



I T H E A



International Journal
INFORMATION **MODELS**
&
ANALYSES



2013 **Volume 2** **Number 4**



**International Journal
INFORMATION MODELS & ANALYSES
Volume 2 / 2013, Number 4**

Editor in chief: **Krassimir Markov** (Bulgaria)

Adil Timofeev	(Russia)	Levon Aslanyan	(Armenia)
Albert Voronin	(Ukraine)	Luis Fernando de Mingo	(Spain)
Aleksey Voloshin	(Ukraine)	Liudmila Cheremisinova	(Belarus)
Alexander Palagin	(Ukraine)	Lyudmila Lyadova	(Russia)
Alexey Petrovskiy	(Russia)	Martin P. Mintchev	(Canada)
Alfredo Milani	(Italy)	Nataliia Kussul	(Ukraine)
Anatoliy Krissilov	(Ukraine)	Natalia Ivanova	(Russia)
Avram Eskenazi	(Bulgaria)	Natalia Pankratova	(Ukraine)
Boris Tsankov	(Bulgaria)	Nelly Maneva	(Bulgaria)
Boris Sokolov	(Russia)	Olga Nevzorova	(Russia)
Diana Bogdanova	(Russia)	Orly Yadid-Pecht	(Israel)
Ekaterina Detcheva	(Bulgaria)	Pedro Marijuan	(Spain)
Ekaterina Solovyova	(Ukraine)	Rafael Yusupov	(Russia)
Evgeniy Bodyansky	(Ukraine)	Sergey Krivii	(Ukraine)
Galyna Gayvoronska	(Ukraine)	Stoyan Poryazov	(Bulgaria)
Galina Setlac	(Poland)	Tatyana Gavrilova	(Russia)
George Totkov	(Bulgaria)	Valeria Gribova	(Russia)
Gurgen Khachatryan	(Armenia)	Vasil Sgurev	(Bulgaria)
Hasmik Sahakyan	(Armenia)	Vitalii Velychko	(Ukraine)
Iliia Mitov	(Bulgaria)	Vladimir Donchenko	(Ukraine)
Juan Castellanos	(Spain)	Vladimir Ryazanov	(Russia)
Koen Vanhoof	(Belgium)	Yordan Tabov	(Bulgaria)
Krassimira B. Ivanova	(Bulgaria)	Yuriy Zaichenko	(Ukraine)

**IJ IMA is official publisher of the scientific papers of the members of
the ITHEA® International Scientific Society**

IJ IMA rules for preparing the manuscripts are compulsory.

The **rules for the papers** for ITHEA International Journals are given on www.ithea.org.

The camera-ready copy of the paper should be received by ITHEA® Submission system <http://ij.ithea.org>.

Responsibility for papers published in IJ IMA belongs to authors.

General Sponsor of IJ IMA is the **Consortium FOI Bulgaria** (www.foibg.com).

International Journal "INFORMATION MODELS AND ANALYSES" Vol.2, Number 4, 2013

Edited by the **Institute of Information Theories and Applications FOI ITHEA**, Bulgaria, in collaboration with

Institute of Mathematics and Informatics, BAS, Bulgaria,

V.M.Glushkov Institute of Cybernetics of NAS, Ukraine,

Universidad Politecnica de Madrid, Spain,

Hasselt University, Belgium

Institute of Informatics Problems of the RAS, Russia,

St. Petersburg Institute of Informatics, RAS, Russia

Institute for Informatics and Automation Problems, NAS of the Republic of Armenia,

and Federation of the Scientific - Engineering Unions /FNITS/ (Bulgaria).

Publisher: **ITHEA®**

Sofia, 1000, P.O.B. 775, Bulgaria. www.ithea.org, e-mail: info@foibg.com

Technical editor: **Ina Markova**

Printed in Bulgaria

Copyright © 2012-2013 All rights reserved for the publisher and all authors.

© 2012-2013 "Information Models and Analyses" is a trademark of Krassimir Markov

© ITHEA is a registered trade mark of FOI-Commerce Co.

ISSN 1314-6416 (printed)

ISSN 1314-6424 (CD)

ISSN 1314-6432 (Online)

ИНТЕГРАЦИЯ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ НАВИГАЦИИ И УПРАВЛЕНИЯ ДВИЖЕНИЯ МЕХАТРОННЫХ РОБОТОВ

Р.М. Юсупов , А.В. Тимофеев

Аннотация: Обсуждаются проблемы интеграции систем мульти-агентной (групповой) навигации, интеллектуального управления движением и функциональной диагностики мехатронных роботов в динамической среде с препятствиями. Определяются понятия глобальной управляемости и стабилизируемости программных движений мехатронных роботов-агентов. Дается сравнительный анализ путей поколений локальных систем навигации и управления движением и стратегий централизованного, децентрализованного и мульти-агентного управления информационными потоками в сложных робототехнических системах. Описываются методы функциональной диагностики и алгоритмы дефектоустойчивого управления движением роботов в реальном времени при наличии разного рода дефектов и возмущений. Даются оценки времени правильного функционирования роботов, степени их неисправности и допустимых границ (допусков) различных классов дефектов и возмущений. Приводятся примеры использования предлагаемых методов для мехатронных роботов и мульти-агентных систем различного типа и назначения (манипуляционные и мобильные роботы, космические роботы, беспилотные летающие роботы и т.п.).

Keywords: интеграция, интеллектуализация, навигация, управление, робототехника

ACM Classification Keywords: I.2.11 Distributed Artificial Intelligence - Multiagent systems

Введение

В последние годы в России и за рубежом значительное внимание уделялось исследованию проблем интеграции и интеллектуализации локальных систем автономной навигации и управления роботом и разработке глобальных стратегий групповой навигации и управления движением сложных робототехнических систем. В связи с бурным развитием систем навигации и управления движением мехатронных роботов как подвижных агентов с элементами искусственного интеллекта и необходимостью организации их коллективной работы возникла потребность в создании общих принципов построения и внедрения мульти-агентных робототехнических систем (МАРС). Такие МАРС объединяют группу агентов-роботов (например, мобильных роботов или беспилотных летательных аппаратов) для достижения общей (глобальной) цели в динамической среде с препятствиями или запретными зонами и возможным противодействием [1–5].

Фундаментальные и поисковые исследования в области адаптивного и интеллектуального управления роботами и робототехническими системами (РТС) активно проводились в России в ряде научных школ (МГТУ, СПИИРАН, СПбГУ, МГУ, МИРЭА, НИИ МВС, ЦНИИ РТК, ИПМ РАН и др.), начиная с 60-х годов XX века. Важная роль в становлении и развитии этих исследований принадлежит академикам И.М.Макарову, Д.Е.Охочимскому, Е.П.Попову, Г.С.Поспелову и Ф.Л.Черноусько, членам-корреспондентам РАН В.А.Якубовичу, В.А.Лопоте, Г.Г.Сербрякову, Е. Д.Теряеву, Р.М.Юсупову и И.А.Каляеву и профессорам Е.И.Юревичу, В.М.Лохину, А.В.Тимофееву и др.

Проблемы и методы адаптивной навигации и интеллектуального управления движением роботов, а также задачи группового управления РТС обсуждались на ряде Международных и Российских конференций,

проведённых в последние десятилетия в России, в том числе на 2-х Международных конференциях "Мехатроника и робототехника", 22-х конференциях "Экстремальная робототехника", 7-и конференциях "Мехатроника, автоматизация, управление", 6-и конференциях "Управление и информационные технологии" и 3-х Мультиконференциях по проблемам управления.

Параллельно в России и за рубежом проводились поисковые исследования и опытно-конструкторские разработки по созданию интегрированных систем навигации и управления движением объектов и роботов различного типа и назначения для экстремальных сред и частично неопределённых условий эксплуатации. Достижения специалистов из России и зарубежный опыт в этой области обсуждались на 19-ти Международных конференциях по интегрированным навигационным системам, организованных ЦНИИ «Электроприбор», и на 33-х научных сессиях Международной академии навигации и управления движением, президентом которой является академик В.Г.Пешехонов.

Первый международный проект по исследованию MAPC на тему "Multi-Agent Robot Systems for Industrial Applications in the Transport Domain" был выполнен в 1997–1999 годах по Европейской программе COPERNICUS. Координатором этого проекта был Prof., Dr.-Ing. U.Rembold (University of Karlsruhe, Institute for Process Control and Robotics, Germany), а его участниками – Saint-Petersburg Institute for Informatics and Automation of Russian Academy of Sciences (Russia), Unite de Recherche INRIA Rhone-Alpes (France), Technical University of Budapest (Hungary), Technical University of Poznan (Poland), Belorussian State University, (Belarus), Ufa State Aviation Technical University (Russia) и корпорации Daimler-Benz AG (Germany) и Beta Computer Automation (Germany).

Развиваемые в настоящем докладе новые информационные и телекоммуникационные технологии и интегрированные интеллектуальные системы навигации и управления движением для MAPC в значительной степени базируются на оригинальных научных результатах поисковых и ориентированных фундаментальных исследований, полученных авторами доклада в последние годы при поддержке грантов Российского фонда фундаментальных исследований (РФФИ), программ Президиума РАН и государственных заказов Министерства образования и науки РФ (см., например, [5–14]).

Важным средством интеллектуализации интегрированных систем навигации и управления движением роботов и РТС является встроенные средства функциональной диагностики в реальном масштабе времени. [5,21,22]. Поэтому в докладе обсуждаются методы функциональной диагностики и алгоритмы дефектоустойчивого управления вместе с оценками допустимых границ для разных классов дефектов и возмущений..

1. Глобальная управляемость и оптимизация программных движений РТС

Рассмотрим некоторую робототехническую систему (РТС) R , состоящую из n роботов $r_i, i = 1, 2, \dots, n$. Обычно роботы и РТС функционируют в динамической среде E с препятствиями или запретными зонами $O_j \in E, j = 1, 2, \dots, q$. В роли динамических препятствий для одних роботов r_i из РТС могут выступать другие роботы $r_j, j \neq i$.

РТС будем называть гомогенной, если она состоит из однотипных роботов. Сложную РТС будем называть гетерогенной, если её структура включает в себя несколько разнотипных роботов или гомогенных РТС.

Архитектура гомогенных и гетерогенных РТС по существу является сетевой и территориально распределённой, т.е. включает в себя множество роботов, взаимодействующих между собой с помощью физических, сенсорных (информационных), управляющих и коммуникационных каналов прямой и обратной связи. Поэтому возникает потребность в разработке новых принципов сетевой организации, управления движением, навигации, обработки и передачи информации в сложных РТС в динамической

окружающей среде с препятствиями или запретными зонами. Необходимость в сетевом управлении и групповой навигации возникла прежде всего в глобальных инфотелекоммуникационных сетях (например, в Internet), в робототехнике и гибких автоматизированных производственных системах, а также в сложных автоматизированных системах вооружений (например, в сетевых и мульти-агентных системах оборонного назначения).

В сложных гомогенных и гетерогенных РТС под действием вектора управляющих воздействий

$$u(t) = \{u_i(t)\}_{i=1}^n, t \in [t_0, t_T] \quad (1)$$

каждый робот $r_i(u_i)$ выполняет некоторый набор локальных технологических операций $l_j(r_i)$, $j = 1, 2, \dots, p$, $p \geq n$ на заданном интервале времени $T = t_T - t_0$. В результате выполнения этих локально управляемых операций изменяются состояния как самой РТС $R(t)$, так и окружающей её среды $E(t)$.

Предположим, что эволюцию РТС $R(t)$ и окружающей среды $E(t)$ под действием вектора управления $u(t)$ формально можно описать дифференциальными уравнениями вида

$$\dot{R} = F(R, u, E), R(t_0) = R_0 \quad (2)$$

$$\dot{E} = \Phi(R, E), E(t_0) = E_0 \quad (3)$$

где F и Φ – некоторые операторы, а R_0 и E_0 – начальные состояния РТС и среды.

Общим (обобщённым) динамическим состоянием РТС R и окружающую её среды E в текущий момент времени t будем называть вектор-функцию вида

$$x(t) = \begin{Bmatrix} R(t) \\ E(t) \end{Bmatrix}, t \in [t_0, t_T]. \quad (4)$$

Глобальная (общая) цель управления РТС заключается в том, чтобы синтезировать такой закон группового управления локальными технологическими операциями роботов $L = \{l_i(r)\}_{i=1}^p$, который обеспечивает перевод РТС R и окружающую её среду E из заданного начального состояния в желаемое конечное (целевое) состояние, т.е. в процессе групповой навигации и управления должны быть выполнены граничные условия вида

$$x(t_0) = x_0, x(t_T) = x_T \quad (5)$$

с учётом заданных ограничений на вектор сетевого управления (1) и вектор состояний (4).

РТС (2) будем называть глобально управляемой в динамической среде (3), если существует закон группового управления (вообще говоря, зависящий от текущего состояния РТС и среды) вида

$$u_p(t) = U(t, x(t)) \in Q_u, t \in [t_0, t_T] \quad (6)$$

обеспечивающий выполнение граничных условий (5) с учетом имеющихся ограничений. Соответствующее этому групповому управлению (6) движение $x_p(t)$, $t \in [t_0, t_T]$, будем называть программным движением (ПД) РТС $R(t)$ в динамической среде $E(t)$ с препятствиями или запретными зонами.

Среди множества групповых управлений (6) и ПД РТС можно выделить наилучшее (оптимальное) управление и соответствующее ему оптимальное ПД. Для этого зададим некоторый интегральный функционал качества групповых ПД вида

$$K(x_p) = \int_{t_0}^{t_T} \mathcal{L}(x_p(t), \dot{x}_p(t)) dt. \quad (7)$$

Тогда оптимальным ПД $x_p^{opt}(t), t \in [t_0, t_T]$, и соответствующим ему оптимальным групповым управлением $u_p^{opt}(t)$ будем называть то групповое ПД и глобальный закон управления РТС, при которых достигается экстремум функционала качества (7).

Следует отметить, что критерии глобальной управляемости роботов с нелинейной динамикой, а также методы аналитического синтеза и оптимизации ПД мехатронных роботов с учётом заданных ограничений при наличии препятствий были предложены и подробно описаны в работах [1,5,12,19,20].

2. Стабилизация, декомпозиция и синхронизация программных движений роботов

Предположим, что ПД РТС $x_p(t)$ и соответствующий ему закон программного группового управления $u_p(t)$ существуют. Тогда ПД называется практически (или асимптотически) устойчивым по отношению к начальным возмущениям $e(t_0) = x_0 - x_p(t_0)$ (или, возможно, к другим возмущениям), если существует закон группового управления роботами такой, что в замкнутой этим управлением РТС и окружающей её динамической среде выполняются условия вида

$$\|e(t)\| = \|x_p(t) - x(t, u)\| \leq \varepsilon t \geq t_p \geq t_0 \quad (8)$$

где $\varepsilon \geq 0$ – параметр, определяющий желаемую точность осуществления ПД, а $T_p = t_p - t_0$ – время переходного процесса.

Закон группового управления РТС (6), обеспечивающий выполнение целевого условия (8), будем называть стабилизирующим. Этот закон управления является декомпозирующим, если целевые условия выполняются независимо для каждого робота из РТС, т.е. $e_i(t)$ не зависит от $e_j(t), j \neq i$. Это означает, что локальное управление каждым роботом r_i осуществляется независимо от локального управления другими роботами $r_j, j \neq i$, с компенсацией возможных перекрёстных динамических связей между роботами РТС [1,5].

В некоторых случаях (например, на конвейерах) от закона группового управления требуется обеспечить полную или частичную синхронизацию движений различных роботов из РТС. В этих случаях целевые условия (8) приобретают вид

$$\|e_{ij}(t)\| = \|x_{pi}(t) - x_j(t, u_j)\| \leq \varepsilon_{ij}, i \neq j, t \geq t_{pi,j} \quad (9)$$

На практике возникают также задачи группового управления роботами, связанные с достижением консенсуса (consensus), когда каждый робот из РТС стремится, чтобы его ПД было близко с ПД своих соседей, или рандеву (rendezvous), когда все роботы и РТС должны попасть в желаемое состояние в заданный момент времени $T = t_T - t_0$.

Следует отметить, что для локальных систем навигации и управления движением мехатронных роботов с нелинейной динамикой законы стабилизирующего, модального (спектрального) и декомпозирующего управления, а также алгоритмы планирования поведения и маршрутизации безопасных маршрутов движения роботов в среде с препятствиями были предложены и детально описаны в работах [1–7, 19–24].

3. Пять поколений локальных систем навигации и управления движением роботов

В работах [1–5] была предложена классификация локальных систем навигации и управления движением (СНУД) роботов на пять поколений:

- 1) программные СНУД;

- 2) адаптивные СНУД;
- 3) интеллектуальные СНУД;
- 4) нейросетевые СНУД;
- 5) креативные СНУД.

Программные СНУД роботов основаны на классических принципах программного или оптимального управления движением, если модель динамики роботов или РТС (2) полностью известна, а среда (или связанные с ней возмущения) известны и неизменны.

Адаптивные СНУД базируются на современных принципах робастного или адаптивного управления, когда модель динамики РТС и среды E (или связанных с ней возмущений) полностью или частично неизвестны. В этом случае программные СНУД дополняются специальными средствами адаптации или идентификации факторов неопределённости или нестационарности, к числу которых можно отнести неизвестные возмущения или препятствия [1–4].

Интеллектуальные СНУД дополняются не только средствами адаптации, но и некоторыми элементами (алгоритмами) искусственного интеллекта (например, способностью распознавать речевые команды или идентифицировать неизвестные препятствия). Как правило, эти элементы искусственного интеллекта описываются логическими, алгебраическими и нечёткими алгоритмами. Поэтому они могут быть программно реализованы на традиционных микропроцессорах и компьютерах [1–3].

Нейросетевые СНУД основываются на обучении, самоорганизации и распараллеливании процессов обработки информации, навигации и управления на нейронных сетях или нейрокомпьютерах [6,27]. В этом заключается их основное отличие и преимущество по сравнению с СНУД роботов первых трёх поколений.

Наконец, креативные СНУД ещё слабо изучены. Они могут базироваться на когнитивных и мультиагентных технологиях, системах виртуальной или дополненной реальности, процессах моделирования сознания и т.п. [24,25].

4. Глобальные стратегии навигации и управления движением сложных робототехнических систем

Важно отметить, что роботы редко используются изолированно, т.е. в полуавтономном или автономном режиме эксплуатации. Обычно они входят в состав РТС и предназначены для группового (коллективного) выполнения некоторой сложной общей (глобальной) задачи, которую каждый робот самостоятельно (автономно) решить не может [1].

Сложная гетерогенная РТС может включать в себя десятки или сотни роботов или гомогенных РТС, функционирующих в динамической среде с препятствиями. В связи с этим важное значение приобретают проблемы системного анализа и сетевого синтеза интегрированных систем групповой навигации, управления движением и обработки сенсорной информации. Однако проектирование и создание интегрированных систем невозможно без организации и координации взаимодействия роботов в РТС с помощью коммуникационной сети [1–6].

Математической моделью такой, вообще говоря, динамической коммуникационной сети РТС может служить коммуникационный граф вида [10,24,25]:

$$G(t) = \{A(t), C(t), w(t)\}, t \in [t_0, t_T] \quad (10)$$

где A – множество узлов, соответствующих локальным системам навигации, управления движением и обработки информации роботов $r_i, i = 1, 2, \dots, n$, C – множество каналов связи между ними, а W –

множество весов (например, длина или пропускная способность) каналов связи.

В процессе решения общей (глобальной) задачи структура и параметры РТС могут изменяться (например, некоторые роботы или каналы связи между ними могут выйти из строя). Тогда будут изменяться структура (топология узлов и каналов связи) или параметры (веса) коммуникационного графа (10). Этот граф характеризует информационное и коммуникационное взаимодействие между роботами r_i из РТС $R(t)$. Поэтому его структура тесно связана со стратегией групповой навигации и управления ПД роботов и РТС, обеспечивающей достижение общей (глобальной) цели управления в динамической среде с препятствиями.

Первоначально (начиная с 70-х годов XX века) стратегии группового управления роботами в РТС разделялись на два класса [1,16,17]:

- стратегии централизованного (глобального) управления РТС из общего командного центра;
- стратегии децентрализованного (распределённого) управления, основанные на локальном управлении каждым роботом из РТС и их взаимодействии.

Преимущества стратегии централизованного (глобального) управления РТС заключаются в простоте иерархической организации и программно-аппаратной реализации систем группового (в том числе оптимального) управления на центральном компьютере. Однако её недостатки проявляются в низкой живучести (надёжности) и возможности искажений и временных задержек при передаче команд программного управления от верхнего уровня иерархии к нижнему.

Преимущества стратегии децентрализованного (локального распределённого) управления роботами из РТС заключаются в высокой локальной производительности (малом времени принятия локальных решений) и параллелизме локального управления роботами, а также в повышенной живучести (надёжности) РТС. В тоже время ей свойственны такие недостатки, как повышенные требования к надёжности узлов и каналов информационной связи между роботами и невысокое качество группового управления РТС, т.к. даже из локальной оптимальности управления отдельными роботами, вообще говоря, не следует глобальная оптимальность управления РТС в целом.

Сравнительный анализ преимуществ и недостатков традиционных стратегий централизованного и децентрализованного управления территориально распределённых РТС привёл к необходимости разработки (начиная с 90-х годов XX века) новой гибридной (смешанной) стратегии мульти-агентной навигации и группового (коллективного) управления РТС на современном уровне развития робототехники, мехатроники и инфотелекоммуникационных технологий [3,5,9–11]. Этот уровень развития характеризуется всё более широким внедрением роботов с интеллектуальным и нейросетевым управлением, глобальных систем навигации типа ГЛОНАСС и (или) GPS и инфотелекоммуникационных систем типа Internet и Grid [24,25].

5. Функциональная диагностика мехатронных роботов и дефектоустойчивое управление

Проблемы технической диагностики играют важную роль в процессе проектирования, испытания и эксплуатации автономных роботов и сложных РТС. Среди этих проблем большое значение имеют задачи функциональной диагностики, когда контроль дефектов и диагностика неисправностей осуществляются в реальном времени непосредственно в процессе эксплуатации роботов и РТС.

Функционирование роботов и РТС в значительной степени зависит не только от возможных неисправностей и дефектов, но и от используемых алгоритмов управления [5,21,22].

Цель навигации и управления движением роботов и РТС в таких системах заключается в построении желаемого (программного) движения и синтезе алгоритмов управления, обеспечивающих фактическое

осуществление этого программного движения (ПД) при наличии известных или неизвестных препятствий, возмущений и дефектов из определённого класса, заданных своими границами (допусками) [5–21]. Для контроля гарантированного достижения этой цели необходимы функциональное диагностирование в процессе навигации и управления движением, а также, по возможности, обнаружение, локализация и компенсация неисправностей (дефектов).

Функциональная диагностика роботов и РТС основывается на адекватных моделях динамики объекта управления и моделях возможных неисправностей (дефектов) в замкнутой системе. Такие модели можно построить в аналитической форме для линейных и нелинейных обратимых моделей динамики роботов, описываемых дифференциальными уравнениями динамики, разрешимыми относительно управления на некотором подпространстве [1,5,20].

В докладе рассматриваются методы функционального диагностирования роботов и алгоритмы дефектоустойчивого управления их движением. В работах [5,21,22] показано, что эти алгоритмы являются робастными или адаптивными по отношению к различным типам возможных дефектов и неконтролируемым возмущениям. Даются оценки допустимых границ (допусков) различных дефектов и возмущений [21,22].

Предлагаемые методы основываются на прямых и обратных (на подпространстве) моделях динамики мехатронных роботов [5,21,22]. В докладе даётся классификация моделей дефектов и приводятся оценки основных показателей правильного функционирования и степени неисправности [21,22]. Большое внимание уделяется синтезу и анализу программного, стабилизирующего, робастного и адаптивного управления движением мехатронных роботов, обеспечивающих инвариантность переходных процессов (динамической ошибки) в различных (узких и широких) классах возможных дефектов. Показано, что использование алгоритмов идентификации неизвестных параметров роботов и внешних возмущений с конечным временем адаптации [1–5] позволяет не только обнаружить и локализовать дефекты из широкого класса неопределённости, но и автоматически их компенсировать.

6. Принципы интеграции систем навигации, управления движением и диагностики в МАРС

В современной робототехнике, а также в информационных технологиях навигации и теории управления движением сложных мехатронных и аэрокосмических систем, существует ряд проблем, связанных с проектированием, созданием и интеграцией систем адаптивной (локальной) и мульти-агентной (групповой) навигации, функциональной диагностики и интеллектуального управления движением мехатронных роботов в динамической среде с известными или неизвестными препятствиями. Некоторые из этих проблем, а также перспективные подходы и новые методы их решения рассмотрены в работах [1–25]. В настоящем докладе обсуждаются принципы проектирования интегрированных систем мульти-агентной навигации, интеллектуального управления и функциональной диагностики применительно к мобильным роботам и РТС.

Мобильный робот рассматривается как интеллектуальный подвижный мехатронный агент, имеющий на борту [1, 24]:

- сенсорную систему датчиков внутренней и внешней информации;
- интеллектуальную систему навигации, управления движением и функциональной диагностики;
- двигательную систему с традиционным (например, гусеничным или четырёхколёсным) или нетрадиционным (например, с одноосным двухколёсным или ползающим, т.е. перемещающимся по поверхностям произвольной ориентации) шасси;

– коммуникационную систему для информационного взаимодействия с человеком-оператором и другими роботами-агентами.

Важными задачами при создании мобильных мехатронных роботов как подвижных агентов является интеллектуализация и интеграция бортовых интегрированных систем навигации, управления движением и функциональной диагностики, обеспечивающая возможность автоматического планирования и оптимизации движений для обхода известных или неизвестных препятствий, адаптации к динамическим факторам неопределённости (например, к возмущениям и неизвестным подвижным препятствиям) и к возможным дефектам или отказам, а также распознавания образов (например, стерео-изображений препятствий), моделирования и анализа сложных 3D-сцен и диагностики состояний мобильного робота-агента в реальном масштабе времени.

Новые актуальные проблемы возникают при групповом (коллективном) использовании мобильных мехатронных роботов-агентов в составе МАРС для решения общей (глобальной) задачи в динамической среде с препятствиями. Группа (коллектив) мобильных мехатронных роботов-агентов, объединённых общей (глобальной) целью и функционирующая в едином информационном и коммуникационном пространстве на базе соответствующих стандартов, команд и форматов передачи потоков данных, называется мобильной МАРС.

Ключевыми проблемами при создании интегрированных систем навигации и управления движением мобильных МАРС являются [1–6, 21–25]:

– декомпозиция общей (глобальной) задачи, возлагаемой на группу (коллектив) мобильных роботов, на ряд локальных задач для каждого робота-агента,

– мульти-агентное (групповое) планирование маршрутов движения и навигация мобильных роботов-агентов в динамической среде с известными или неизвестными препятствиями,

– функциональная диагностика и отказоустойчивое управление движением мобильных роботов-агентов,

– интеллектуализация и интеграция систем навигации, управления движением и диагностики на базе бортовых средств распознавания сложных образов (мульти-изображений, речевых команд и т.п.), мобильных навигаторов (спутниковые системы ГЛОНАСС и GPS и т.п.) и функциональной диагностики в составе мобильных МАРС.

Важными задачами являются также организация интеллектуального человеко-машинного интерфейса и комплексирование перспективных решений этих проблем для создания интегрированных систем мульти-агентной навигации, интеллектуального управления движением и функциональной диагностики мобильных роботов как агентов мехатронных МАРС, способных функционировать в реальном времени в динамической среде с препятствиями.

Заключение

Актуальность проведённых исследований связана с повышенным интересом во всём мире к интеллектуальным мехатронным роботам как подвижным агентам в МАРС ввиду их способности автоматически функционировать под контролем человека как в традиционных отраслях производства (машиностроение, приборостроение, микроэлектроника и т.п.), так и в нетрадиционных областях и экстремальных средах (космическая и подводная робототехника, атомная энергетика, ликвидация техногенных аварий и террористических угроз и т.п.). Важно отметить, что круг задач, решаемых мехатронными роботами-агентами и мобильными МАРС, непрерывно расширяется и усложняется

(необходимость эффективно работать в экстремальных средах в условиях неопределённости или противодействия при наличии препятствий или запретных зон и т.п.).

Большую научную значимость и актуальность имеют поисковые исследования в области создания интегрированных систем мульти-агентной (групповой) навигации, интеллектуального управления движением и функциональной диагностики мехатронных роботов-агентов и мобильных МАРС в динамических (изменяющихся) средах с препятствиями. Сегодня этим инновационным исследованиям и опытно-конструкторским разработкам во всём мире уделяется приоритетное внимание.

Работа выполнена при частичной поддержке грантов РФФИ № 09–08–00767–а, № 12–08–01167–а и РФФИ–ГФЕН Китая № 10–08–91159–ГФЕН–а и Программы № 14 "Проблемы создания национальной научной распределенной информационно-вычислительной среды на основе GRID-технологий, облачных вычислений и современных телекоммуникационных сетей" Президиума РАН.

Библиография

- [1]. Тимофеев А.В. Адаптивные робототехнические комплексы. Л.: Машиностроение, 1988.
- [2]. Тимофеев А.В., Юсупов Р.М. Интеллектуализация систем автоматического управления. – Известия АН. Техническая кибернетика, 1994, № 5.
- [3]. Тимофеев А.В., Юсупов Р.М. Принципы построения интегрированных систем мульти-агентной навигации и интеллектуального управления мехатронными роботами. – International Journal "Information Technologies and Knowledge", Vol. 5, Number 3, 2011, pp. 237–244.
- [4]. Тимофеев А.В. Мульти-агентные системы и нейросетевые технологии в робототехнике и информатике. – Известия Кабардино-Балкарского научного центра РАН, № 1 (39), 2011, с. 249–252.
- [5]. Тимофеев А.В. Методы высококачественного управления, интеллектуализации и функциональной диагностики автоматических систем. Часть I, Часть II. – Мехатроника, автоматизация, управление, 2003, № 5, 2004, № 2.
- [6]. Timofeev A. V. Intelligent Control Applied to Non-Linear Systems and Neural Networks with Adaptive Architecture //International Journal on Intelligent Control, Neurocomputing : Fuzzy Logic. 1996. P. 1–18.
- [7]. Тимофеев А.В., Дерин О.А., Гуленко И.Е., Андреев В.А. Распознавание объектов в сложных мультиизображениях и методы и средства видеозахвата движений. - Мехатроника, автоматизация, управление, № 6 (111), 2010.
- [8]. Тимофеев А. В., Сырцев А. В. Мультиагентная и нейросетевая маршрутизация потоков данных в телекоммуникационных сетях // Труды 10-й международной конференции "Knowledge–Dialogue–Solution" (16–26 июня. 2003, Варна) 2003. С. 187–190.
- [9]. Тимофеев А. В. Модели мульти-агентного диалога и информационного управления в глобальных телекоммуникационных сетях // Труды 10-й международной конференции "Knowledge–Dialogue–Solution" (16-26 июня, 2003, Варна). 2003. С. 180–186.
- [10]. Тимофеев А.В., Сырцев А.В. Модели и методы маршрутизации потоков данных в телекоммуникационных системах с изменяющейся динамикой. – М.: Новые технологии, 2005.
- [11]. Timofeev A. V. Adaptive Control and Multi-Agent Interface for Infotelecommunication Systems of New Generation // International Journal "Information Theories & Applications". 2004. Vol. 11.
- [12]. Зотов Ю.К., Тимофеев А.В., Шишкин Д.С. Информационные технологии навигации и управления полетом мало-размерных летательных роботов корабельного базирования. – Информационно-измерительные и управляющие системы, № 8, т. 6, 2008.
- [13]. Тимофеев А.В., Юсупов Р.М., Нагоев З.В. Проблемы и методы интеллектуального управления движением мультиагентных робототехнических систем. – Материалы Второй Международной конференции, посвящённой 15-летию ИИПРУ КБНЦ РАН «Автоматизация управления и интеллектуальные системы и среды» (15–23 декабря 2011 г.), 2011, с. 100–108.

