qquad (1)$$

где $P(z)$ – вероятность условия z, а ε – как угодно малое положительное число.

Для доказательства теоремы рассмотрим случайную выборку $\{X_1/g_1,\ldots,X_N/g_N\}$, полученную при фиксированной последовательности условий $g_1,\ldots,g_N \in G$. Рассчитанное по ней выборочное среднее $m_{x/\vec{g}}^* = \frac{1}{N}\sum_{n=1}^{N} X_n/g_n$ при устремлении объема выборки N к бесконечности согласно теореме Чебышева сходится по вероятности к среднему условных математических ожиданий $m_{x/\vec{g}}$:

$$\lim_{N\to\infty} P(\left|m_{x/\vec{g}}^* - m_{x/\vec{g}}\right| > \varepsilon) = 0 .$$

Сходимость случайной величины $m_{x/\vec{g}}^*$ к $m_{x/\vec{g}}$ для всех $\vec{g} = (g_1,...,g_N),\ g_n \in G,\ n = \overline{1,N}$ означает сходимость по вероятности гиперслучайного выборочного среднего m_x^* к множеству чисел m_x.

При любой фиксированной последовательности условий и $N \to \infty$ выборочное среднее $m^*_{x/\bar{g}}$ ограничено интервалом $[\inf m^*_x,\ \sup m^*_x]$, а среднее условных математических ожиданий $m_{x/\bar{g}}$ – интервалом $[m_{ix},\ m_{sx}]$. При этом минимальному среднему значению m_{ix} среднего математического ожидания соответствует нижняя граница $\inf m^*_x$ выборочного среднего, а максимальному среднему значению m_{sx} – верхняя граница $\sup m^*_x$ выборочного среднего, т. е. справедливо равенство (1).

Теорема о сходимости оценок границ выборочного среднего

Теорема 2. Пусть гиперслучайная величина X имеет границы математического ожидания m_{ix}, m_{sx} и конечную верхнюю границу дисперсии D_{sx}. Из генеральной совокупности гиперслучайной величины X в неконтролируемо меняющихся статистических условиях формируется L непересекающихся гиперслучайных выборок $(X_{11},...,X_{N1}),...,(X_{1L},...,X_{NL})$ объемом N каждая ($L \geq 2$). Пусть $m^*_{x_1},...,m^*_{x_L}$ – соответствующие этим выборкам гиперслучайные выборочные средние:

$$m^*_{x_1} = \frac{1}{N}\sum_{n=1}^{N} X_{n1},...,m^*_{x_L} = \frac{1}{N}\sum_{n=1}^{N} X_{nL},$$

а m^*_{ix}, m^*_{sx} – гиперслучайные оценки границ выборочных средних $m^*_{x_1},...,m^*_{x_L}$:

$$m^*_{ix} = \underline{\inf}_{l=\overline{1,L}} m^*_{x_l},\quad m^*_{sx} = \sup_{l=\overline{1,L}} m^*_{x_l}. \tag{2}$$

Тогда при неограниченном увеличении количества выборок и объема каждой выборки оценки границ выборочного среднего m^*_{ix}, m^*_{sx} сходятся по вероятности к соответствующим границам m_{ix}, m_{sx} математического ожидания гиперслучайной величины X:

$$\lim_{L\to\infty}\lim_{N\to\infty} m^*_{ix} = m_{ix},\qquad \lim_{L\to\infty}\lim_{N\to\infty} m^*_{sx} = m_{sx}.$$

Доказательство теоремы состоит в следующем. Согласно теореме 1, для любой l-ой выборки при устремлении объема выборки к бесконечности границы гиперслучайного выборочного среднего $m^*_{x_l}$ стремятся по вероятности к границам математического ожидания m_{ix}, m_{sx} гиперслучайной величины X. Это означает, что область значений гиперслучайного выборочного среднего $m^*_{x_l}$ ограничена интервалом, стремящимся к интервалу $[m_{ix}, m_{sx}]$.

Принимая во внимание равенства (2), при $L \to \infty$ и $N \to \infty$ оценки границ выборочного среднего m^*_{ix}, m^*_{sx} сходятся по вероятности к тем же границам m_{ix}, m_{sx}.

Из теоремы следует, что состоятельные оценки границ математического ожидания гиперслучайной величины могут быть получены путем вычисления средних по множеству отсчетов для множества выборок и расчета по ним искомых границ.

Нетрудно убедиться, что таким же путем можно вычислять состоятельные оценки границ $m^*_{ix\nu} = \underline{\inf}_{l=\overline{1,L}} m^*_{x_l\nu}$, $m^*_{sx\nu} = \sup_{l=\overline{1,L}} m^*_{x_l\nu}$ начальных моментов любого порядка ν, где $m^*_{x_l\nu}$ – оценка начального момента ν-го порядка, соответствующая l-ой выборке.

Следует заметить, что рассчитывать границы центральных моментов μ^*_{ixv}, μ^*_{sxv} по описанной выше схеме нельзя из-за отсутствия необходимой для этого информации об оценках условных математических ожиданий.

Теорема, аналогичная теореме Бернулли

Теорема 3. Пусть в неконтролируемо меняющихся статистических условиях проводится серия N независимых опытов, в каждом из которых может произойти некоторое событие A, рассматриваемое как гиперслучайное событие, представляемое множеством случайных событий A/g в условиях $g \in G$: $A = \{A/g \in G\}$.

Вероятность появления события A/g в фиксированных условиях $g \in G$ равняется $p_{a/g}$. Нижняя и верхняя границы вероятности события A соответственно равны P_{Ia}, P_{Sa}. Частота появления события A в рассматриваемой серии опытов описывается выражением $\Xi = \frac{N_a}{N}$, где N_a – число опытов, в которых произошло событие A. Эта частота – гиперслучайная величина, представляемая множеством случайных величин Ξ/g ($g \in G$): $\Xi = \{\Xi/g \in G\}$.

Тогда границы $F_{I\xi}(\xi)$, $F_{S\xi}(\xi)$ функции распределения $F_{\xi}(\xi)$ частоты Ξ при $N \to \infty$ сходятся по вероятности к функциям единичного скачка в точках P_{Ia}, P_{Sa}.

Доказательство основано на теореме 1. Гиперслучайное событие A можно рассматривать как гиперслучайную величину X, принимающую значение, равное единице, когда происходит событие A, и значение, равное нулю, когда событие не происходит. Условное математическое ожидание $m_{x/g}$ случайной величины X/g равно $p_{a/g}$, а условная дисперсия

$$D_{x/g} = (1-p_{a/g})^2 p_{a/g} + (0-p_{a/g})^2 (1-p_{a/g}) = p_{a/g}(1-p_{a/g})$$

есть величина ограниченная. Границы математического ожидания гиперслучайной величины X равны P_{Ia}, P_{Sa}. Выборочное среднее гиперслучайной величины X представляет собой частоту $\Xi = \frac{N_a}{N}$ появления события A. Тогда на основании теоремы 1 справедливо утверждение доказываемой теоремы. Из теоремы следует, что при увеличении числа опытов $(N \to \infty)$ частота появления события A вне зависимости от количества условий G (если $G \neq 1$) находится в диапазоне $[P_{Ia}, P_{Sa}]$, не стремясь к какому-то конкретному значению.

Теорема о соотношении границ математического ожидания и математического ожидания границ

Теорема 4. Пусть m_{ix} и m_{sx} – соответственно нижняя и верхняя границы математического ожидания гиперслучайной величины $X = \{X/g \in G\}$ с функциями распределения $F_{ix}(x)$, $F_{sx}(x)$, а m_{Sx} и m_{Ix} – математические ожидания соответственно верхней $F_{Sx}(x)$ и нижней $F_{Ix}(x)$ границ функции распределения этой гиперслучайной величины.

Тогда $m_{ix} \geq m_{Sx}$ и $m_{sx} \leq m_{Ix}$.

Для доказательства теоремы рассмотрим функции распределения $F_{ix}(x)$ и $F_{Sx}(x)$. На одних участках эти функции могут совпадать, на других – нет. На интервалах, где они не совпадают, кривая $F_{ix}(x)$ располагается правее кривой $F_{Sx}(x)$. Поэтому $\int_0^1 x\mathrm{d}F_{ix}(x) \geq \int_0^1 x\mathrm{d}F_{Sx}(x)$, т.е. $m_{ix} \geq m_{Sx}$. Аналогично доказывается неравенство $m_{sx} \leq m_{Ix}$.

Выводы

1. На базе известной теоремы Чебышева для случайных величин доказаны для гиперслучайных величин две теоремы, определяющие условия сходимости выборочного среднего гиперслучайной величины к границам ее математического ожидания и алгоритм расчета состоятельных оценок границ начальных моментов гиперслучайных величин при быстрых изменениях статистических условий.

2. Доказана для гиперслучайных событий теорема, аналогичная теореме Бернулли для случайных событий, определяющая границы функции распределения частоты гиперслучайных событий при устремлении объема выборки к бесконечности.

3. Доказано, что нижняя граница математического ожидания гиперслучайной величины не меньше математического ожидания верхней границы распределения, а верхняя граница математического ожидания гиперслучайной величины не больше математического ожидания нижней границы распределения.

Благодарности

Работа опубликована при финансовой поддержке проекта **ITHEA XXI** Института информационных теорий и приложений FOI ITHEA Болгария www.ithea.org и Ассоциации создателей и пользователей интеллектуальных систем ADUIS Украина www.aduis.com.ua.

Литература

[Гнеденко, 1961] Гнеденко Б.В. Курс теории вероятностей. – М.: Изд-во физ.-мат. литературы, 1961. – 406 с.

[Горбань, 2006] Горбань И.И. Оценки характеристик гиперслучайных величин // Математичні машини і системи. – 2006. – № 1. – С. 40 –48.

[Горбань, 2007] Горбань И.И. Теория гиперслучайных явлений К.: Інститут проблем математичних машин и систем НАН Украины, ГП «УкрНИУЦ» Госпотребстандарта Украины, 2007. – 184 с (http://ifsc.ualr.edu/jdberleant/intprob/).

[Королюк, 1985] Королюк В.С. и др. Справочник по теории вероятностей и математическая статистика. – М.: Наука, 1985. – 637 с.

Сведения об авторе

Игорь Горбань *– главный научный сотрудник ИПММС НАН Украины, доктор технических наук, профессор, Украина, 03187, Киев, пр. академика Глушкова. 42; e-mail: igor.gorban@yahoo.com.*

КРАТКИЙ МЕТОДОЛОГИЧЕСКИЙ МЕМОРАНДУМ –ЧАСТЬ ПЕРВАЯ

Анатолий Крисилов, Екатерина Соловьева, Авенир Уемов

– Ищи карту, Джим! – говорил Билли Бонс. – Они здорово ее запрятали.
Но вся важная информация – на ней!
Р.Л. Стивенсон

"Лучше найти системное объяснение, чем занять царский престол".
Демокрит, Ю. Урманцев

***Аннотация**: Работа посвящена методологии системного исследования сложных и слабоформализованных объектов. Показаны преимущества системологического подхода, приведено содержание этапов системного анализа; затронуты вопросы инженерии и менеджмента знаний.*

***Ключевые слова**: системологическое исследование, системный анализ, описание плохо поддающихся структуризации и слабоформализуемых объектов, системные параметры, когнитивные задачи.*

***ACM Classification Keywords**: H.1.1 – Systems and Information Theory – General Systems Theory.*

***Conference**: The paper is selected from XVth International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.*

1. Вводные замечания

Англичане говорят, что если хочешь быть понятым, когда ты о чем-то говоришь, то сделай это в три этапа: сперва скажи, о чем ты будешь говорить, потом это самое толково расскажи, а потом объясни, о чем ты, собственно, говорил.