- [14]. Cai Z., He H., Timofeev A. Navigation and Control of Mobile Robots in Unknown Environment: A Survey// Proceedings of 10-th International Conference on Integrated Navigation Systems (June 27–29, St-Petersburg). 2003. Vol. 1. P. 158–166.
- [15]. Охтилев М.Ю., Соколов Б.В., Юсупов Р.М. Интеллектуальные технологии управления структурной динамикой сложных технических объектов. М.: Наука, 2006.
- [16]. Макаров И.М., Лохин В.М., Манько С.В., Романов М.П. Искусственный интеллект и интеллектуальные системы управления. М.: Наука, 2006.
- [17]. Каляев И.А., Гайдук А.Р., Капустян С.Г. Модели и алгоритмы коллективного управления в группах роботов. М.: Наука, 2009.
- [18]. Зотов Ю.К., Тимофеев А.В., Шишкин Д.С. Информационные технологии навигации и управления полетом мало-размерных летательных роботов корабельного базирования. – Информационно-измерительные и управляющие системы, № 8, т. 6, 2008.
- [19]. Тимофеев А.В. Синтез адаптивных регуляторов с помощью функций Ляпунова // ДАН СССР, 1984, т.237, № 2, с. 276-279.
- [20]. Zotov Yu.K., Timofeev A.V. Controllability and Stabilization of Programmed Motions of Reversible Mechanical and Electromechanical Systems - Journal of Applied Mathematics and Mechanics, 1992, vol. 56, № 6, pp. 873-880 (Pergamon Press Ltd.).
- [21]. Тимофеев А.В. Функциональный анализ неисправностей динамических систем и дефектоустойчивость стабилизирующего управления.- Доклады АМАН, т.6, № 1, 2002, с. 62-71.
- [22]. Timofeev A.V. Physical Diagnostics and Fault Relevant Feedback Control. – CD-ROM Proceedings of International Conference "Physics and Control" (August 20-22,2003, Saint-Petersburg).
- [23]. Timofeev A.V., Madani K. Algorithms of Adaptive and Robust Control on the Basis of Neural Networks and Technology of Their Realization on DSP. – Transactions of French-Russian A.M.Liapunov Institute for Applied Mathematics and Computer Science, 2003, pp. 100-107.
- [24]. Тимофеев А.В. Методы нейросетевого и мульти-агентного управления в робототехнике и мехатронике. // Нелинейная теория управления и её приложения. Динамика, управление, оптимизация. – М.: Физматлит. 2003, с. 101-126.
- [25]. Тимофеев А.В. Адаптивное управление и интеллектуальный анализ информационных потоков в компьютерных сетях. – СПб.: Анатолия, 2012, 280 с.

Сведения об авторах



Юсупов Рафаэль Мидхатович – член-корреспондент РАН, директор Санкт-Петербургского института информатики и автоматизации РАН, доктор технических наук, профессор, Заслуженный деятель науки РФ, 199178, Россия, Санкт-Петербург, 14-я линия, д. 39, СПИИРАН, yusupov@ias.spb.su



Тимофеев Адиль Васильевич – Заведующий лабораторией информационных технологий в управлении и робототехнике Санкт-Петербургского института информатики и автоматизации РАН, заведующий кафедрой нейроинформатики и робототехники Санкт-Петербургского университета аэрокосмического приборостроения, профессор кафедры информатики математико-механического факультета Санкт-Петербургского университета, профессор кафедры мехатроники и робототехники Балтийского государственного университета "Военмех" им. Маршала Д.Ф.Устинова, доктор технических наук, профессор, Заслуженный деятель науки РФ, 199178, Россия, Санкт-Петербург, 14-я линия, д. 39, СПИИРАН, tav@ias.spb.su

ВЫБОР ВЫЧИСЛИТЕЛЬНОГО КЛАСТЕРА В СОКРАЩЕННОМ ПРОСТРАНСТВЕ КРИТЕРИЕВ

**Алексей Петровский, Василий Лобанов,
Алла Заболева-Зотова, Татьяна Шитова**

Аннотация: В работе описан подход к многокритериальному выбору вычислительного кластера с использованием многоэтапной технологии последовательного снижения размерности пространства критериев. С помощью методов вербального анализа решений построена иерархическая система составных критериев, агрегирующих исходных характеристики кластера. Предложенный подход позволяет упростить процедуру многокритериальных сравнения и выбора вариантов, уменьшить трудоемкость процедуры, провести анализ полученных результатов.

Ключевые слова: многокритериальный выбор, снижение размерности признакового пространства, агрегирование признаков, вербальный анализ решений, вычислительный кластер

ACM Classification Keywords: A.0 General Literature - Conference proceedings

Введение

В настоящее время при решении различных прикладных и научных задач все большее применение находят высокопроизводительные вычислительные кластеры. Причиной растущей распространенности вычислительных кластеров является увеличение числа задач, решение которых на персональном компьютере все чаще становится нецелесообразным. В качестве примеров можно указать обнаружение и отслеживание целей в радиолокации и гидроакустике, моделирование быстропротекающих процессов в физике и ряд других задач, требующих обработки больших массивов информации за короткое время. Архитектура современных кластеров предполагает использование стандартных комплектующих изделий, стоимость которых постоянно снижается вследствие массового выпуска и возрастающей конкуренции производителей. Все это ведет к уменьшению итоговой стоимости кластеров. Современные стандартные и относительно недорогие микропроцессоры, сетевые технологии и периферийные устройства позволяют создавать разнообразные конфигурации вычислительных кластеров, которые можно последовательно и гибко изменять и наращивать за счет добавления новых вычислительных модулей, обеспечивая требуемую производительность и соответствие потребностям пользователей [Фортон и др., 2002].

Сравнение разных вариантов конфигураций кластеров, выбор наиболее предпочтительных аппаратных и программных решений представляет собой достаточно сложную, слабо структурируемую и плохо формализованную задачу. Это обусловлено тем, что вычислительные кластеры, как и другие сложные технические системы, характеризуются большим числом показателей, а выбор осуществляется по многим критериям, среди которых могут быть и количественные, и качественные. Вместе с тем, как правило, вариантов конфигураций бывает немного. Поэтому такие варианты обычно оказываются несравнимыми друг с другом по своим показателям, что не позволяет применять известные методы принятия решений для выбора лучшего варианта кластера.

В работе рассматривается задача многокритериального выбора наиболее предпочтительной конфигурации вычислительного кластера, обладающего требуемыми для прикладных применений

параметрами. Для решения задачи применена многоэтапная технология ПАКС (Последовательное Агрегирование Классифицируемых Состояний) [Петровский, Ройзензон, 2012]. Подходы к решению задачи многокритериального выбора вычислительных кластеров были предложены в работах [Ройзензон, 2005], [Лобанов, Петровский, 2012]. Процедура снижения размерности признакового пространства в задачах многокритериальной классификации описана в работе [Петровский, Ройзензон, 2008].

Многоэтапная технология снижения размерности признакового пространства

При решении задач многокритериального выбора с помощью многоэтапной технологии ПАКС размерность признакового пространства последовательно сокращается, исходя из предпочтений лица, принимающего решение (ЛПР), или знаний эксперта [Петровский, Ройзензон, 2012]. Используя несколько методов вербального анализа решений и/или их комбинаций, большое число исходных характеристик рассматриваемых объектов агрегируется в небольшое число критериев, имеющих порядковые шкалы оценок (количественные и качественные). Методы группового вербального анализа решений дают возможность при решении задач выбора одновременно учитывать различные интересы многих участников, разнообразие и несовпадение их целей и способов выражения их предпочтений. Процедура решения задачи многокритериального выбора с применением технологии ПАКС включает три этапа.

На первом этапе, основываясь на предпочтениях ЛПР и/или знаниях эксперта, проводится снижение размерности признакового пространства путем построения иерархической системы составных критериев. Составным критерием называется обобщенный показатель, который характеризует выбранное ЛПР/экспертом свойство объектов, агрегирующее исходные характеристики. Каждая градация шкалы составного критерия является комбинацией градаций оценок исходных показателей. Процедура агрегирования исходных показателей и конструирования составных критериев включает: разбиение всех исходных показателей на группы критериев, внутри которых критерии близки друг другу по смыслу или являются неотъемлемыми характеристиками какого-либо составного критерия; построение дерева агрегирования для каждой сформированной группы критериев.

На втором этапе последовательно формируются шкалы оценок для каждого составного критерия с указанием диапазона изменения для каждой градации на шкале. Для этого используют различные методы вербального анализа решений [Ларичев, 2006], [Петровский, 2009]. Построение шкалы каждого составного критерия рассматривается как задача порядковой классификации, где в качестве классифицируемых объектов выступают комбинации градаций оценок исходных показателей, а классами являются градации оценок составного критерия. Тем самым каждая комбинация градаций оценок будет соответствовать некоторой градации оценок на шкале комплексного критерия.

На третьем этапе с помощью того или иного метода принятия решений выполняется окончательное решение рассматриваемой задачи выбора в полученном пространстве составных критериев меньшей размерности, которое определяется построенной системой составных критериев.

Формально задача снижения размерности признакового пространства имеет следующий вид:

$$X_1 \times \dots \times X_m \rightarrow Y_1 \times \dots \times Y_n, n < m,$$

где X_1, \dots, X_m – исходный набор признаков, Y_1, \dots, Y_n – новый набор признаков, m – размерность исходного признакового пространства, n – размерность нового признакового пространства. Каждый из признаков имеет свою собственную шкалу $X_i = \{x_i^1, \dots, x_i^p\}$, $i = 1, \dots, m$, $Y_j = \{y_j^1, \dots, y_j^h\}$, $j = 1, \dots, n$ с упорядоченной градацией качественных (символьных, вербальных) оценок. В ходе снижения размерности признакового пространства различные комбинации исходных признаков (кортежи оценок) последовательно агрегируются в меньшие наборы новых признаков (составных критериев), имеющих для ЛПР или эксперта

вполне определенный смысл. Итогом является иерархическая система критериев, верхний уровень которой определяется содержанием практической проблемы.

Представим процесс построения шкал составных критериев в виде однотипных блоков. Блоки содержательно выделяются ЛПР в зависимости от специфики решаемой задачи. Каждый блок классификации i -го уровня иерархии состоит из некоторого набора признаков и одного составного критерия. В качестве объектов классификации выступают все комбинации градаций оценок на шкалах признаков. Классами решений i -го уровня служат градации оценок на шкале составного критерия.

В блоке классификации $(i+1)$ -го уровня иерархии составные критерии i -го уровня считаются признаками, множество комбинаций градаций оценок которых представляет собой новые объекты классификации в сокращенном признаковом пространстве, а классами решений будут теперь градации оценок на шкале составного критерия $(i+1)$ -го уровня. Процедура повторяется до тех пор, пока не останется единственный составной критерий верхнего уровня, являющийся искомым интегральным показателем, шкала оценок которого образует упорядоченные классы решений D_1, \dots, D_q . Тем самым устанавливается соответствие между классами решений D_1, \dots, D_q и совокупностью исходных показателей – множеством $X_1 \times \dots \times X_m$ всех возможных комбинаций градаций оценок на шкалах критериев $X_i = \{x_i^1, \dots, x_i^{g_i}\}$, $i=1, \dots, m$, а также находятся границы классов. Это позволяет легко построить упорядочение или классификацию реальных вариантов A_1, \dots, A_p , оцененных по многим критериям K_1, \dots, K_m .

Шкалы оценок по составным критериям можно сформировать, воспользовавшись несколькими способами из арсенала средств вербального анализа решений. Наиболее простым и легко воспринимаемым ЛПР способом конструирования порядковой шкалы составного критерия является метод стратификации кортежей, в котором используются однотипные (например, с одинаковым числом градаций) наборы порядковых вербальных шкал исходных показателей. Метод стратификации кортежей основан на сечении многомерного дискретного признакового пространства параллельными «гиперплоскостями». Каждый слой (страта) состоит из однотипных комбинаций градаций оценок на шкалах критериев X_i , а число таких сечений определяет ЛПР из содержательных соображений. Максимально возможное число слоев можно рассчитать по формуле $s = 1 - m + \sum_{i=1}^m g_i$. Каждый слой образуется как комбинация кортежей градаций оценок, сумма номеров которых фиксирована. Число классов $r \leq s$. В более сложных процедурах построения шкал критериев используются методы вербального анализа решений ЗАПРОС или ОРКЛАСС [5, 6], которые оперируют на множестве всех возможных кортежей оценок в признаковом пространстве, образованном декартовым произведением градаций оценок на шкалах критериев $X_1 \times \dots \times X_m$. В этих случаях число возможных комбинаций оценок (альтернатив) равно $t = \prod_{i=1}^m g_i$.

Рассмотрим небольшой иллюстративный пример. Пусть необходимо построить шкалу составного критерия D из градаций оценок исходных показателей A , B и C . Допустим, что все критерии A , B , C и D имеют шкалы с тремя вербальными градациями порядковых оценок $A = \{a^0, a^1, a^2\}$, $B = \{b^0, b^1, b^2\}$, $C = \{c^0, c^1, c^2\}$ и $D = \{d^0, d^1, d^2\}$, где e^0 – высокая оценка, e^1 – средняя оценка, e^2 – низкая оценка.

Применяя для построения порядковой шкалы составного критерия метод стратификации кортежей, можно, к примеру, объединить градации оценок исходных критериев в обобщенные градации составного критерия по такому принципу: все высокие оценки по исходным показателям образуют одну высокую оценку по составному критерию, все средние оценки по исходным показателям – одну среднюю оценку, все низкие оценки по исходным показателям – одну низкую оценку по составному критерию (рис. 1).

С помощью метода ОРКЛАСС строится полная непротиворечивая порядковая классификация многопризнаковых объектов, в качестве которых в рассматриваемом случае выступают наборы градаций

оценок по исходным показателям, образующим составной критерий (рис. 2). Получаемые классы решений соответствуют градациям шкалы составного критерия.

$a^0b^0c^1$	$a^0b^1c^1$	$a^0b^0c^2$	$a^0b^1c^2$	$a^2b^1c^0$	$a^0b^2c^2$	$a^1b^1c^2$	$a^1b^2c^2$	Страты кортежей		
$a^0b^0c^0$	$a^0b^1c^0$	$a^1b^1c^0$	$a^0b^2c^0$	$a^1b^0c^2$	$a^1b^1c^1$	$a^1b^2c^0$	$a^2b^0c^2$		$a^1b^2c^1$	$a^2b^2c^2$
$a^1b^0c^0$	$a^1b^0c^1$	$a^2b^0c^0$	$a^0b^2c^1$	$a^2b^0c^1$	$a^2b^2c^0$	$a^2b^1c^1$	$a^2b^2c^1$			
d^0			d^1			d^2				
Составной критерий D										

Рис.1. Конструирование шкалы составного критерия с помощью метода стратификации кортежей.

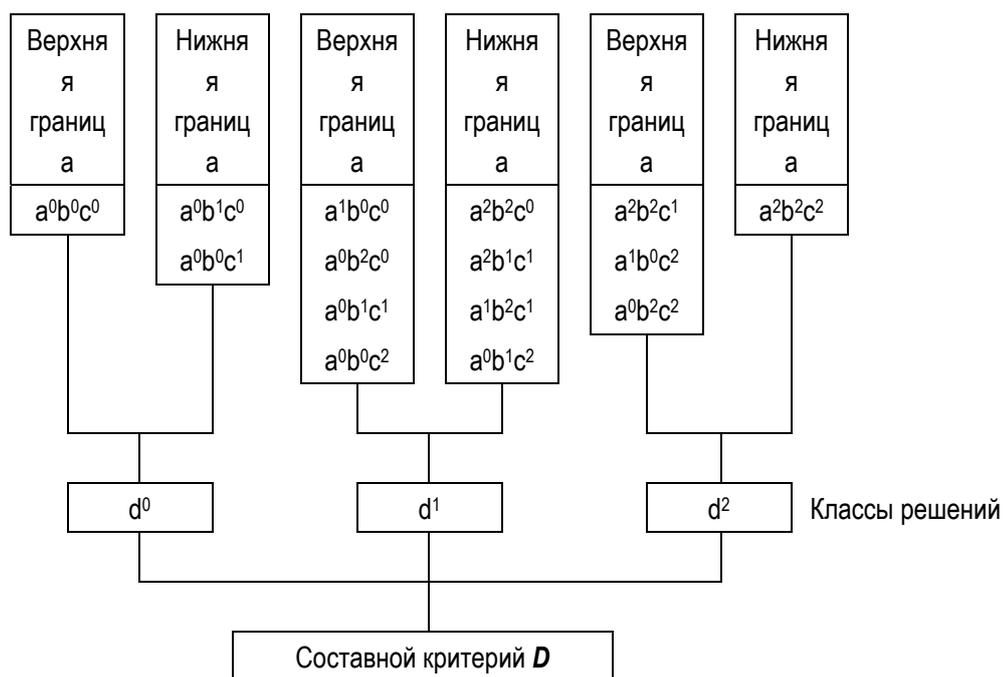


Рис. 2. Конструирование шкалы составного критерия методом ОРКЛАСС.

Новый методологический подход к снижению размерности пространства качественных признаков обладает определенной универсальностью, так как в общем случае можно оперировать как символьной (качественной), так числовой информацией. Сокращенное описание объектов дает возможность упростить процедуру решения первоначальной задачи выбора, сформировать разные наборы промежуточных критериев и воспользоваться разными способами конструирования их шкал. Сопоставление результатов, получаемых для разных иерархических систем критериев, позволяет сравнить их между собой и выбрать наиболее предпочтительную систему критериев, а также оценить качество сделанного выбора.

Агрегирование характеристик вычислительного кластера

Опишем основные этапы построения нескольких иерархических систем критериев, которые агрегируют исходные характеристики вычислительных кластеров [Лобанов, Петровский, 2012]. По результатам анализа характеристик рассматриваемых вариантов конфигураций кластеров и другой информации [Фортов и др., 2002] в качестве исходных были выбраны показатели, объединенные в следующие группы.

ХМ. Технические характеристики модуля: частота ядра процессора; разрядность ядра процессора; количество потоков; количество ядер процессора; объем поддерживаемой процессором оперативной памяти; количество процессоров в модуле; объем оперативной памяти модуля; наличие ускорителя универсальных вычислений; дисковая память модуля; наличие в модуле оптического накопителя данных.

ВХ. Вычислительные характеристики кластера: число модулей в кластере; скорость обмена между модулями; наличие встроенных средств ввода-вывода; наличие бесперебойного питания; программные характеристики кластера; возможность модернизации технических и программных средств кластера.

КХ. Конструкционные характеристики кластера: размеры кластера (высота, глубина, ширина); масса кластера; защищенность от помех.

ЭХ. Эксплуатационные характеристики кластера: энергопотребление; уровень шума; тепловыделение; условия эксплуатации (температура, влажность); наработка на отказ.

ПК. Производительность кластера.

СИ. Стоимость изготовления кластера.

Для каждого исходного показателя была сформирована своя вербальная шкала оценок с 3 или 2 градациями. Например, производительность кластера оценивалась как ПК⁰ – высокая (>2000 Гфлопс); ПК¹ – средняя (2000-500 Гфлопс); ПК² – низкая (<500 Гфлопс).

Иерархическая система критериев строится на основе предпочтений и знаний ЛПР. Для сравнения и выбора лучшего варианта кластера в рассматриваемом примере были построены две иерархические системы критериев, которые имели разную степень агрегирования исходных показателей. Прежде всего, были определены исходные показатели, которые рассматривались в качестве итоговых критериев. В качестве таковых были выделены два исходных показателя: производительность кластера и стоимость изготовления кластера. Остальные исходные характеристики были объединены в составные критерии.

В результате агрегирования все исходные характеристики были сведены в первом случае к следующим пяти критериям: ПК. Производительность кластера; СИ. Стоимость изготовления кластера; ВХ. Вычислительные характеристики кластера; КХ. Конструкционные характеристики кластера; ЭХ. Эксплуатационные характеристики кластера. Во втором случае были определены три итоговых критерия: ПК. Производительность кластера; СИ. Стоимость изготовления кластера; ОХ. Обобщенные характеристики кластера. Последний критерий объединяет вычислительные, конструкционные и эксплуатационные характеристики кластера.

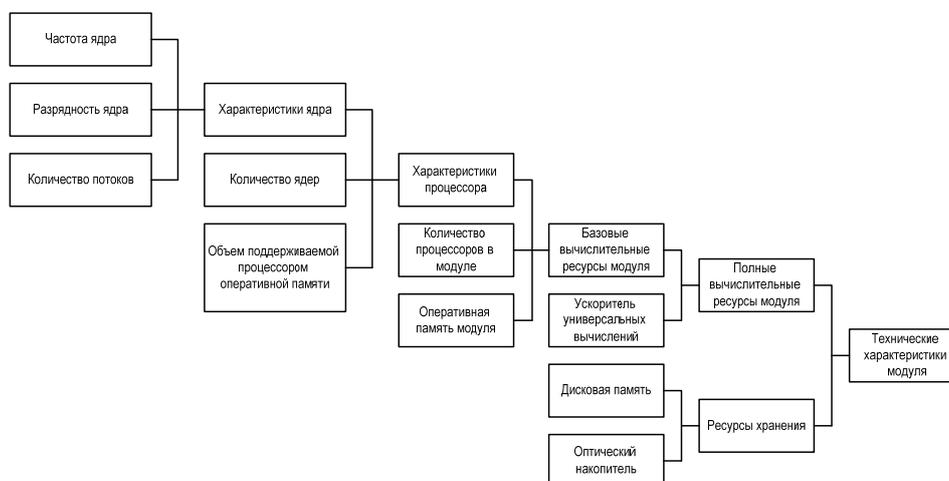


Рис. 3. Технические характеристики модуля.



Рис. 4. Вычислительные характеристики кластера.

Для каждой сформированной группы, образующей итоговый критерий, построено дерево агрегирования исходных показателей в составные критерии. Технические характеристики модуля были включены как составная часть в вычислительные характеристики кластера. Дерево технических характеристик модуля представлено на рис. 3, дерево вычислительных характеристик кластера – на рис. 4.

Для каждого составного критерия с помощью метода стратификации кортежей формировалась шкала оценок с назначением диапазона изменения характеристики критерия для каждой градации на шкале оценок. Например, шкала оценок итогового критерия ОХ. Обобщенные характеристики кластера имеет вид, аналогичный приведенному на рис. 1. В данной работе шкалы оценок исходных показателей и составных критериев формировались таким образом, чтобы количество градаций в любой шкале не превышало трех. Это связано, главным образом, с тем, что увеличение числа градаций шкалы оценок увеличивает количество комбинаций оценок при их агрегировании в составной критерий. А это обстоятельство в свою очередь усложняет процедуру агрегирования критериев и затрудняет объяснение полученных результатов для ЛПР.

Отметим, что определение комбинаций градаций оценок (кортежей оценок) по критериям предыдущего уровня иерархии, которые образуют соответствующую градацию оценок на шкале критерия следующего уровня иерархии, является субъективной неформализованной процедурой, которая выполняется ЛПР/экспертом. Конструировать каждую градацию шкалы составного критерия можно различными способами, используя разные методы для конструирования самой шкалы составного критерия, выбирая разные числа градаций и диапазоны изменения переменных на градациях шкал критериев предыдущего или данного уровня иерархии исходя из предпочтений ЛПР или знаний эксперта. При этом даже незначительное изменение диапазона переменных в какой-нибудь градации одного из критериев может оказать заметное влияние на совокупности кортежей оценок, входящих в те или иные градации оценок на шкалах составных критериев следующих уровней и итогового критерия. Именно эти обстоятельства и приводят к построению нескольких иерархических систем критериев, различающихся между собой.

Удобно рассматривать каждую такую систему критериев как выражение точки зрения (предпочтений, знаний) некоторого ЛПР/эксперта. Тем самым задача выбора вычислительного кластера трансформируется в задачу группового многокритериального выбора, которая должна решаться с помощью методов группового вербального анализа решений [Петровский, 2009].

Многокритериальный выбор конфигурации вычислительного кластера

Пусть имеется три варианта конфигурации вычислительного кластера: ВК1, ВК2 и ВК3, из которых требуется выбрать наиболее предпочтительную по многим критериям. При агрегировании исходных показателей были построены три различные иерархические системы итоговых критериев, на основе которых проводилось сравнение вариантов кластеров и выбор наилучшего. Иными словами, можно считать, что каждый вариант кластера независимо оценивался тремя экспертами по многим критериям.

Рассмотрим сначала первый случай агрегирования, в котором имелось 5 итоговых критериев: ПК, СИ, ВХ, КХ, ЭХ. Результаты многокритериальной оценки кластеров, представленные в виде кортежей вербальных оценок, приведены в табл. 1. Как следует из таблицы, кластеры остаются несравнимыми между собой, однако их различие по итоговым критериям становится более понятным. Вместе с тем такое представление результатов экспертной оценки пока не позволяет выбрать лучший вариант кластера.

Таблица 1. Оценки кластеров, представленные в виде кортежей (1 случай).

	ВК1	ВК2	ВК3
Эксперт 1	(ПК ¹ , СИ ¹ , ВХ ² , КХ ⁰ , ЭХ ⁰)	(ПК ² , СИ ⁰ , ВХ ¹ , КХ ² , ЭХ ¹)	(ПК ⁰ , СИ ² , ВХ ⁰ , КХ ² , ЭХ ²)
Эксперт 2	(ПК ¹ , СИ ¹ , ВХ ² , КХ ⁰ , ЭХ ⁰)	(ПК ² , СИ ⁰ , ВХ ⁰ , КХ ² , ЭХ ¹)	(ПК ⁰ , СИ ² , ВХ ⁰ , КХ ² , ЭХ ²)
Эксперт 3	(ПК ¹ , СИ ¹ , ВХ ² , КХ ⁰ , ЭХ ⁰)	(ПК ² , СИ ⁰ , ВХ ¹ , КХ ² , ЭХ ¹)	(ПК ⁰ , СИ ² , ВХ ⁰ , КХ ² , ЭХ ²)

Для дальнейшего сравнения кластеров и выбора наилучшей конфигурации экспертные оценки кластера ВК_{*i*}, *i*=1, 2, 3 были представлены как мультимножество или множество с повторяющимися элементами:

$$A_i = \{k_{A_i}(x_1^{e_1}) \circ x_1^{e_1}, \dots, k_{A_i}(x_1^{h_1}) \circ x_1^{h_1}, \dots; k_{A_i}(x_5^{e_5}) \circ x_5^{e_5}, \dots, k_{A_i}(x_5^{h_5}) \circ x_5^{h_5}\}.$$

Порождающим для мультимножества A_i множеством служит множество $X = X_1 \cup \dots \cup X_5$ всех градаций оценок на шкалах X_s критериев ПК, СИ, ВХ, КХ, ЭХ. Число $k_{A_i}(x_s^{e_s})$ указывает, сколько раз градация оценки $x_s^{e_s} \in X_s$, $s=1, \dots, 5$, $e_s=0, 1, 2$ встречается в описании варианта кластера ВК_{*i*}, знак \circ обозначает кратность градации оценки $x_s^{e_s}$ [Петровский, 2003]. Агрегированные многокритериальные оценки вычислительных кластеров ВК1, ВК2 и ВК3, записанные как мультимножества, имеют в первом случае следующий вид:

$$A_1 = \{0 \circ x_1^0, 3 \circ x_1^1, 0 \circ x_1^2; 0 \circ x_2^0, 3 \circ x_2^1, 0 \circ x_2^2; 0 \circ x_3^0, 0 \circ x_3^1, 3 \circ x_3^2; 3 \circ x_4^0, 0 \circ x_4^1, 0 \circ x_4^2; 3 \circ x_5^0, 0 \circ x_5^1, 0 \circ x_5^2\},$$

$$A_2 = \{0 \circ x_1^0, 0 \circ x_1^1, 3 \circ x_1^2; 3 \circ x_2^0, 0 \circ x_2^1, 0 \circ x_2^2; 1 \circ x_3^0, 2 \circ x_3^1, 0 \circ x_3^2; 0 \circ x_4^0, 0 \circ x_4^1, 3 \circ x_4^2; 0 \circ x_5^0, 3 \circ x_5^1, 0 \circ x_5^2\},$$

$$A_3 = \{3 \circ x_1^0, 0 \circ x_1^1, 0 \circ x_1^2; 0 \circ x_2^0, 0 \circ x_2^1, 3 \circ x_2^2; 3 \circ x_3^0, 0 \circ x_3^1, 0 \circ x_3^2; 0 \circ x_4^0, 0 \circ x_4^1, 3 \circ x_4^2; 0 \circ x_5^0, 0 \circ x_5^1, 3 \circ x_5^2\}.$$

Выбор наиболее предпочтительного варианта кластера проводился с помощью трех методов группового многокритериального выбора: метода АРАМИС, метода лексикографического упорядочивания по градациям критериальных оценок и метода взвешенных сумм рангов [Петровский, 2009].

Метод АРАМИС (Агрегирование и Ранжирование Альтернатив около Многопризнаковых Идеальных Ситуаций) позволяет упорядочивать многокритериальные варианты по близости к опорной точке, когда варианты представлены в виде мультимножеств. В качестве опорных точек были взяты две идеальные (возможно, гипотетические) конфигурации: кластер ВК₊, получивший по суждениям всех экспертов высокие оценки по всем критериям, и кластер ВК₋, получивший по суждениям всех экспертов низкие оценки по всем критериям. Эти варианты кластеров представляются соответственно мультимножествами

$$A_+ = \{3 \circ x_1^0, 0 \circ x_1^1, 0 \circ x_1^2; 3 \circ x_2^0, 0 \circ x_2^1, 0 \circ x_2^2; 3 \circ x_3^0, 0 \circ x_3^1, 0 \circ x_3^2; 3 \circ x_4^0, 0 \circ x_4^1, 0 \circ x_4^2; 3 \circ x_5^0, 0 \circ x_5^1, 0 \circ x_5^2\},$$

$$A_- = \{0 \circ x_1^0, 0 \circ x_1^1, 3 \circ x_1^2; 0 \circ x_2^0, 0 \circ x_2^1, 3 \circ x_2^2; 0 \circ x_3^0, 0 \circ x_3^1, 3 \circ x_3^2; 0 \circ x_4^0, 0 \circ x_4^1, 3 \circ x_4^2; 0 \circ x_5^0, 0 \circ x_5^1, 3 \circ x_5^2\}.$$

Близость конфигурации кластера BK_i к наилучшему кластеру определяется показателем относительной близости $L(BK_i)$ к опорной точке BK_+ , который вычисляется по формуле

$$L(BK_i) = d^+(BK_i) / [d^+(BK_i) + d^-(BK_i)].$$

Здесь $d^+(BK_i) = d(A_i, A_+)$, $d^-(BK_i) = d(A_i, A_-)$ – расстояния между мультимножествами A_i , A_+ и A_i , A_- в метрическом пространстве мультимножеств [Петровский, 2003], которые задаются выражением

$$d(A_i, A_{\pm}) = m(A_i \Delta A_{\pm}) = \sum_{s=1}^m w_s \sum_{e_s=1}^{h_s} |k_{A_i}(x_{s^e_s}) - k_{A_{\pm}}(x_{s^e_s})|,$$

где s – номер критерия, e^s – номер градации критериальной оценки.

Варианты кластеров упорядочиваются по возрастанию значения показателя относительной близости $L(BK_i)$ к наилучшему варианту. Наиболее предпочтительный вариант имеет меньшее значение $L(BK_i)$.

Метод лексикографического упорядочивания вариантов основан на последовательном сравнении вариантов по общему числу соответствующих градаций экспертных оценок. Сначала варианты сравнивают по числу высоких оценок или числу первых мест, затем по числу средних оценок или вторых мест, далее по числу низких оценок или третьих мест. Наиболее предпочтительным вариантом становится тот, у которого наибольшее число высоких экспертных оценок. Если вариантов с одинаковым количеством высоких оценок несколько, то сравнение этих вариантов происходит по средним оценкам, соответствующим второму месту и т.д.

Метод взвешенных сумм рангов основан на том, что разным градациям экспертных оценок ставится в соответствие определенный ранг и вес, соответствующий рангу градации. В данной работе высокой градации оценки был присвоен вес 3, средней градации – вес 2, низкой градации – вес 1. Наиболее предпочтительный вариант определяется по максимальной сумме произведений числа градаций оценок на вес ранга.

Результаты сравнения кластеров методами АРАМИС, лексикографического упорядочивания и взвешенных сумм рангов в первом случае представлены в табл. 3.

Таблица 2. Сравнение кластеров (1 случай).

	ПК ⁰	ПК ¹	ПК ²	СИ ⁰	СИ ¹	СИ ²	ВХ ⁰	ВХ ¹	ВХ ²	КХ ⁰	КХ ¹	КХ ²	ЭХ ⁰	ЭХ ¹	ЭХ ²	$L(BK_i)$	1м	2м	3м	Сумма
BK1	0	3	0	0	3	0	0	0	3	3	0	0	3	0	0	0,43	6	6	3	33
BK2	0	0	3	3	0	0	1	2	0	0	0	3	0	3	0	0,55	4	5	6	28
BK3	3	0	0	0	0	3	3	0	0	0	0	3	0	0	3	0,60	6	0	9	27

Как следует из таблицы 2, по методу АРАМИС кластер BK1 предпочтительнее кластера BK2, а кластер BK2 предпочтительнее кластера BK3: $BK1 > BK2 > BK3$; по методу лексикографического упорядочивания кластер BK1 предпочтительнее кластера BK3, а кластер BK3 предпочтительнее кластера BK2: $BK1 > BK3 > BK2$; по методу взвешенных сумм рангов кластеры BK2 и BK3 отличаются по сумме рангов незначительно и примерно равноценны, а кластер BK1 предпочтительнее их обоих: $BK1 > BK2 \approx BK3$.

Итоговое обобщенное упорядочение вариантов кластеров было построено при помощи процедуры Борда [Петровский, 2009] и имеет вид: $BK1 > BK2 \approx BK3$. Таким образом, в первом случае агрегирования критериев получаем, что кластер BK1 предпочтительнее кластеров BK2 и BK3, которые можно считать примерно равноценными.

Рассмотрим теперь второй случай агрегирования, в котором имелось 3 итоговых критерия: ПК, СИ, ОХ. Результаты многокритериальной оценки кластеров, представленные в виде кортежей вербальных оценок,

приведены в табл. 3. Как следует из таблицы 3, кластеры по-прежнему остаются несравнимыми между собой, однако их различие по итоговым критериям становится еще более понятным и наглядным.

Таблица 3. Оценки кластеров, представленные в виде кортежей (2 случай).

	ВК1	ВК2	ВК3
Эксперт 1	(ПК ¹ , СИ ¹ , ОХ ⁰)	(ПК ² , СИ ⁰ , ОХ ²)	(ПК ⁰ , СИ ² , ОХ ²)
Эксперт 2	(ПК ¹ , СИ ¹ , ОХ ⁰)	(ПК ² , СИ ⁰ , ОХ ¹)	(ПК ⁰ , СИ ² , ОХ ²)
Эксперт 3	(ПК ¹ , СИ ¹ , ОХ ⁰)	(ПК ² , СИ ⁰ , ОХ ²)	(ПК ⁰ , СИ ² , ОХ ²)

Агрегированные многокритериальные оценки вычислительных кластеров ВК1, ВК2 и ВК3, записанные как мультимножества, имеют во втором случае такой вид:

$$A_1 = \{0 \circ x_1^0, 3 \circ x_1^1, 0 \circ x_1^2; 0 \circ x_2^0, 3 \circ x_2^1, 0 \circ x_2^2; 3 \circ x_3^0, 0 \circ x_3^1, 0 \circ x_3^2\},$$

$$A_2 = \{0 \circ x_1^0, 0 \circ x_1^1, 3 \circ x_1^2; 3 \circ x_2^0, 0 \circ x_2^1, 0 \circ x_2^2; 0 \circ x_3^0, 1 \circ x_3^1, 2 \circ x_3^2\},$$

$$A_3 = \{3 \circ x_1^0, 0 \circ x_1^1, 0 \circ x_1^2; 0 \circ x_2^0, 0 \circ x_2^1, 3 \circ x_2^2; 0 \circ x_3^0, 0 \circ x_3^1, 3 \circ x_3^2\}.$$

Результаты сравнения кластеров методом АРАМИС, методом лексикографического упорядочивания и методом взвешенных сумм рангов во втором случае даны в табл. 4.

Таблица 4. Сравнение кластеров (2 случай).

	ПК ⁰	ПК ¹	ПК ²	СИ ⁰	СИ ¹	СИ ²	ОХ ⁰	ОХ ¹	ОХ ²	L(ВК _i)	1м	2м	3м	Сумма
ВК1	0	3	0	0	3	0	3	0	0	0,40	3	6	0	21
ВК2	0	0	3	3	0	0	0	1	2	0,60	3	1	5	16
ВК3	3	0	0	0	0	3	0	0	3	0,67	3	0	6	15

И во втором случае агрегирования критериев, как следует из таблицы 2, по методу АРАМИС кластер ВК1 предпочтительнее кластера ВК2, а кластер ВК2 предпочтительнее кластера ВК3: ВК1>ВК2>ВК3; по методу лексикографического упорядочивания кластер ВК1 предпочтительнее кластера ВК3, а кластер ВК3 предпочтительнее кластера ВК2: ВК1>ВК3>ВК2; по методу взвешенных сумм рангов кластеры ВК2 и ВК3 отличаются по сумме рангов незначительно и примерно равноценны, а кластер ВК1 предпочтительнее их обоих: ВК1>ВК2≈ВК3.

Итоговое обобщенное упорядочение вариантов кластеров во втором случае агрегирования критериев также строилось при помощи процедуры Борда. Оно совпадает с упорядочением, полученным в первом случае агрегирования критериев, и имеет вид: ВК1>ВК2≈ВК3. Таким образом, в обоих случаях агрегирования критериев получаем одинаковый результат: кластер ВК1 предпочтительнее кластеров ВК2 и ВК3, которые считаются примерно равноценными.

Обсудим полученные результаты. В первом случае кластер ВК1 обладает высокими конструкционными и эксплуатационными характеристиками, имеет среднюю стоимость и среднюю производительность, но низкие вычислительные характеристики. Кластер ВК2 имеет самую низкую стоимость изготовления среди трех кластеров, средние вычислительные и эксплуатационные характеристики, но и самую низкую производительность. Кластер ВК3 является лучшим по производительности, но имеет высокую стоимость, низкие конструкционные и эксплуатационные характеристики. Преобладание высоких и средних оценок по

большинству критериев выводит кластер ВК1 на первое место по предпочтительности в сравнении с кластерами ВК2 и ВК3. Совокупности оценок по критериям, в целом, делают кластеры ВК2 и ВК3 примерно равноценными, но уступающими по предпочтительности кластеру ВК1.

Аналогичным образом во втором случае кластер ВК1 обладает высокими обобщенными характеристиками, имеет среднюю стоимость и среднюю производительность. Кластер ВК2 имеет самую низкую стоимость изготовления среди трех кластеров, скорее низкие обобщенные характеристики и самую низкую производительность. Кластер ВК3 является лучшим по производительности, но имеет высокую стоимость, низкие обобщенные характеристики. Преобладание высоких и средних оценок по большинству критериев выводит кластер ВК1 на первое место по предпочтительности в сравнении с кластерами ВК2 и ВК3. Совокупности оценок по критериям, в целом, делают кластеры ВК2 и ВК3 примерно равноценными, но уступающими по предпочтительности кластеру ВК1.

Заключение

Итак, несмотря на первоначальную сложность задачи выбора, обусловленную большим числом исходных характеристик вычислительных кластеров, удалось снизить размерность признакового пространства описания за счет агрегирования показателей. С помощью методов группового многокритериального принятия решений найдена предпочтительная конфигурация кластера, описываемая конкретными и понятными значениями признаков. Совпадение результатов выбора лучшего варианта кластера для двух разных иерархических систем критериев, различающихся степенью агрегирования исходных показателей, можно расценивать как свидетельство достаточно высокой надежности предложенного подхода.

Вместе с тем построение систем критериев показало, что результаты агрегирования сильно зависят от знаний экспертов и предпочтений ЛПР. Более того, даже если задача выбора решается одним человеком, то и в этом случае на разных этапах процедуры агрегирования критериев результаты могут в большей или меньшей степени отличаться друг от друга в зависимости от суждений человека, как обусловленных взглядом на проблему с разных сторон, так и продиктованных жизненным опытом.

Новая многоэтапная технология решения задач многокритериального выбора в сочетании со снижением размерности пространства признаков обладает определенной универсальностью, так как в общем случае можно оперировать и с количественной (числовой), и с качественной (символьной, вербальной) информацией. Технология обеспечивает систематизацию имеющейся информации, облегчает выбор окончательного варианта решения, позволяет проанализировать и обосновать итоговые результаты. Важной особенностью технологии является возможность её совместного использования с другими методами принятия решений и технологиями обработки информации.

Разработанная технология может быть успешно применена в различных проблемных областях, где необходимо получить интегральный показатель оценки деятельности на основе исходной слабо структурируемой качественной информации. В частности, примером такой задачи может служить многоаспектная оценка результативности научных проектов, поддержанных Российским фондом фундаментальных исследований [Петровский и др., 2009].

Благодарности

The paper is published with partial support by the project ITHEA XXI of the ITHEA ISS (www.ithea.org) and the ADUIS (www.aduis.com.ua).

Библиография

- [Ларичев, 2006] Ларичев О.И. Вербальный анализ решений / Под ред. А.Б.Петровского. – М.: Наука, 2006.
- [Лобанов, Петровский, 2012] Лобанов В.Н., Петровский А.Б. Агрегирование критериев в задаче выбора вычислительного кластера. // Теория и практика системного анализа: Сборник трудов Всероссийской молодежной конференции. – Белгород: ИД «Белгород», 2012. – С.107-111.
- [Петровский и др., 2009] Петровский А.Б., Ройзензон Г.В., Тихонов И.П. Построение интегральных показателей оценки результативности научных проектов // Intelligent Support of Decision Making / Ed. by K. Markov, A. Voloshyn, K. Ivanova, I. Mitov. – No. 10. – Sofia: FOI ITHEA, 2009. – P. 59–66.
- [Петровский, 2003] Петровский А.Б. Пространства множеств и мультимножеств. – М: Едиториал УРСС, 2003.
- [Петровский, 2009] Петровский А.Б. Теория принятия решений. – М.: Издательский центр «Академия», 2009.
- [Петровский, Ройзензон, 2008] Петровский А.Б., Ройзензон Г.В. Интерактивная процедура снижения размерности признакового пространства в задачах многокритериальной классификации // Поддержка принятия решений: Труды Института системного анализа Российской академии наук / Под ред. А.Б. Петровского. – М.: Едиториал УРСС, 2008. – Т.35. – С.43-53.
- [Петровский, Ройзензон, 2012] Петровский А.Б., Ройзензон Г.В. Многокритериальный выбор с уменьшением размерности пространства признаков: многоэтапная технология ПАКС. // Искусственный интеллект и принятие решений. 2012. – № 4. – С.88-103.
- [Ройзензон, 2005] Ройзензон Г.В. Многокритериальный выбор вычислительных кластеров. // Методы поддержки принятия решений. Труды Института системного анализа Российской академии наук. / Под ред. С.В. Емельянова, А.Б. Петровского. – М.: Едиториал УРСС, 2005. – Т.12. – С.68-94.
- [Фортов и др., 2002] Создание и применение системы высокопроизводительных вычислений на базе высокоскоростных сетевых технологий / В.Е. Фортов, Г.И. Савин, В.К. Левин и др. // Информационные технологии и вычислительные системы. 2002. – № 1. – С.3-9.

Сведения об авторах

Петровский Алексей Борисович – д.т.н., профессор, заведующий лабораторией Института системного анализа РАН, Россия, Москва 117312, пр-т 60-летия Октября, 9, e-mail: pab@isa.ru

Лобанов Василий Николаевич – ведущий инженер ОАО «НИИ вычислительных комплексов им. М.А. Карцева», Россия, Москва 117437, ул. Профсоюзная, 108, e-mail: fisher_1980@mail.ru

Заболеева-Зотова Алла Викторовна – д.т.н., профессор, профессор Волгоградского государственного технического университета, Россия, Волгоград 400005, пр-кт Ленина, 28, e-mail: zabzot@gmail.com

Шитова Татьяна Алексеевна – экономист Института системного анализа РАН, Россия, Москва 117312, пр-т 60-летия Октября, 9, e-mail: tanya-petrovskay@yandex.ru

МЕТОДЫ МОДЕЛИРОВАНИЯ ВРЕМЕННЫХ ЗАВИСИМОСТЕЙ В ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМАХ С ИСПОЛЬЗОВАНИЕМ ТЕМПОРАЛЬНЫХ ПРЕЦЕДЕНТОВ

Александр Еремеев, Иван Куриленко

Аннотация: Работа посвящена методам моделирования временных зависимостей в интеллектуальных системах типа систем поддержки принятия решений (ИСППР) на основе прецедентов. Описываются и анализируются методы построения рассуждений на основе темпоральных прецедентов (СВР-систем). Анализируются преимущества применения систем на основе темпоральных прецедентов для поддержки принятия решений в открытых и динамических предметных областях в условиях наличия неопределенности в исходных данных и достаточно жестких временных ограничений.

Ключевые слова: интеллектуальные системы поддержки принятия решений, моделирование временных зависимостей, временные логики, задачи согласования временных ограничений, темпоральные прецеденты, СВР-системы.

ACM Classification Keywords: H.4.2 [Information systems applications]: Types of systems – Decision support; I.2.3 [Artificial intelligence]: Deduction and Theorem Proving – Uncertainty, "fuzzy," and probabilistic reasoning; I.2.4 [Artificial intelligence]: Knowledge Representation Formalisms and Methods – Temporal logic.

Введение

Важной задачей в области создания перспективных интеллектуальных систем типа интеллектуальных систем поддержки принятия решений реального времени (ИСППР РВ) является задача моделирования правдоподобных рассуждений [Вагин и др., 2008].

Наряду с широко используемыми в области искусственного интеллекта (ИИ) методами правдоподобных рассуждений на основе индукции, абдукции, аргументации и аналогии активно применяются методы на основе прецедентов (СВР-методы) [Варшавский и др., 2009]. Вывод на основе прецедентов позволяет принимать решения в новых, неизвестных ситуациях, используя или адаптируя решение, принятое ранее в уже известной ситуации, т.е. используя уже накопленный к моменту принятия решения опыт [Watson et al., 1994, Schank, Abelson, 1977].

Обычно СВР-методы включают четыре этапа, образующих цикл обучения по примерам (прецедентам) - СВР-цикл. Информация о новой проблемной ситуации используется для извлечения из базы прецедентов (БП) прецедентов, удовлетворяющих заданному порогу подобия [Варшавский и др., 2006]. Они используются повторно при получении решения задачи (определяемой проблемной ситуацией). В случае необходимости, предложенное решение может быть адаптировано к особенностям новой ситуации и применено на практике. Если решение успешно применено оно совместно с описанием проблемной ситуации образует новый прецедент, который сохраняется в БП, что позволяет системе накапливать опыт (в виде прецедентов).

Таким образом, CBR-система должна обеспечивать: поиск прецедентов, наиболее соответствующего наблюдаемой ситуации с определением меры сходства (извлечение прецедента), пересмотр и адаптацию полученного решения, сохранение вновь принятого решения в качестве нового прецедента или для изменения имеющегося (обучение) [Eremeev et al., 2009]. На практике в CBR-системах обычно предусматривается база решающих правил, база известных прецедентов и алгоритмы обобщения, которые используются для построения решающих правил, на основе которых распознаются новые проблемные ситуации.

Извлечение прецедентов напрямую связано со способом представления прецедентов и соответственно со способом организации БП. Известны различные методы извлечения прецедентов из БП системы: метод ближайшего соседа и его модификации; метод поиска на деревьях решений; метод извлечения на основе знаний; метод извлечения с учетом применимости прецедентов.

CBR-методы хорошо проработаны и используются на практике, однако, как правило во многих случаях при извлечении прецедентов используются «мгновенные» снимки ключевых параметров контролируемого объекта или системы, а история их изменения не учитывается [Еремеев и др., 2011].

С учетом того, что природа физических процессов такова, что к одному и тому же значению ключевых параметров контролируемый процесс зачастую может перейти разными путями, которые и будут во многом определять дальнейшую динамику их изменения, остро встает вопрос об улучшении CBR-методов за счет надления их возможностью учета временных и причинно-следственных зависимостей [Еремеев и др., 2013]. Учитывая историю изменения состояний наблюдаемого объекта или процесса, можно находить более качественные решения, чем на основании анализа только по текущему состоянию. В связи с этим актуальной задачей является разработка методов рассуждений на основе прецедентов, позволяющих учитывать историю изменения параметров, а также адаптированных к учету фактора времени алгоритмы извлечения прецедентов.

В рамках данной работы предлагается расширение CBR-методов, позволяющее учитывать поведение контролируемого процесса или объекта во времени. При учете фактора времени появляется возможность рассматривать проблемную ситуацию в динамике, что позволяет отслеживать процесс изменения значений. Использование темпоральных прецедентов дает возможность строить предположения, используя не только критерий сходства, но и учитывать временные и причинно-следственные зависимости.

Построение темпорального расширения CBR-методов

Известно несколько методов внедрения фактора времени в существующие способы описания прецедентов – в частности методы, на основе подмены «мгновенного» снимка параметров историей изменения параметров, метод учета истории принимаемых решений и метод временных аргументов. Метод подмены «мгновенного» снимка ключевых параметров – историей их изменения является наиболее простым, но при этом и более затратным с точки зрения объема хранимой информации [Смирнова и др., 2012, Еремеев и др., 2013].

Эта история может быть представлена в виде матрицы значений, составленной на определенный период наблюдения с определенным временным шагом (тактом). Для работы с такими прецедентами могут использоваться имеющиеся алгоритмы и методы, так как легко перейти к обычному параметрическому представлению прецедента путем переупаковки матрицы в вектор через последовательную запись в него элементов ее столбцов. При этом эксперту или ЛПР в интерфейсе CBR-системы может представляться

более удобная для интерпретации и заполнения матричная или графическая форма, а ее трансформация в массив СВР-система может выполняться автоматически.

Методы учета истории принимаемых решений позволяют учитывать решения, выводимых ранее на прошлых этапах, при выборе решения на текущем этапе. Т.е. решения, принятые на прошлом шаге влияют на выбор решения на текущем шаге.

В первую очередь определяется отрезок времени для анализа проблемной ситуации (глубина анализа истории изменения параметров). Далее рассматриваемый промежуток времени разбивается на N равных интервалов. После этого происходит извлечение прецедентов с заданным порогом подобия на основе сравнения значений параметров ситуации в каждый момент времени i , где $i = 0, 1, \dots, N$ с прецедентами, имеющимися в БП. Таким образом, для всех моментов времени i получаются варианты прецедентов с некоторыми оценками подобия, которые могут быть выбраны как соответствующие ситуации в этой точке для текущего значения порога подобия. На следующем этапе по интегральным оценкам в точках i выбирается наиболее вероятный прецедент или группа прецедентов, удовлетворяющих поисковому пороговому условию. Каждому результирующему прецеденту в соответствие ставится усредненная мера сходства (вычисляемая как отношение суммы мер сходства для этого прецедента во всех точках i от 0 до N). Далее среди получившегося множества прецедентов с оценками сходства выбираются удовлетворяющие заданному порогу [Eremeev et al., 2012].

Следует отметить, что с использованием параметров в точке i с помощью интерполяции можно построить прогноз развития проблемной ситуации (на уровне параметров). Также для такого прогноза, могут использоваться данные из имеющихся прецедентов, в случае комбинации этого метода, с рассмотренным ранее.

На результат поиска прецедентов, оказывают влияние выбор метрики, выбор шага, с которым разбивается на промежутки рассматриваемый отрезок времени, и выбор порогового значения, определяющего степень сходства. Выбор метрики в каждом конкретном случае производится по-разному, в зависимости от целей ЛПР, физической и статистической природы используемой информации при управлении сложным объектом и других ограничений и факторов. Вариация размеров шага оказывает влияние не только на точность выбора прецедента, но и на скорость поиска, а так же на величину погрешности, с которой может быть построен прогноз.

Далее в сформированной БП, где каждый прецедент представляется в параметризованном виде, происходит сравнение значений параметров ситуации в каждый момент времени t_i , где $i = 1, \dots, N$. Значения параметров прецедентов сравниваются между собой по методу ближайшего соседа с учетом выбранной метрики (например, Евклидовой) и соответствующего порогового значения [Варшавский и др., 2006, 2009]. Каждому результирующему прецеденту в соответствие ставится мера сходства (расстояние Хэмминга), которая определяется по правилу: «При совпадении всех параметров в описании прецедента и текущей ситуации степень сходства будет равна 1, а каждый совпавший параметр дает вклад равный $1/K$, где K – число параметров в описании прецедента и текущей ситуации».

Следует отметить, что на результат поиска прецедентов, удовлетворяющих проблемной ситуации, оказывает влияние несколько факторов. Так, влияние оказывает выбор шага, с которым разбивается на промежутки рассматриваемый отрезок времени. Вариация размеров шага дает различное количество дискретных точек, по которым происходит сравнение значений параметров прецедентов. Это оказывает влияние не только на точность выбора прецедента, но и на скорость выполнения поискового алгоритма, а так же на величину погрешности, с которой может быть определено соответствие. Также важным оказывается выбор порогового значения, определяющего степень сходства.

Предложенные методы могут быть использованы как по отдельности, так и в комбинации друг с другом. В практических приложениях важно правильно определить величину N и размер такта. Если следует учитывать вероятность скачкообразного изменения поведения параметров контролируемого объекта или процесса, то предпочтительнее использовать первый из рассмотренных выше методов.

Следует отметить, что у рассмотренных методов есть несколько недостатков. Первым является необходимость хранения большого объема информации. Ведь даже в случае, если в такты i и $i+1$ ничего не изменялось, будут повторно сохранены все контрольные параметры. Вторым недостатком может рассматриваться неявная форма учета временных характеристик (через изменения), что не позволяет в явной форме задавать сложные временные зависимости. Оба эти недостатка могут быть устранены, если для построения представления прецедента будет использована временная логика. В настоящее время известно большое количество временных логик, характеризующихся разным уровнем выразительности и разным уровнем сложности алгоритмов вывода. С учетом фактора, что разрабатываемые методы создаются для применения в составе современных ИС реального времени, следует рассматривать логики, для которых существуют алгоритмы вывода с полиномиальными оценками сложности. Исходя из этого, для реализации механизма вывода на основе темпоральных прецедентов могут быть применены качественная точечная временная логика и метрическая временная логика [Еремеев и др., 2007].

Использование темпоральной логики

Рассмотрим метод представления прецедентов с использованием выражений метрической точечной временной логики. Данный выбор обусловлен тем, что для этой логики существуют алгоритмы вывода с полиномиальной оценкой сложности [Еремеев и др., 2010]. Эти алгоритмы основываются на переходе к задаче согласования временных ограничений (ЗСВО).

Метрическая точечная ЗСВО (МЗСВО) определяется как $Z=(V,D,C)$, где $V = \{V_1..V_k\}$ - конечное множество временных переменных, соответствующих моментам времени; D - область значений временных переменных (является либо множеством вещественных чисел, либо множеством целых чисел); C - конечное число бинарных временных ограничений вида $C_{ij}=\{[a_1, b_1], \dots, [a_k, b_k]\}$, где интервалы попарно не пересекаются. Для решения задачи требуется скорректировать все ограничения так, чтобы они не противоречили друг другу. Если подобная корректировка возможна МЗСВО считается согласованной, иначе – несогласованной.

Каждое бинарное ограничение C_{ij} определяет для временных переменных V_i и V_j разрешенную дистанцию между соответствующими моментами времени и интерпретируются как дизъюнктивное ограничение: $(a_1 \leq V_j - V_i \leq b_1) \vee \dots \vee (a_k \leq V_j - V_i \leq b_k)$.

Над метрическими точечными ограничениями определяются следующие операции:

- *отрицание ограничения* $C_{ij}=\{[a_1, b_1], \dots, [a_k, b_k]\}$ ($\sim C_{ij}$) определяется как $\sim C_{ij}=\{[-b_1, -a_1], \dots, [-b_k, -a_k]\}$;
- *пересечение ограничений* T и S ($T \cap S$) содержит значения, которые одновременно присутствуют в обоих ограничениях;
- *композиция ограничений* T и S ($T \cdot S$) содержит значения r , для которых существуют значения $t \in T$ и $s \in S$, такие, что $r = t+s$.

МЗСВО может быть применена как основа для описания наблюдаемых ситуаций и прецедентов [Куриленко, 2012]. При этом есть возможность фиксации как самого факта наступления тех или иных событий, их порядка, так и времени их появления (метрики). Далее будем предполагать, что эти МЗСВО согласованы и преобразованы к минимальной форме.

Для задания структуры темпорального прецедента на основе МВЛ вводится определение ситуации как $S = \langle V, D, C, P \rangle$, где Z - МЗСВО; $P = \{P^1, \dots, P^k\}$ – наборы параметров, $P^1 \dots P^k$, где $P^i = \{p_{i1} \dots p_{im}\}$ - задают набор параметров, которые характеризовали состояние контролируемого объекта или процесса в моменты времени $V_1 \dots V_k$ соответственно.

Темпоральный прецедент $U = \langle V, D, C, P, R \rangle$ формально определяется как ситуация, дополненная диагнозом и рекомендациями ЛПП - R . Кроме этого компонент R может содержать статус прецедента (норма, ошибка, предупреждение), описание, ссылки, и т.д.

Организация прецедентов на основе МЗСВО позволяет учитывать как последовательности событий, так и их длительности. Для определения прецедента на основе выборки похожих ситуаций (обучения) используется смягчение временных ограничений C_{ij} в одном из вариантов так, чтобы выполнялись условия его подобия другим ситуациям.

Рассмотрим алгоритм построения темпорального прецедента по истории изменения параметров. Он содержит несколько этапов. На первом этапе используется алгоритм 1 (алг. 1) для сжатия истории изменения параметров наблюдаемого объекта в серию событий $S = \{e_1 \dots e_j\}$, где $e_i = (t_i, P_i)$ – событие, $t_i \in \mathbb{R}$ – время наблюдения события, $P_i = (p_{i1} \dots p_{ik})$ – параметрическое описание объекта в момент времени t_i . На втором шаге если регистрируется новый прецедент, то полученная серия событий нормализуется (по времени) и трансформируется в МЗСВО с помощью алг. 2.

Алгоритм 1. Формирование сжатого описания ситуации по истории изменения параметров наблюдаемого объекта

Входные данные: $H = \{H_i\}$, где $H_i = (p_{i1} \dots p_{ik})$ - параметрическое описание объекта на i -м такте, τ – число записанных тактов

Выходные данные: $S = \{e_j\}$ – сжатое описание ситуации, где $e_j = (t_j, P_j)$ – событие, $t_j \in \mathbb{R}$ – время наблюдения события, $P_j = (p_{j1} \dots p_{jk})$ – параметрическое описание объекта в момент времени t_j

01: $k \leftarrow 0$

02: $S \leftarrow S \cup \{(k, H_k)\}$

03: **for** ($i \leftarrow 1$ **to** $\tau - 1$)

04: {

05: **if** ($H_i \neq H_k$)

06: {

07: $S \leftarrow S \cup \{(i, H_i)\}$

08: $k \leftarrow i$

09: }

10: }

11: **return** S

При этом получается МЗСВО в подклассе STP (Simple Temporal Problems) с точными ограничениями. Если же уточняется уже имеющийся прецедент, то после выполнения алг. 2 используется алг. 3 для получения объединенного прецедента.

Отметим, что в строке 8 алг.3 могут применяться различные стратегии. Простое объединение ограничений может привести к увеличению уровня фрагментации, чтобы ее избежать и при этом смягчить получаемые ограничения можно применить метод верхне-нижнего сжатия. В этом случае получаемая МЗСВО останется в подклассе STP и на уровне оценки сходства наблюдаемой ситуации и прецедента по времени будут разрешены любые значения времени наступления рассматриваемого события между нижней и верхней границами (рис. 1).

Отметим, что в алг. 3 предполагается, что состояние контролируемого объекта в объединяемых прецедентах идентично в соответствующие моменты времени. Однако на практике, такое условие является достаточно жестким.

Алгоритм 2. Формирование первичного темпорального прецедента по описанию ситуации

Входные данные: $S=\{e_i\}$, $i=0..n$, где $e_i=(t_i, P_i)$ – событие, $t_i \in R$ – время наблюдения события, $P_i=(p_1..p_k)$ – параметрическое описание объекта в момент времени t_i , Q – диагноз и рекомендации ЛПР

Выходные данные: $U = (V, P', D, C, W)$ – темпоральный прецедент, где $V=\{V_i\}$, $i=0..n$, – множество переменных (моментов времени), $V_i \in R$, $P' = \{P'_i\}$ – параметрическое описание объекта в момент времени V_i , $P'_i=(p_1..p_k)$, D - область значений временных переменных (множество вещественных чисел); C - конечное число бинарных временных ограничений вида $C_{ij}=\{[a_1, b_1], \dots, [a_k, b_k]\}$, где интервалы попарно не пересекаются, Q – диагноз и рекомендации ЛПР

01: $e_i \leftarrow e_i=(t_i, P_i) \in E : \forall j \neq i \ t_j > t_i$, где $e_j=(t_j, P_j)$

02: **foreach** ($e_k \in S$)

03: {

04: $e_k \leftarrow (t_k - t_i, P_k)$

05: $V \leftarrow V \cup \{V_k\}$

06: $P' \leftarrow P' \cup \{P_k\}$

07: }

08: **for** ($i \leftarrow 0$ to $n+1$)

09: **for** ($j \leftarrow 0$ to $n+1$)

10: {

11: **if** ($i \neq j$)

12: $C_{ij} \leftarrow \{[t_i - t_j, t_j - t_i]\}$, где t_i и t_j определяются из $e_i=(t_i, P_i)$ и $e_j=(t_j, P_j)$

13: **else**

14: $C_{ij} \leftarrow \emptyset$

15: }

16: **return** $U = (V, P', D, C, Q)$

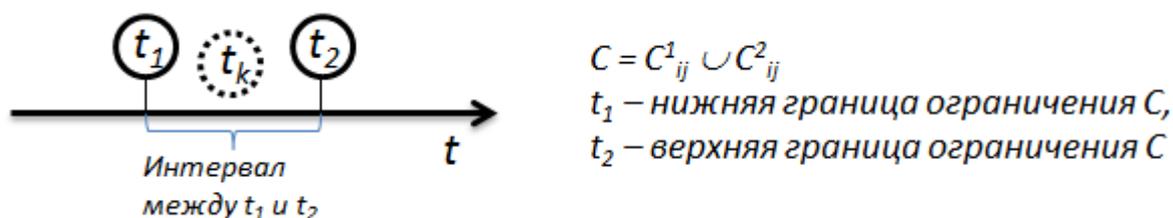


Рис. 1. Иллюстрация стратегии для уменьшения фрагментации метрических временных ограничений при формировании темпоральных прецедентов слиянием

Алгоритм 3. Слияние темпоральных прецедентов

Входные данные: U^1, U^2 – прецеденты для слияния, где $U^i = (V^i, P^i, D^i, C^i, Q^i)$ – где $V^i = \{V_j\}, j=0..n$, – множество переменных (моментов времени), $V_j \in R, P^i = \{P_j\}$ – параметрическое описание объекта в момент времени $V_j, P_j = (p_1, \dots, p_k), D^i$ – область значений временных переменных (множество вещественных чисел); C^i – конечное число бинарных временных ограничений вида $C_{ij} = \{[a_1, b_1], \dots, [a_k, b_k]\}$, где интервалы попарно не пересекаются, Q – диагноз и рекомендации ЛПР.