В настоящей работе в сжатом виде будут изложены основные соображения о смысле и содержании методологического исследования, системного (системологического) подхода, и о том, почему полезно его придерживаться. Затем будет вкратце описано содержание основных шагов системного анализа и высказан ряд соображений по тому, что можно назвать системным синтезом.

Практически весь материал изложен в конспективном виде: почти каждое положение нуждается в расширении, обсуждении, детализации. Однако при этом неизбежны «расползание» материала и утрата целостности. С другой стороны – работа ограничена принятым форматом. Слабая надежда – на ч. II.

Системный анализ возник (и формируется) как результат стремления конкретно применить идеологию общей теории систем (ОТС) в *практике управления* сложными системами. Одним из фундаментальных положений теории систем является представление сложного объекта (то есть такого, который с трудом поддается наблюдению и описанию, формализации) – в виде целенаправленной системы или, что весьма близко по смыслу, – имеющей внешнюю функцию. По существу своему системный анализ представляет собой совокупность методов исследования таких свойств и отношений в системе, которые именно так (как целенаправленную) ее характеризуют, и изучению связей в этой системе как взаимоотношений между ее целями и способами их реализации.

Отметим, что интенсивное развитие излагаемых в настоящем материале положений приходится на последние полтора-два десятка лет, однако очень серьезные, можно сказать, основополагающие результаты имели место и гораздо ранее (см., напр., [1 – 4] и др.) В настоящее время можно указать на два «встречных» фактора, обеспечивающих эту интенсификацию. С одной стороны, разворачивающийся ноосферный этап развития выдвигает все более жесткие, императивные требования в описании и понимании происходящих (проявленных и непроявленных!!) чрезвычайно сложных и не всегда и/или полно по́нятых процессов и событий. С другой, – имеет место появление все более сильных, все более разработанных и операциональных приемов, методов, теорий, аналитических и «конструирующих» средств, – составляющих в совокупности и глубокую аппаратную базу для такой работы, и новую атмосферу для нового понимания. (Есть восточная поговорка: «Сколько Аллах придумал болезней, столько премудрый сотворил и лекарств», важно лишь не пропустить их). Тем не менее, вряд ли следует утверждать, что процесс формирования такого совершенного инструментария можно считать завершенным, хотя фундаментальные и прикладные результаты налицо – от эффективного решения производственных и иных управленческих задач до CASE-технологий, реинжиниринга бизнес-процессов (BPR) и введения на Украине профессии «аналитик консолидированной информации» (см. [5 – 6] и др.).

Несомненно, многие задачи можно решать с позиций здравого смысла. Но игнорировать инструментарий, специально разработанный для анализа и описания сложных ситуаций, для поиска грамотных решений в условиях значительной неопределенности, для работы со слабоформализуемыми объектами и ситуациями – представляется сегодня просто несерьезным и несвоевременным.

Не так уж редки случаи, когда достаточно некоторому автору выяснить, что объектом его исследования является сложная система, – как все, что он с этим объектом делает, он называет системным подходом. На деле же в этом его исследовании системным подходом и не пахнет. Эпитет «системный» есть характеристика инструмента, с которым субъект исследования подходит к своему объекту, он (эпитет) отражает состав и структуру этого подхода, а не свойства объекта. Когда Наполеон говорил: «Большие батальоны всегда правы», – это было характеристикой инструмента Наполеона, а не того объекта, к которому он с этим инструментом «подходил».

Здесь следует отметить (в контексте нашего «меморандума»), что пугающая разноголосица с определением системы резко уменьшится, если введено дуальное представление о системе – в зависимости от выбранной или наблюдаемой системообразующей характеристики: а) система – это произвольный объект, на котором реализуется какое-то отношение, обладающее некоторым определенным свойством; б) система – это произвольный объект, обладающий какими-то свойствами, которые находятся в некоторых определенных отношениях. Для ноосферных (и, практически, для всех *управленческих и организационных*) задач следует рассматривать систему вместе с ее окружением, надсистемой; в этом случае под системой понимается функциональный объект, свойства которого обусловлены функцией этого объекта, сводящейся, в конечном счете, к поддержанию определенных свойств надсистемы. Как видим, на один из верхних уровней выходит задача *анализа функций* объекта.

Какие доводы следует высказать, настаивая на пользе системологического подхода, в частности, для ряда направлений и работ, представляемых на конференциях KDS?

2. Некоторые аргументы

В первую очередь обратим внимание на то, что системологический аппарат представляет собой не рамку, в которую вписывается данное исследование, не внешнюю форму, в которой будет представлена некоторая структура или процесс. Говоря метафорически, этот инструментарий есть некий стержень, опорно-двигательная система исследования, т. е. совокупность мышц, скелета, всей нервной системы,

которые совместно и управляют процессом, и осуществляют само движение. Более того, – этот аппарат позволяет выявить содержание, сущность, т. е. является инструментом познания.

Уместно здесь указать разницу между методикой и методологией. Если методика, методическое руководство отвечает на вопрос: «**Как** нечто надо делать?», то методология отвечает на вопрос: «**Что** в этом нечто надо делать?» Это означает, что методология определяет, какие задачи следует решить, чтобы описать/представить твой объект (кем бы или чем бы он ни был) как целостное, не «комплексное», а единое образование, с его структурой (то есть совокупностью элементов и связей между ними), с его окружением (то есть чтó представляет собой мир, в котором он живет), с его, объекта, свойствами, отношениями между этими свойствами и т. д.

Основное свойство системологии – ее выраженная (и универсальная!) методологическая направленность. Будучи интерпретированной в терминах конкретной науки, системология может, как указывалось выше, очень эффективно выполнять в этой науке методологические функции. Одновременно системология может использоваться для содержательной и конструктивной обработки создаваемого данной наукой знания, вплоть до синтеза; в этой роли она выступает одним из мощных средств «организации» ноосферы (как сферы знания).

По сути, речь идет о получении (построении) системных знаний об исследуемом объекте – в данной предметной области. И системологический инструментарий является, т. о., средством для формирования таких знаний; отметим – *data mining* и *knowledge discovery* в чистом виде.

Далее, поскольку объектом исследований, скажем, в рамках проблематики KDS (и, разумеется, многих других) являются, как правило, именно сложные, плохо поддающиеся структурированию и формализации процессы, ситуации, явления, то понятно, что **адекватным** этим задачам будет именно их системный анализ. Мало того, можно назвать случаи, когда методологический аппарат давал и в применении его к относительно простым объектам новую информацию о них, открывал в них новые свойства, выявлял не замеченные ранее исследователями их связи с внешним миром и т. д.

К числу весомых примеров полезности применения аппарата системологии можно также отнести:

- нацеленность на выявление контекста (без привлечения контекста крайне трудно дать смысловое описание задачи) и получение глубинных, объективных знаний;
- оценка соответствия сформулированных целей – системе требований [7];
- построение валидных прогнозов;
- выявление типа отношений в исследуемой системе (элементы данной системы находятся в субъект-объектных отношениях или в субъект-субъектных? – наше руководство считает себя субъектом, а все предприятия – объектами, серьезное заблуждение) – и ряд других.

Во многих практических сферах интуитивно пришли к системологии, например, попытка учета контекста в объектном подходе, анализ бизнеса для повышения конкурентоспособности, др.

Немаловажным аспектом является и такой. Порой разные исследователи работают с одним и тем же или с очень сходными объектами, однако описывают их по-разному, используют свою терминологию, строят свои описания на разных языках. Ясно, что сравнить результаты этих исследований или составить общий портрет этого объекта становится весьма затруднительным. Знание основ системологии и применение методологических средств резко облегчает эту проблему, давая разработчикам общий язык и общий аппарат, решая таким образом задачу организации междисциплинарных связей и т. д.

Наконец, есть еще одна группа причин, объясняющих актуальность методологических, системологических постановок, – это видимая (и невидимая!) катастрофичность нынешних характеристик социогенеза в целом. Некоторые из этих вопросов будут кратко рассмотрены в заключении.

3. Как формируются системные представления об объектах

Картину формирования целостных знаний об объекте исследования, получения системного представления о нем, – можно описать с помощью определенной *геометрической аналогии* [8].

Представим себе некоторую плоскость, в которой располагается содержательная информация из определенных наук (такую плоскость называют онтологической). Там лежат предметные представления о географии (где и какие находятся континенты и океаны, горные хребты и реки, где какой климат, где и какая обитает биота и т. д.), об экономике (уклады, ресурсы, предприятия, транспорт, рынки,..), о социальной истории (хронология событий, императоры и военачальники, становление и падение цивилизаций,..), о биологии, о сопротивлении материалов и так далее.

Перпендикулярно к этой, скажем, горизонтальной плоскости существует другая плоскость, вертикальная. В ней обретаются совсем другие понятия и категории: принципы, закономерности, механизмы, критерии, понятие о структуре и инфраструктуре и т. д., – это методологическая плоскость.

Так вот, настоящее знание, целостное и системное представление об объекте (процессе, ситуации, явлении) формируется в объеме, образованном двумя указанными плоскостями. В этот объем (скажем, в гиперсферу, репрезентирующую объект данного исследования) из онтологической плоскости берутся конкретные предметные данные об этом объекте (например, имя предприятия, его персонал, выпускаемая продукция, доходы, культура,..), а из методологической плоскости – понятия о его целях и/или функциях, о характере его структуры, о степени удовлетворения им запросов охватывающей его системы и т. д. Из этих организованных определенным способом данных (отчет, монография, видеофильм) и формируется целостное представление об объекте, системное знание о нем; оно и является настолько ценным, что, по мнению компетентных специалистов, его следует предпочесть царскому трону, как об этом сказано в одном из эпиграфов.

Некоторым может показаться, что эти соображения как-то далеки от практики, что они, мягко говоря, заумны, что в них многовато абстрактного…

Со всей определенностью следует сказать, что это глубоко ошибочное мнение. Замечательная английская исследовательница мыслительной деятельности детей Маргарет Доналдсон, ученица и последователь Жана Пиаже, говорит: «Именно интеллектуальная работа имеет самое непосредственное отношение к реальной практической жизни… Чтобы разбираться в реальном мире с максимальной компетенцией (здесь и далее выделение авторов), необходимо учитывать (знать!) структуру вещей, необходимо овладеть умениями работать с системами и научиться абстрагировать формы и схемы (то есть, проводить глубинный системный анализ, научиться выходить на мета-уровень, – авторы). Парадокс в том-то и состоит, что наиболее значимые успехи в практической деятельности, например, в технике, были бы невозможны, если бы мы стали игнорировать трудную задачу – действовать без опоры на мир знакомых (известных, видимых) явлений». И далее: «Вот истина, которую человек как биологический вид только начинает постигать. Если бы нам когда-нибудь пришлось отказаться от этой деятельности, то плата была бы слишком велика» [9].

Добавим, что сегодня эти соображения актуальны как никогда ранее. Именно «абстрагирование форм и схем», «учет структуры вещей» и событий, «умение работать с системами» – лежат в основе формирования ноосферной парадигмы, более того – в разработке механизмов ее внедрения в массовое сознание. Роль методологии в этой работе трудно переоценить.

4. Основные этапы системного анализа

В последнее время стала модной идиома «дорожная карта». В это понятие вкладывают разные смыслы – последовательность ориентиров движения, план работы, перечень заданий,… Обратим внимание на то, что эти понятия носят, в основном, номинативный характер, они лишь обозначают жанр документа, являются именами, почти не раскрывая содержания и смысла (работы, действия и т. п.)

По отношению к системному анализу можно провести аналогию лишь с такой картой местности, которую топографы называют «поднятой», на которой не только изображены детали рельефа, географические и хозяйственные объекты, но и показаны связи и расстояния между ними, описано их окружение, приведены их детальные характеристики. Но даже и в этом случае, хоть образ и хорош, речь может идти лишь о достаточно далекой аналогии, о самом начальном разъяснении.

Одно из первых кратких изложений этапов системного анализа содержится в книге В. М. Глушкова «Введение в АСУ», конец 60-х – начало 70-х годов. За истекшие 30 – 40 лет имели место определенные коррективы и усовершенствования, разными авторами уточнялось содержание некоторых этапов, появлялись новые и т. д. В приведенном ниже описании шаги системного анализа сцеплены друг с другом, однако и здесь возможны варианты. Например, анализ ресурсов (для исследования или для выполнения работы) можно проводить после постановки задачи или перед моделированием, проводить рефлексию после каждого этапа, некоторые шаги выполнять параллельно и т. д. Многое зависит от содержания задачи, объема известной информации, характеристик исследователя и пр.

Содержание системного анализа описывается, в основном, следующими этапами:

1. Самоопределение.
2. Целеполагание; постановка задачи.
3. Первичная структуризация.
4. Декомпозиция или вторичная структуризация.
5. Выявление основных принципов и закономерностей, действующих в исследуемой системе.
6. Основные качественные и количественные характеристики данной системы.
7. Анализ основных процессов, протекающих в системе, и действующих в ней механизмов.
8. Анализ ресурсов.
9. Моделирование.
10. Анализ результатов; формулировка новых целей.

Весьма часто моделирование включает в себя и синтез; можно сказать, что в системном исследовании не всегда легко проложить четкую границу между анализом и синтезом.

5. Первая фаза системного анализа

Самоопределение. Это чрезвычайно важный и один из наиболее сложных этапов. Здесь надо разобраться – кто мы есть по существу, на самом деле? где мы находимся вообще и в данное время? во что мы погружены? с чем имеем дело?

У Руссо есть такая фраза: «Истиной сознания является самосознание, но чтобы достичь этой истины, сознание должно охватить то, что лежит вне его», и, добавим, оперировать этим. Помните, как Мюнхгаузен, современник Руссо, вытягивал себя из болота? То, что должен проделать исследователь на этом этапе, очень похоже на действия знаменитого барона, по крайней мере – по сложности. Мысль Ж.-Ж. Руссо хорошо объясняет эту сложность. Отметим, что и II теорема Геделя о неполноте говорит о том же.

Правильная и полная реализация этого шага в очень большой степени определяет успех работы.

Каких качеств требует выполнение самоопределения? Объективность, логичность, умение непредвзято видеть себя со стороны, широта охвата и последовательность, самокритичность и «самоответственность» – вот примерный букет необходимых качеств.

Техника, бизнес, история знают десятки примеров правильного, адекватного самоопределения и сотни тысяч – неправильного, неадекватного задаче и ситуации, безграмотного. Подавляющее число ошибок в нашем родном управлении является следствием того, что наши «управители» не проводят, не имеют адекватного самоопределения, а порой – вообще никакого.

Если при выполнении последующих этапов вы увидите или почувствуете, что что-то не ладится, не вяжется – вернитесь к этому этапу, скорей всего окажется, что еще что-то важное не учтено.

Целеполагание. О значимости этого этапа сказано и написано много хорошего. Сколько-то времени назад сформировалась даже такая наука, которую называют телеологией или аксиологией (с использованием греческого и латинского корня соответственно). Ограничимся несколькими замечаниями.

На самом деле в своей работе нужно сформулировать (выявить, сконструировать, определить) две цели. Одна – это цель той системы, к которой принадлежит исследуемый объект (по отношению к нему эта охватывающая система называется внешней или надсистемой). Вторая – это собственно цель твоего объекта, в системологии ее еще называют внешней детерминантой исследуемой системы, внешним запросом, определяющим функциональное назначение данной системы; этим названием сразу выстраивается представление о иерархическом характере отношений между исследуемой системой и той, что ее охватывает; эти отношения далеко не так просты, как может показаться с первого взгляда. Естественно, указанные две цели не должны быть конкурентными и, тем более, конфликтными.

Какими свойствами должна обладать формулируемая цель? Этих свойств, по крайней мере, четыре: цель должна быть **конкретна**, **понятна**, **измерима** и **не быть процессом** [7].

На верхнем уровне может быть не одна цель, а две, три. Например, для муниципалитета можно назвать такие две цели: повышение благосостояния населения города и повышение эффективности общественного производства. Поскольку город является открытой системой, то понятно, что вторая цель не является способом достижения первой. Мало того, практически сразу видно, что эти две цели являются конкурентными, а подчас и конфликтными. Конкурируют они за все виды ресурсов, включая кадровый ресурс и внимание управляющей сферы [10]. Таким образом, возникает третья цель, не так уж очевидная ранее: сформировать систему гармоничного, сбалансированного сочетания двух названных выше целей; к сожалению, ни одна из партий, из политических команд не ставит перед собой в явном виде эту третью цель, управление ведется стихийно, конъюнктурно, результаты – очевидны.

На этом этапе (так же, впрочем, как и на предыдущем) возникает необходимость в применении элементов системного синтеза. Здесь это происходит потому, что цели следует формулировать исходя из определенной совокупности требований. Возникает необходимость в инженерии требований к исследуемой или проектируемой системе, требований количественных и качественных, более сложных и более простых, в оценке их иерархии и т. д. И систему целей следует формулировать/синтезировать в соответствии с этой осмысленной выверенной совокупностью требований.

Первичная структуризация. На этом этапе объект исследования рассматривается как «черный ящик», как непрозрачная система, – исследуется вся доступная разработчику гамма его, объекта, отношений с охватывающей системой, надсистемой. Какие операции здесь следует проделать?

а) выяснить родо-видовые характеристики исследуемого объекта – к какому роду (и виду) он относится, какие вообще существуют роды (классы) и виды в его систематике; как он с ними соотносится?

б) провести все мыслимые его границы; например, для города – территориальные, демографические (кого включать в состав населения этого города – имеющих прописку, живущих без нее, прописанных, но отсутствующих и т. д., это зависит от поставленных целей – для обеспечения жильем **или** хлебом, транспортом **или** школами; может, придется вернуться к целеполаганию и провести там ревизию); далее – границы культурные, экономические, производственные, экологические и т. п.

в) выписать вектор **X** = f (φ, t) входных воздействий на анализируемую систему (φ – характери-стики внешней среды, надсистемы; t – время); здесь выявляется, какие ресурсы потребляет данная система и какие действуют на нее факторы извне;

г) выписать вектор **Y** = f (x, A, t) выходных характеристик рассматриваемой системы; здесь A – некоторое обобщенное (например, структурно-функциональное) представление этой системы.

Важная информация о системе будет получена, если мы ответим на вопросы о трех аспектах системы, называемых системными дескрипторами: концепт, субстрат и структура (причем, как это видно из дальнейшего, такой поиск уже есть переход к следующему шагу – вторичной структуризации). Исходным дескриптором является концепт (от слова «концепция», чтó лежит в основе определения системы?), далее – субстрат (из чего система «сделана»?) и структура системы (как она построена?).

Дескриптор «концепт» тесно связан с понятием цели, с выявлением внешней функции системы. В зависимости от конкретного системообразующего фактора, концепт может быть атрибутивным или реляционным, то есть описывать свойства системы или связи, отношения в ней.

Для большого класса сложных задач (а в нашей работе подразумевается анализ гибридных систем, например, с социальным компонентом) первичная структуризация – очень важный этап.

Вторичная структуризация или декомпозиция. Настоящий этап можно отнести к одному из наиболее емких в системном анализе. Здесь объект исследования перестает быть «черным ящиком», изучению и описанию подлежит, собственно, его структура.

Морфологический анализ объекта может быть выполнен различными путями. По-видимому, целесообразно начать эту работу, исходя из наиболее общих представлений и оперируя такими понятиями, как «вещи», «свойства» и «отношения». Как организован объект исследования, какими свойствами обладают его элементы и какими связаны отношениями, – вот начальный набор вопросов, ответы на которые могут быть получены в рамках ЯТО, языка тернарного описания [4].

Результатом исследования структуры объекта будет описание его состава (т. е. из каких элементов он состоит) **плюс** описание отношений между этими элементами **плюс** описание свойств этих отношений и свойств самих элементов, – здесь большой объем работы. Следует разобраться, имеют ли элементы одинаковую природу (система гомогенна) или они обладают разной природой (система гетерогенна), имеют они одинаковые масштабы или различны и т. д. Отметим, что в двух приведенных примерах (о свойствах природы элементов и о сходстве/различии их масштабов) мы оперировали *атрибутивным системным параметром*: основание деления систем на гомогенные и гетерогенные, открытые и замкнутые, стабильные и нестабильные и т. д. – предполагает системное рассмотрение соотносящихся и соотносимых вещей. В нашем случае использовался *бинарный* атрибутивный параметр.

Очень важно определить тип структуры – она может быть иерархической (как структура армейских подразделений), гетерархической (иметь несколько более или менее равноправных центров управления), или вообще – анархической, не иметь выделенных центров управления. Наиболее простой иерархической структурой является древовидная, можно назвать системы звездчатые, кольцевые и др.

Примером *линейного* атрибутивного параметра (в отличие от бинарных, где возможны только два значения) является простота/сложность рассматриваемой системы. Применение здесь системных

дескрипторов дает исследователю значительные преимущества, позволяя изучить (и затем использовать в модели!) *различные типы* простоты/сложности: концептуальную, структурную, субстратную, структурно-субстратную, оценить меру сложности и т. д. [11].

При завершении данного этапа мы уже располагаем довольно весомой информацией.

6. Вторая фаза системного анализа

Последующие шаги значительно содействуют выявлению, описанию и анализу функций системы.

Выявление основных принципов и закономерностей. Действует ли в рассматриваемой системе закон сохранения (вещества, энергии, информации)? По каким законам происходят изменения в системе? Какие вообще законы и закономерности имеют место в системе? (Например, соблюдаются ли и в какой мере II и III законы Барри Коммонера: «все должно куда-то деваться» и «за все нужно платить»?).

Действует ли принцип конкуренции? Принцип взаимопомощи? В какой мере соблюдается согласованность между элементами или синхронность в их действиях? Как осуществляется баланс между принципами централизации (центростремительные тенденции) и децентрализации – автономизации?

Анализ основных процессов и действующих в системе механизмов. Очень важно выявить природу происходящих в системе процессов – какие из элементов развиваются, какие неизменны, какие деградируют; какова природа этих изменений – непрерывно или скачками, причины бифуркаций и т. д.

Далее, следует определить, какие механизмы в системе действуют (и для каких элементов):

- как выглядит и чем запускается механизм изменчивости, какими факторами определяется возникновение мутаций (элементов, их свойств, связей между ними);
- как происходит отбор изменений и каковы его критерии, что и почему наследуется при этом, а что и почему отвергается, какова роль памяти системы в этом и каков ее механизм;
- как закрепляются новации в системе, как и в какой мере производится различение между «унаследованными» и «благоприобретенными» инновациями и кто/что оценивает их качество – и т. д.