Предполагается что $|V^1| = |V^2|$ и $P^1 = P^2$

Выходные данные: $M = (V, P^*, D, C, Q)$ – темпоральный прецедент, где $V = \{V_j\}, j=0..n$, – множество переменных (моментов времени), $V_j \in R, P = \{P_j\}$ – параметрическое описание объекта в момент времени $V_j, P_j = (p_1, \dots, p_k), D$ – область значений временных переменных (множество вещественных чисел); C – конечное число бинарных временных ограничений вида $C_{ij} = \{[a_1, b_1], \dots, [a_k, b_k]\}$, где интервалы попарно не пересекаются

01: **if** ($|V^1| \neq |V^2|$)

02: **return** // Не удалось построить объединенный прецедент

03: **for** ($i \leftarrow 0$ to $n+1$) // Смягчение временных ограничений

04: **for** ($j \leftarrow 0$ to $n+1$)

05: {

06: **if** ($i \neq j$) $C^*_{ij} \leftarrow C^1_{ij} \cup C^2_{ij}$

07: }

08: **return** $M = (V, P^*, D, C^*, Q)$

Поэтому имеет смысл реализовать алгоритм, производящий усреднение значений параметров при объединении событий (как это делается в алг. 4) или осуществляющий переход от точного значения параметра к допустимому диапазону изменения его значения. Следует отметить, что усреднение не единственный выход в данной ситуации. Например, можно реализовать также учет качественных характеристик для каждого параметра, таких как рост, постоянство, уменьшение и так далее.

При поиске прецедента для наблюдаемой ситуации используется алгоритм, учитывающий временные характеристики и состояние контролируемого объекта или процесса в момент наблюдения тех или иных событий (алг. 5). Для вычисления соответствия временной структуры наблюдаемой ситуации и прецедента могут быть предложены несколько методов [Egemenov et al., 2012]. Самым простым из них является метод с «жесткими» ограничениями. Он предполагает точное соответствие наблюдаемых событий и событий, имеющих в прецеденте, а также для каждого метрического ограничения C_{ij} в

прецеденте его пересечение с ограничением C_{ij}^* в наблюдаемой ситуации должны равняться C_{ij} ($C_{ij} \cap C_{ij}^* = C_{ij}$). При этом для того, чтобы события в прецеденте и анализируемой ситуации соответствовали друг другу предполагается использовать для их нумерации числа, получаемые в результате сортировки по имени параметра и времени. В этом случае для определения подобия параметров контролируемого объекта или процесса в момент наступления того или иного события может использоваться подходящий на практике метод. Этот метод полезен для систем, в которых временные характеристики протекающих процессов (время наступления событий и их порядок) достаточно стабильны. Однако столь жесткое соответствие временной структуры необходимо далеко не всегда. Смягчение условия $C_{ij} \cap C_{ij}^* = C_{ij}$ может рассматриваться в нескольких вариантах. В первом снимается условие, что для каждого метрического ограничения C_{ij} в прецеденте его пересечение с ограничением C_{ij}^* в наблюдаемой ситуации должны равняться C_{ij} заменяется на то, что это пересечение не пусто и C_{ij}^* вложено в C_{ij} ($C_{ij} \cap C_{ij}^* \neq \emptyset$ и $C_{ij}^* \subseteq C_{ij}$). Во втором случае снимается требование вложенности ограничения C_{ij}^* в C_{ij} . Также это условие можно заменить на условие близости границ в ограничениях ($C_{ij} = \{[a_1, b_1], \dots, [a_k, b_k]\}$ и $C_{ij}^* = \{[a_1^*, b_1^*], \dots, [a_k^*, b_k^*]\} : (|a_1 - a_1^*| < \varepsilon) \wedge \dots \wedge (|a_k - a_k^*| < \varepsilon) \wedge (|b_1 - b_1^*| < \varepsilon) \wedge \dots \wedge (|b_k - b_k^*| < \varepsilon)$, где $\varepsilon > 0$). Дальнейшее смягчение условий возможно при анализе степени сходства с учетом изменения истории изменений каждого параметра по отдельности. Тут возможны разные стратегии (например, учитывающие ограничения между событиями изменения значений разных параметров с большим весом, чем ограничения, наложенные на события полученные из-за изменения одного параметра) [Куриленко, 2012].

Алгоритм 4. Слияние темпоральных прецедентов

Входные данные: U^1, U^2 – прецеденты для слияния, где $U^i = (V^i, P^i, D^i, C^i)$ – где $V^i = \{V_{ij}, i=0..n$, – множество переменных (моментов времени), $V_{ij} \in R$, $P^i = \{P_{ij}\}$ – параметрическое описание объекта в момент времени V_{ij} , $P_{ij} = (p_1, \dots, p_k)$, D^i – область значений временных переменных (множество вещественных чисел); C^i – конечное число бинарных временных ограничений вида $C_{ij} = \{[a_1, b_1], \dots, [a_k, b_k]\}$, где интервалы попарно не пересекаются. Предполагается что $|V^1| = |V^2|$.

Выходные данные: $M = (V, P, D, C)$ – темпоральный прецедент, где $V = \{V_{ij}, i=0..n$, – множество переменных (моментов времени), $V_{ij} \in R$, $P = \{P_{ij}\}$ – параметрическое описание объекта в момент времени V_{ij} , $P_{ij} = (p_1, \dots, p_k)$, D – область значений временных переменных (множество вещественных чисел); C – конечное число бинарных временных ограничений вида $C_{ij} = \{[a_1, b_1], \dots, [a_k, b_k]\}$, где интервалы попарно не пересекаются

01: if ($|V^1| \neq |V^2|$)

02: **return** // Не удалось построить объединенный прецедент

03: **for** ($i \leftarrow 0$ to $n+1$) // Усреднение значений параметров

04: {

05: $P_i = (P_i^1 + P_i^2) / 2$

05: }

08: **for** ($i \leftarrow 0$ to $n+1$)

09: **for** ($j \leftarrow 0$ to $n+1$)

10: {

11: **if** ($i \neq j$)

12: $C_{ij}^* \leftarrow C_{ij}^1 \cup C_{ij}^2$

13: }

14: **return** $M = (V1, P', D, C^*)$

В случае если анализ длительностей не принципиален, а важен порядок наступления событий можно воспользоваться качественной временной логикой – например, точечной временной логикой, для которой имеются быстродействующие алгоритмы вывода.

Для определения подобия (сходства) по параметрам, описывающих состояние контролируемого объекта в конкретный момент времени, может быть применен подходящий для решаемой задачи метод.

При сравнении с альтернативными подходами к представлению темпоральных прецедентов предложенные методы являются достаточно сбалансированными и позволяют с большей легкостью осуществлять CBR-цикл, чем, например, при использовании аппарата искусственных нейронных сетей для организации вывода на основе прецедентов с учетом фактора времени [Варшавский и др., 2012].

Алгоритм 5. Извлечение прецедента

Входные данные: S – наблюдаемая ситуация, где $S = \{e_i\}, i=0..n$, где $e_i = (t_i, P_i)$ – событие, $t_i \in R$ – время наблюдения события, $P_i = (p_1, .., p_k)$ – параметрическое описание объекта в момент времени t_i , Q – диагноз и рекомендации ЛПР,

$U = \{U_i\}$ – база прецедентов, где $U_k = (V^k, P^k, D^k, C^k)$ – темпоральный прецедент, где $V^k = \{V_j^k\}, i=0..n$, – множество переменных (моментов времени), $V_i \in R$, $P^k = \{P_j^k\}$ – параметрическое описание объекта в момент времени V_i , $P_j^k = (p_1, .., p_k)$, D^k – область значений временных переменных (множество вещественных чисел); C^k – конечное число бинарных временных ограничений вида $C_{ij} = \{[a_1, b_1], .., [a_k, b_k]\}$, где интервалы попарно не пересекаются.

Выходные данные: U_k – прецедент, соответствующий ситуации S или \emptyset

```

01:  $U^s =$  Алгоритм 2 ( $S, \emptyset$ )
02: foreach ( $U^k = (V^k, P^k, D^k, C^k) \in U$ ) {
03:   bool isFailed  $\leftarrow$  false
04:   foreach ( $C_{ij}^k \in C^k$ ) // Сравнение временных характеристик
05:     if ( $C_{ij}^k \cap C_{ij}^s = \emptyset$ , где  $C_{ij}^k \in C^k, C_{ij}^s \in C^s$ )
06:       {
07:         isFailed  $\leftarrow$  true;
08:         break;
09:       }
10:   if (!isFailed)
11:     {
12:       for ( $i \leftarrow 0$  to  $|V^s|$ )
13:         if (!СравнитьПодобие( $\alpha(V^i), \alpha(V^k)$ ))
14:           {
15:             isFailed  $\leftarrow$  true;
16:             break;

```

```
17: }
18:  if (!isFailed) return Uk
19: }
20: }
21: return ∅
```

Программная реализация прототипа

Рассмотренные выше методы реализованы в прототипе CBR-системы, используемой для анализа нетиповых (нештатных) ситуаций на примере ИС для управления крупными парковочными комплексами (ИС УП), построенной по модульному принципу с учетом требований расширяемости и простоты интеграции [Куриленко, 2009].

ИС УП связана с большим количеством датчиков и исполнительных устройств для обеспечения жизненного цикла парковки и сведение к минимуму функций обслуживающего персонала. В задачи ИС УП входит управление движением на территории комплекса, учет владельцев транспортных средств, предотвращение угона и актов вандализма. Возможность анализа последовательностей действий, наблюдаемых системой, позволяет реализовать в ней более надежные механизмы контроля функционирования парковочного комплекса. Рассмотрены ситуации, в которых учет временных характеристик и расположения наблюдаемых событий позволяют реализовать реакцию системы на ситуации, не подлежащие контролю обычных систем управления движением, реализованных аппаратно (без применения программного обеспечения).

Благодаря хранению истории проведения операций на точках проезда и точках доступа в СВР возможна организация модуля поиска аномальных ситуаций и автоматической классификации сбоев. При построении этого модуля использовано то, что операции, протекающие на точках доступа и проезда в штатном режиме, формируют во времени достаточно стандартные последовательности событий. При этом процесс работы точки въезда/выезда может быть представлен в виде последовательности однотипных операций, каждая из которых также есть совокупность наблюдаемых стандартных событий. Таким образом, аномальные ситуации могут быть выделены за счет анализа этой последовательности путем проверки эквивалентности наблюдаемых в процессе конкретной операции событий и эталонных моделей штатного или нештатного развития ситуации.

На базе предлагаемых моделей, методов и алгоритмов реализован и апробирован прототип блока анализа нештатных ситуаций для промышленной системы управления крупными парковочными комплексами, которая разрабатывается и поставляется компанией ООО «ААМ Автоматик». Также разработанный прототип использован в учебно-научном процессе кафедры Прикладной математики НИУ МЭИ.

Программная реализация выполнена на базе платформы Microsoft .NET Framework на языке C# в среде разработки Microsoft Visual Studio 2010 и является кросс-платформенной. Прототип реализует следующие функции: поиск и выделение сбоев в работе парковочного комплекса, которые относятся к типовым проблемным ситуациям, а также определение нештатных ситуаций, которые возникли в процессе эксплуатации ИС УП, но не были учтены при разработке управляющих правил. Прецеденты, включая темпоральные, хранящиеся в БП, представляют собой операции (ситуации). В общем случае каждый прецедент имеет имя, код (номер), описание (диагноз), последовательность действий (что необходимо предпринять при выявлении этого прецедента), а также структуру событий.

В блоке сопоставления для выявления типовых и нестандартных ситуаций применяются алгоритмы сопоставления. Алгоритм сопоставления ситуаций опирается на их представление. В зависимости от представления прецедентов строится та или иная модель прецедента. Решатель на основе прецедентов соотносит временные диаграммы в последовательностях операций за определенный период с моделями, хранящимися в этих базах. В случае, если выявляется типовая нестандартная ситуация, ее данные заносятся в базу нестандартных ситуаций. Если же выделяется ситуация, которая не описана ни в базе штатных типовых ситуаций, ни в базе нестандартных типовых ситуаций, то эта ситуация заносится в отдельную базу, содержимое которой анализируется разработчиками (экспертами) ИС УП.

Заключение

В работе рассмотрена проблема моделирования правдоподобных рассуждений (рассуждений «здравого смысла») на основе темпоральных прецедентов. Предложенные методы моделирования временных зависимостей в CBR-системах на основе темпоральных прецедентов и соответствующие алгоритмы могут применяться для расширения возможностей и повышения эффективности современных ИСППР реального времени, позволяя учитывать в процессе поиска решения (процессе рассуждения) временные зависимости и динамику развития ситуации на объекте. Подход на основе использования темпоральных прецедентов позволяет упростить процесс принятия решений в нестандартных (аномальных) ситуациях, которые могут возникнуть по причине различного рода «НЕ-факторов» (неполноты, противоречивости, неопределенности и др.) в исходных данных и экспертных знаниях, а также перспективен для создания систем мониторинга технических систем и объектов, функционирующих по алгоритмам, описываемым автоматными моделями. К таким системам относится широкий класс систем автоматизации, систем контроля доступа, систем управления движением и транспортом.

Использование темпоральных прецедентов в CBR-системах, ориентированных на использование в составе ИСППР реального времени для мониторинга и управления различными сложными техническими объектами и процессами позволяет ЛПП своевременно выбрать необходимую стратегию поведения для устранения проблемной ситуации за счет сокращения поискового пространства и уменьшения времени реакции системы.

Благодарности

Работа поддержана РФФИ (проекты 11-01-00140, 12-07-31239, 12-07-00508).

Список литературы

- [Eremeev et al., 2009] Alexander Eremeev, Ivan Kurilenko, Pavel Varshavskiy. Application of Temporal Reasoning and Case-based Reasoning in Intelligent Decision Support Systems // International Book Series «Information science & computing», Number 10, Supplement to IJ «Information technologies & knowledge» Volume 3/2009 – 2009. – Pp. 9–16.
- [Eremeev et al., 2012] Alexander Eremeev, Ivan Kurilenko, Pavel Varshavskiy. Моделирование временных зависимостей в интеллектуальных системах поддержки принятия решений на основе прецедентов // Information technologies and knowledge Vol. 6, 2012, № 3, с. 279-294.
- [Schank, Abelson, 1977] Schank R.C., Abelson R.P. Scripts, Plans, Goals and Understanding. Erlbaum, Hillsdale, New Jersey, US, 1977, - 248 p.
- [Watson et al., 1994] Watson I.D., Marir F. Case-based reasoning: A review. The Knowledge Engineering Review, Vol. 9, No. 4, 1994. – Pp. 355-381.

- [Вагин и др., 2008] Вагин В.Н., Головина Е.Ю., Загорянская А.А., Фомина М.В. Достоверный и правдоподобный вывод в интеллектуальных системах. 2-е изд. // Под редакцией В.Н. Вагина, Д.А. Поспелова. – М.: ФИЗМАТЛИТ, 2008.
- [Варшавский и др., 2006] Варшавский П.Р., Еремеев А.П. Методы правдоподобных рассуждений на основе аналогий и прецедентов для интеллектуальных систем поддержки принятия решений // Новости искусственного интеллекта. 2006. №3. – С. 39-62.
- [Варшавский и др., 2009] Варшавский П.Р., Еремеев А.П. Моделирование рассуждений на основе прецедентов в интеллектуальных системах поддержки принятия решений // Искусственный интеллект и принятие решений, №2, 2009, с. 45-57.
- [Варшавский и др., 2012] Варшавский П.Р., Куриленко И.Е., Макашова М.Б. Реализация нейромодуля для поиска решения на основе прецедентов в интеллектуальных системах поддержки принятия решений // Труды XIII национальной конференции по искусственному интеллекту с международным участием КИИ-2012. Т.3. – Белгород: Изд-во БГТУ, 2012. - С.80-87.
- [Еремеев и др., 2007] Еремеев А.П., Куриленко И.Е. Реализация механизма временных рассуждений в современных интеллектуальных системах // Известия РАН. Теория и системы управления, 2007, № 2, с. 120–136.
- [Еремеев и др., 2010] Еремеев А.П., Куриленко И.Е. Применение темпоральных моделей в интеллектуальных системах / Интеллектуальные системы. Коллективная монография. Выпуск четвертый. / Под. Ред. В.М. Курейчика. – М.: Физматлит, 2010, 300 с., 222-252.
- [Еремеев и др., 2011] Еремеев А.П., Куриленко И.Е., Смирнова А.Е. Разработка темпорального расширения методов рассуждений на основе прецедентов // Труды конгресса по интеллектуальным системам и информационным технологиям (IS&IT'11). Научное издание в 4-х томах. - Т.1. – М.: Физматлит, 2011. - С.50-59.
- [Еремеев и др., 2013] Еремеев А.П., Куриленко И.Е. Расширение возможностей моделирования временных зависимостей в интеллектуальных системах на основе применения темпоральных прецедентов / Интеллектуальные системы. Коллективная монография. Выпуск 6. / Под. Ред. В.М. Курейчика. – М.: Физматлит, 2013, 300 с., 89-118.
- [Куриленко, 2009] Куриленко И.Е. Применение временной логики при построении интеллектуальной системы управления крупными парковочными комплексами // Сб. док. научно-практ. конф. «Интегрированные модели и мягкие вычисления в искусственном интеллекте» 2009 в 2 т. – Т.2 – М.:ФизМатЛит, 2009. С. 171 -180.
- [Куриленко, 2012] Куриленко И.Е. О применении метрической временной логики при построении механизма вывода на основе прецедентов // Труды XIII национальной конференции по искусственному интеллекту с международным участием КИИ-2012. Т.1. – Белгород: Изд-во БГТУ, 2012. - С.25-33.
- [Смирнова и др., 2012] Куриленко И.Е., Смирнова А.В. Реализация прототипа системы анализа типовых ситуаций для интеллектуальной системы // Труды XX международной научно-технической конференции Информационные средства и технологии – М.:Издательский дом МЭИ, 2012. - С.152-158.

Сведения об авторах

Еремеев Александр Павлович – д.т.н., проф., заведующий кафедрой Прикладной математики Московского энергетического института, 111250, Россия, Москва, Красноказарменная ул., 14; e-mail: eremeev@appmat.ru

Область научных интересов: искусственный интеллект, принятие решений, экспертные системы

Куриленко Иван Евгеньевич – к.т.н., доцент кафедры Прикладной математики Московского энергетического института, 111250, Россия, Москва, Красноказарменная ул., 14; e-mail: ivan@appmat.ru

Область научных интересов: искусственный интеллект, принятие решений, темпоральные логики, методы правдоподобных рассуждений

ТЕМПОРАЛЬНЫЕ СЕТИ ПЕТРИ И ИХ ПРИМЕНЕНИЕ В ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМАХ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ РЕАЛЬНОГО ВРЕМЕНИ

Александр Еремеев, Юрий Королев

Аннотация: Работа посвящена методам моделирования сложных процессов в динамических системах на основе модифицированных (расширенных) сетей Петри (СП), позволяющих учитывать временные зависимости и неопределенность в имеющейся информации. Анализируются различные расширения и модификации СП и предлагается расширение класса цветных СП – цветные СП с поддержкой темпоральной логики Аллена - в плане использования данного аппарата в интеллектуальных системах поддержки принятия решений (ИСППР), включая перспективные ИСППР реального времени (ИСППР РВ).

Ключевые слова: моделирование, сети Петри, цветные сети Петри, представление временных зависимостей, принятие решений, реальное время, интеллектуальные системы поддержки принятия решений.

ACM Classification Keywords: H.4.2 [Information systems applications]: Types of systems – Decision support; I.2.3 [Artificial intelligence]: Deduction and Theorem Proving – Uncertainty, "fuzzy," and probabilistic reasoning; I.2.4 [Artificial intelligence]: Knowledge Representation Formalisms and Methods – Temporal logic.

Введение

На кафедре прикладной математики Национального исследовательского университета «МЭИ» более двадцати лет активно проводятся исследования по разработке математического и программного обеспечения интеллектуальных систем поддержки принятия решений (ИСППР), в том числе наиболее сложных ИСППР реального времени [Вагин, 1988][Еремеев и др., 1994].

Большинство имеющихся объектов управления относятся к слабо структурированным или плохо формализуемым объектам, которые обладают рядом нетипичных для традиционного управления свойств, таких как: уникальность, отсутствие формализуемой цели существования и количественного критерия оптимальности, высокая динамичность, неполнота описания объекта и, наконец, индивидуальность и субъективность поведения лица, принимающего решения (ЛПР).

В последние годы технология решения информационных задач значительно изменилась. Основное назначение информационных систем – это своевременное представление необходимой информации ЛПР для принятия им адекватных и эффективных решений при управлении процессами, ресурсами, финансовыми транзакциями, персоналом или организацией в целом. Однако, в процессе развития информационных технологий, исследования операций и технологий моделирования, а также с возрастанием потребителей информационно-аналитической поддержки появляется потребность в системах, не только представляющих информацию, но и выполняющих некоторый ее предварительный анализ, способных давать некоторые советы и рекомендации, осуществлять прогнозирование развития ситуаций, отбирать наиболее перспективные альтернативы решений, т.е. поддерживать решения ЛПР, взяв на себя значительную часть рутинных операций, а также функции предварительного анализа и

оценок. Новая информационная технология основывается на интеллектуальных технологиях и теории искусственного интеллекта.

Возможны различные способы интеграции интеллектуальных систем, основанных на знаниях, с математическим моделированием. Разработка ИСППР с применением математических моделей является сегодня актуальным и перспективным направлением исследований.

Темпоральные модификации сетей Петри

Математическое моделирование предполагает эксперименты с математическими моделями систем. В отличие от физической модели, которая материальна, математическая модель является логическим объектом. Подобный подход позволяет разрабатывать системы, опираясь на уже разработанные теоретические выкладки, что существенно облегчает процесс. В зависимости от дальнейшего использования модели можно отнести к одному из следующих типов:

- *функциональные модели*, отображающие закономерности функционирования объектов и используемые при проектировании систем для исследования принципов их работы и характера протекающих процессов;
- *структурные модели*, отражающие структурные характеристики и используемые при конструировании систем в ситуациях, когда важными являются связи отдельных частей объектов между собой.

Очевидно, что гибридные модели, сочетающие особенности обоих перечисленных, могут обладать существенно большей выразительностью, что необходимо учитывать при выборе инструментов моделирования. Другим принципиально важным критерием являются характерные свойства моделируемой системы.

ИСППР РВ в общем случае является динамической нестационарной нелинейной системой. При выборе математической модели для моделирования таких систем необходимо учесть эти свойства. Также важно отразить не только закономерности функционирования, но и структурные характеристики системы. Очевидно, что для выявления конструктивных (структурных) закономерностей системы необходимо добиться максимальной визуальной выразительности. Поэтому для дальнейших исследований в качестве базовой модели был выбран один из графоориентированных инструментов моделирования: *сети Петри (СП)*.

СП – признанный инструмент моделирования и функционального анализа параллельных и распределенных вычислительных систем и процессов. СП позволяют естественно описывать синхронизацию, параллелизм, конфликт и причинную зависимость, а также наглядно представлять структуру и поведение систем, что делает их отличным потенциальным инструментом для разработки ИСППР. Основы теории обычных СП изложены, в частности, в работе [Котов, 1984]. Можно вводить ряд дополнительных правил и условий в алгоритмы моделирования, получая ту или иную разновидность СП. В частности, теория *цветных сетей Петри (ЦСП)* разрабатывается более 20 лет рабочей группой под руководством профессора Курта Йенсена [Jensen, 1992-1997]. ЦСП – это графоориентированный язык для проектирования, описания, имитации и контроля распределенных и параллельных систем. Графическими примитивами показывается течение процесса, а конструкциями специального языка имитируется необходимая обработка данных. Показано [Jensen, 1992-1997], что для каждой ЦСП можно построить обычную СП, и наоборот. Таким образом, использование ЦСП не дает теоретических преимуществ, но на практике такие сети представляют собой более компактный и удобный язык моделирования, чем обычные СП.

Существенным недостатком как классических СП, так и ЦСП является отсутствие учета фактора времени. Это не позволяет эффективно моделировать те реальные процессы, в которых от текущего времени зависит состояние анализируемой системы. О важности наличия средств представления времени и темпоральных зависимостей в ИС говорится практически с момента появления таких систем. Особенно актуальна эта проблема встала именно в связи с появлением и развитием динамических ИС, типичным представителем которых являются ИСППР РВ [Куриленко, 2010], предназначенными для помощи ЛПР при управлении сложными объектами и процессами в условиях жестких временных ограничений и наличия различного рода неопределенности как в поступающей извне информации, так и в заложенных в систему знаниях. Важнейшей задачей при создании перспективных ИСППР РВ является задача представления и оперирования временными зависимостями.

Цветные сети Петри реального времени (ЦСП РВ) [Szpyrka et al., 2006] являются функциональным подклассом ЦСП, ориентированным на моделирование и анализ систем реального времени. По сравнению с ЦСП, в ЦСП РВ используется другая модель времени и приоритеты переходов, на них наложены некоторые структурные ограничения. Эти характеристики ЦСП РВ позволяют разработчикам производить прямое моделирование элементов, типичных для параллельного программирования, таких как приоритеты задач, таймауты и т.д.

Временные зависимости подразделяются на [Троицкий и др., 2003]:

- *количественные* (метрические) – когда для представления времени используются количественные меры на временной оси (например: «сбой произошел в 17.00»);
- *качественные* – когда используется только относительное положение во времени событий или действий (например: «сначала произошел сбой в системе питания, затем аварийное отключение»).

Аппарат ЦСП РВ предполагает работу с количественными временными зависимостями, однако ясно, что выразительность представления увеличится при наличии средств, позволяющих представление и оперирование также и качественными временными зависимостями. В числе одного из возможных путей решения этой проблемы предлагается модифицировать ЦСП РВ, добавив возможность использования аппарата временных логик, а именно, *темпоральную интервальную логику Аллена* [Allen, 1983], которая характеризуется достаточной выразительностью и наличием полиномиальных алгоритмов вывода, что позволяет использовать ее в интеллектуальных системах типа ИСППР РВ [Королев и др., 2011]. В качестве временных примитивов в ней используются интервалы, что важно при моделировании сложных систем. Точечные (оперирующие моментами) временные логики недостаточно чувствительны к смысловым различиям между глагольными предикатами.

Формальное определение ЦСП РВ с поддержкой логики Аллена приведено в [Королев и др., 2012]. Чтобы проиллюстрировать главные аспекты, приведем в качестве примера модель автоматической остановки поезда [Szpyrka et al., 2006]. В кабине машиниста каждые 60 секунд загорается световой сигнал, чтобы проверить, контролирует ли он идущий поезд. Если машинист проигнорирует световой сигнал, то через 6 секунд включается звуковой сигнал. Затем, если машинист не деактивирует его в течение 3 секунд, срабатывает механизм аварийного торможения. Модель ЦСП РВ с поддержкой логики Аллена для данного примера, представленная на рис 1., содержит шесть мест:

- LightSig и SoundSig (контроль состояния индикаторов),
- Brake (механизм торможения),
- Driver (машинист поезда),
- Timer1 и Timer2 (таймеры);

и шесть переходов:

- TurnOnLS (включение светового сигнала),
- TurnOnSS (включение звукового сигнала),
- TurnOnBr (запуск механизма торможения),
- DisactLS и DisactSS (дезактивация машинистом соответствующих сигналов),
- Activity (моделирование действий машиниста).

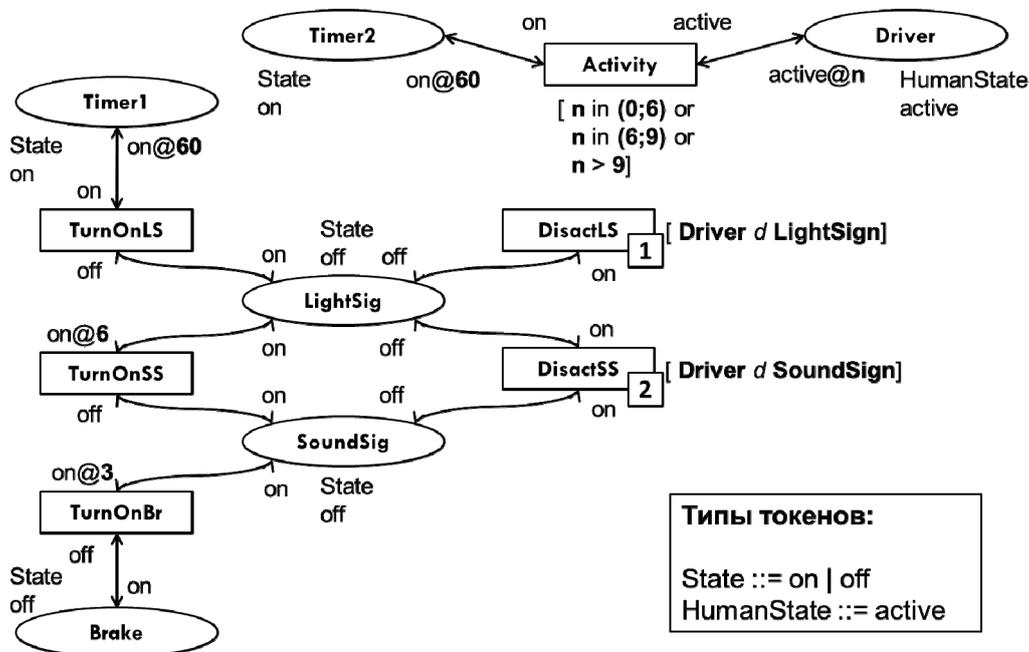


Рис. 1. Модель системы остановки поезда (ЦСП РВ с поддержкой логики Аллена)

Формулы логики Аллена применены в данном случае как защитные функции переходов DisactLS и DisactSS, обозначающих своевременную реакцию машиниста (ЛПР) на световой и звуковой сигнал соответственно:

- **Driver d LightSig** – переход DisactLS работает, если токен в месте Driver появится в тот момент, когда в месте LightSig будет находиться токен on;
- **Driver d SoundSig** – переход DisactSS работает, если токен в месте Driver появится в тот момент, когда в месте SoundSig будет находиться токен on.

Главным преимуществом в данном случае является возможность задавать не конкретное время реакции, как в случае ЦСП РВ, а интервалы, на которых ЛПР может дезактивировать систему и каждый из которых определяет дальнейшее поведение модели.

Таким образом, включение в модель аппарата интервальных темпоральных логик позволило корректно отразить неопределенность, присущую задаче, что является одним из основополагающих принципов при проектировании перспективных ИСППР РВ.

Разработка базового инструментария моделирования

Средства компьютерного моделирования в составе ИСППР РВ актуально как для теоретических исследований, так и для практического применения. Создание инструментария для решения подобной задачи представляет серьезную проблему, прежде всего потому, что среда разработки должна поддерживать концепцию реального времени, а также из-за существенной универсальности разрабатываемых моделей. Однако некоторые системы, в частности, инструментальный комплекс

конструирования систем реального времени G2 (Gensym Corp., США) позволяют реализовывать подобные проекты.

G2 – это объектно-ориентированная интегрированная среда для разработки и сопровождения приложений – интеллектуальных систем реального времени, использующих базы знаний. В отличие от систем, ориентированных на какую-то одну методологию или на конкретную предметную область, G2 интегрирует в себе множество взаимодополняющих методов искусственного интеллекта, что упрощает и ускоряет процесс разработки приложений и позволяет делать их достаточно универсальными [Чибизова и др., 1998]. Эти особенности позволили разработать в данной среде эффективный инструментарий для моделирования систем на основе ЦСП РВ с поддержкой логики Аллена. Модель системы остановки поезда, созданная средствами данного инструментария, приведена на рис. 2.

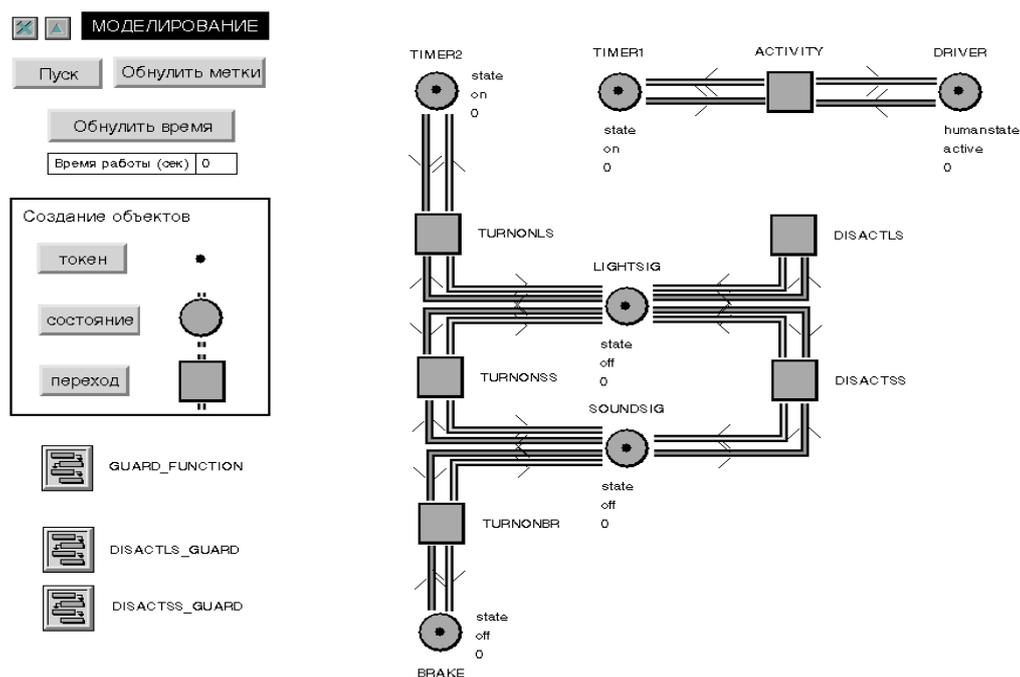


Рис. 2. Модель системы остановки поезда (среда G2)

Проведем моделирование при 9-тисекундной задержке реакции машиниста. В начальный момент одновременно параллельно срабатывают 2 независимых перехода TurnOnLS и Activity, изменив состояние сети на новое (рис. 3).

После этого в течение 6 секунд ни один переход не становится допустимым. По истечении этого времени срабатывает переход TurnOnSS, меняя цвет токена в месте SoundSig (рис. 4).

Через 3 секунды после этого становятся допустимыми 2 перехода: TurnOnBr и DisactSS. Но поскольку изначально приоритет последнего был установлен более высоким, сработает переход деактивации. При этом поменяются токены в местах LightSig и SoundSig, что не позволит сработать переходу TurnOnBr (рис. 5).

Далее в течение 51 секунды система останется стабильной, единственным изменением будет непрерывное уменьшение временных меток с установленным шагом в 1 секунду. После чего модель снова вернется к исходному состоянию, и сработают переходы при таймерах.

Изменяя значение временной метки дуги Activity-Driver, можно исследовать другие возможные варианты работы системы, таким образом, всесторонне изучив данную модель. Модель на основе ЦСП РВ с поддержкой логики Аллена системы остановки поезда отвечает требованиям корректности.

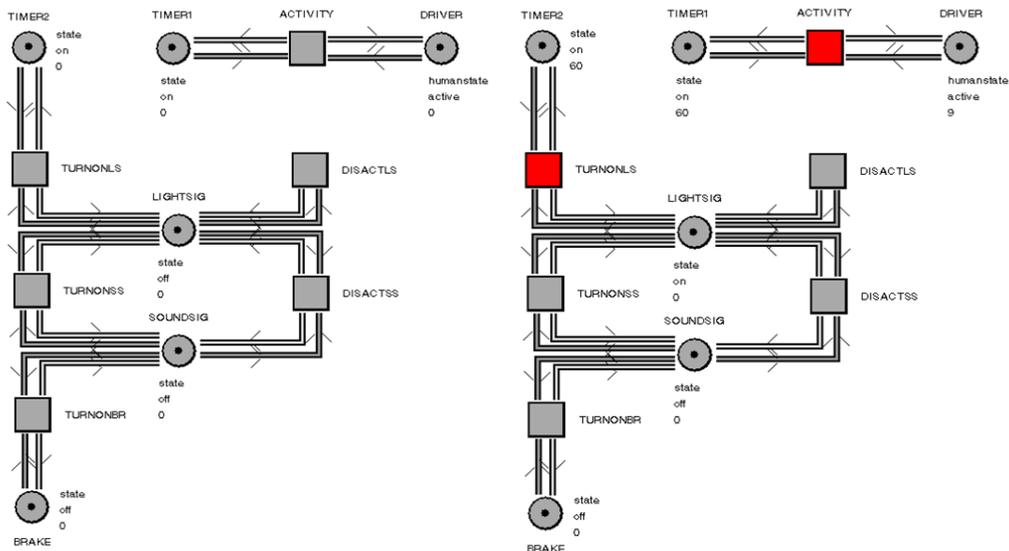


Рис. 3. Срабатывание переходов TurnOnLS и Activity

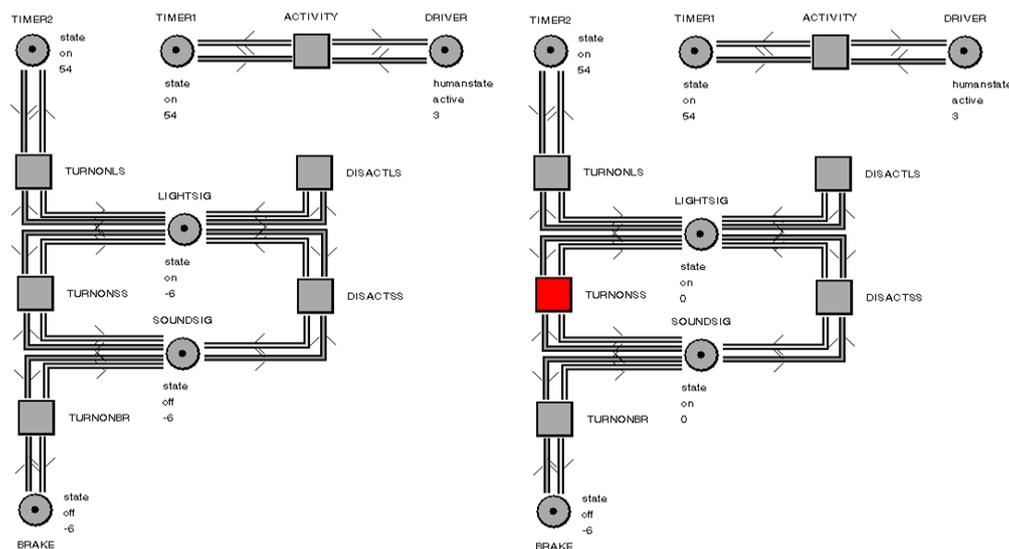


Рис. 4. Срабатывание перехода TurnOnSS

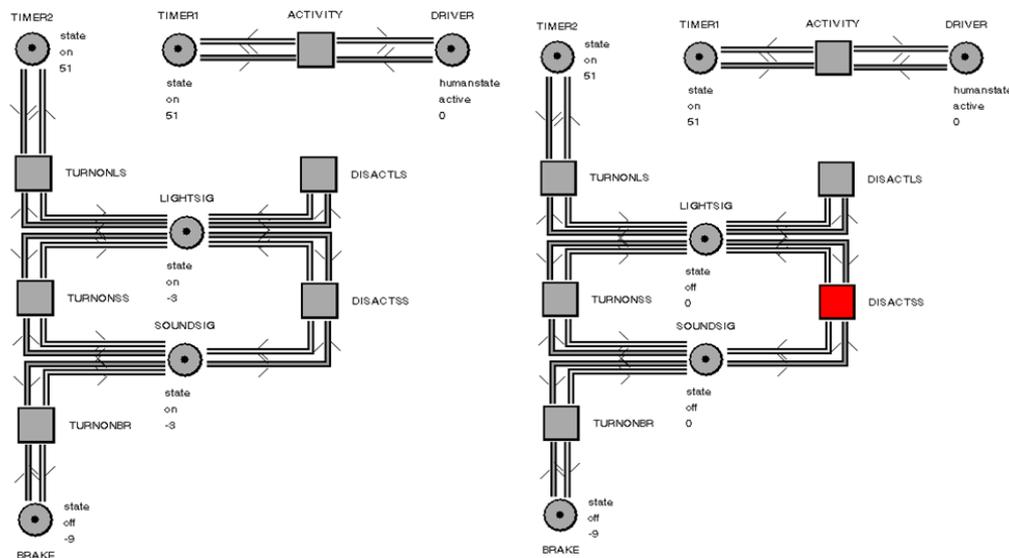


Рис. 5. Срабатывание перехода DisactSS

Методология создания моделей на основе цветных сетей Петри реального времени

Для повышения эффективности моделирования предлагается разбиение исходной ЦСП РВ на несколько подсетей, которые называют страницами. Особая форма иерархической ЦСП РВ, называемая канонической формой, была определена для ускорения и облегчения разработки моделей. ЦСП РВ в канонической форме состоят из четырех типов подсетей с жестко определенной структурой (рис. 6):

- исходные страницы мест,
- исходные страницы переходов,
- страницы связи,
- D-сети.

Такая модель описывает структуру соответствующей системы, а также ее поведение и функциональные аспекты (рис. 7). Исходные страницы мест (рис. 7, а) используются для представления активных объектов (т.е. объектов, осуществляющих деятельность) и их деятельности. Они ориентированы на представление объектов и находятся на высшем уровне иерархии страниц. Исходные страницы переходов (рис. 7, б) ориентированы на представление деятельности и расположены на втором уровне иерархии. Страницы связи (рис. 7, с) принадлежат функциональному уровню модели и используются, чтобы при необходимости представлять алгоритм, описывающий деятельность в деталях. Кроме того, страницы связи используются в качестве интерфейса для вставки соответствующих D-сетей в модель. D-сети (рис. 7, d) используются для представления систем, основанных на продукционных правилах, в форме сетей Петри. D-сеть содержит два места: условия и результата. Каждое правило вывода представлено переходом и его входной и выходной дугами. Токен, размещенный в месте условия, определяет последовательность условий правил. Аналогично, токен места результата определяет последовательность атрибутов результата. D-сети принадлежат к нижнему уровню модели.

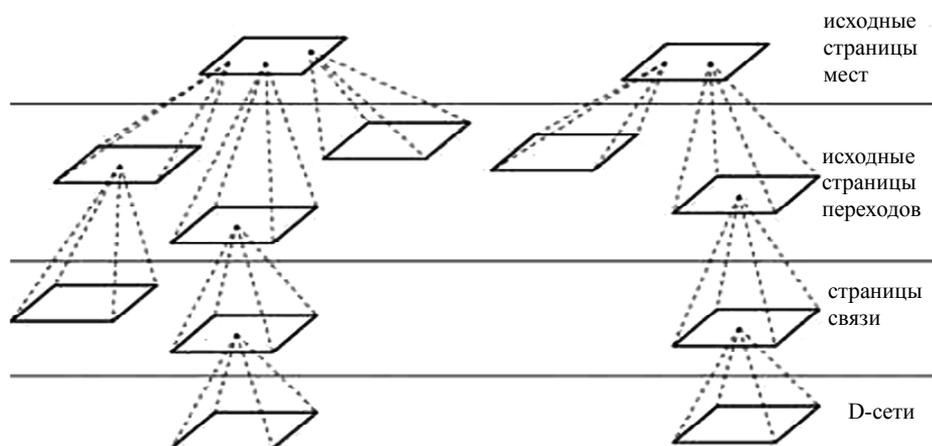


Рис. 6. Общая структура ЦСП РВ с поддержкой логики Аллена в канонической форме

Связи между отдельными страницами модели задаются на странице графа иерархии. Вершина такого графа представляет одну страницу, а дуга – связь с подстраницей с учетом заменяемого перехода.

Если рассматриваются временные свойства, то можно сосредоточиться на исходных страницах переходов и пропустить их подстраницы. Любая деятельность страниц связи начинается срабатыванием входного перехода и заканчивается срабатыванием выходного. Кроме того, каждое срабатывание входного перехода должно сопровождаться такой последовательностью срабатываний переходов, где

последним является выходной, а все промежуточные – вспомогательными. Такая модель деятельности подобна вызову процедуры в языках программирования.

Декомпозиция системы – первый шаг при разработке модели. Она начинается с определения типов объектов, которые входят в систему. Объекты делятся на активные, т. е. осуществляющие деятельность, и пассивные, которые не выполняют каких-либо собственных действий. Объект представляется основным местом. Для каждого объекта определяются список атрибутов и их типы. Декартово произведение типов задает тип соответствующего места. Создание исходных страниц мест завершает разработку на этой стадии.

Следующий этап разработки связан с описанием модели действий, что особенно важно для систем реакции. Переходы, размещенные в исходных страницах мест, как правило, являются заменяемыми переходами. Для каждого из них создается исходная страница перехода. Проектирование исходной страницы перехода подобно объявлению процедуры в языке программирования. Необходимо описать входные и выходные параметры. Если переход на исходной странице перехода не является заменяемым, то эта страница представляет собой полное определение деятельности, описываемой этим переходом. После завершения этого этапа, ЦСП РВ представляет все элементы, которые составляют моделируемую систему, и все ее действия.

Последний этап связан с разработкой функционального уровня системы. Для этого используются страницы связи и D-сети.

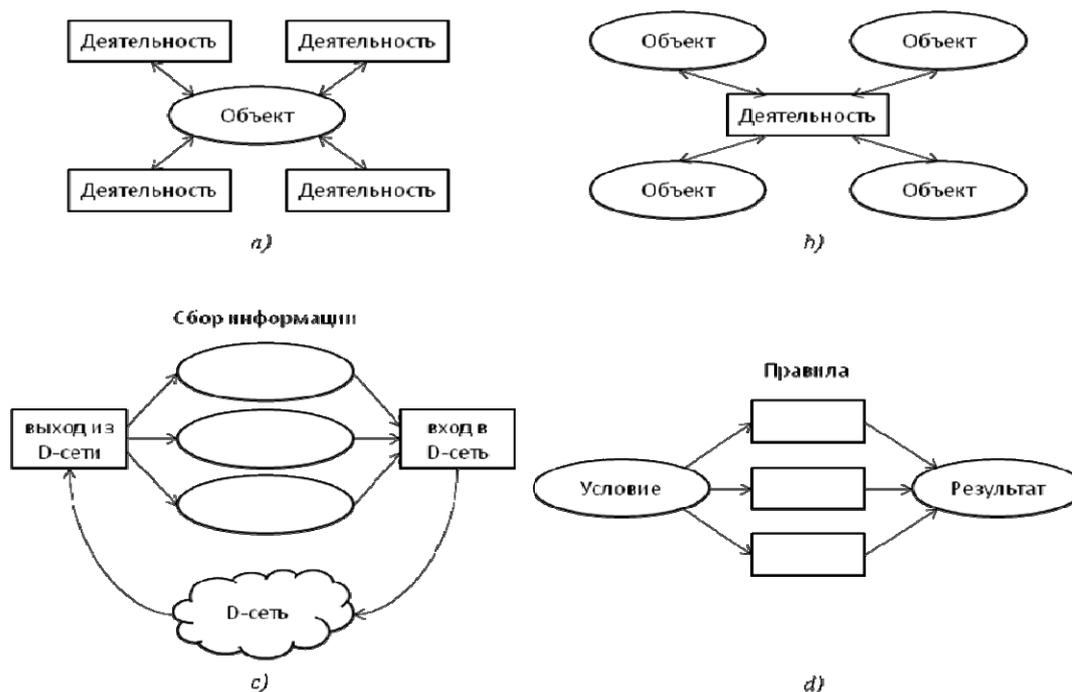


Рис. 7. Типы подсетей: а) исходные страницы мест; б) исходные страницы переходов; в) страницы связи; д) D-сети.

Заключение

В представленной работе исследованы временные подклассы сетей Петри в применении к моделированию интеллектуальных систем поддержки и принятия решений, в том числе реального времени. Рассмотрен аппарат цветных сетей Петри реального времени, приведен пример анализа модели, построенной на его основе. Предлагается новое расширение подкласса – цветные сети Петри с

поддержкой темпоральной логики Аллена, обуславливается его функциональность как аппарата, позволяющего моделировать как количественные, так и качественные временные задержки. Исследуется возможность моделирования сетей Петри с временными ограничениями в среде конструирования интеллектуальных систем реального времени G2. На основе результатов анализа разработаны базовые программные средства для моделирования с использованием сетей Петри, дано их описание и пример использования.

Благодарности

Работа выполнена при финансовой поддержке РФФИ, проекты №№ 11-01-00140, 11-07-00038.

Список литературы

- [Allen, 1983] Allen J.F. Maintaining knowledge about temporal intervals. – Communications of the ACM, 1983. – v. 26. – n. 11. – pp. 832-843. – ISBN:1-55860-095-7.
- [Jensen, 1992-1997] Jensen K. Coloured Petri Nets. Basic Concepts, Analysis Methods and Practical Use. – 1992-1997. – vol. 1-3. – ISBN: 3-540-62867-3.
- [Szyrka et al., 2006] Szyrka M., Szmuc T. Integrated approach to modelling and analysis using RTCPnets. – IFIP International Federation for Information Processing. – New York, 2006. – vol. 227, pp. 115-120. – ISBN: 978-0-387-39387-2.
- [Вагин, 1988] Вагин В.Н. Дедукция и обобщение в системах принятия решений. – М.: Наука, 1988. – 384 с. – ISBN 5-02-013887-8.
- [Еремеев и др., 1994] Башлыков А.А., Еремеев А.П. Экспертные системы поддержки принятия решений в энергетике. – М.: Издательство МЭИ, 1994. – 216 с. – ISBN 5-7046-0044-1
- [Королев и др., 2011] Еремеев А.П., Королев Ю.И. Применение модифицированных сетей Петри при создании интеллектуальных систем поддержки принятия решений. Интегрированные модели и мягкие вычисления в искусственном интеллекте. Сб. научных трудов VI-й Международной научно-практической конференции (Коломна, 16-19 мая 2011 г.). В 2-х томах. Т2. – М.: Физматлит, 2011. –с. 744-754. - ISBN 978-5-9221-1328-1.
- [Королев и др., 2012] Еремеев А.П., Королев Ю.И. Средства моделирования на основе темпоральных сетей Петри для интеллектуальных систем поддержки принятия решений. Тринадцатая национальная конференция по искусственному интеллекту с международным участием КИИ-2012 (16-20 октября 2012 г., г. Белгород, Россия): Труды конференции. Т.3. – Белгород: Изд-во БГТУ, 2012. – Стр. 105-112 – ISBN 978-5-361-00183-5.
- [Котов, 1984] Котов В.Е. Сети Петри – М.: Наука, 1984. – 160 с.
- [Куриленко, 2010] Еремеев А.П., Куриленко И.Е. Средства темпорального вывода для интеллектуальных систем реального времени. – Интеллектуальные системы. Колл. моногр. – Вып. 4. – М.: Физматлит, 2010. – с. 222-252.
- [Троицкий и др., 2003] Еремеев А.П., Троицкий В.В. Модели представления временных зависимостей в интеллектуальных системах поддержки принятия решений. – Изв. РАН. Теория и системы управления, 2003. – № 5. – с. 75-88.
- [Чибизова и др., 1998] Еремеев А.П., Чибизова Н.В. Проектирование экспертных систем реального времени на основе инструментального комплекса G2. Методологические указания – М.: Издательство МЭИ – 1998.

Сведения об авторах

Еремеев Александр Павлович – д.т.н., проф., заведующий кафедрой Прикладной математики Московского энергетического института, 111250, Россия, Москва, Красноказарменная ул., 14; e-mail: eremeev@appmat.ru. Область научных интересов: искусственный интеллект, принятие решений, экспертные системы

Королев Юрий Ильич – м.н.с., аспирант кафедры Прикладной математики Московского энергетического института, 111250, Россия, Москва, Красноказарменная ул., 14; e-mail: korolevyu@gmail.com. Область научных интересов: искусственный интеллект, принятие решений, моделирование, сети Петри

МЕТОД ОПИСАНИЯ ИЗМЕНЕНИЯ КОЛИЧЕСТВА ПОЛЬЗОВАТЕЛЕЙ ТЕЛЕКОММУНИКАЦИОННОЙ СЕТИ

Галина Гайворонская, Олег Домаскин

Аннотация: *предложен подход к разработке математической модели, адекватно описывающей изменение требований пользователей на подключение к сети и обслуживание этих требований исследуемой сетью. Решение задачи включает исследование модели требований на развитие сети за счет увеличения ее емкости, объема и видов предоставляемых инфокоммуникационных услуг.*

Ключевые слова: *телекоммуникационная сеть, математическая модель, требования к сети, развитие телекоммуникационной сети.*

Ключевые слова классификации АСМ: *C.2. Computer-communication networks, H. Information Systems - H.1 Models and Principles, K. Computing Milieux - K.6 Management of computing and information system*

Введение

Оптимизация телекоммуникационных сетей (ТС) включает разнообразные задачи: от математических и технических, до экономических и общественно-политических. Значительные усилия исследователей направлены на изучение способов использования средств передачи и коммутации, чтобы найти приемлемый компромисс между необходимыми характеристиками сети и затратами на ее сооружение. Опубликованные работы, посвященные оптимизации ТС [Рогинский, 1981 - Boorstyn, 1977], ограничиваются сетями, режим работы которых характеризуется стационарностью в следующем смысле: техническая база сети и номинальные эксплуатационные параметры ее элементов остаются неизменными; неизменны технология обслуживания и алгоритмы действия сети; случайные внешние воздействия на элементы сети, вызывающие их отказы, стационарны в вероятностном смысле; потребность в связи неизменна; сеть как стохастический объект обладает эргодичностью и рассматривается в состоянии статистического равновесия. Переходные и аномальные режимы работы, когда ТС подвергается модернизации и существенно изменяется общая потребность в связи или интенсивность воздействия внешней среды на ее элементы, в известных публикациях не рассматривались, хотя сети, для которых такие режимы работы являются нормальными, представляют большой интерес [Давыдов, 1984].

Для определения закономерностей, описывающих процессы развития ТС, в конкретных условиях используются различные математические модели [Гайворонская, 2005]. Проектирование ТС с учетом динамики ее развития требует решения задач синтеза сети для некоторого заданного отрезка времени и заключается в том, чтобы указать вариант или несколько вариантов будущей сети, удовлетворяющих прогнозируемым к этому времени требованиям с заданной эффективностью. Эти задачи должны определять процесс развития сети во времени с учетом ее исходного состояния, для того, чтобы обеспечить оптимальный переход от существующей сети к сети в конце исследуемого периода через некоторые фиксированные промежуточные состояния. Теоретическая сложность решения таких задач обусловлена структурной сложностью сети и ее отдельных частей, большим количеством и разнообразием исходных данных, необходимостью их прогнозирования и другими факторами. Особая сложность задач характерна для исследования динамики развития сети от заданного исходного состояния [Рогинский, 1977]. В [Гайворонская, 2005] сформулирована общая проблема синтеза развивающихся информационных сетей, а в [Gayvoronska, 2011, Гайворонская, 2006] предложены некоторые подходы к ее решению. Одним из важных аспектов этой проблемы является формализация процесса изменения количества пользователей, подключенных к ТС, во времени. Решение этой задачи практически не встречается в работах, посвященных исследованию ТС, поскольку в ней просто не было необходимости.

В условиях, когда спрос на подключение к сети практически всегда опережал предложение, изменений емкости ТС описывалось только линейно зависимостью. В современных условиях возможны различные ситуации, при которых емкость ТС может не только увеличиваться, но и уменьшаться. Некоторые подходы к созданию моделей развития ТС рассмотрены в работах [Gayvoronska, 2012 - Гайворонская, 2013], где анализируется возможность использования математических функций для описания процессов изменения емкости ТС, и указаны условия, при которых целесообразно использование той или иной функции.

Постановка задачи

При проектировании сетей требование высокой точности анализа и строгой оптимальности синтеза присутствует далеко не всегда. Зачастую возникает необходимость в достаточно грубом оценочном анализе качества работы сети или предварительной оценке структуры будущей сети, предшествующей детальному проектированию. В представленной авторами работе [Гайворонская, 2013], где предлагаемая модель развития ТС основана на математических методах, описывающих распространение эпидемии, показано, что если процесс развития ТС трактовать как процесс чистого размножения, то для временной зависимости вероятности подключения конкретного числа пользователей к сети в конкретный момент времени получим достаточно громоздкие и неудобные для вычислений выражения. Поэтому в данном исследовании поставлена задача предложить асимптотические формулы, дающие простые выражения для вероятностей состояний сети, с точки зрения количества пользователей, обслуживаемых сетью, и моментов распределения функции, описывающей количество пользователей сети в момент времени.

Асимптотические формулы для вероятностей состояний сети

Пусть N – число всех потенциальных пользователей ТС на рассматриваемой территории, а $m(t)$ – число пользователей, подключенных к ТС к моменту t . Если рассмотреть в качестве случайной величины не $m(t)$, а момент τ_n подключения к сети n -го пользователя ($n = 2, \dots, N$), то можно получить простые асимптотические формулы. Между распределениями $m(t)$ и τ_n существует простая связь. Значение $m(t)$ не превосходит n тогда и только тогда, когда $(n+1)$ -й пользователь в момент t еще не подключен к сети, т. е. когда $\tau_{n+1} > t$. Следовательно, $P(m(t) \leq n) = P(\tau_{n+1} > t) = 1 - P(\tau_{n+1} \leq t)$ (1)

Обозначим через α_v промежуток $\tau_n - \tau_{n-1}$ ($n = 2, \dots, N$). Тогда $\tau_n = \sum_{v=2}^n \alpha_v$. (2)

α_v является случайной величиной с экспоненциальным законом распределения. Действительно,

$$P(\alpha_v \leq x) = 1 - e^{-\lambda \frac{v-1}{N(N-v+1)} x}. \quad (3)$$

Отсюда легко получить моменты для τ_n : $E(\tau_n) = \sum_{v=2}^n \mu_v = \frac{N}{\lambda} \sum_{k=1}^{n-1} \frac{1}{k(N-k)}$, (4)

$$\sigma^2 = \sum_{v=2}^n \sigma_v^2 = \frac{N^2}{\lambda^2} \sum_{k=1}^{n-1} \frac{1}{k^2(N-k)^2}. \quad (5)$$

Для $E(\tau_n)$ и $\sigma^2(\tau_n)$ можно получить наглядные асимптотические выражения.

Предположим, что $1 < n < N$. Используя равенство и обозначение, приведенные ниже,

$$\frac{1}{k(N-k)} = \frac{1}{N} \left(\frac{1}{k} + \frac{1}{N-k} \right), \quad (6)$$

$$h_n = \sum_{v=1}^n \frac{1}{v}, \quad (7)$$

а так же формулы (4) и (7) получим

$$E(\tau_n) = \frac{1}{\lambda} (h_{n-1} + h_{N-1} - h_{N-n}). \quad (8)$$

Учитывая, что величина h_n отличается от l_n на величину C_n , которая при $n \rightarrow \infty$ стремится к постоянной Эйлера C :

$$h_n = \ln n + C_n, \quad \lim_{n \rightarrow \infty} C_n = C = 0,5772... \quad (9)$$

(8) и (9) дают

$$E(\tau_n) = \frac{1}{\lambda} \ln \frac{(n-1)(N-1)}{(N-n)} + \frac{1}{\lambda} C_{n,N}, \quad (10)$$

где $\lim_{\substack{n \rightarrow \infty \\ N \rightarrow \infty, N-n \rightarrow \infty}} C_{n,N} = C$. Положим теперь $n | N = \alpha$ и допустим, что $\delta < \alpha < 1 - \delta$ (11)

при фиксированном $\delta (0 < \delta < 1)$. Тогда из (11) следует

$$\lim_{N \rightarrow \infty} \left(E(\tau_n) - \frac{1}{\lambda} \ln N \right) = \frac{1}{\lambda} \left(\ln \frac{\alpha}{1-\alpha} + C \right). \quad (12)$$

Для вычисления простого приближенного выражения для $\sigma^2(\tau_n)$ используем тождество

$$\frac{1}{k^2(N-k)^2} = \frac{1}{N^2} \left(\frac{1}{k^2} + \frac{1}{(N-k)^2} \right) + \frac{2}{N^3} \left(\frac{1}{k} + \frac{1}{N-k} \right). \quad (13)$$

Обозначив

$$s_n = \sum_{v=1}^n \frac{1}{v^2} \quad (14)$$

и используя (5), (8) и (14), получаем

$$\sigma^2(\tau_n) = \frac{1}{\lambda^2} (s_{n-1} + s_{N-1} - s_{N-n}) + \frac{2}{N\lambda^2} (h_{n-1} + h_{N-1} - h_{N-n}) = \frac{1}{\lambda^2} (s_{n-1} + s_{N-1} - s_{N-n}) + \frac{2}{N\lambda} E(\tau_n). \quad (15)$$

Известно, что

$$\lim_{k \rightarrow \infty} s_k = \frac{\pi^2}{6}. \quad (16)$$

Так как, кроме того, имеет место равенство

$$\lim_{k \rightarrow \infty} \frac{E(\tau_n)}{N} = 0, \quad (17)$$

то из (14), (16) и (17) получаем

$$\lim_{N \rightarrow \infty} \sigma^2(\tau_n) = \frac{\pi^2}{6\lambda^2}. \quad (18)$$

Путем ряда преобразований легко доказать, что для достаточно больших N дисперсия τ_n при условии выполнения (11) не зависит от α $E(\tau_n) = t(\alpha) = t$, α является временем, после которого средняя доля пользователей, подключенных к сети, достигает значения α .

Заключение

В работе рассмотрен один из подходов к созданию математической модели развития ТС. А, именно один из аспектов разработки математических моделей, адекватно описывающих изменение требований пользователей на подключение к сети и обслуживание этих требований исследуемой сетью. При этом была поставлена задача, предложить асимптотические формулы, дающие простые выражения для вероятностей состояний сети, с точки зрения количества пользователей, обслуживаемых сетью, и моментов распределения функции, описывающей количество пользователей сети в конкретный момент времени. Исследования выполнены при условии достаточно больших значений числа всех потенциальных пользователей на рассматриваемой территории N . Для дальнейших исследований поставлена задача применения логистической функции для описания процесса подключения пользователей при прогнозировании развития ТС. Логистическая функция позволяет описывать разнообразные процессы, с точностью, более высокой, чем аппроксимирующие функции.