Качественные и количественные характеристики системы В этой части предметом анализа выступают количественные и качественные признаки основных элементов системы, их свойств и отношений, динамика этих характеристик во времени и т. д. Первый пример: качественная и количественная оценка жизненного цикла самой системы и ее основных элементов. Второй пример: если в системе присутствует адаптация, то рассмотреть, в чем именно и каково ее качество – патологическая (наступает дисфункция), компенсаторная (за счет других связей или элементов) или нормальная, функциональная адаптация; имеет ли место самовоспроизводство и т. д.

Анализ ресурсов. Этот этап работы (выше говорилось, что он может располагаться и в другом месте, дело зависит от содержания задачи, условий работы и т. д.) включает в себя анализ и оценку, по крайней мере, двух групп ресурсов: а) какие ресурсы необходимы исследователю для достижения поставленных целей (время, кадры, исходная информация, знания о предметной области, измерительная техника и т. д.) и б) какие ресурсы использует или должна использовать рассматриваемая система для выполнения своих функций, своевременно ли они поступают, какие обменные соотношения между ними существуют, каковы ограничения и пр. Немаловажным является вопрос о том, какие и сколько тратить в системе ресурсов на развитие, а сколько – на поддержание стабильного функционирования. Кто отвечает за ресурсы, предоставляет их, кто распределяет, каков у этого персонала уровень профессионализма?

7. Моделирование (построение системной модели)

Значительная часть предыдущего проводилась, по сути, для того, чтобы построить возможно более полную модель нашей системы и на ней практически исследовать все необходимое.

Сегодня арсенал моделей очень велик, «модельеры» имеют очень широкий выбор: от моделей феноменологических, где моделируется лишь явление, феномен, до сложных полиситуационных имитационных моделей. Виды моделей и технологии системного моделирования представлены в разных источниках. Весьма емкий и последовательный материал именно по системным моделям содержится в работе [6]; здесь же укажем на некоторые важные, в том числе – и в методологическом отношении, **объекты** моделирования. К их числу относятся:

- Моделирование взаимодействия сложных объектов, с разными целями и интересами; исчисление этого взаимодействия, оценка синергического эффекта; исследование факторов влияния.
- Анализ зоны взаимных интересов и приемлемых оценок, поиск открытых компромиссов.
- Моделирование мотиваций, коллективных и индивидуальных, их природы и зависимостей.
- Положительные и отрицательные обратные связи, их влияние на согласованность и разлад.

Некоторые из этих вопросов предполагается рассмотреть в продолжении настоящей работы.

Наконец, по результатам анализа – определение новых целей, например, в рамках синтеза.

8. Некоторые вопросы системного синтеза

Не рассматривая в этой части работы методов и средств синтеза сложных объектов (собственно в рамках системологического инструментария), приведем для примера ряд задач синтеза, решение которых представляется весьма насущным с разных точек зрения (включая и методологическую!), с другой стороны, – этого можно достичь именно с помощью системологического арсенала.

Вот этот примерный перечень, без ранжирования, – для разных условий ранги будут разными.

- Разработка механизмов перехода от конфликта к конкуренции, от конкуренции к кооперации.
- Формирование конкуренции за качество, уровень жизни, аттрактивность, сотворчество, за степень соответствия системе требований… Законы формирования новых требований и мотиваций.
- Формирование новой системы (и шкалы!) потребностей.
- Генезис объекта, его оценка и метризация (пространства измерений, единицы, методы, процедуры, приборы, таксономия, прогноз, валидизация,..)
- Разработка такой системы агрегирования, свертки, при которой в объектах более высокого уровня сохранялась бы основная информация об объектах нижних уровней (их свойства, отношения,..)
- Конструирование тенсегритных (напряженных, скрепляющих) связей в системе.
- Методы социальной рефлексии (что это было? + синтез социальной самоответственности).
- Наконец, новое научное направление – социофенетика. Оно включает: изучение процессов и закономерностей формирования социально осваиваемого ненаследственного нового (социальная инноватика) и конструирование методов закрепления такого полезного нового; тьма проблем – оценки, критерии полезности, уровни закрепления, социальная память, учителя,..

9. Вместо заключения

Наверно, при очень прагматичном подходе можно для каких-то задач применять лишь определенные отдельные этапы системного анализа. Многое зависит от задачи и умения исследователя. Однако в подавляющем числе случаев изучения сложных объектов такое применение системного анализа будет похоже на ситуацию, в которой бы мы, собирая космонавта на работу в открытом космосе, защитили бы ему, скажем, только голову и руки.

Важным аспектом системного подхода является «рекуррентное» поведение разработчика: не следует бояться рефлексии, ретроспекции; не следует думать, что каждый этап – это законченный готовенький набор операций, и, выполнив их, ты обязательно попадаешь в намеченную точку. Каждому из нас может понадобиться вернуться, заново проделать определенную часть работы, дополнительно осмыслить и по-новому понять сделанное. Видимо, это один из путей преодоления парадокса образования/воспитания, который (парадокс) был отмечен в третьем тезисе Маркса о Фейербахе: если делить мир на учителей и учеников, то кто же научил учителей? Мы обречены на непрерывное обучение и самообучение, самообразование. По-видимому, методолог теперь будет учиться всегда.

Продолжением настоящей работы представляется изложение некоторых системологических приемов и методов исследования сложных объектов и построения (синтезирования) их системных моделей; по-видимому, речь будет идти об оценке неопределенности в системе, об использовании системных параметров и дескрипторов для описания широкого класса гибридных систем, имеющих естественно-искусственную природу (Е-И-систем), о конструктивных свойствах системологических методов, например, на базе естественной классификации (см. [12] и др.), о выявлении (и формировании!) эмерджентных характеристик системы. Перечисленные здесь работы являются, наверно, лишь небольшой долей всего ансамбля средств, предоставляемых в распоряжение разработчика современной методологией. Можно надеяться, что универсальный характер этого инструментария со временем сможет помочь науке в целом, формируя и отслеживая ноосферную действительность, вернуть себе функцию социального критического органа, в значительной мере подрастерянную в последние годы.

Одной из наиболее важных социальных функций само́й системной методологии является на сегодняшний день проведение возможно более широкого анализа сложившейся критической мирохозяйственной ситуации в целом (т. е. социально-экономической, социально-политической, социально-экологической), – и *конструирование стратегии осмысленного и приемлемого выхода* из нее. Эта системная, системологическая деятельность, крайне необходимая с точки зрения ноосферных представлений (и, безусловно – не только для стран «бедного» Юга!), требует квалифицированной работы, высокой культуры, высоких и естественных моральных принципов, социального и духовного уважения и толерантности *ко всему живому*. Здесь бездна методологических и онтологических задач – создание концепций выхода из «эколого-капиталистического» тупика (выражение А. Субетто [13]), поиск методов перехода от либеральной, конкурентной и индивидуализированной доминанты к кооперативной деятельности и просто экономной жизни, формирование институтов согласия и отказа от бесконечного и все увеличивающегося (очень асимметрично) потребления, вопреки желаниям мировых ТНК.

На этом поле очень много работы для методологии – от формулировки задач, поиска критериев, процедур, скрытых связей и т. д., до построения системных моделей, экспериментирования с ними (а не со странами), способов демонстрации результатов и пропаганды – путей формирования массового сознания, социальной и мировоззренческой адаптации и многое другое (см., напр., [14]).

В заключение хотелось бы сказать вот о чем. Применение системологического инструментария является сложной и трудоемкой задачей, однако результаты могут быть очень высокими, должны лишь быть

соблюдены некоторые условия. Весьма верной аналогией, причем по всем трем приведенным ниже позициям, является, по нашему мнению, высказывание известного советского литературоведа Ю. М. Лотмана, сделанное вроде бы совсем по другому поводу: «Понимание такого произведения, как «Евгений Онегин» – это серьезная задача. Задача, требующая труда, любви и культуры». Вот так.

Благодарности

Работа опубликована при финансовой поддержке проекта **ITHEA XXI** Института информационных теорий и приложений FOI ITHEA Болгария www.ithea.org и Ассоциации создателей и пользователей интеллектуальных систем ADUIS Украина www.aduis.com.ua.

Литература

1. В. И. Вернадский. Несколько слов о ноосфере. // Успехи соврем. биологии. – 1944, №18, вып. 2. Философские мысли натуралиста. – М.: «Наука», 1988. Научная мысль как планетное явление. – М.: «Наука», 1991.
2. Г. П. Мельников. Системология и языковые аспекты кибернетики. – М.: «Сов радио», 1978.
3. Ю. А. Шрейдер. Равенство, сходство, порядок. – М.: «Наука», 1978.
4. Уемов А. И. Системный подход и общая теория систем. – М.: «Мысль», 1978.
5. М. Ф. Бондаренко, Е. А. Соловьева, С. И. Маторин. Основы системологии. – Харьков, ХТУРЭ, 1998.
6. А. І. Катренко. Системний аналіз об'єктів та процесів комп'ютеризації. – Львів, „Новий світ – 2000", 2003
7. В. А. Крисилов. Оценка сложных объектов – одна из основных задач интеллектуальной обработки информации. Сб. трудов Од. политехн. Университета, №17, 1997.
8. А. Д. Крисилов, В. А. Крисилов. Формирование целеориентированной векторной модели для построения агрегированных оценок сложных объектов. // Моногр. «Методы решения экологических проблем». Под ред. проф. Л. Мельника. – Сумы: «Козацький вал», 2005
9. М. Доналдсон. О мыслительной деятельности детей. – М.: «Мысль», 1990.
10. А. Д. Крисилов. Модельное описание процессов развития: механизмы, система целей, индикаторы. Proc. XIV Intern. Conf. "Knowledge – Dialogue – Solution", vol. # 1, June 21 – 26, 2008, Varna (Bulgaria). – ITHEA, Sofia, 2008.
11. А. И. Уемов. Системные аспекты философского знания. – Одесса, «Негоциант», 2000.
12. Е. А. Соловьева. Естественная классификация. – Харьков, ХНУРЭ, 2003 и др.
13. А. И. Субетто. «Ноосферизм и вернадскианская революция: к модели выхода человечества из эколого-капиталистического тупика истории», ж. «Социальная экономика», №1 - 2, 2004 и др.
14. А. Д. Крисилов. Некоторые компоненты структурной модели формирующейся ноосферной парадигмы. Proc. XV Intern. Conf. "Knowledge – Dialogue – Solution", vol. # 1, June 20 – 25, 2009, Varna (Bulgaria). – ITHEA, Sofia, 2009.