Литература

- [Boorstyn, 1977] Boorstyn K. R., Erank H. Large-scale network topological optimization // IEEE Trans. Commun. – 1977. – Vol. 25, №1. – P. 29–47.
- [Camerini, 1979] Camerini P. M., Fratta L., Maffioli F. Some results on the design of tree structured communication networks // IX International Teletraffic Torremolinos. – 1979. – P. 214.
- [Damaskin, 2012] Galyna Gayvoronska Formalization of Variation Process of Information Networks' Users' Quantity / Galyna Gayvoronska, Oleg Damaskin // "Modern problems of radio engineering, telecommunications and computer science". Proceedings of the XIth International Conference TCSET'2012. – 2012. – Lviv: Publishing House of Lviv Polytechnic. – P.338-339.
- [Frank, 1971] Optimal design of centralized computer networks / H. Frank, I. Frisch, W. Chou, W. Van Slyke // Networks, Vol. 1. – 1971. – P. 43-57.
- [Gayvoronska, 2011] Gayvoronska G.S. Formalization of telecommunication networks' evolution's model / G.S. Gayvoronska // Applicable Information Models. – Sofia: ITHEA, 2011. – № 22. – P. 155-169.
- [Gayvoronska, 2012] Gayvoronska G. Analysis of mathematical models describing the requirements for network development / Galyna Gayvoronska, Oleg Domaskin // Natural Information Technologies.–Madrid:ITHEA,2012.–P.52-58.
- [Kershenbaum, 1975] Kershenbaum A., Boorstyn R. Centralized teleprocessing network design // Nat. Telecommun. Conf. New Orleans, La, 1975. – Vol. 2. – N. Y. – 1975, P. 27/11– 27/13.
- [Гайворонская, 2005] Гайворонская Г.С. Проблема синтеза развивающихся информационных сетей / Г.С. Гайворонская // Вісник ДУИКТ. – №3. – 2005. – Киев С.14-21
- [Гайворонская, 2006] Гайворонская Г. С. Исследование модели требований на развитие информационной сети / Г.С. Гайворонская, Д.А. Сомсиков // Холодильна техніка і технологія.– 2006.– №3 (101). – С. 99-104
- [Гайворонская, 2013] Гайворонская Г.С. Один из подходов к созданию математической модели развития телекоммуникационной сети / Г.С. Гайворонская, О.М. Домаскин // Збірник матеріалів VI МНТК «Проблеми телекомунікацій». – Київ. – НТУУ «КПІ». – 2013.
- [Давыдов, 1984] Давыдов Г.Б. Информация и сети связи. – М.: Наука, 1984. – 128 с.
- [Давыдов, 1985] Давыдов Г.Б. Некоторые проблемы оптимизации развития сетей // Электросвязь.–1985.–№2.– С.1-5
- [Нейман, 1983] Нейман В. И. Структуры систем распределения информации. – М.: Радио и связь, 1983. – 216 с.
- [Рогинский, 1977] Давыдов Г. Б., Рогинский В. Н., Толчан А. Д. Сети электросвязи. – М.: Связь, 1977. – 380 с.
- [Рогинский, 1981] Теория сетей связи / Под ред. В. Н. Рогинского. – М.: Радио и связь, 1981. – 192 с.
- [Советов, 1990] Советов Б. Я., Яковлев С. А. Построение сетей интегрального обслуживания. – Л.: Машиностроение, 1990. – 332 с.
- [Франк, 1982] Франк Г., Чжоу В. Топологическая оптимизация сетей ЭВМ // ТИИЭР. – 1982. – №11 (60). – С. 147-162.
- [Харкевич, 1975] Харкевич А. Д., Ковалева В. Д. Некоторые результаты моделирования сети связи // Методы теории телетрафика в системах распределения информации. – М.: Наука, 1975. – С. 122-127.

Благодарности

Статья опубликована при частичной поддержке проекта ITHEA XXI ITHEA ISS (www.ithea.org) и ADUIS (www.aduis.com.ua).

Информация об авторах



Галина Гайворонская – Факультет Информационных технологий и кибербезопасности ОНАПТ, д.т.н., профессор, зав. кафедрой информационно-коммуникационных технологий, советник ректора по инфокоммуникациям; Украина, Одесса, 65026, ул. Дворянская, 1/3; тел. (048)-720-91-48; e-mail: gsgayvoronska@gmail.com

Области научных исследований: оптимизация переходных периодов при эволюции телекоммуникационных сетей. Поток вызовов, нагрузка и межзвонное тяготение в сетях. Проблемы создания сетей доступа. Проблема построения полностью оптических сетей и систем коммутации



Олег Домаскин – Факультет Информационных технологий и кибербезопасности ОНАПТ, аспирант кафедры информационно-коммуникационных технологий; Украина, Одесса, 65026, ул. Дворянская, 1/3; тел. (048)-720-91-48

Главные области научных исследований: оптимизация переходных периодов при эволюции телекоммуникационных сетей.

УПРАВЛЕНИЕ КАЧЕСТВОМ УСЛУГ В ТЕЛЕКОММУНИКАЦИОННЫХ СЕТЯХ

Анастасия Кальченко

Аннотация: В работе рассмотрены особенности управления в телекоммуникационных сетях в условиях возрастающего спектра предоставляемых сетью услуг. Целью исследования является определение эффективных методов управления качеством услуг в NGN с использованием подходов нечеткой логики и нейронных сетей.

Ключевые слова: сети следующего поколения, искусственный интеллект

Ключевые слова классификации АСМ: С.2. Computer-communication networks, Н. Information Systems - Н.1 Models and Principles, К. Computing Milieux - К.6 Management of computing and information system

Введение

Стремительное развитие и распространение коммуникационных технологий приобретает сегодня характер глобальной информационной революции. У абонентов появилась потребность в услугах связи, которые можно было бы гибко настраивать в соответствии с их требованиями. Эти услуги должны позволять маршрутизировать вызовы на разные телефонные номера согласно заданным условиям, например, в зависимости от дня недели (рабочий или выходной) или времени суток, обеспечивать мобильность терминала и абонента, когда доступ к корпоративной сети осуществляется через телефонную сеть или Интернет из любой географической точки мира, и т.п. Сети следующего поколения (NGN) представляют собой единую транспортную платформу, на базе которой объединяются различные виды услуг. Применение сетей следующего поколения позволяет значительно расширить спектр предоставляемых интеллектуальных услуг. Одним из основных аспектов, который должен приниматься во внимание при проектировании сетей NGN, является обеспечение соответствующего качества обслуживания. С внедрением мультисервисных сетей преобладающим становится подход к заданию уровня обслуживания на основании требований самих абонентов к качеству услуг. Таким образом, создание усовершенствованной системы управления качеством услуг в сетях следующего поколения является актуальной задачей.

Этапы развития сетей связи

В историческом развитии сетей и услуг связи можно выделить следующие этапы: PSTN, IDN, ISDN, IN, NGN, FN.

Первый этап – построение телефонной сети общего пользования PSTN (Public Switched Telephone Network). Телефонная связь отождествлялась с единственной услугой – передачей речевых сообщений. В дальнейшем по телефонным сетям с помощью модемов стала осуществляться передача данных.

Второй этап – цифровизация телефонной сети, были созданы интегральные цифровые сети IDN (Integrated Digital Network), предоставляющие также в основном услуги телефонной связи на базе цифровых систем коммутации и передачи.

Третий этап – интеграция услуг: появилась концепция цифровой сети с интеграцией служб ISDN (Integrated Service Digital Network). В процессе развития сетей связи особое внимание стало уделяться

дополнительным услугам. Именно поэтому интеграция услуг начинает заменяться концепцией интеллектуальной сети.

Четвертый этап – интеллектуальная сеть IN (Intelligent Network). Эта сеть предназначена для быстрого, эффективного и экономичного представления информационных услуг массовому пользователю. Принципиальное отличие интеллектуальной сети от предшествующих сетей – в гибкости и экономичности предоставления услуг.

Интеллектуальные услуги приобретают все большую популярность. К подобным сервисам относятся Freephone (звонки за счет вызываемой стороны), Premium Rate Service (звонки с начислением дополнительной оплаты, например за доступ к информационным ресурсам или за участие в телефонных лотереях, голосованиях и т.п.), Prepaid Calling (звонки по предоплате с доступом абонентов по паролям), Least Cost Routing (маршрутизация по наиболее выгодному маршруту) и ряд других.

Кроме того, в настоящее время распространение в мире получают новые услуги связи, относящиеся за рубежом к услугам с добавленной стоимостью – VAS (Value Added Services). К этим услугам относятся интеллектуальные услуги, реализуемые на базе архитектурной концепции интеллектуальной сети связи.

Вопросы управления интеллектуальными услугами рассматриваются в работах Гольдштейна Б.С., Лихтцендера Б.Я., Ехриеля И. М. и др. [1].

В рамках интеллектуальной сети известны два способа создания инфраструктуры для предоставления интеллектуальных услуг:

- на базе "классической" интеллектуальной сети, предполагающей наличие таких элементов архитектуры, как узел коммутации услуг (SSP), узел управления услугами (SCP) и др.;
- сетевые решения на основе узлов услуг (Service Node).

Базовыми элементами классической архитектуры IN являются: узел коммутации услуг (SSP); узел управления услугами (SCP); узел управления IN (SMP); среда создания услуг (SCE) и так называемая интеллектуальная периферия (IP). Другой вариант архитектуры IN, «выросший» из «компьютерной телефонии» и построенный на базе так называемого узла услуг (SN), представляет собой единую систему, объединяющую функции SSP, SCP и IP и подключенную к одной из цифровых АТС телефонной сети.

Реализация интеллектуальных услуг (ИУ) осуществляется на базе узлов служб (SN) и/или узлов управления услугами (SCP):

- 1) узел управления услугами (Service Control Point – SCP) – специализированный узел связи, осуществляющий управление предоставлением услуг в соответствии с концепцией интеллектуальной сети связи и принадлежащий оператору сети связи;
- 2) узел служб (Service Node – SN) – специализированный узел сети связи, осуществляющий предоставление ИУ и принадлежащий поставщику услуг.

В классических интеллектуальных сетях связи IN узел управления услугами SCP (Service Control Point) отделен от узла коммутации услуг SSP (Service Switching Point). Такая архитектура IN позволила операторам связи предоставлять услуги своим абонентам, даже если они совершают звонки из другой сети (роуминг услуг).

Узлы упрощенной схемы IN размещены на трех уровнях иерархии: узел коммутации услуг SSP с интеллектуальной периферией IP; узел управления услугами SCP с узлом данных услуги (базой данных) SDP; узел менеджмента услуг SMP с узлом создания услуг SCEP.

Для получения ИУ пользователь сети набирает номер той АТС, которая обладает функциями SSP, а также код услуги и номер услуги. Пользуясь протоколом INAP (Intelligent Network Application Part), АТС с функциями SSP общается с узлом SCP и получает необходимую информацию для предоставления услуги и обслуживания вызова. В обслуживании вызова принимает участие IP (для передачи голосовых команд пользователю, сбора дополнительной информации и т. д.). Общение между SCP, SSP и IP происходит в режиме реального масштаба времени с учетом жестких временных ограничений на обслуживание телефонного вызова. Подготовка новых услуг происходит в узле SСЕР, а за введение новых услуг отвечает узел SMP. Эти два центра действуют в условиях относительного масштаба времени, и для передачи информации о новых услугах в узел SCP используется, например, протокол X.25 или Frame Relay.

Такие системы позволяют интегрировать в единое целое автоматические и полуавтоматические (операторские) службы, телефонную сеть и Интернет, обеспечивая абонентам возможности доступа практически к любой необходимой информации. При этом расширение системы (как в плане производительности, так и в плане функционального наполнения) сводится лишь к включению (потенциально - в любой точке сети Интернет) необходимого количества дополнительных модулей, взаимодействующих по стандартным протоколам.

На рис. 1 представлена концептуальная модель IN. Процесс подключения телефонных соединений осуществляется на транспортном уровне, включающем сетевые узлы и коммутационные станции. Логика предоставления ИУ реализуется в соответствующих узлах интеллектуального уровня. Для взаимодействия интеллектуального и транспортного уровней используется сеть передачи данных (СПД), в качестве которой чаще всего используется сеть общеканальной сигнализации ОКС №7 со специальной прикладной подсистемой пользователя интеллектуальной сети INAP.

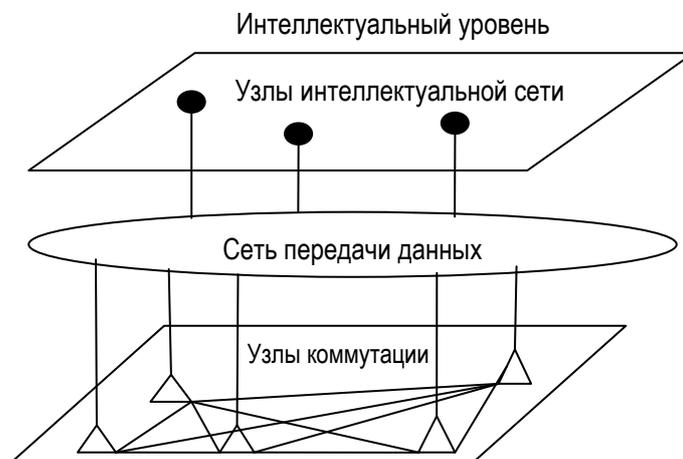


Рис. 1. Концептуальная модель IN

В компьютерах SCP наряду с базой данных была запрограммирована и так называемая логика услуги, состоящая из сценариев, описывающих данную услугу. Именно с этого исторического момента логика услуги начала перемещаться за пределы АТС, что и составило суть концепции IN. В IN стандартизирован набор услуг ITU Q.1211 «Intelligent Network – Introduction to Intelligent Network Capability Set 1» (далее – CS-1). Он представляет собой описание голосовых услуг, перечень которых представлен в таблице 1.

Однако концепция IN не принесла желаемого многообразия услуг из-за сложности протокола между SCP и SSP [1].

Дальнейшим развитием стало появление сетей связи следующего поколения (NGN – Next Generation

Network). Основу сети NGN составляет мультипротокольная сеть – транспортная сеть связи, входящая в состав мультисервисной сети, обеспечивающая перенос разных типов информации с использованием различных протоколов передачи. Концепция NGN отделяет не только услуги SCP от управления соединением SSP, но и управление соединением SSP от транспорта [2]. Представлена на рис. 2 [3].

Таблица 1

	Название из CS-1	Русское название
ABD	Abbreviated dialing	Сокращенный набор
ACC	Account card calling	Карты предоплаты
CF	Call forwarding	Перенаправление вызова на номер, определенный абонентом
FPH	Freephone	Бесплатный вызов за счет вызываемого абонента
PRM	Premium rate	Вызов с дополнительной стоимостью
VOT	Televoting	Телеголосование
VPN	Virtual private network	Виртуальная частная сеть
и т.д., всего более 20 услуг		

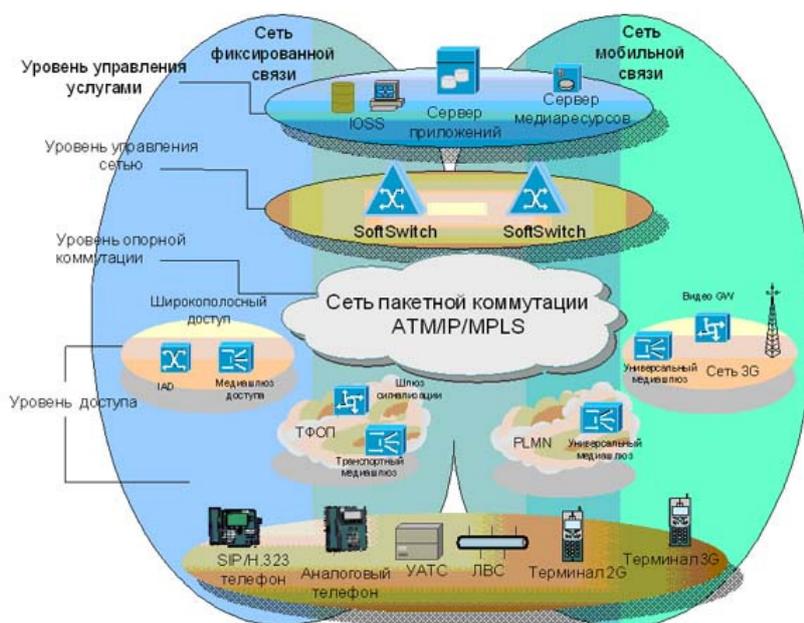


Рис. 2. Архитектура сети NGN

Вводятся новые элементы сети: программный коммутатор Softswitch, или функция управления вызовами и сессиями CSCF (Call Session Control Function) в IMS (IP Multimedia Subsystem), которые, с одной стороны, управляют соединением, а с другой – взаимодействуют с серверами предоставления услуг по SIP протоколу (Session Initiation Protocol).

В терминах NGN платформа предоставления интеллектуальных услуг называется SDP (Service Delivery Platform). Основа идеологии NGN – это открытые стандарты консорциума 3GPP (3rd Generation Partnership Project). Открытые стандарты позволяют оператором использовать уже приобретенное оборудование, поддерживающее данные стандарты, и при необходимости менять только требуемые узлы

сети, дают возможность не привязываться к одному поставщику оборудования, а для каждого узла сети выбирать наиболее подходящего.

Есть две концепции перехода к сетям NGN со своими положительными и отрицательными сторонами:

- операторы строят сразу NGN сети на базе IMS (IP Multimedia Subsystem);
- операторы сначала внедряют в существующие сети программные коммутаторы (Softswitch), потом осуществляют постепенный переход на IMS.

Основное отличие этих подходов в том, что IMS – это полностью IP сеть и доступ к ней происходит с помощью 3G телефонов, с поддержкой Wi-Fi, широкополосного Интернета, GPRS. Программный коммутатор Softswitch предназначен для конвергентных сетей и его основная функция - это управления медиашлюзами. Доступ к таким сетям происходит по обычным телефонным линиям, которые подключаются к медиашлюзам, а маршрутизация вызова, управление соединением, предоставление дополнительных голосовых услуг происходит в IP сети.

В архитектуре NGN системы предоставления услуг находятся в IP домене и предназначены для предоставления интеллектуальных (без участия оператора) услуг. Это могут быть, как голосовые услуги, такие как Автоинформатор (IVR) или Голосовая Почта (Voice Mail), так и видео услуги и мультимедийные сервисы.

Архитектура IMS определяет три типа приложений:

- 1) IMS Applications - входят в архитектуру IMS и называются серверами приложений (Application Servers). Различают: SIP AS – для предоставления услуг на основе SIP протокола, IM-SSF (IP Multimedia – Service Switching Function) – для предоставления услуг IN сети, и OSA-SCS (Open Service Access – Service Capability Server) – услуги предоставляются с помощью OSA/Parlay интерфейса;
- 2) Applications - это приложения, не входящие в структуру IMS, но работающие внутри инфраструктуры оператора связи;
- 3) 3rd party Applications - приложения сторонних производителей, работающие вне инфраструктуры оператора.

По сравнению с интеллектуальными сетями связи IN, новым элементом сети IMS является HSS (Home Subscriber Server) – сервер домашних абонентов, в котором хранятся данные абонентов.

Идеология построения NGN обеспечивает возможность предоставления абонентам услуг Triple-Play (передача речи, данных и видео) на базе мультисервисных сетей, создаваемых путем модернизации существующих сетей электросвязи.

Особенности управления в сети NGN

Концепция NGN во многом опирается на технические решения, уже разработанные международными организациями стандартизации. Так, взаимодействие серверов в процессе предоставления услуг предполагается осуществлять на базе протоколов, специфицированных IETF (MEGACO), ETSI (TIPHON), Форумом 3GPP2 и т.д. Для управления услугами используются протоколы H.323, SIP и подходы, применяемые в интеллектуальных сетях связи [2].

На рис. 2 отчетливо видна иерархия сетевой инфраструктуры: уровень опорной коммутации, уровень управления коммутацией и передачей информации, уровень управления услугами, уровень доступа. Задача уровня опорной коммутации — коммутация соединений и прозрачная передача информации. Уровень управления коммутацией и передачей служит для обработки сигнальных команд, маршрутизации

вызовов и управления потоками. Уровень управления услугами содержит в себе логику предоставления услуг и доступа к приложениям. Уровень доступа предоставляет широкий набор интерфейсов для подключения к услугам сети.

Уровень услуг включает функции управления услугами, включая функции профилей услуг пользователей; функции поддержки приложений и функции поддержки услуг.

Функции управления услугами (service control functions, SCF) включают управление ресурсами, функции регистрации, аутентификации и авторизации для различных услуг на уровне услуг. Они также могут включать функции управления медиаресурсами, такими как специализированные устройства и шлюзы на сигнальном уровне. Функции управления услугами поддерживают профили услуг пользователей, которые представляют собой комбинацию пользовательской информации и других данных управления, образующую индивидуальный профиль каждого пользователя и объединенные в функциональные базы данных.

Функции административного управления (management functions) обеспечивают возможность управлять сетью NGN для предоставления услуг с заданным уровнем качества, безопасности и надежности. Эти функции распределяются децентрализованно по всем функциональным блокам (FE) и они взаимодействуют с функциональными блоками управления сетевыми элементами, управления сетью и управления услугами. Функции административного управления используются на транспортном уровне и уровне услуг и для каждого этого уровня они реализуют следующие задачи:

- управление процессом устранения отказов (Fault Management);
- управление конфигурацией сети (Configuration Management);
- управление расчётами с пользователями и поставщиками услуг (Accounting Management);
- контроль производительности сети (Performance Management);
- обеспечение безопасности работы сети (Security Management).

Некоторые аспекты управления услугами рассматриваются в работах В. Стеклова, Л. Беркман и др.. Нарботки в данной сфере сделаны также Е. Штейнбергом. В его работах внимание уделяется распределенным системам управления в интеллектуальных сетях. Однако все эти работы посвящены совершенствованию централизованной системы управления (ЦСУ). В [4] на основании моделирования процессов функционирования ЦСУ и децентрализованной системы управления (ДСУ) проведен сравнительный анализ ЦСУ с ДСУ.

При ЦСУ существует единый центр управления, в котором и находится вся необходимая информация и сервисная логика. При ДСУ предусмотрено существование отдельной интеллектуальной надстройки, представляющей собой по сути сеть сигнализации. В эту сеть входят несколько центров управления, каждый из которых содержит только логику обслуживания и необходимые данные для определенного класса или нескольких классов услуг.

В современных сетях NGN используется централизованная система управления. Фрагмент сети с такой архитектурой представлен на рис.3. Интеллектуальная надстройка, которая отвечает за управление дополнительными услугами, может представлять собой, например, сеть TMN. В нее входит часть Softswitch, выполняющая функцию коммутации услуги SSF и сервер, который выполняет в первую очередь функцию обслуживания услуги SCF. Предполагается, что существует несколько территориально разнесенных районов. Управление фрагментом сети этого района осуществляет установленный Softswitch. Он одновременно управляет транспортной частью сети и является точкой коммутации дополнительных услуг. Каждый район имеет свою сеть передачи данных и сеть сигнализации.

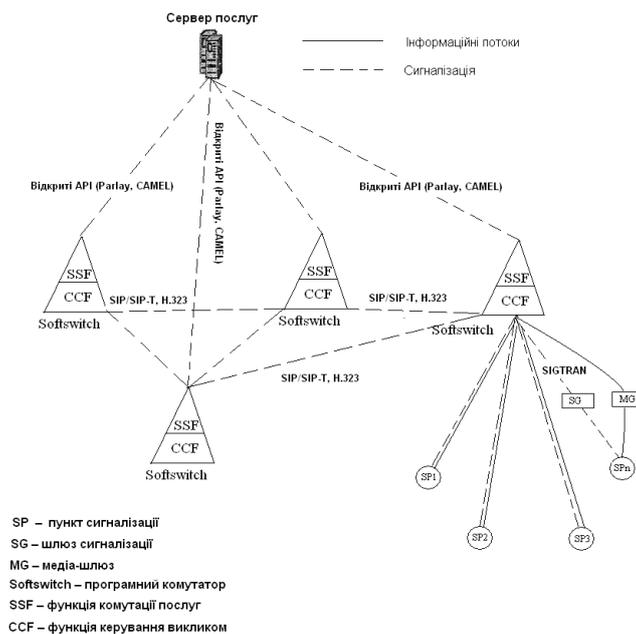


Рис. 3. Архитектура NGN с ЦСУ дополнительными услугами

При интенсивности поступления заявок, меньшей интенсивности их обслуживания на сервере, более качественно функционирует ЦСУ. При интенсивности поступления заявок, равной или большей, чем интенсивность их обслуживания на сервере, лучше применять ДСУ.

Управление качеством в сети NGN

Одним из основных аспектов, который должен приниматься во внимание при проектировании сетей NGN, является обеспечение качества обслуживания. Вопросы обеспечения качества услуг связи являются актуальными, и в настоящее время этим вопросом занимается порядка 12 крупных международных организаций, включая: МСЭ-Т, ETSI, 3GPP, DSL Forum, CableLab и др [5].

В Рекомендации ITU-T E.800 приведены следующие определения [6].

Качество услуги (Quality-of-service, QoS) - совокупность характеристик телекоммуникационной услуги, относящихся к способности удовлетворить установленные и предполагаемые потребности пользователя услугой (определение заимствовано из стандарта ISO 8402).

Параметры функционирования сети (Network Performance, NP) – способность сети предоставить функциональность, обеспечивающую взаимодействие пользователей.

Параметры функционирования сети (NP) [7]: IPTD – задержка передачи информации, IPDV – девиация задержки, IPLR – доля потерь информации, IPER – доля ошибок.

На рис. 4 представлена взаимосвязь Quality of Service и Network Performance [8].

Рекомендация Y.1541 устанавливает соответствие между классами качества обслуживания и приложениями:

- Класс 0 – приложения реального времени, чувствительные к джиттеру, характеризующиеся высоким уровнем интерактивности (VoIP, видеоконференции);
- Класс 1 – приложения реального времени, чувствительные к джиттеру, интерактивные (VoIP, видеоконференции);

- Класс 2 – транзакции данных, характеризуемые высоким уровнем интерактивности (например, сигнализация);
- Класс 3 – транзакции данных, интерактивные;
- Класс 4 – приложения, допускающие низкий уровень потерь (короткие транзакции, массивы данных, потоковое видео);
- Класс 5 – традиционные применения сетей IP.

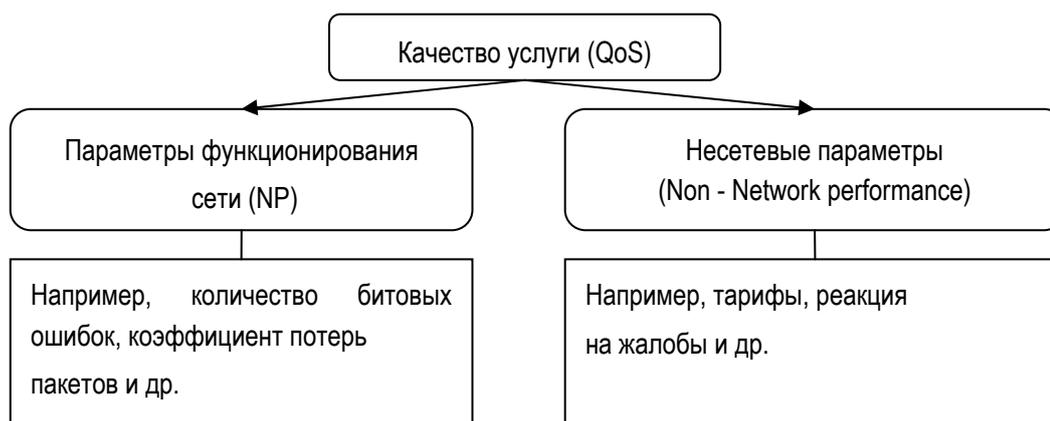


Рис. 4. Взаимосвязь Quality of Service и Network Performance

Нормы для характеристик сетей IP с распределением по классам качества обслуживания приведены в таблице 2 [2].

Таблица 2

Сетевые характеристики	Классы QoS					
	0	1	2	3	4	5
Задержка доставки пакета IP, IPTD	100 мс	400 мс	100 мс	400 мс	1 с	Н
Вариация задержки пакета IP, IPDV	50 мс	50 мс	Н	Н	Н	Н
Коэффициент потери пакетов IP, IPLR	1×10^{-3}	Н				
Коэффициент ошибок пакетов IP, IPER	1×10^{-4}	Н				
Примечание: Н – не нормировано						

Качество восприятия (Quality-of-experience, QoE) – приемлемость услуги или приложения в целом, субъективно воспринимаемая конечным пользователем (P10/G.100. Annex 1. Quality of Experience). Качество восприятия учитывает влияние всех аспектов и участников предоставления услуги (пользователь, терминал, сеть и т.д.). Приемлемость услуги может зависеть от ожиданий пользователя.

Параметры, влияющие на качество восприятия услуги [8]:

- коэффициент сетевой эффективности (NER),
- задержка предоставления услуги,
- качество предоставляемой медиа-информации (MOS/R-фактор).

На рис. 5 представлена взаимосвязь точек зрения пользователя и оператора на качество услуги.

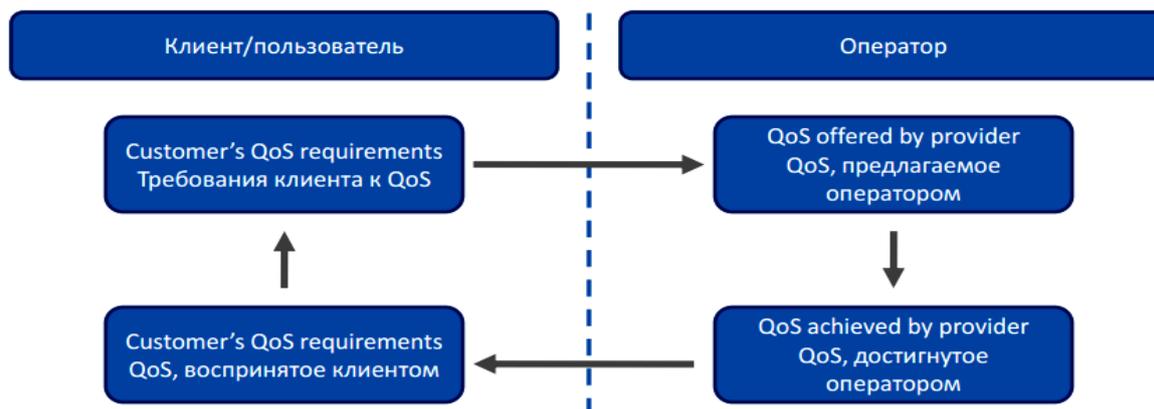


Рис. 5. Точки зрения на качество услуги

Требования клиента к QoS (Customer's QoS requirements) определяют уровень качества услуги, требуемый клиентом. Критерии и параметры, определившие этот уровень, находят отражение в требованиях.

QoS, предлагаемое оператором (QoS offered by provider), – это перечень четких однозначно определенных требований, которые могут быть использованы: как основа для формирования SLA (Service Level Agreement); для декларирования оператором уровня качества доступного пользователям; как основа для планирования и поддержания услуги на заданном уровне; как основа для пользователей при выборе оператора, обеспечивающего наиболее приемлемый уровень качества услуги.

QoS, достигнутое оператором (QoS achieved by provider), – это уровень качества услуги, фактически предоставленный оператором. Может использоваться: клиентами, регулятором как основа для сравнения предлагаемого оператором уровня качества услуги и фактически предоставляемого и проверки выполнения SLA; оператором как основа для корректив.

QoS, воспринятое клиентом (QoE, Customer's QoS experience), – это качество услуги, воспринимаемое клиентом и выражаемое в виде оценки. Основывается на опросах клиентов и характеризует мнение клиента о качестве полученных услуг. Эти данные могут быть использованы для: сравнения с предлагаемым уровнем качества услуги и определения причин отклонений; планирования корректив.

В целях обеспечения требуемого качества передачи речевой информации каждый из классов обслуживания определяется тремя количественными характеристиками: общей оценкой качества передачи (R); качеством речи, воспринимаемым слушателем (качеством односторонней неинтерактивной передачи речи из конца в конец); задержкой из конца в конец (односторонней).