Сведения об авторах

Анатолий Крисилов *– Институт информационных технологий Одесской государственной Академии холода, доц. кафедры информационных и коммуник. технологий; ул. Дворянская, 1/3, Одесса-26, 65026, Украина; тел. (0482)-632-598; моб. (38097)-291-33-24; E – m: adkrissilov@list.ru*

Екатерина Соловьева *– Харьковский Национальный университет радиоэлектроники, профессор, зав. кафедрой социальной информатики, зав. научно-учебным Центром приобретения знаний; пр-т Ленина, 14, Харьков-166, 61166, тел. (38057)-409-591; E – m: nulpz@kture.kharkov.ua; gt_ekasolo@yahoo.com*

Авенир Уемов *– Одесский Государственный университет им. И. И. Мечникова, профессор кафедры философии естественных факультетов, член Академии истории и философии науки (Калифорния, США); ул. Дворянская, 2, Одесса-26, 65026, Украина; тел. (0482)-636-817.*

ЗАЧЕМ И ПОЧЕМУ В ЗРИТЕЛЬНОЙ СИСТЕМЕ ИМЕЮТ МЕСТО ИНВЕРСИЯ СЕТЧАТКИ И ПЕРЕКРЕСТЫ ЗРИТЕЛЬНЫХ ВОЛОКОН

Геннадий Воронков

***Аннотация**: В результате анализа литературных данных получены свидетельства о том, что известные анатомо-физиологические феномены, перекресты зрительных волокон и инверсия сетчатки, хорошо изученные в нейроанатомическом плане, но остающиеся до сих пор без общепринятого объяснения в функциональном плане, являются механизмами зеркальных преобразований. Благодаря этим механизмам обеспечивается формирование такой нейронной модели (карты, проекции), оси координат которой на воспринимающей стороне направлены незеркально (а на отдающей – зеркально) таковым исходного оригинала – "лица" единого поля зрения и "лица" его составляющих, полей зрения. Без этих преобразований оси координат на воспринимающей стороне нейронной модели были бы направлены зеркально, как это имеет место у оптической проекции.*

***Ключевые слова**: инверсия сетчатки, перекресты зрительных волокон, зеркальные преобразования (соответствия), проекции полей зрения.*

***Conference**: The paper is selected from XV[th] International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.*

Введение

Инверсия сетчатки (ИС) и перекрест в хиазме (ПХ) позвоночных – хорошо известные морфофизиологические феномены. Однако какой цели они служат, до сих пор остается неизвестным [Хьюбел, 1990]. У беспозвоночных сетчатка, как правило, не инвертирована. В то же время, у них имеет место перекрест рецепторных аксонов на ипсилатеральной глазу стороне [Cajal, 1917; Заварзин, 1950], отсутствующий у позвоночных. Биологическая роль этого "ипсилатерального" перекреста (ИП) тоже не имеет удовлетворительного объяснения [Bullock, Horridge, 1965].

Целью данного сообщения является: 1) показать, что в последовательности экранных структур зрительного пути проекция (модель) поля зрения (ПЗ) после ИП (у беспозвоночных) или после ИС (у позвоночных), а также проекция единого поля зрения (ЕПЗ) после ПХ являются зеркальными[1] по отношению к непосредственно предшествующей модели ПЗ и ЕПЗ соответственно. Это будет свидетельствовать о том, что ИП, ИС и ПХ являются механизмами зеркальных преобразований (соответствий) [2], 2) показать, что эти механизмы – ПХ, а также ИП и ИС – направлены на формирование такой нейронной модели ЕПЗ и такой модели каждой его составляющей (ПЗ), отдающая сторона каждой из которых была бы зеркальна (а воспринимающая – незеркальна) соответственно "лицу" исходного ЕПЗ и "лицу" исходного ПЗ.

Обсуждается необходимость зеркальных преобразований.

---

[1] Зеркальные друг другу объекты, например двумерный объект и его отображение в зеркале, отличаются друг от друга направлением координатных осей на их «лицевых» сторонах. При сравнении «лиц» таких объектов видно, что при одинаковой направленности осей в одной паре одноименных осей в другой паре осей оси имеют обратные друг другу направления.

[2] Ниже в тексте термин "соответствие" везде опускается – как пока мало употребляемый в физиологической литературе. В то же время именно термин "механизм зеркального соответствия" наиболее адекватно отражает функциональную суть описываемых феноменов – ИС, ИП и ПХ.

Краткое описание феноменов

Инверсия сетчатки (ИС). У всех позвоночных сетчатка инвертирована. Это означает, что ее рецепторный слой обращен воспринимающей стороной от хрусталика, от света, а отдающей - навстречу свету. Такой слой можно сравнить (с точки зрения зеркальности его воспринимающей стороны) с фотопленкой, обращенной эмульсионной стороной от объектива, от света. У большинства беспозвоночных сетчатка не инвертирована, т. е. ее слой рецепторов обращен воспринимающей стороной к свету, а отдающей – от света. Такой слой можно сравнить с фотопленкой, обращенной эмульсионной стороной к объективу.

Ипсилатеральный перекрест (ИП). Наиболее известен перекрест зрительных волокон в хиазме позвоночных. Менее известно, что у большинства беспозвоночных (тех, сетчатка которых не инвертирована) имеет место еще перекрест, который можно назвать "ипсилатеральным". Это перекрест аксонов фоторецепторов (фоторецепторы у беспозвоночных, в отличие от таковых у позвоночных, имеют длинный аксон), он находится на той же стороне, где находится глаз. Особенность его состоит еще в том, что точки нейронного слоя, в которые проецируются вертикальные (у моллюска осьминога) или горизонтальные (в фасеточном глазу у насекомых) ряды точек рецепторного слоя располагаются в обратном (зеркальном) порядке, по сравнению с таковыми рецепторного слоя. При этом проекция по типу точка в точку сохраняется.

Перекрест в хиазме (ПХ) позвоночных. У ряда позвоночных (например, некоторые птицы) ПХ является полным (ПХп); у многих других позвоночных (например, приматы) он частичный – (ПХч). Позвоночные с полным перекрестом имеют неперекрывающиеся ПЗ, тогда как у позвоночных с частичным перекрестом ПЗ перекрываются. Другими словами, у позвоночных с полным ПХ ПЗ не пересекают (не "перехлестывают" за) вертикаль – ось симметрии ЕПЗ, по левую и правую стороны от которой симметрично располагаются ПЗ. Поэтому каждое ПЗ этих животных является полностью ипсилатеральным глазу. У позвоночных с частичным ПХ ПЗ пересекают ("перехлестывают" за) ось симметрии ЕПЗ. Поэтому каждое ПЗ этих позвоночных делится осью симметрии ЕПЗ на две, обычно неравные, ипси- и контралатеральную части. Перекрест, т. е. переход в контралатеральное полушарие, осуществляют волокна глаза, "видящие" только ипсилатеральную часть ЕПЗ, т.е. "видящие" либо полностью ипсилатеральное ПЗ (в случае ПХп), либо только его ипсилатеральную часть (в случае ПХч).

ИП, ИС и ПХ как механизмы зеркальных преобразований

ИП. На рис.1 представлена схема проекций ПЗ на рецепторный и нейронный слои в зрительной системе осьминога Octopus. Она построена на основе нейроанатомических данных, суммированных в фундаментальном труде Баллока и Хорриджа [Bullock, Horridge, 1965].

Хрусталик создает на неинвертированном рецепторном слое, как на экране, оптическую проекцию ПЗ (и объекта в нем). Или, другими словами, устанавливает соответствие по типу точка в точку (но по законам оптики) всех точек ПЗ с точками рецепторного слоя. Оптическая проекция ПЗ (оптическая модель ПЗ) на воспринимающей стороне является повернутой "с ног на голову" и зеркальной в отношении оригинала (ПЗ). (В литературе зеркальный характер оптической проекции, как правило, не отмечается или ему не придается значения.) Поэтому создающаяся рецепторная модель ПЗ с воспринимающей стороны идентична оптической проекции – она повернутая и зеркальная.

В свою очередь с помощью аксонов точки отдающей стороны рецепторной модели устанавливают однозначное соответствие с точками следующего, нейронного слоя. Однако, благодаря ИП, вертикальная ось координат нейронной модели ПЗ меняет, как можно видеть на этом рисунке (рис. 1), свое направление на обратное, зеркальное (в сравнении с таковой у рецепторной модели). Таким образом, нейронная модель

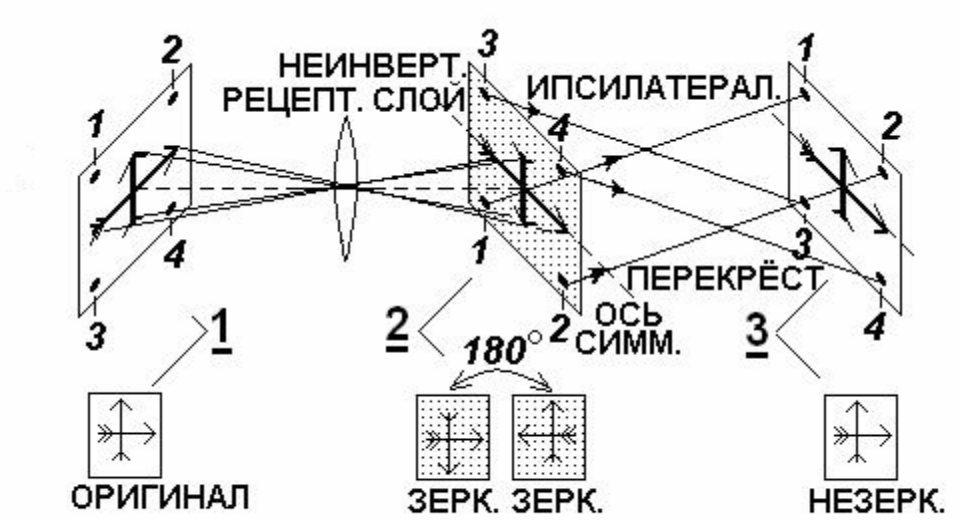


Рис. 1. Ипсилатеральный перекрест (ИП) в зрительной системе беспозвоночного (осьминог Octopus) как механизм зеркального преобразования.

1 – поле зрения (ПЗ). **2** – рецепторная модель ПЗ; она рассматривается также как обычный экран для оптической проекции. **3** – первая нейронная модель ПЗ.

Стрелки в ПЗ и моделях – принятое направление осей координат в ПЗ и таковое, "создающееся" в моделях, соответственно; они же – объект в ПЗ и в моделях.

Нижний ряд изображений – "лицо" ПЗ и виды на воспринимающую сторону его моделей; зерк. (и незерк) – зеркальная (и незеркальная) модель оригинала ("лица" ПЗ).

1- 4 – точки ПЗ и поставленные им в соответствие точки в рецепторной (**2**) и нейронной (**3**) моделях. Точечной заливкой окрашена та сторона каждой модели, со стороны которой модель зеркальна "лицу" ПЗ

ПЗ (и нейронная модель изображения в нем) является зеркальной в отношении рецепторной модели. Из этого следует вывод, что ИП есть механизм зеркального преобразования. Кроме того, можно также видеть, что нейронная модель незеркальна исходному оригиналу ("лицу" ПЗ) [1]. Из этого следует вывод (наблюдение), что механизм ИП направлен на формирование нейронной модели ПЗ, воспринимающая сторона которой незеркальна оригиналу.

На рис. 2 представлена схема проекций ПЗ на рецепторный и нейронный слои в зрительной системе насекомого. Она построена на основе известных нейроанатомических данных [Мазохин-Поршняков, 1983].

Поскольку фасеточный глаз насекомого не имеет для рецепторных единиц-омматидиев общего хрусталика, "составное" ПЗ глаза как бы "отпечатывается" на омматидиевом (рецепторном) слое. Как можно видеть, создаваемая при этом рецепторная модель ПЗ оказывается неперевернутой, но тоже (как и у осьминога) зеркальной (с воспринимающей стороны) своему оригиналу – ПЗ. Благодаря ИП следующая, нейронная модель оказывается, как можно видеть, зеркальной предшествующей, рецепторной, модели и незеркальной исходному оригиналу – ПЗ. Таким образом, как и у осьминога, у насекомых механизм ИП является механизмом зеркального преобразования, и направлен этот механизм на формирование нейронной модели ПЗ, которая с воспринимающей стороны незеркальна оригиналу.

[1] Оригинал всегда обращен "лицом" (источником света) к воспринимающей системе, а модель, рецепторная или нейронная, всегда здесь, кроме оговоренных случаев, рассматривается с ее воспринимающей (дендритной) стороны.

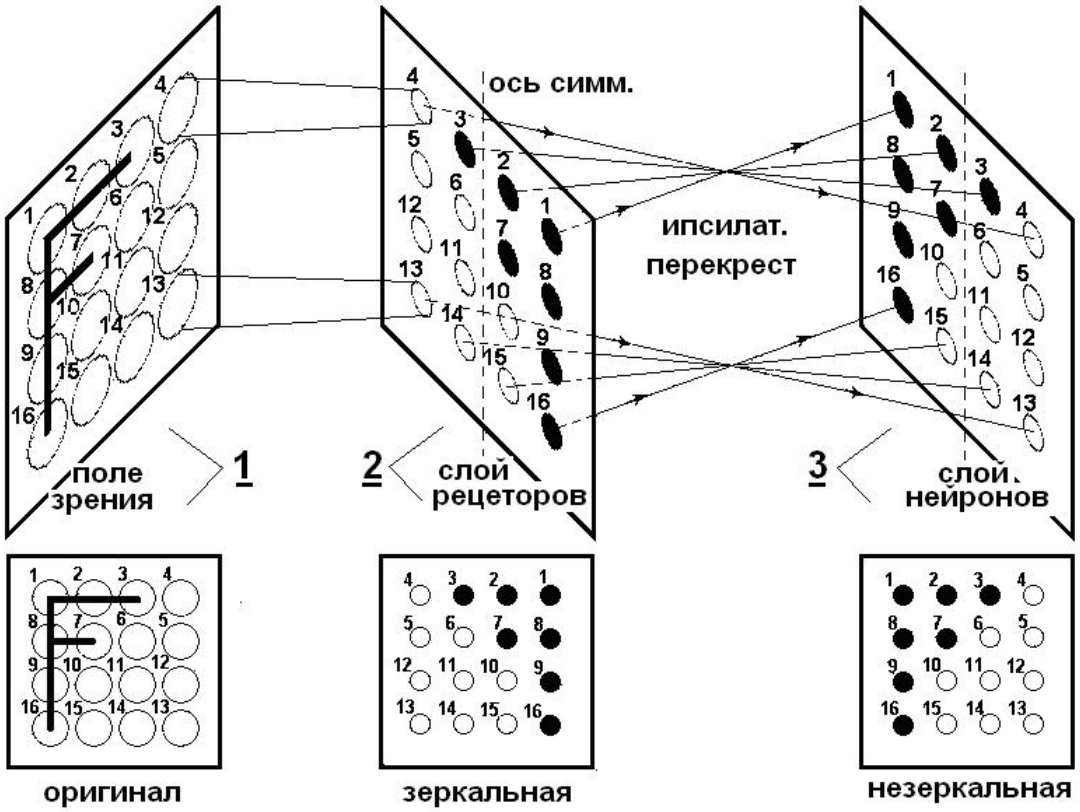


Рис. 2. Ипсилатеральный перекрест (ИП) в зрительной системе беспозвоночного (насекомое с фасеточными глазами) как механизм зеркального преобразования.

1 – поле зрения (ПЗ). **2** – рецепторная модель ПЗ. **3** – первая нейронная модель ПЗ.

1-16 – поля зрения отдельных омматидиев и поставленные им в соответствие точки-омматидии рецепторной и точки-участки нейронной моделей ПЗ. **F** объект в ПЗ и соответствующие ему модели (из темных, "активированных" точек) в рецепторном и нейронном слоях. Нижний ряд изображений – "лицо" ПЗ и виды на "воспринимающую" сторону его моделей.

ИС позвоночных. На рис.3 представлена схема рецепторной и нейронных проекций ПЗ в глазу позвоночных (в том числе человека). Как и у осьминога, хрусталик создает на рецепторном слое, как на экране, оптическую проекцию ПЗ, повернутую и зеркальную в сравнении с оригиналом. С обратной стороны экрана эта оптическая проекция, хотя и остается повернутой "с ног на голову", является, естественно, зеркальной той, что создается на воспринимающей его стороне[1]. Однако именно эта, обратная, сторона экрана является воспринимающей стороной инвертированного рецепторного слоя. Поэтому создаваемая рецепторная модель (с ее воспринимающей стороны) является зеркальной (а с отдающей – незеркальной) в отношении оптической модели на воспринимающей стороне экрана. Таким образом благодаря ИС рецепторная модель (с ее воспринимающей стороны) в инвертированной сетчатке оказывается сформированной сразу как незеркальная (а с отдающей – как зеркальная) в отношении исходного оригинала, ПЗ.

[1] Зеркальность сторон плоскости – атрибутивное свойство любой плоскости.

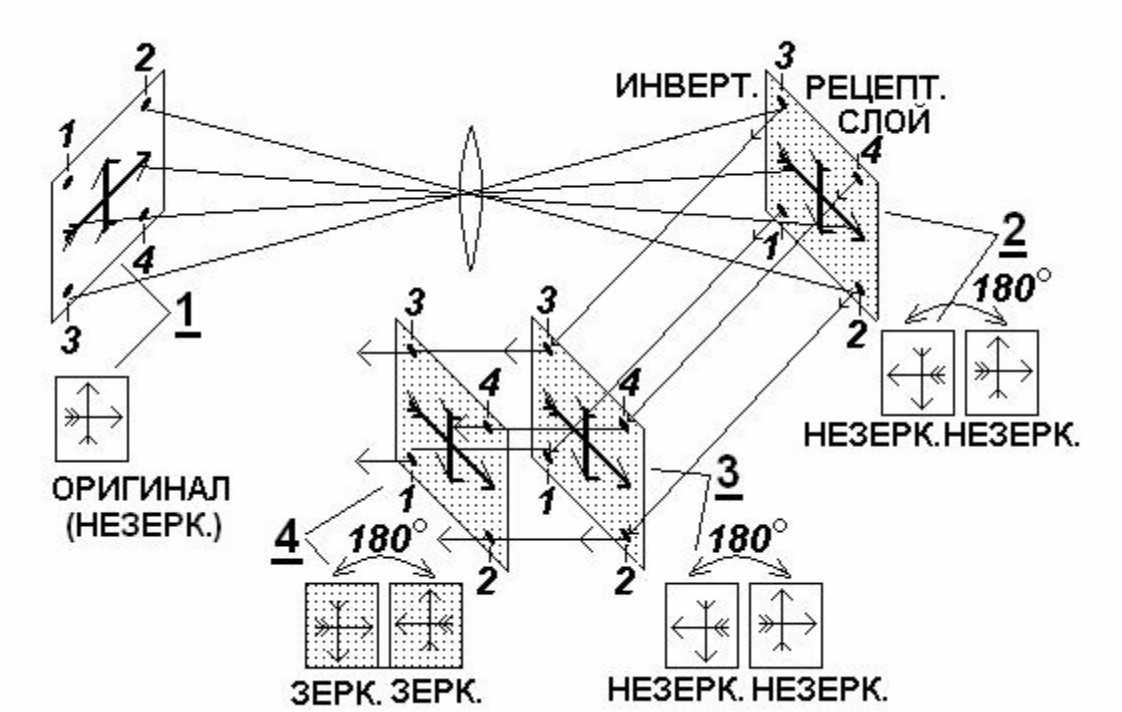


Рис. 3. Инверсия сетчатки (ИС) позвоночных животных как механизм зеркального преобразования.

1 – поле зрения (ПЗ); внизу – "лицо" ПЗ. **2** – рецепторная модель ПЗ в инвертированной сетчатке; внизу – вид на воспринимающую сторону модели. **3** и **4** – нейронные модели ПЗ (биполярная и ганглиозная соответственно); внизу – виды на воспринимающую и отдающую сторону моделей **3** и **4** соответственно. Остальные обозначения те же, что на рис. 1

Таким образом, рецепторная модель до инверсии (такая модель характеризует большинство беспозвоночных) и рецепторная модель после инверсии (такая модель характеризует всех позвоночных) зеркальны друг другу. Это свидетельствует о том, что ИС есть механизм зеркального преобразования[1] (точнее, механизм, *эквивалентный* механизму зеркального преобразования – см. сноску на стр. 7).

Известно, что вследствие плотного прилегания друг к другу каждый предшествующий слой сетчатки как бы отпечатывается на следующем слое, и что отпечаток всегда зеркален в отношении печати. В силу того, что отдающая сторона рецепторной модели ПЗ у позвоночных зеркальна исходному ПЗ (как показано выше), и что она "отпечатывается" на воспринимающей стороне нейронного биполярного слоя сетчатки, последний с воспринимающей стороны оказывается зеркальным в отношении отдающей стороны рецепторной модели и, следовательно, незеркальным в отношении исходного оригинала, "лица" ПЗ. Очевидно, что то же справедливо и по отношению к слою ганглиозных клеток – ганглиозная модель ПЗ также незеркальна (с ее воспринимающей стороны) исходному оригиналу. Таким образом, можно видеть, что ИС, как и ИП, тоже направлена на формирование незеркальной нейронной модели ПЗ.

Полученные здесь данные о направленности ИС и ИП на формирование незеркальной (с воспринимающей стороны) физиологической модели каждого ПЗ в отдельности позволили

[1] Здесь можно отметить, что механизм ИС является, очевидно, более экономным и продуктивным механизмом, чем ИП. Ибо, как и ИП, направленный на формирование незеркальной модели ПЗ, он решает еще, по крайней мере, две задачи: 1) устраняет необходимость в сложном, громоздком ИП и 2) благодаря ИС становится незеркальной уже самая первая физиологическая (не оптическая), рецепторная модель ПЗ. (Есть основания предполагать, что незеркальность самой первой физиологической (рецепторной) модели могла быть важным фактором, способствующим эволюции зрительной системы у позвоночных.)

сформулировать еще более широкое предположение, своего рода **постулат** – что незеркальной своему оригиналу должна быть и модель поля зрения в целом (единого поля зрения – ЕПЗ). Из **постулата** вытекает следствие о необходимости в зеркальном преобразовании модели ЕПЗ в отношении полей зрения (ПЗ) как единичных элементов, составляющих ЕПЗ. Естественно было предположить, что механизмом этого преобразования является перекрест зрительных волокон в хиазме – ПХ. Приводимые ниже результаты анализа в отношении ПХ подтверждают это предположение.

Полный ПХ позвоночных (ПХп). Отметим два важных момента в описании известных данных о ПХп. 1) ПЗ у животных с полным перекрестом являются полностью ипсилатеральными – они не пересекают ось симметрии ЕПЗ; другими словами, точка бификсации взора у этих животных отсутствует (является мнимой). 2) При переходе волокон в контралатеральное полушарие имеет место прямая (не зеркальная) топическая проекция выходных нейронов сетчатки на нейронные слои НКТ, т. е. пространственные отношения между внутренними элементами (точками) модели ПЗ сохраняются в НКТ такими же, как между таковыми в выходном слое сетчатки (при рассматривании обеих моделей с воспринимающей стороны). Другими словами, полный ПХ обеспечивает преобразование ЕПЗ в отношении всего двух его элементов – двух ПЗ, выступающих как единичные элементы.