В соответствии с методикой MOS (Mean Opinion Score) качество речи, получаемое при прохождении сигнала от говорящего (источник) через систему связи к слушающему (приемник), оценивается как арифметическое среднее от всех оценок, выставляемых экспертами после прослушивания тестируемого тракта передачи. Экспертные оценки определяются в соответствии со следующей пятибалльной шкалой: 5 — отлично, 4 — хорошо, 3 — приемлемо, 2 — плохо, 1 — неприемлемо. Оценки 3,5 балла и выше соответствуют стандартному и высокому качеству, 3,0 — 3,5 — приемлемому качеству, 2,5 — 3,0 — синтезированному звуку. Для передачи речи с хорошим качеством целесообразно ориентироваться на значения MOS не ниже 3,5 баллов.

Взаимосвязь между общей оценкой качества передачи (R) и восприятием качества пользователем определяется в соответствии с таблицей 3 (Рекомендация МСЭ-T G.109).

Таблица 3

Оценка QoS на основе R-фактора и оценок MOS		
Значение R-фактора	Категория качества и оценка пользователя	Значение оценки MOS
90<R<100	Самая высокая (отлично)	4,34 — 4,50
80<R<90	Высокая (хорошо)	4,03 — 4,34
70<R<80	Средняя (приемлемо: часть пользователей оценивает качество как неудовлетворительное)	3,60 — 4,03
60<R<70	Низкая (плохо: большинство пользователей оценивает качество как неудовлетворительное)	3,10 — 3,60
50<R<60	Неприемлемая (не рекомендуется)	2,58 — 3,10

Взаимосвязь QoE, параметров функционирования и параметров производительности сети представлены в таблице 4.

Таблица 4

Характеристика услуги (Профиль)	Условия эталонной модели услуги	Предельные значения функционирования сети (NP)	Предельные значения производительности сети
VAD	MOS -> 3.5	IPTD	Полоса пропускания (BW)
Codec	Call time -180 сек	IPDV	De-jitter buffer
G.168	Signalling timer < lim	IPLR	Алгоритмы маршрутизации и приоритетизации трафика
Packetization time и packet formation time		IPER	

Использование искусственного интеллекта в управлении качеством услуг

Изменение парадигмы в концепции услуг, которое было связано с общим изменением концепции сетей NGN, выражается в первую очередь в том, что роли оператора и пользователя существенно поменялись. Теперь пользователь и оператор выступают как союзники в едином процессе информатизации, и это сотрудничество можно считать лейтмотивом эволюции современных услуг.

Следовательно, разрабатывая систему управления услугами рационально использовать системный подход: проблему обеспечения качества нужно решать не изолированно, а в единстве связей с окружающей средой – пользователем. Удовлетворение требований пользователя включает в себя как технические аспекты (параметры качества функционирования сети), так и нетехнические (обслуживание пользователей) [6]. В процессе управления услугами необходимо отслеживать как соответствие характеристик услуг нормативным показателям, так и производить при необходимости коррекцию нормативов.

Формирование качества услуги включает в себя как объективную оценку сетевых характеристик, так и субъективную экспертную и пользовательскую оценку. И в то время как параметры работы сети можно определить при помощи соответствующего оборудования, учет мнения клиентов о качестве полученных услуг осуществляется путем соотношения QoS, предлагаемого оператором, и QoS, воспринятого клиентом, или QoE. При этом в сети должно функционировать устройство, которое сравнивает разницу между требуемым уровнем качества и фактическим, и, если она превышает допустимое значение – определяет, какие изменения конфигурации сети необходимы, и формирует соответствующие управляющие сигналы. Важным аспектом является время отклика сети на мнение пользователя, сократить которое возможно при постоянном мониторинге QoE и способности системы управления сетью

прогнозировать отклик пользователей. Сеть должна запоминать и анализировать состояние сети и соответствующую оценку качества услуг клиентом и уметь корректировать конфигурацию сети на основании полученного опыта. Данный подход может быть реализован путем введения искусственного интеллекта (ИИ) в систему управления (СУ) услугами.

Главная идея использования ИИ состоит в изменении парадигмы сетевой инфраструктуры: теперь не пользователь со своим приложением подстраивается под возможности сети, а сеть меняет свои настройки с учетом требований пользователя. Конфигурация сети и функциональность сетевого оборудования автоматически изменяются в зависимости от требований пользователя. Сеть не только реагирует на текущие запросы пользователя, но также анализирует его предпочтения и текущее окружение, предоставляя соответствующую информацию СУ.

Одним из эффективных подходов к реализации ИИ в управлении ИУ является применение современных методов – нечеткой логики и нейронных сетей. Важнейшим достоинством нейронных сетей является возможность их обучения и адаптации, а также то, что не требуются полные знания об объекте управления (например, его математическая модель). На основе входных и заданных (эталонных) сигналов нейронная сеть может научиться управлять объектом. Нейронные сети могут включать огромное количество взаимосвязанных простых обрабатывающих элементов (нейронов), что в результате дает громадную вычислительную мощность при использовании параллельной обработки информации.

Нечеткое управление (Fuzzy Control, Fuzzy-управление) в настоящее время является одной из перспективнейших интеллектуальных технологий, позволяющих создавать высококачественные системы управления [9]. Основным достоинством метода нечеткой логики является возможность представления субъективных категорий в математической форме. Система принимает решения на основе правил, записанных в форме импликации IF-THEN. Простейший подход к проектированию таких систем заключается в формулировании правил управления и функции принадлежности по результатам наблюдения за процессом управления, осуществляемым человеком либо уже существующим регулятором, с последующим оцениванием корректности функционирования такой системы. Если проект оказывается неудачным, то функцию принадлежности и/или правила управления можно легко модифицировать [10].

Реализация управления качеством осуществляются на уровне услуг и транспортном уровне. При этом необходимо обеспечить взаимодействие с системой управления как новых поставщиков услуг, поставщиков информации, так и пользователей.

Использование данных подходов позволит, с одной стороны, привнести в СУ способность к обучению и вычислительную мощность нейронных сетей, а с другой стороны – использовать свойственные «человеческому» способу мышления нечеткие правила выработки решений.

Заключение

Сеть NGN обеспечивает предоставление неограниченного набора услуг с гибкими возможностями по их управлению, персонализации и созданию новых услуг за счет унификации сетевых решений, предполагающая реализацию универсальной транспортной сети с распределенной коммутацией, вынесение функций предоставления услуг в оконечные сетевые узлы и интеграцию с традиционными сетями связи. Одной из важнейших характеристик работы сети является качество предоставляемых услуг. Качество услуги связи (Quality-of-service, QoS) трактуется Рекомендацией E.800 МСЭ-T как степень удовлетворения этой услугой пользователя.

Существующие международные нормативы по показателям качества имеют рекомендательный и методологический характер и в основном ориентированы на качество услуг. Применение искусственного

интеллекта позволит максимально эффективно осуществлять управление услугами в условиях возрастающего спектра предоставляемых сетью услуг, а также производить коррекцию логики предоставления различных услуг, непрерывно анализируя степень удовлетворенности пользователей сети.

Литература

1. Гольдштейн Б.С., Ехриель И.М., Рерле Р.Д. Интеллектуальные сети. – М.: Радио и связь. – 2000. – 500 с.
 2. Семенов Ю.В. Проектирование сетей связи следующего поколения. – СПб: Наука и техника. – 2005. – 240 с.
 3. Материалы Интернет-сайта "УКРИНФОРМСВЯЗЬ". — Режим доступа: <http://www.informsviaz.co.ua/> (дата обращения 25.03.2013).
 4. Князева Н. О., Шестопапов С.В. Підвищення якості управління послугами при застосуванні децентралізованої системи управління.// «Вісник ДУІКТ» Т.8, №1. – 2010.
 5. "Международный союз электросвязи (ITU)", официальное Интернет-представительство. — Режим доступа: <http://www.itu.int> (дата обращения 15.03.2013 г.).
 6. Тихвинский В.О., Терентьев С.В. Управление и качество услуг в сетях GPRS/UMTS: учеб. пособие. - М.: Эко-Трендз, 2007. - 395с.
 7. Андреев Д.В. Методология тестирования параметров функционирования сети в целях обеспечения качества услуг.// Семинар ФГУП ЦНИИС «Подходы по тестированию параметров функционирования сети в целях обеспечения качества услуг связи». – Москва. – 2011. – 45 с.
 8. Иткин А. Стандартизация QoS и понятие качества услуг.// Семинар ФГУП ЦНИИС «Подходы по тестированию параметров функционирования сети в целях обеспечения качества услуг связи». – Москва. – 2011. – 20 с.
 9. Рутковская Д., Пилиньский М., Рутковский Л. Нейронные сети, генетические алгоритмы и нечеткие системы: Пер. с польск. И. Д. Рудинского. - М.: Горячая линия -Телеком, 2006. - 452 с.
 10. Усков А.А. Принципы построения систем управления с нечеткой логикой // Приборы и системы. Управление, Контроль, Диагностика. 2004. № 6. С. 7-13.
-

Благодарности

Настоящая работа была выполнена при поддержке интернационального проекта ITHEA XXI Института информационных теорий и их приложений FOI ITHEA и Ассоциации ADUIS Украина (Ассоциация разработчиков и пользователей интеллектуальных систем)., в связи с чем автор и выражает свою искреннюю благодарность.

The paper is partially financed by the project ITHEA XXI of the Institute of Information Theories and Applications FOI ITHEA and the Association of Developers and Users of Intelligent Systems ADUIS Ukraine. www.ithea.org , www.aduis.com.ua

Информация об авторах

Анастасия Кальченко – Учебно-научный институт холода, криотехнологий и экоэнергетики им. Мартьяновского В.С. Одесской государственной академии пищевых технологий, Факультет информационных технологий и кибербезопасности, аспирант кафедры информационных систем и сетей; ул. Дворянская, 1/3, Одесса-26, 65026, Украина; тел. (048)-7209173; моб. (38097)-284-32-86; E – m: ja_anastasija@mail.ru

Главные области научного исследования: Управление качеством услуг в сетях следующего поколения

PROBLEM OF RESTORING THE FUNCTIONS-SIGNALS BY FINITE SET OF DATA WITH ERRORS

Veda Kasyanyuk, Iryna Volchyna

Abstract: This paper researches the problem of restoring the functions-signals which allow a spectral representation by a finite set of data which is corrupted by errors. The authors consider the cases when the errors are random values and which certain statistical information is either known or not (the errors are considered as fuzzy values). The pseudoinverse solution, solution in the form of an expansion by eigenfunctions as well as Pareto-optimal and the maximum plausible estimations are constructed. A connection between them is established.

Keywords: function-signal, pseudoinversion, maximum plausible estimations, Pareto-optimization, fuzzy values.

ACM Classification Keywords: I.6 Simulation and Modeling.

Introduction

Let's consider the class of signals that can be represented by the following structural formula:

$$f(z) = \int_Q g(z,t)u(t)dt, \quad z \in D, \quad (1)$$

where $u(t)$ and $g(z,t)$ (complex-valued functions) are a spectrum and a kernel (transfer function) of a signal; $Q \subset R^{n_1}$, $D \subset R^{n_2}$; $u(t) \in L_2[Q]$, $g(z,t)$ is a continuous function at $D \times Q$. The function $f(z)$ admitting the integral representation should be named a function-signal.

We understand the problem of restoring the function-signal (1) by the finite set of data as the following interpolation problem: it is necessary to restore (to estimate) the value $f(z)$ at the arbitrary point $z \in D$ by the values $f(z)$ at the points $z_1, \dots, z_m \in D$, known with random errors v_k , $k = 1, \dots, m$. In other words, it is necessary to estimate the linear functional (1) at the arbitrary point $z \in D$ by a set of values m of the linear functionals:

$$y_k = \int_Q g_k(t)u(t)dt + v_k, \quad k = 1, \dots, m, \quad (2)$$

We assume that statistical characteristics (mathematical expectation and correlation matrix) of random errors are known, and $M(v) = 0$, $M(vv^*) = R$, $\det R \neq 0$, $v = (v_1, \dots, v_m)^T$, $*$ – is a symbol of Hermitian conjugation. It obviously follows from this that the problem of restoring the function-signal by values of derivatives.

This problem has many important applications in the theory of communication, spectral estimation, control theory, radio- and hydrolocation, radioastronomy, optics, radio astronomy, medicine and other applied spheres.

The stated problem is incorrect in the mathematical sense because of incomplete data (the finite set of values $f(z_k)$) and the presence of errors in them, and therefore requires special approaches to its solving.

Traditional approaches, similar to Tikhonov regularization which are based on minimizing the residual norm and on narrowing the set of possible solutions up to a compact, don't take into consideration the statistical information

on errors at the data. Furthermore, it isn't always possible to estimate the error of the regularized solution. Therefore, the authors have set themselves the task to consider other possible approaches to solving the defined problem, both in terms of the availability of the statistical information on errors and in its absence (in this case the errors were considered as the fuzzy variables) and to establish a connection between the obtained results and to offer an optimal solution of this problem.

Pseudoinverse solution

As the first approach, let's consider the solution of the system of integral equations (2) which is obtained by means of pseudoinversion tools [Альберт, 1977].

Now we rewrite the system (2) in the operator form:

$$y = Gu + v \quad (3)$$

$$y = (y_1, \dots, y_m)^T, G = \int_Q g(t) dt : L_2[Q] \rightarrow C^m, g(t) = (g_1(t), \dots, g_m(t))^T.$$

Let's denote the unknown vector by f : $f = Gu = (f_1, \dots, f_m)^T$.

By definition [Альберт, 1977], the operator of the following form is named as the pseudoinverse to the operator G :

$$G^- = \lim_{\alpha \rightarrow 0} G^*(GG^* + \alpha E)^{-1}, \alpha > 0.$$

G^* , presented here is an operator conjugated with G relative to scalar products:

$$(G^*r, u)_{L_2[Q]} = (r, Gu)_{C^m}, r = (r_1, \dots, r_m)^T,$$

$u = u(t)$, E is an identity matrix.

In this condition, G^* is m -dimensional vector-function of the following form: $G^* = g^*(t) = (\bar{g}_1(t), \dots, \bar{g}_m(t))$.

The operator GG^* is a covariance matrix of kernels of the system (2):

$$K = (k_{ij})_{i,j=1}^m = \left(\int_Q g_i(t) \bar{g}_j(t) dt \right)_{i,j=1}^m,$$

which characterizes the "quality" of the points z_k for restoring (the matrix K , depending on the choice of $z_k, k = 1, \dots, m$ may be better or worse conditioned). Then the estimation of the spectrum $u(t)$ will be as follows ($\alpha > 0$):

$$\hat{u}(t) = G^- y = \lim_{\alpha \rightarrow 0} g^*(t)(K + \alpha E)^{-1} y \quad (4)$$

and

$$\hat{f}_0(z) = \lim_{\alpha \rightarrow 0} \int_Q g(z, t) g^*(t) dt (K + \alpha E)^{-1} y \quad (5)$$

The estimation (4) can be written as:

$$\hat{u}(t) = \lim_{\alpha \rightarrow 0} \left(-\frac{1}{\Delta_\alpha} \right) \begin{vmatrix} 0 & \bar{g}_1(t) & \dots & \bar{g}_m(t) \\ y_1 & k_{11} + \alpha & \dots & k_{1m} \\ \dots & \dots & \dots & \dots \\ y_m & k_{m1} & \dots & k_{mm} + \alpha \end{vmatrix}, \alpha > 0 \quad (6)$$

$$\Delta_\alpha = \begin{vmatrix} k_{11} + \alpha & \dots & k_{1m} \\ \dots & \dots & \dots \\ k_{m1} & \dots & k_{mm} + \alpha \end{vmatrix}, \alpha > 0.$$

Relying on the pseudoinversion properties at Hilbert spaces, there's known that for the linear operator equation $Aw = v$ the element $w = A^{-1}v$ minimizes the residual norm $\|v - Aw\|$, and among all the elements that have this property $w = A^{-1}v$ has a minimum norm. Therefore, for this problem one can interpret $\hat{u}(t) = G^{-1}y$ as the energy spline (or spline that smoothes the values $y_k, k = 1, \dots, m$) (depending on whether $\min\|y - G\hat{u}(t)\|_{C^m}$ is reduced to zero or not), because:

$$\|u(t)\|_{L_2[Q]}^2 = \int_Q |u(t)|^2 dt$$

is physically interpreted as the energy of spectrum of the signal (1).

The signal restoration by the energy spline is optimal in the following sense. As shown in [Белов, 1986], the set of values of the function $f(z)$ of the form (1) at the point $z \in D$, where $u(t)$ ranges over the class of functions $L_2[Q]$ satisfying the following inequality:

$$\int_Q |u(t)|^2 dt \leq I^2,$$

is a circle $|f(z) - \hat{f}_0(z)| \leq S(z)$ centered at the point $\hat{f}_0(z)$. It turns out, $\hat{f}_0(z)$ is the same as the energy spline (5). The radius $S(z)$ is determined by the dissipation of energy $I^2 - \int_Q |\hat{u}(t)|^2 dt$ and the choice of the points z_k . The energy spline is optimal in the sense of:

$$\inf_{\hat{f}(z)} \sup_{f(z)} |f(z) - \hat{f}(z)| \geq |f_u(z) - \hat{f}_0(z)|,$$

where $f_u(z)$ is the true signal and $\hat{f}(z)$ is the signal estimation.

Using the representation (6) for $\hat{u}(t)$, it is easy to show that the signal restoration by means of pseudoinversion is stable to small changes in the input information y . The value of deviation is:

$$\delta^2 = \int_Q |\hat{u}(t, y + \Delta y) - \hat{u}(t, y)|^2 dt = \lim_{\alpha \rightarrow 0} \frac{1}{|\Delta_\alpha|^2} \int_Q \begin{vmatrix} 0 & \bar{g}_1(t) & \dots & \bar{g}_m(t) \\ \dots & \dots & \dots & \dots \\ \Delta y_m & k_{m1} & \dots & k_{mm} + \alpha \end{vmatrix}^2 dt, \alpha > 0$$

and $\delta^2 \rightarrow 0$ at $\Delta y_i \rightarrow 0, i = 1, \dots, m$.

As a remark, we want to note that as a special case, the obtained results lead to Kotelnikov-Shannon theorem [Харкевич, 2007] (if there is no noises and $K = E$) and to various generalizations of this theorem for a finite number of counts. So, we can select as the counts not only the values of the function at the points z_k , but the values of its various derivatives either at these or other points that further should lead to a change of the vector $g(t)$.

Expansion by eigenfunctions of self-conjugate operator

When considering the pseudoinverse solution we don't use any information about the noise ν . Therefore, for the estimation (4)

$$\hat{u}(t) = G^{-1}f + G^{-1}\nu$$

the component $G^{-1}\nu$ can uncontrollably effect upon the estimation. One can specify a way to reduce this effect, so, let's consider the estimation of the solution of the system (3) in the form of expansion on some basis. By multiplying (3) by G^* we transform the equation (3) to the following form:

$$G^*y = G^*Gu + G^*\nu \quad (7)$$

Then let's find the eigenvalues and eigenfunctions of the operator G^*G from the following ratio:

$$G^*Ge_k(t) = \lambda_k e_k(t) \quad (8)$$

After denoting $Ge_k(t) = c_k$, we transform (8) to the form $GG^*c_k = \lambda_k Ge_k(t)$ or $(K - \lambda_k E)c_k = 0$.

Therefore, the problem of finding the eigenvalues and eigenfunctions of the operator G^*G is reduced to the problem of finding the eigenvalues and eigenvectors of the matrix K , which is Hermitian and exactly has m eigenvalues λ_k (taking multiplicity into account). Let's renumber them in nonincreasing order. The eigenvectors of the matrix K (the orthonormal system, $(c_j, c_k) = \delta_{jk}$) define the eigenfunctions of the operator G^*G by the following equality:

$$e_k(t) = \lambda_k^{-1} g^*(t)c_k, \quad k = 1, \dots, p \leq m, \quad \lambda_k \neq 0.$$

Clearly, that

$$(e_k(t), e_j(t)) = e_j^*(t)e_k(t) = (\lambda_k \lambda_j)^{-1} c_j^* G G^* c_k = (\lambda_k \lambda_j)^{-1} c_j^* K c_k = \lambda_j^{-1} c_j^* c_k = \begin{cases} 0, & j \neq k \\ \lambda_k^{-1}, & j = k \end{cases}$$

i.e. the system of eigenfunctions $e_k(t)$ is orthogonal. It is easy to make it orthonormal by multiplying it by a scalar $\lambda_k^{1/2}$:

$$\tilde{e}_k(t) = \lambda_k^{1/2} g^*(t)c_k = \lambda_k^{1/2} e_k(t).$$

Let's find the solution of the system (3) in the following form:

$$u(t) = \sum_{k=1}^p \gamma_k \tilde{e}_k(t) \quad (9)$$

Then, to find the coefficients γ_k we substitute (9) at (7):

$$G^*(y - \nu) = \sum_{k=1}^p \gamma_k G^*G \tilde{e}_k(t) = \sum_{k=1}^p \gamma_k \lambda_k \tilde{e}_k(t)$$

whence, it follows by multiplying on the scalarwise left:

$$\gamma_j = \lambda_j^{-1} (G^*(y - \nu), \tilde{e}_j(t)), \quad j = 1, \dots, p.$$

Then

$$u(t) = \sum_{k=1}^p \lambda_k^{-1} (y - \nu, G \tilde{e}_k(t)) \tilde{e}_k(t),$$

and the estimation of the solution of the equation (3) takes the following form:

$$\hat{u}(t) = \sum_{k=1}^p \lambda_k^{-1} (f, G\tilde{e}_k(t)) \tilde{e}_k(t) + \sum_{k=1}^p \lambda_k^{-1} (v, G\tilde{e}_k(t)) \tilde{e}_k(t) \quad (10)$$

The first summand in (10) determines the estimation of the signal spectrum in the absence of noise (the useful component), the second one represents the contribution to the solution of the noise component. Since λ_k are arranged in nonincreasing order $\lambda_1 \geq \dots \geq \lambda_p$, the noise component after achieving a sufficiently large index k begins to dominate over the useful component of the estimation of the solution.

The problem of choosing the value $p_1 \leq p$ can be solved as follows. Let's consider the normalized parameter:

$$\mu(k) = \left(\frac{\lambda_1^2 + \dots + \lambda_k^2}{\lambda_1^2 + \dots + \lambda_p^2} \right)^{1/2}, \quad 1 \leq k \leq p.$$

When approaching k to p , we have $\mu(k) \rightarrow 1$.

For some situations $\mu(k)$ is close to unity even for k significantly lower than p . The desired value p_1 can be defined by the minimum value k under which one can consider $\mu(k)$ sufficiently close to unity. The question of what is meant by sufficient proximity to unity is subjective and the answer to it should be determined by means of the computational experiment taking into account the specifics of the problem being solved.

Let's return to the estimation (10):

$$\hat{u}(t) = \sum_{k=1}^p \lambda_k^{-1} (y, G\tilde{e}_k(t)) \tilde{e}_k(t) = \sum_{k=1}^p (y, G e_k(t)) e_k(t) = \sum_{k=1}^p e_k(t) c_k^* y = \sum_{k=1}^p \lambda_k^{-1} G^* c_k c_k^* y.$$

Using the property of pseudoinverse solution of minimizing the square of the residual norm $\|Gu - y\|_{C^m}^2$ it is easy to show that:

$$G^- = \sum_{k=1}^p \lambda_k^{-1} G^* c_k c_k^* \quad (11)$$

Thus, defining G^- in the form (11) by selecting the appropriate value p makes it possible to reduce the impact of component G^-v at the estimation $\hat{u}(t)$.

Pareto-optimal estimations

It is possible to go into further details of the influence of the noise component on the restoring quality using the idea [Пытьев, 1989] about reduction of measurement up to the output from the specified device. Unlike to the above-mentioned approaches, we shouldn't look for estimations of solutions $\hat{u}(t)$ and should estimate $f(z)$ at once. Following the idea of Pitiev, we consider the system (2) as a linear scheme of measurement by devices with characteristic functions $g_k(t)$, $k = 1, \dots, m$, and we consider the data y_k , $k = 1, \dots, m$ as the output from these devices, which is distorted by noise. Then the problem of estimating the value of the functional $f(z)$ at the arbitrary point $z \in D$ can be considered as the problem of estimating the output from the device Π_z :

$$\Pi_z u = \int_Q \pi(t) u(t) dt$$

with the characteristic function $\pi(t) = g_z(t) = g(z, t)$ by the data (2). Let's look for the desired estimation at the class of estimations that are linear relatively the mentioned estimations, i.e. we should construct such vector

$B \subset C^m$ that $\hat{\Pi}_z u = By$. For that, we transform (3) to the following form:

$$By = \Pi_z u + (BG - \Pi_z)u + Bv,$$

which shows that the By is different from $\Pi_z u$ by two summands: the displacement $(BG - \Pi_z)u$ and the noise background Bv . Since

$$\|By - \Pi_z u\|^2 \leq \|BG - \Pi_z\|^2 \|u\|^2 + M\|Bv\|^2,$$

and $u(t)$ is unknown, then for finding the vector B we should solve the following two-criteria problem of minimization as follows:

$$\begin{cases} \varphi(B) = \|BG - \Pi_z\|^2 \rightarrow \min_B \\ h(B) = M\|Bv\|^2 \rightarrow \min_B \end{cases} \quad (12)$$

The first criterion (the operator residual), presented here characterizes the proximity of the characteristic function of the synthesized device BG , whose input should be adopted as the desired estimation $\hat{\Pi}_z u$, to the characteristic function $\pi(t)$, the second criterion characterizes the level of noise background (dispersion) of the desired estimation. Let's consider two-criteria decision of the problem in the sense of Pareto optimization [Ногин, 2005], i.e. we should look for such vector B that both criteria (12) are nondecreased at the same time. In view of convexity of the optimization criteria the problem (12) can be reduced [Глинкин, 1981] to a one-criterion problem of the following form:

$$\begin{aligned} \lambda \varphi(B) + (1 - \lambda)h(B) &\rightarrow \min_B, \lambda \in (0,1) \\ \varphi(B) &= \text{tr} \int_Q Bg(t) - \pi(t)(Bg(t) - \pi(t))^* dt, \\ h(B) &= \text{tr} B R B^*. \end{aligned}$$

By solving this problem we should obtain the continuum family of the desired estimations as follows:

$$\hat{\Pi}_z u = By = \Pi G^* (K + \alpha R)^{-1} y, \quad 0 < \alpha < \infty \quad (13)$$

or

$$\hat{f}(z) = \int_Q g(z, t) g^*(t) dt (K + \alpha R)^{-1} y, \quad 0 < \alpha < \infty \quad (14)$$

which are naturally named the Pareto-optimal estimations.

In a view of the solution of the problem (12) all the estimations (13) are equal.

Every estimation from (13) has both the nondecreased value of the operator residual:

$$\varphi(\alpha) = \text{tr} \Pi_z G^* ((K + \alpha R)^{-1} K (K + \alpha R)^{-1} - 2(K + \alpha R)^{-1}) (\Pi_z G^*)^*$$

and the level (dispersion) of noise background:

$$h(\alpha) = \text{tr} \Pi_z G^* (K + \alpha R)^{-1} R (K + \alpha R)^{-1} (\Pi G^*)^*.$$

Analyzing these functions shows that $h(\alpha)$ monotonically decreases and $\varphi(\alpha)$ monotonously increases with increasing α . Besides, $h'(\alpha)$ and $\varphi'(\alpha)$ are connected by the "conservation law"

$$\alpha h'(\alpha) + \varphi'(\alpha) = 0.$$

To select the specific estimation $\hat{\Pi}_z u$ from the family (13) one can use, for example, the principle of the

guaranteed result, according to which $\alpha = \underset{\alpha}{\operatorname{argmin}} \max_{(h,\varphi)}(h(\alpha), \varphi(\alpha))$. As a result of the monotonicity of the optimization criteria relative to α and because of opposite trends to change these criteria such value of the parameter is unique. To realize a certain compromise between the optimization criteria one can use other general principles of multi-criteria optimization, namely, the principle of uniform optimality $\alpha = \underset{\alpha}{\operatorname{argmin}}(\varphi(\alpha) + h(\alpha))$, the principle of reasonable compromise $\alpha = \underset{\alpha}{\operatorname{argmin}}(\varphi(\alpha)h(\alpha))$, "Eldorado" principle $\alpha = \underset{\alpha}{\operatorname{argmin}}(\varphi^2(\alpha) + h^2(\alpha))$.

In fact, choosing the specific estimation of the family (13) needs clearly understanding of what is preferable: to minimize the noise dispersion, which leads to increasing the operator residual generated by the data incompleteness, or contrariwise. Therefore, to select the appropriate value of the parameter α let's construct the Pareto set (parametric dependence $\varphi(h)$, $\alpha \in (0, \infty)$), where the decision-maker, being acquainted with the ratio between the noise level and the operator residual value should select a suitable point which is corresponding to one or another value of the parameter α .

As a remark, we want to note that the set of the Pareto-optimal estimations of the functions-signals (14) contains the estimation, obtained by the pseudoinversion tools (5), if the system is previously decorrelated by multiplying by $R^{-1/2}$, then it will coincide with the estimation (14) at $\alpha = +0$.

Maximally plausible estimations

Let's briefly dwell on another approach to restoring the function-signal $f(z)$, namely, the maximally plausible restoration. In that case, we should find the estimation $\hat{u}(t)$ as a function that maximizes the functional of plausibility.

Let's suppose the random noise vector is distributed by the normal law with the following probability density distribution:

$$p(v) = ((2\pi)^m \det R)^{-1/2} \exp(-1/2 v^* R^{-1} v),$$

and then we obtain the functional of plausibility as follows:

$$\Phi(u) = ((2\pi)^m \det R)^{-1/2} \exp\left\{-1/2 \left(y - \int_Q g(t)u(t)dt\right)^* R^{-1} \left(y - \int_Q g(t)u(t)dt\right)\right\}, \quad (15)$$

The problem of maximizing the functional (15) is equivalent to the problem of minimizing the following quadratic functional:

$$J(u) = \left(y - \int_Q g(t)u(t)dt\right)^* R^{-1} \left(y - \int_Q g(t)u(t)dt\right), \quad (16)$$

By calculating the variation of the functional (16) and equating it to zero, we obtain the Euler equation:

$$\int_Q g^*(t)R^{-1}g(\tau)u(\tau)d\tau = g^*R^{-1}y.$$

This is a Fredholm equation of the 1st kind. The problem of finding of its solutions belongs to a class of incorrect ones and requires regularization. For the Tikhonov regularizer:

$$T_\alpha(u) = J(u) + \alpha \int_Q |u(t)|^2 dt, \quad \alpha > 0$$

the Euler equation that determines the desired estimation $\hat{u}(t)$, is as follows:

$$\int_Q g^*(t)R^{-1}g(\tau)u(\tau)d\tau = g^*(t)R^{-1}y - \alpha u(t),$$

where $\alpha > 0$ is a regularization parameter, which implies that:

$$\hat{u}(t) = g^*(t)(k + \alpha R)^{-1}y, \quad \alpha > 0$$

and in that case, $\hat{f}(z)$ coincides with Pareto-optimal estimations (14).

Restoration under fuzzy data errors

Let's consider the case when there is no statistical information about the vector of errors. In this case, we can assume that the data y are distorted by fuzzy noise v . At the same time, a problem formulation should be added by information on distribution of the fuzzy element $v - \varphi^v(\cdot) : C^m \rightarrow L$ (it is assumed that an expert should specify this distribution). One believes that this distribution is such that a bigger noise corresponds to a smaller possibility, and this is a natural assumption for many measurement processes. In accordance with specifying the distribution of fuzzy value [Пытьев, 2000], the scale L is defined at the segment $[0,1]$ in natural order, specified by the inequality \leq and by two rules of composition: addition and multiplication (max and min, respectively).