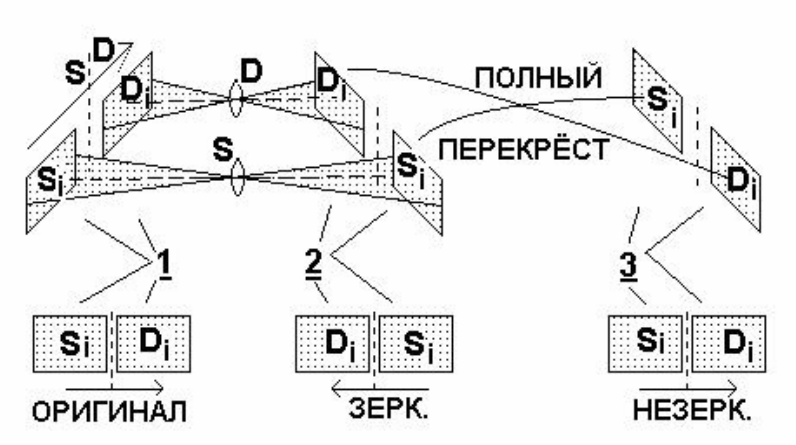


Рис. 4. Полный перекрест в хиазме (ПХп) позвоночных животных как механизм зеркального преобразования.

1 - единое поле зрения (ЕПЗ), состоящее из двух ПЗ (непересекающихся, симметрично расположенных относительно оси симметрии ЕПЗ). **2** – модель ЕПЗ, состоящая из моделей ПЗ, на уровне сетчаток. Изображены только выходные слои обеих сетчаток. **3** – модель ЕПЗ, состоящая из моделей ПЗ, на уровне НКТ. Изображен только первый нейронный слой НКТ.

Нижний ряд изображений – "лицо" ЕПЗ и виды на воспринимающую сторону моделей ЕПЗ.

S и D – левая и правая стороны пространства относительно оси симметрии ЕПЗ, а также левый и правый глаз соответственно.

Si и Di – полностью ипсилатеральные глазу левое и правое ПЗ и их модели.

Стрелки – принятое направление горизонтальной оси координат в ЕПЗ (слева направо) и таковое, "создающееся" в его моделях

Точечная заливка – метка ипсилатеральной данному (D или S) глазу части ПЗ (здесь ипсилатеральным является каждое ПЗ полностью).

Схема на рис. 4 построена с учетом этих моментов. Можно видеть на этом рисунке, что после полного ПХ нейронная модель ЕПЗ, состоящего двух ПЗ, зеркальна таковой до перекреста, а также что после полного ПХ она становится незеркальной исходному оригиналу, ЕПЗ. Из этого следуют два вывода: 1) полный ПХ есть механизм, обеспечивающий зеркальное преобразование в отношении элементов ЕПЗ – его ПЗ как целых единиц; 2) направленность этого механизма отвечает **постулату**.

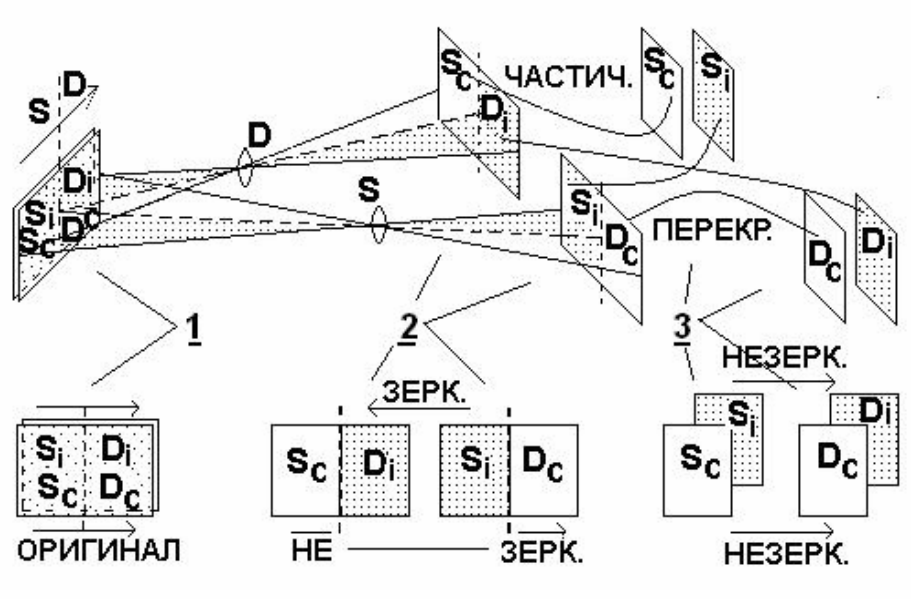


Рис. 5. Частичный перекрест в хиазме (ПХч) у позвоночных
как механизм зеркального преобразования.

1 - единое поле зрения (ЕПЗ), состоящее из двух ПЗ (полностью перекрывающихся и каждое из которых разделено осью симметрии ЕПЗ пополам). **2** – модель ЕПЗ (состоящая из моделей ипси- и контралатеральной частей ПЗ) на уровне сетчаток. **3** – модель ЕПЗ (состоящая из моделей ипси- и контралатеральной частей ПЗ) на уровне НКТ. Изображены только первые слои НКТ, получающие проекции из сетчаток.

Si – ипсилатеральная глазу часть ПЗ левого глаза (она находится в левой части ЕПЗ) и ее модели. Di – ипсилатеральная глазу часть ПЗ правого глаза (она находится в правой части ЕПЗ) и ее модели. Sc – контралатеральная глазу часть ПЗ правого глаза (она находится в левой части ЕПЗ) и ее модели. Dc – контралатеральная глазу часть ПЗ левого глаза (она находится в правой части ЕПЗ) и ее модели.

Точечной заливкой окрашены ипсилатеральные глазу части каждого ПЗ и их модели. Остальные обозначения такие же, как на рис. 4

Частичный ПХ позвоночных (ПХч). У ряда позвоночных (например приматы и человек) оба ПЗ перекрываются почти полностью (т. е. наполовину "перехлестывают" за ось симметрии ЕПЗ), и поэтому каждое их ПЗ состоит из двух почти равных частей, ипси- и контралатеральной (а все ЕПЗ – из четырех частей). Эти части разделяет ось симметрии ЕПЗ, проходящая вертикально через точку бификсации взора (рис. 5). Модели этих четырех частей тоже, как и модели ПЗ в случае с ПХп, выступают в преобразованиях, связанных с ПХч, как единичные элементы. Теоретически в согласии с **постулатом**, а также как это следует из сделанного ранее (выше) описания полного ПХ, в контралатеральное полушарие должны проецироваться только ипсилатеральные части-элементы ЕПЗ, тогда как контралатеральные должны проецироваться в ипсилатеральное полушарие. Именно такая нейроанатомическая организация частичного ПХ имеет место в действительности (Хьюбел, 1990). Таким образом, рис.5 отражает одновременно и теоретические, и экспериментальные данные. Они полностью совпадают.

Можно видеть (рис.5), что модель доли ЕПЗ, состоящей из его ипсилатеральных частей, после ПХч зеркальна таковой до перекреста, и что вся модель ЕПЗ после ПХч становится полностью незеркальной исходному оригиналу, ЕПЗ. Отсюда следуют два вывода: 1) переход волокон в частичном ПХ является механизмом зеркального преобразования (в отношении доли ЕПЗ, состоящей из двух ипсилатеральных частей ПЗ); 2) в целом же весь механизм ПХч направлен на формирование модели, незеркальной исходному оригиналу – ЕПЗ, состоящему из четырех (2 ипси- и 2 контралатеральных) частей ПЗ, выступающих как единичные элементы.

Заключение

Таким образом, результаты проведенного анализа свидетельствуют, что ИС, ИП и ПХ являются механизмами зеркальных преобразований[1]. ИС (у позвоночных) и ИП (у беспозвоночных) осуществляют зеркальные преобразования моделей ПЗ в отношении их элементов, соответствующих точкам ПЗ. Полный ПХ и частичный ПХ осуществляют зеркальные преобразования модели ЕПЗ, соответственно в отношении двух и четырех составляющих его частей, выступающих как единичные элементы. В итоге формируется единая нейронная модель ЕПЗ, которая с отдающей стороны становится полностью (в целом и во всех ее частях) зеркальной (а с воспринимающей – незеркальной) "лицу" исходного оригинала. Без этих преобразований "зеркальность" сторон модели была бы обратной – такой, как у оптической проекции.

Естественно задать вопрос: "зачем и почему мозгу нужно это сложное зеркальное преобразование?". Ответ-гипотеза на заданный вопрос возможен, на наш взгляд, с самых общих позиций. Мозг и среда (мир) взаимодействуют: зрительная среда воздействует на мозг, мозг воздействует (в итоге) на зрительную среду. По А.А. Заварзину [Заварзин,1950], мозг состоит, по большому счету, из последовательности проекционных экранных структур (сенсорных и эффекторных). Каждая из них имеет, естественно, зеркальные друг другу, воспринимающую и отдающую стороны. Среда всегда обращена "лицом" к экранным структурам – как при воздействии ее на мозг, так и при воздействии мозга на среду. Другими словами, она всегда "одинаковая", незеркальная по определению. Ответ на вопрос появляется, если принять, что по направленности координатных осей среда находится всегда в отношениях «комплементарности» с входной и выходной экранными структурами, т. е. в отношениях принципиальной совместимости – подобно тому, как это имеет место, например, либо в отношениях гомотетичных фигур, либо "лицевой" стороны печати с её оттиском, либо объекта с его отображением в зеркале. Действительно, для выполнения этого условия необходимо, чтобы и воспринимающая сторона первой входной экранной структуры (модели)[1], и отдающая сторона последней выходной экранной структуры были бы зеркальны "лицу" мира. Это в свою очередь требует, как очевидно, чтобы где-то в последовательности экранных структур имело место зеркальное преобразование. Именно это и обнаруживается в действительности, что показано в настоящей работе.

Поскольку в литературе ИС, ИП и ПХ и обеспечиваемые ими преобразования не были ранее идентифицированы в качестве механизмов зеркальных преобразований, а зеркальные преобразования не были идентифицированы в ЦНС именно как вид функционально значимого преобразования, их можно, по-видимому, рассматривать соответственно как новый вид механизмов и новый вид преобразований в зрительной системе и мозге в целом.

Отметим еще один момент, связанный с механизмами зеркальных преобразований. Возможно, знание о них имеет отношение к волнующей, но еще совершенно неразработанной проблеме "субъективного" – с какими структурами непосредственно связано ощущение видения. Действительно, если принять, что ощущение предметного видения есть проявление активности нейронов нейронной ЕПЗ-модели (что общепринято), то, по причине очевидного сходства (по параметру «зеркальность», по крайней мере), зрительного ощущения (=образа видимого мира) и "лица" видимого мира, следует, по-видимому,

[1] Будет правильнее именовать механизм ИС, в отличие от ИП и ПХ, *эквивалентным* механизму зеркального преобразования – ибо в случае с ИС не происходит зеркального перемещения точек ни в самой плоскости, ни в следующей за ней проекционной плоскости (что имеет место в случае с ИП и ПХ). Случай с ИС аналогичен, в принципе, случаю с обращением фотопленки к объективу обратной, неэмульсионной стороной (в результате чего "негатив" с эмульсионной стороны оказывается незеркальным "лицу" мира).

признать, что субъективный образ (ощущение) сходен с "картиной", "видимой" на *воспринимающей* стороне нейронной модели.

Зеркальные преобразования имеют, видимо, достаточно широкое представительство в мозге – их механизмы (по крайней мере перекресты) характеризуют помимо зрительной также некоторые другие мозговые системы. Так, среди сенсорных систем перекресты имею место, например, в соматосенсорной и слуховой системах (в тех системах, стимул которых имеет пространственные характеристики, но не в тех системах, стимул которых имеет только качественные характеристики, например в обонятельной системе).

Кроме того, известны функциональные проявления "зеркальности" в мозговой деятельности в норме и при мозговых расстройствах. Например, зеркальное письмо и рисование, а также некоторые феномены мышления. В теоретическом аспекте важным моментом, по-видимому, является также уже сама нейроанатомическая природа механизмов зеркальных преобразований, указывающая, по сути, на способ (нейроанатомический) мозгового представительства и обработки информации о пространственных координатах зрительного мира.

Благодарности

Работа опубликована при финансовой поддержке проекта **ITHEA XXI** Института информационных теорий и приложений FOI ITHEA Болгария www.ithea.org и Ассоциации создателей и пользователей интеллектуальных систем ADUIS Украина www.aduis.com.ua.

Литература

[Bullock, Horridge, 1965] Bullock T.H., Horridge G.A. Structure and function in the nervous systems of invertebrates. 1965. Vol. 2. San Francisco;London. P. 801-1719.

[Cajal, 1917] Cajal S.R. Contribucion al conocimiento de la retina y centros opticos de los Cefalopodos.// Trab. Lab. Invest. biol. Univ. Madr.. 1917. 15.

[Заварзин, 1950] Заварзин А.А. Избранные труды. Том 3. Очерки по эволюционной гистологии нервной системы. Москва-Ленинград. 1950. С. 420.

[Мазохин-Поршняков, 1983] Мазохин–Поршняков Г.А. (Ред.). Руководство по физиологии органов чувств насекомых. 1983. Москва. С. 261.

[Хьюбел, 1990] Хьюбел Д. Глаз, мозг, зрение. 1990. Москва. С. 240.

Информация об авторе

Геннадий С. Воронков *– ведущий научный сотрудник, доктор биологических наук, Московский государственный университет им. М.В. Ломоносова, 119992, Москва, Россия;*
e-mail: av13675@yandex.ru

[1] Первой моделью ПЗ является оптическая модель. С воспринимающей стороны экрана она зеркальна оригиналу.

In memoriam:

профессор дфмн Николай Фёдорович Кириченко

19 декабря 2008 года на 69-ом году ушёл из жизни известный специалист в области механики, математического моделирования робото-технических систем, теории оптимального управления, компьютерной графики, обработки сигналов и распознавания образов, доктор физико-математических наук, профессор Николай Фёдорович Кириченко.

В научной карьере Н.Ф. Кириченко явственно выделяются несколько этапов, связанных с деятельностью в той или иной конкретной прикладной области, в каждой из которых им получены значимые результаты.

Начало научной деятельности Н.Ф. Кириченко связано с работой в Институте гидромеханики АН Украины, где он защитил кандидатскую, а в 1973 году в возрасте 33-х лет – докторскую диссертацию, посвящённую оценкам устойчивости и стабильности при ограниченных возмущениях.

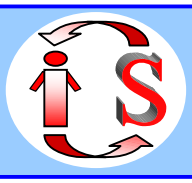
С 1970 по 1986 год научная и педагогическая деятельность Н.Ф. Кириченко проходит на факультете кибернетики Киевского университета им. Т.Г. Шевченко, где его научные интересы связаны с теорией оптимального управления, а с момента, когда он возглавил кафедру теоретической кибернетики – с обработкой сигналов, распознаванием образов, робототехникой.

Как известный специалист в области обработки информации он был приглашён возглавить кафедру математических проблем управления в Черновицком университете им. Ю. Федьковича. Пятилетняя работа Н.Ф. Кириченко в Черновцах имела результатом создание коллектива, который эффективно работает и поныне.

С 1992 года жизнь Н.Ф. Кириченко снова связана с Киевом, а научная работа – с Институтом кибернетики НАН Украины, с отделом интеллектуальных систем управления динамическими объектами, где он работал ведущим научным сотрудником. Этот этап научной работы Н.Ф. Кириченко определяет разработка фундаментальных математических методов обработки сигналов, а также созданием на этой основе конкретных систем обработки изображений, распознавания текстовой информации, распознавания образов, разработкой интеллектуальных систем принятия решений. Выдающимся достижением этого этапа научной деятельности Н.Ф. Кириченко стала разработка аналитической теории возмущения псевдообращения по Муру-Пенроузу, которая предоставляет широкие аналитические возможности решения важных прикладных задач в самых разных областях: от классификации, кластеризации и распознавания образов до теории управления динамическими системами.

Н.Ф. Кириченко оставил после себя научную школу, в которой только им самим подготовлено 30 кандидатов и 5 докторов наук. Он является автором 4 монографий и 11 учебных пособий.

Светлая память о Николае Фёдоровиче Кириченко: учёном, педагоге, коллеге, друге, учителе, сохранится в сердцах и памяти всех, кто имел счастье работать рядом и вместе с ним, общаться и участвовать в научных исследованиях под его руководством вне зависимости от географии.



ADUIS

ASSOCIATION OF DEVELOPERS AND USERS OF INTELLIGENT SYSTEMS

ADUIS consists of about one hundred members including ten collective members. The Association was founded in Ukraine in 1992. The main aim of ***ADUIS*** is to contribute to the development and application of the artificial intelligence methods and techniques. The efforts of scientists engaged in ***ADUIS*** are concentrated on the following problems: expert system design; knowledge engineering; knowledge discovery; planning and decision making systems; cognitive models designing; human-computer interaction; natural language processing; methodological and philosophical foundations of AI.

Association has long-term experience in collaboration with teams, working in different fields of ***research and development***. Methods and programs created in Association were used for revealing regularities, which characterize chemical compounds and materials with desired properties. Some thousands of high precise prognoses have been done in collaboration with chemists and material scientists of Russia and USA.

Association can help ***businessmen*** to find out conditions for successful investment taking into account region or field peculiarities as well as to reveal user's requirements on technical characteristics of products being sold or manufactured.

Physicians can be equipped with systems, which help in diagnosing or choosing treatment methods, in forming multi-parametric models that characterize health state of population in different regions or social groups.

Sociologists, politicians, managers can obtain the Association's help in creating generalized multi-parametric "portraits" of social groups, regions, enterprise groups. Such "portraits" can be used for prognostication of voting results, progress trends, and different consequences of decision making as well.

Association provides a useful guide in technical diagnostics, ecology, geology, and genetics.

ADUIS has at hand a broad range of high-efficiency original methods and program tools for solving analytical problems, such as knowledge discovery, classification, diagnostics, and prognostication.

ADUIS unites the creative potential of highly skilled scientists and engineers

Since 1992 ***ADUIS*** holds regular conferences and workshops with wide participation of specialists in AI and users of intelligent systems. The proceedings of the conferences and workshops are published in scientific journals. ***ADUIS*** cooperates through its foreign members with organizations that work on AI problems in Russia, Belarus, Moldova, Georgia, Bulgaria, Czech Republic, Germany, Great Britain, Hungary, Poland, etc. ***ADUIS*** is the collective member of the European Coordinating Committee for Artificial Intelligence (ECCAI).

www.aduis.com.ua

For contacts: V.M.Glushkov Institute of Cybernetics; National Academy of Science of Ukraine;

Prospect Akademika Glushkova, 40, 03680 GSP Kiev-187, Ukraine;

Phone: (380+44) 5262260; Fax: (380+44) 5263348; E-mail: glad@aduis.kiev.ua

ITHEA INTERNATIONAL SCIENTIFIC SOCIETY

The ITHEA International Scientific Society (**ITHEA ISS**) a successor of the international scientific co-operation organized within 1986-1992 by international workgroups (**IWG**) researching the problems of data bases and artificial intelligence. As a result of tight relation between these problems in 1990 in Budapest appeared the international scientific group of Data Base Intellectualization (**IWGDBI**) integrating the possibilities of databases with the creative process support tools. The leaders of the IWGDBI were Prof. Victor Gladun (Ukraine) and Prof. Rumyana Kirkova (Bulgaria). Starting from 1992 till now the international scientific co-operation has been organized by the Association of Developers and Users of Intellectualized Systems (**ADUIS**), Ukraine. It has played a significant role for uniting the scientific community working in the area of the artificial intelligence.

To extend the possibilities for international scientific collaboration in all directions of informatics by wide range of concrete activities, in 2002 year, the Institute for Information Theories and Applications FOI ITHEA (**ITHEA**) has been established as an international nongovernmental organization. **ITHEA Official representatives** are Krassimir Markov – President, Krassimira Ivanova and Ilia Mitov - Vice-presidents.

ITHEA Scientific Council includes:

- *Honorable Chair Persons:* Victor Gladun (Ukraine) and Rumyana Kirkova (Bulgaria);
- *Chairman:* Levon Aslanyan (Armenia);
- *Members:* Adil Timofeev (Russia), Alexander Kleshchev (Russia), Alexander Kuzemin (Ukraine), Alexander Palagin (Ukraine), Alexey Voloshyn (Ukraine), Arkadij Zakrevskij (Belarus), Constantin Gaindric (Moldova), Hasmik Sahakyan (Armenia), Ilia Mitov (Bulgaria), Juan Castellanos (Spain), Koen Vanhoof (Belgie), Krassimira Ivanova (Bulgaria), Krassimir Markov (Bulgaria), Larisa Zainutdinova (Russia), Laura Ciocoiu (Romania), Luis Fernando de Mingo (Spain), Martin Mintchev (Canada), Nikolay Zagoruiko (Russia), Peter Stanchev (USA), Stefan Karastanev (Bulgaria), Tatyana Gavrilova (Russia), Vladimir Lovitskii (GB), Vladimir Ryazanov (Russia).

ITHEA Scientific Sections: Business Informatics (Koen Vanhoof), Conferences and Publishing (Krassimira Ivanova), Cooperative Scientific Projects (Hasmik Sahakyan), Decision Support (Alexey Voloshyn), Information Interaction (Stefan Karastanev). Intelligent Engineering Systems (Martin Mintchev), Logical Inference and Design (Arkadij Zakrevskij), Modern (e-) Learning (Larisa Zaynutdinova), Natural Computing (Juan Castellanos), Pattern Recognition (Vladimir Ryazanov), Philosophy and Methodology of Informatics (Krassimir Markov), Scientific Innovation Management (Luis Fernando de Mingo), Software Engineering (Ilia Mitov), Theoretical Informatics (Levon Aslanyan).

ITHEA is aimed to support international scientific research through international scientific projects, workshops, conferences, journals, book series, etc. The achieved results are remarkable. The ITHEA became world-wide known scientific organization. One of the main activities of the ITHEA is building an International Scientific Society aimed to unite researches from all over the world who are working in the area of informatics.

Now, the **ITHEA International Scientific Society** is joined by **1516** members from **44** countries all over the world: Armenia, Azerbaijan, Belarus, Brazil, Belgium, Bulgaria, Canada, Czech Republic, Denmark, Egypt, Estonia, Finland, France, Germany, Greece, Hungary, India, Iran, Ireland, Israel, Italy, Japan, Jordan, Kyrgyz Republic, Latvia, Lithuania, Malaysia, Malta, Mexico, Moldova, Netherlands, Poland, Portugal, Romania, Russia, Scotland, Senegal, Serbia and Montenegro, Spain, Sultanate of Oman, Turkey, UK, Ukraine, and USA.