In that case, to construct the optimal estimation let's use the results from the paper [Zavorotnyy, 2008]: we should find the estimation $\Pi_z u$ in the form of $B y$, where B is a solution of the problem of simultaneous minimizing a norm of the operator residual and maximizing the need of correctness of fuzzy value estimating, or the integral of necessity by $I(\Pi_z u, d(x))$, written in terms of the theory of possibilities.

$$\begin{cases} \|BG - \Pi_z\| \rightarrow \min_B \\ \theta \sup_x \sup_u \min(\varphi^y(x, u), \theta I(\Pi_z u, d(x))) \rightarrow \max_B \end{cases} \quad (17)$$

In (9) $I(\Pi_z u, d(x))$ is a possibility of lack of the error that occurs under selecting the estimating strategy $d(x)$ as a value $\Pi_z u$ for each value u , θ is an involution [Пытьев, 2000], $\varphi^y(z, u)$ is the joint distribution of the values y and u . Solving the problem leads to the following result: the solution of the problem (2) under fuzzy errors in the data y coincides with (4), and the corresponding estimation of the function-signal $f(z)$ is in the form of (5).

Conclusion

In this paper we constructed and thoroughly investigated (relatively the data), linear estimations of functions-signals on a finite set of data containing errors. The authors showed that all constructed estimations (namely, pseudoinverse, maximally plausible at normal distribution of the vector of noise, in the form of an expansion by eigenfunctions of self-conjugate operator) are contained in the set of Pareto-optimal estimations (14). One can control the effect of errors by selecting the parameter value $\alpha \in (0, \infty)$, the smaller α the smaller a noise background of the estimation. However, the decrease of α leads to increasing the operator residual, generated by data incompleteness, and the expert has to seek a suitable compromise between these two components which distort the desired estimation. The authors also showed the following fact: if there is no statistical information about noise and one can naturally consider the errors as fuzzy values, the desired estimation coincides with the Pareto-optimal estimation (14) on the predecorrelated data, which corresponds to $\alpha = +0$.

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

- [Zavorotnyy, 2008] Zavorotnyy A., Kasyanyuk V. Reduction measurements for calculation on fuzzy experiment scheme. // Decision making and business intelligence and techniques/ Supplement to international journal "Information, technologies and knowledge". – 2008. – Vol. 2, № 3. – P. 29-34.
- [Альберт, 1977] Альберт А.. Регрессия, псевдоинверсия и рекуррентное оценивание. – Москва: Наука, 1977. – 224 с.
- [Белов, 1986] Белов Ю.А., Касьянюк В.С. Об оптимальной оценке линейного функционала по значениям n других линейных функционалов в гильбертовом пространстве. // Доклады АН СССР. – 2000. – Т. 289, № 5. – С.1033-1035.
- [Глинкин, 1981] Глинкин И.А. Об одновременном поиске экстремумов нескольких функций. – Москва: Издательство Московского университета, 1981. – С. 46-54.
- [Ногин, 2005] Ногин В.Д. Принятие решений в многокритериальной среде. – Москва: ФИЗМАТЛИТ, 2005. – 176 с.
- [Пытьев, 1989] Пытьев Ю.П. Математические методы интерпретации эксперимента. – Москва: Высшая школа, 1989. – 315 с.
- [Пытьев, 2000] Пытьев Ю.П. Возможность. Элементы теории и применения. – Москва: Эдиториал УРСС, 2000. – 192 с.
- [Харкевич, 2007] Харкевич А.А. Спектры и анализ. – 4-е издание. – Москва: URSS: ЛКИ, 2007. – 89 с.

Authors' Information

Veda Kasyanyuk – PhD, Head of Scientific Sector, theoretical cybernetics sector, theoretical cybernetics department, faculty of cybernetics, Kyiv National Taras Shevchenko University, Glushkov av. 4-D, Kyiv-03127, Ukraine, e-mail: veda.sia@mail.ru

Iryna Volchyna – Engineer; theoretical cybernetics sector, theoretical cybernetics department, faculty of cybernetics, Kyiv National Taras Shevchenko University, Glushkov av. 4-D, Kyiv-03127, Ukraine, e-mail: volchyn@voliacable.com

ИССЛЕДОВАНИЕ МУРАВЬИНЫХ АЛГОРИТМОВ ОПТИМИЗАЦИИ В ЗАДАЧЕ КОММИВОЯЖЕРА

Юрий Зайченко, Николай Мурга

Abstract. *Ant colony optimization (ACO) algorithms are considered and discussed. New ACO algorithm for travelling salesman problem (TSP) is suggested. The experimental investigations of the suggested algorithm for solution of TSP were performed and presented in this paper.*

Keywords: *Ant colony optimization algorithm, travelling salesman problem, global and local optimization*

Введение

В последние годы для решения комбинаторных задач оптимизации все шире применяются алгоритмы вычислительного интеллекта. Эти алгоритмы обладают рядом несомненных достоинств: достаточны просты в реализации, гибкие в настройке, весьма эффективны и позволяют найти глобальное или близкие к нему решения за полиномиальное время. К числу таких алгоритмов относятся и так называемые муравьиные алгоритмы (Ant colony optimization algorithms), которые моделируют поведение колонии муравьев при решении общей задачи. Целью настоящей работы является исследование эффективности алгоритмов муравьиных колоний в решении важной практической задачи - коммивояжера. В работе описаны общие идеи известных муравьиных алгоритмов и предлагается специальный алгоритм для решения задачи коммивояжера. При водятся результаты его экспериментальных исследований и дается сравнительный анализ эффективности его различных модификаций.

Общая характеристика муравьиных алгоритмов оптимизации

Муравьи появились 100 млн. лет назад и их земная полная популяция насчитывает примерно 10^{16} индивидов. Произведенная оценка общего веса муравьев имеет тот же порядок величин, что и общий вес людей. Большинство муравьев – социальные животные, живущие в колониях общей численностью до 30 млн. особей [Deneubourg, 1990]. Комплексное поведение, возникающее в колониях муравьев, всегда интриговало человечество и было выполнено много исследований муравьиных колоний, направленных на лучшее понимание их коллективного поведения. Коллективное поведение, которое исследовалось, включало поиск продовольствия, разделение труда, организацию кладбищ и забота о потомстве, конструкция гнезд.

Метаэвристики оптимизации муравьиной колонии. Одним из первых типов поведений, который был исследован энтомологами, является способность муравьев находить кратчайшие пути между их гнездом и источником пищи. В результате исследований поведения муравьев при поиске пищи были разработаны алгоритмы. Они носят названия *мета-эвристик, оптимизации колонии муравьев (ACO-MH)* [Dorigo, 1999].

Цель исследования поведения колоний муравьев при обеспечении продовольствием заключалась в выяснении механизма, как муравьи отыскивают кратчайший путь между своим гнездом (муравейником) и источником пищи без каких-либо видимых активных механизмов координации.

Исследование поведения нескольких видов муравьев вскрыли первоначально случайную или хаотическую деятельность муравьев в поисках пищи. Но как только источник пищи обнаружен, образцы

поведения становятся более организованными и все больше и больше муравьев следует к пище тем же самым кратчайшим путем. Вскоре все почти муравьи почти магически следуют по одному и тому же кратчайшему пути. Механизмы регулирования отличаются для разных видов муравьев и могут проходить в форме непосредственно контакта или косвенно. Большинство разновидностей муравьев используют косвенную форму контакта через феромоновые следы. Когда муравей находит источник пищи, он тащит на себе часть пищи в муравейник и откладывает феромоны по следам.

Фуражные муравьи решают, по какому пути следовать на основе концентрации феромонов. *Пути с большей концентрацией феромонов имеют большую вероятность быть выбранными.* По мере того, как все больше муравьев следует по определенному пути, желательность этих путей подкрепляется большим числом феромонов, оставленных муравьями, которые привлекают все больше муравьев двигаться по этому пути. Коллективное поведение, которое здесь возникает, является формой так называемого «автокаталитического» поведения, в котором положительная ОС по пище приводит к тому, что все больше муравьев будут следовать по данному пути.

Поведение муравьев базируется на так называемом механизме «стигмергии».

Вообще говоря, стигмергия – это класс механизмов, которые служат медиаторами взаимодействия животное-животное. Индивиды наблюдают сигналы, которые вызывают специфические реакции или действия. Эти действия могут подкреплять или изменять сигналы, которые влияют на действия других индивидов. Было определено 2 формы стигмергии [Dorigo, 2000]: сематектоническая и на основе сигналов (сигнальная). Сематектоническая стигмергия относится к коммуникациям посредством изменений в физических характеристиках окружения. Примерами деятельности, которая осуществляется посредством сематектонической стигмергии являются: строительство общих жилищ, сортировка потомства. Стигмергия, основанная на сигналах, способствует коммуникациям через сигнальный механизм, который реализуется посредством химических компонент, откладываемых муравьями. Как пример, поведение поиска пищи появляется у муравьев, которые движутся по следам феромонов, отложенных другими муравьями.

Муравьиные алгоритмы – системы из популяций, создание которых основано на результатах наблюдения за колониями муравьев. Кооперация между индивидами в муравьиных алгоритмах достигается путем использования коммуникационных механизмов, которые наблюдаются в колониях живых муравьев.

Алгоритмическое моделирование поведения муравьев таким образом базируется на концепции искусственной стигмергии, которая определялась Дориго и Каро [Dorigo, 1999], как «косвенная» коммуникация, медиатором которой выступают многочисленные модификации состояний среды, которые доступны только локально коммуникационным муравьям.

Сущность моделирования поведения муравьев- это найти математическую модель, которая точно описывает стигмергетические характеристики соответствующих особей муравьев. В контексте поведения по заготовке пищи искусственные феромоны играют роль стигмергетических переменных.

Простой алгоритм оптимизации колонии муравьев

Первый муравьиный алгоритм, который был разработан, был моделью муравьиной системы, и с тех пор было разработано несколько его модификаций.

Рассмотрим сначала простой алгоритм оптимизации колонии муравьев SACO (Simple Ant Colony Optimization). SACO является алгоритмической реализацией эксперимента двойного моста Денинборга [Deneubourg, 1990], и используется в данном разделе для иллюстрации основных компонентов и поведения ACO-MH [Dorigo, 1997]

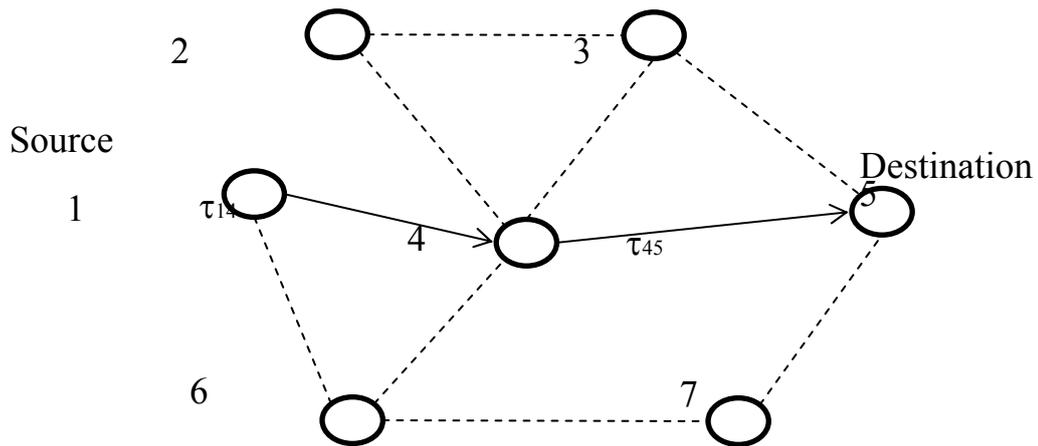


Рис. 1. График проблемы нахождения кратчайшего пути

Рассмотрим общую задачу нахождения кратчайшего пути между двумя узлами в графе $G = (V, E)$, где V - это множество вершин (узлов) и E является матрицей, представляющей соединения между узлами. Граф имеет $n_g = |V|$ узлов. Длина пути L^k , построенного муравьем, рассчитывается как количество переходов на пути от начального узла к узлу назначения. Пример графа и выбранного пути показаны на рисунке 1. Длина указанного маршрута равна 2 (хопа- перехода). Концентрация феромона τ_{ij} связанная с каждым ребром (i, j) графа.

В алгоритме SACO, каждому ребру присваивается небольшая случайная величина, чтобы указать начальную концентрацию феромона $\tau_{ij}(0)$. Строго говоря, ребра не имеют никакой концентрации феромонов на первом шаге. Муравей случайным образом выбирает, по какому ребру следовать дальше.

Для упрощения реализации алгоритма инициализируем концентрацию феромонов в каждом звене небольшой случайной величиной. Количество муравьев, размещенных в исходном узле $k = 1, \dots, n_k$. На каждой итерации SACO каждый муравей постепенно строит путь (решение), к узлу назначения. На каждом узле, каждый муравей принимает решение, чтобы определить следующее звено пути. Если муравей k в настоящее время находится в узле i , он выбирает следующий узел $j \in N_i^k$, основываясь на вероятности перехода,

$$p_{ij}^k(t) = \begin{cases} \frac{\tau_{ij}^\alpha(t)}{\sum_{j \in N_i^k} \tau_{ij}^\alpha(t)}, & \text{если } j \in N_i^k \\ 0, & \text{если } j \notin N_i^k \end{cases} \quad (1)$$

где N_i^k - множество возможных узлов, связанных с узлом i , относительно к муравью k . Если, для любого узла i и муравья k , $N_i^k = \emptyset$, то предыдущий узел для i включается в N_i^k . Заметим, что это может привести к возникновению петель в течение построения пути. Эти петли удаляются, как только узел назначения был достигнут.

В уравнении (1), α - положительная константа, которая используется для усиления влияния концентрации феромона. Большие значения α дают чрезмерное значение влияния феромона, особенно

на стадии инициализации начальных случайных концентраций феромонов, которые могут привести к быстрой сходимости к неоптимальным путям.

После того как все муравьи построили полный путь от начального узла к узлу назначения, а все петли были удалены, каждый муравей детерминировано проследивает путь к исходному узлу, и откладывает количество феромона согласно выражению (2)

$$\Delta \tau_{ij}^k(t) \propto \frac{1}{L^k(t)} \quad (2)$$

для каждого звена (i, j) , соответствующего пути; $L^k(t)$ - длина пути, построенного муравьем k в момент времени t . То есть,

$$\tau_{ij}(t+1) = \tau_{ij}(t) + \sum_{k=1}^{n_k} \Delta \tau_{ij}^k(t) \quad (3)$$

где n_k - количество муравьев.

Используя уравнение (3), общая интенсивность (т.е. концентрация) феромонов будет пропорциональна желательности путей, которые содержат звено, на основе длины соответствующих путей. Величина феромона $\Delta \tau_{ij}^k(t)$ рассчитывается по уравнению (3), и выражает качество соответствующего решения.

Для алгоритма SACO, качество решения (построенный путь) определяется как величина, обратная к длине пути (по числу переходов на пути). Может быть использована также любая другая мера, например, стоимость прохождения пути.

В общем, если $x^k(t)$ обозначает решение в момент времени t , то $f(x^k(t))$ выражает качество решения. Если $\Delta \tau^k$ не пропорционально качеству решения и все муравьи откладывают одинаковое количество феромонов (т.е. $\Delta \tau_{ij}^1 = \Delta \tau_{ij}^2 = \dots = \Delta \tau_{ij}^{n_k}$), то тогда это только так называемая «дифференциальная длина пути», которая смещает выбор в сторону кратчайшего пути - очень похоже на наблюдения Дениборга и его соавторов [Deneubourg, 1990]. Это обсуждение приводит к двум основным формам оценки решений, используемых в муравьиных алгоритмах, а именно:

- *неявные оценки*, где муравьи используют дифференциальную длину пути для эффекта смещения направления поиска других агентов, и
- *явные оценки*, где количество феромонов пропорционально некоторому показателю качества построенного решения.

Если количество хранящихся феромонов обратно пропорционально качеству решения, то чем больше $f(x^k(t))$ (то есть, наихудшее построенное решение), тем меньше $1/f(x^k(t))$, следовательно, меньше количество феромонов, откладываемого в звене..

Любое условие остановки может быть использовано в алгоритме (и для остальной части муравьиных алгоритмов рассмотренных ниже), например:

прекращаются, когда максимальное число итераций, n_t , была превышено;

прекращаются, когда приемлемое решение было найдено, и $f(x^k(t)) \leq \varepsilon$;

- прекращается, когда все муравьи (или большинство муравьев) следуют по тому же пути.

Первоначальные эксперименты с алгоритмом муравьиной колонии ACO обнаружили, что муравьи быстро приходят к решению, и мало времени проводят для исследования альтернативных путей. Чтобы

заставить муравьев проводить больше исследований и для предотвращения преждевременной сходимости, интенсивности феромона на звеньях разрешается "испаряться" в каждой итерации алгоритма, прежде, чем они увеличиваются на основе вновь построенных путей. Для каждого звена (i, j) , пусть испарение феромона происходит согласно

$$\tau_{ij}(t) \leftarrow (1 - \rho)\tau_{ij}(t) \quad (4)$$

где $\rho \in [0;1]$. Константа ρ , определяет скорость, с которой феромоны испаряются, заставляя муравьев "забыть" ранее принятые решения. Иными словами, ρ определяет степень влияния истории поиска. Для больших значений ρ , феромон испаряется быстро, в то время как малые значения ρ приводят к замедлению скорости испарения. Чем больше феромоны испаряются, тем более случайным становится поиск, что способствует улучшению разведки. Для $\rho = 1$, поиск совершенно случайный

Простой алгоритм ACO, рассмотренный выше, показал некоторый успех в поиске кратчайшего пути в графах. Эффективность алгоритма может быть значительно улучшена с помощью простого изменения в алгоритме. Эти изменения включают в себя, помимо эвристической информации для определения вероятности выбора звена, память для предотвращения циклов, а также различные правила обновления количества феромона путем использования локальной и / или глобальной информации об окружающей среде. В следующем разделе представлен более эффективный муравьиный алгоритм, который основан на таких изменениях простого алгоритма ACO.

Муравьиная система

Первый муравьиный алгоритм был разработан Дориго и называется системой муравьев (AS) [Dorigo, 1997]. Алгоритм AS улучшает SACO путем изменения вероятности переходов p_{ij}^k , которая использует эвристическую информацию, путем добавления памяти, за счет включения списка табу. В алгоритме AS, вероятность перехода от узла i к узлу j определяется как

$$p_{ij}^k(t) = \begin{cases} \frac{\tau_{ij}^\alpha(t)\eta_{ij}^\beta(t)}{\sum_{u \in N_i^k} \tau_{iu}^\alpha(t)\eta_{iu}^\beta(t)}, & \text{если } j \in N_i^k \\ 0, & \text{если } j \notin N_i^k \end{cases} \quad (5)$$

где τ_{ij} представляет апостериорную эффективность движения от узла i к узлу j , которая выражается через интенсивности феромона соответствующего звена, (i, j) ; η_{ij} представляет априорную эффективность перехода от i к j (то есть привлекательность, или желательности перехода), которая вычисляется с помощью некоторой эвристики.

Вероятность перехода в уравнении (4) отличается от SACO в уравнении (1)

по двум аспектам:

- Вероятность перехода, которая используется в алгоритме AS, является балансом между интенсивностью феромона (т. е. историей предыдущих успешных ходов), τ_{ij} , и эвристической информацией (которая выражает желательность движения), η_{ij} . Это эффективно уравновешивает компромисс между этапами « поиск-применение». Процесс поиска поощряет

действия, которые были найдены в прошлом и доказали свою эффективность, тем самым используя полученные знания о пространстве поиска.

Оптимальный баланс между поиском и использованием достигается за счет надлежащего выбора параметров α и β . Если $\alpha = 0$, информация о феромоне не используется, то есть пренебрегаем предыдущим опытом поиска. Затем поиск распадается на жадный стохастический поиск. Если $\beta = 0$, то привлекательностью (или потенциальной выгодой) ходов пренебрегаем и алгоритм поиска становится похожим на алгоритм SACO и связанные с ним проблемы.

Эвристическая информация добавляет явное смещение в сторону наиболее привлекательных решений, и, следовательно, проблемно-зависимых функций. Например, для задач где расстояние (или стоимость) пути d_{ij} должна быть сведена к минимуму, выбирают

$$\eta_{ij} = \frac{1}{d_{ij}};$$

- Набор N_i^k определяет множество допустимых узлов для муравья k , при его расположении на узле i . Множество допустимых узлов может включать в себя только ближайших соседей узла i . Кроме того, для предотвращения петель, N_i^k может включать в себя все узлы, которые муравей k еще не посетил. Для этого, как правило, ведется список табу для каждого муравья. Как только муравей посетил новый узел, этот узел будет добавлен в список табу муравья. Узлы в списка табу удаляются из N_i^k , убедившись, что узел был посещен более чем один раз.

Маниеззо и Корбонаро [Maniezzo, 1999] использовали другую формулу оценки вероятности для определения следующего узла:

$$p_{ij}^k(t) = \begin{cases} \frac{\alpha\tau_{ij}(t) + (1-\alpha)\eta_{ij}}{\sum_{u \in N_i^k(t)} (\alpha\tau_{iu}(t) + (1-\alpha)\eta_{iu}(t))}, & \text{если } j \in N_i^k \\ 0, & \text{если } j \notin N_i^k \end{cases} \quad (6)$$

Параметр α определяет относительную важность концентрации феромона $\tau_{ij}(t)$ относительно желательности $\eta_{ij}(t)$ звена (i, j) . Вероятность p_{ij}^k выражает компромисс между желательностью использования перехода (для малых α) и интенсивностью феромонов (концентрацией). Эта формула устраняет необходимость параметра β .

Испарение феромона реализовано, как в уравнении (4). После завершения пути каждым муравьем, феромон на каждом звене обновляется как

$$\tau_{ij}(t+1) = \tau_{ij}(t) + \Delta\tau_{ij}(t) \quad (7)$$

$$\Delta\tau_{ij}(t) = \sum_{k=1}^{n_k} \Delta\tau_{ij}^k(t),$$

где

$\Delta\tau_{ij}^k(t)$ - количество феромонов, которые хранит муравей k в звене (i, j) в момент времени t .

Были разработаны три варианта алгоритма AS, каждый из которых отличается тем, как рассчитывается $\Delta\tau_{ij}^k(t)$:

- Алгоритм муравьиный цикл AS(если задача минимизации):

$$\Delta \tau_{ij}^k(t) = \begin{cases} \frac{Q}{f(x^k(t))}, & \text{если звено } (i, j) \text{ встречается в пути } x^k(t) \\ 0, & \text{иначе} \end{cases} \quad (8)$$

Для реализации муравьиного цикла, вклад феромонов обратно пропорционален качеству $f(x^k(t))$ полного пути, построенного муравьем. Поэтому используется глобальная информация для обновления концентрации феромона, Q является положительной константой.

Алгоритм «Муравьиная плотность» AS:

$$\Delta \tau_{ij}^k(t) = \begin{cases} Q, & \text{если звено } (i, j) \text{ встречается в пути } x^k(t) \\ 0, & \text{иначе} \end{cases} \quad (9)$$

Каждый муравей откладывает одинаковое количество феромонов на каждом звене построенного пути. Такой подход существенно упрощает подсчет количества муравьев, которые следуют к звену (i, j) . Чем выше плотность (интенсивность) движения по звену, тем более желательно, чтобы звено оказалось составной частью окончательного решения.

- Алгоритм «Количество муравьев» AS:

$$\Delta \tau_{ij}^k(t) = \begin{cases} \frac{Q}{d_{ij}}, & \text{если звено } (i, j) \text{ встречается в пути } x^k(t) \\ 0, & \text{иначе} \end{cases} \quad (10)$$

В этом случае, для обновления концентрации феромона используется только локальная информация d_{ij} . Звенья с более низкой стоимостью становятся более желательными. Если d_{ij} представляет собой расстояние между узлами, то алгоритм AS «количество муравьев» предпочитает выбор кратчайших связей.

Во время инициализирующего шага, размещение муравьев диктуется задачей, которая решается. Если цель состоит в нахождении кратчайшего пути между исходным узлом и узлом назначения, то все n_k муравья размещены в исходном узле. С другой стороны, если целью является построение кратчайшего Гамильтонова пути (т.е. путь, который соединяет все узлы, только один раз), то n_k муравьев случайным образом распределены по всему графу. Размещая муравьев на случайно выбранных узлах, достигаем улучшения исследовательских свойств алгоритма поиска. Феромоны либо инициализируются константой τ_0 , либо небольшими случайными значениями в диапазоне $[0, \tau_0]$.

Применение муравьиного алгоритма к решению задачи коммивояжера

Проблема коммивояжера (Travelling Salesman Problem –TSP) первая задача, к которой был применен алгоритм муравьиной колонии (ACO) [Dorigo, 1997]. Эта задача является комбинаторной NP- трудной и наиболее часто используемой задачей в литературе по тематике ACO.

Постановка задачи.

Задано множество из n_{π} городов, требуется найти цикл минимальной длины, проходящий через все города по разу. (Гамильтонов цикл). Пусть π представляет решение как перестановку из городов $\{1, 2, \dots, n_{\pi}\}$, в которой $\pi(i)$ номер города, посещенного на i -м шаге. Тогда $\Pi(n_{\pi})$ – множество всех перестановок из

$\{1, 2, \dots, n_\pi\}$ т.е. пространство поиска. Формально, задача TSP определяется как нахождение оптимальной перестановки π^* , такой, что:

$$\pi^* = \operatorname{argmin}_{\pi \in \Pi(n_\pi)} f(\pi)$$

где $f(\pi) = \sum_{i,j=1}^{n_\pi} d_{ij}$.

d_{ij} – расстояние (или стоимость проезда) между городами i и j . Обозначим через $D = \|d_{ij}\|_{i,j=1..n_\pi}$ матрицу расстояний.

Определены две версии задачи коммивояжера в зависимости от матрицы D :

Если $d_{ij} = d_{ji}$, для всех $i, j = 1..n_\pi$ - симметричная задача TSP;

Если $d_{ij} \neq d_{ji}$ - асимметричная задача TSP.

Граф, представляющий задачу, задается тройкой $G = (V, E, D)$ где $V = [v_i]$ - множество узлов (городов) $E = \{(i, j)\}$ - множество связей, (путей) между городами; D – матрица расстояний. Решение представляет собой упорядоченную последовательность $\pi = (i_1, i_2, \dots, i_{n_\pi})$ - которая указывает порядок посещения городов.

Эвристическая желательность посещения города j после города i вычисляется как

$$n_{ij}(t) = \frac{1}{d_{ij}(t)}$$

Указание шага времени t включено в формулу для решения динамических задач, в которых расстояния могут изменяться с течением времени.

Задача TSP включает 2 ограничения:

Все города должны быть посещены;

Каждый город должен быть посещен однажды.

Чтобы обеспечить ограничение (2), создается список табу для каждого частичного решения, содержащий все города, которые мы уже посетили. Пусть Γ_k обозначает список табу k -го муравья. Тогда

$N_i^k(t) = V \setminus \Gamma_k$ – множество городов, еще не посещенных, после достижения города i . и ограничение удовлетворяется введением требования, что каждое решение содержит ровно n городов и соответственно список табу.

Описание муравьиного алгоритма для решения задачи коммивояжера

Рассмотрим предлагаемый муравьиный алгоритм для решения задачи коммивояжера.

Первоначально рассмотрим случай, когда рёбра графа для задачи коммивояжера неориентированные, то есть стоимости прямых и обратных путей из двух соседних вершин одинаковые.

Предусловия алгоритма

Каждая вершина графа содержит три множества: множество муравьёв для дальнейшего передвижения – FD , множество муравьёв для возвращения в гнездо после успешного обхода всего маршрута без тупиков и петель – BK и множество муравьёв, которым не удалось обойти маршрут без тупиков и петель – BD .

Находясь в конкретной вершине, муравью необходимо принять решение, по какой дуге переходить к следующей вершине. Пусть n – общее количество дуг, исходящих из данной вершины и идущих в вершины, которые ещё не посещал муравей. Тогда для принятия решения муравью предлагается следующий показатель:

$$D_i = c_1 \cdot \left(1 - \frac{L_i}{\sum_{j=1}^n L_j} \right) + c_2 \cdot \frac{\tau_i}{\sum_{j=1}^n \tau_j} + c_3 \cdot U(0,1) \quad (11)$$

где $i = \overrightarrow{1, \dots, n}$ – порядковый номер дуги, обладающей обозначенным свойством, L_i – стоимость перехода муравья по данной дуге (либо длина пути), τ_i – концентрация феромона на данной дуге, $U(0,1)$ – генератор случайных чисел, равномерно распределённых на отрезке $[0;1]$.

Величины c_1 , c_2 и c_3 должны зависеть от номера текущей итерации t и в данном алгоритме их предлагается использовать в форме (12), (13) и (14).

$$c_1 = a_1 \frac{M - t}{M} \quad (12)$$

$$c_2 = \frac{t}{M} \quad (13)$$

$$c_3 = a_2 \frac{M - t}{M} \quad (14)$$

В данных формулах использованы следующие обозначения: M – некоторая константа, которая, к тому же, удовлетворяет условию $M > t$. a_1 и a_2 – константы удовлетворяющие условиям $a_1 \geq 0, a_2 \geq 0, a_1 + a_2 = 1$. Не трудно заметить, что $c_1 + c_2 + c_3 = 1$.

Муравей переходит на следующую вершину с самым большим значением показателя D_i .

Последним предусловием алгоритма является помещение муравьёв во все вершины графа путей. Это эквивалентно их помещению во множества FD каждой из вершин графа.

Описание алгоритма

1. Каждый муравей из множества FD в каждой вершине ищет такие рёбра, которые ведут к вершинам, которые он ещё не посещал. Если поиск завершается успешно, то – переход на шаг 3, иначе – на шаг 2.
2. В случае, если уже пройдены все вершины, то муравей перемещается в множество BK данной вершины, иначе – в множество BD .
3. Для дуг, ведущих к вершинам, которые муравей ещё не посещал, рассчитывается показатель D_i из формулы (1). Для дуги с наибольшим значением показателя количество феромона увеличивается на 1 и муравей перемещается в множество FD вершины, на втором конце дуги.

4. Все муравьи из множества BK всех вершин перемещаются в множества BK предыдущих посещённых вершин, при этом увеличивая количество феромона в дугах, по которым они переходят, на 1. Если вершина, в которую должен перейти муравей – его муравейник, то он помещается в её множество FD .
5. Все муравьи из множества BD всех вершин перемещаются в множества BD предыдущих посещённых вершин, при этом уменьшая количество феромона в дугах, по которым они переходят, на 1. Если вершина, в которую должен перейти муравей – его муравейник, то он помещается в её множество FD .
6. Проверка условия превышения допустимого числа итераций. Если число итераций превышено – завершение работы алгоритма, иначе – на шаг 1.

Примечание 1. Как следует из алгоритма, каждый муравей должен запоминать последовательность посещённых им вершин и, при обратном ходе, удалять вершины с конца данной последовательности.

Примечание 2. В случае ориентированных рёбер алгоритм не меняется.

Описание экспериментов.

В процессе экспериментов исследовались три варианта алгоритма:

- 1) При одинаковом изменении феромона во всех ребрах на 1;
- 2) При локальном изменении феромона на ребрах обратно пропорционально длине (стоимости) ребра;
- 3) При глобальном изменении феромона во всех ребрах (обратно пропорционально длине (стоимости) пути).

В экспериментах варьировалось число муравьёв, размещаемых в узлах сети: 1-5, 10, 30, 50. Первоначально муравьи размещались во всех узлах сети. В ходе экспериментов генерировались различные топологии сети, а также варьировалось число итераций и фиксировалось число итераций, при которых достигался оптимальный цикл.

Некоторые из результатов экспериментов приводятся ниже, в приведенных результатах в таблицах цифра 0 означает, что не был найден ни один цикл, цифра 1- что найден неоптимальный путь, 2- что найден оптимальный путь (цикл).

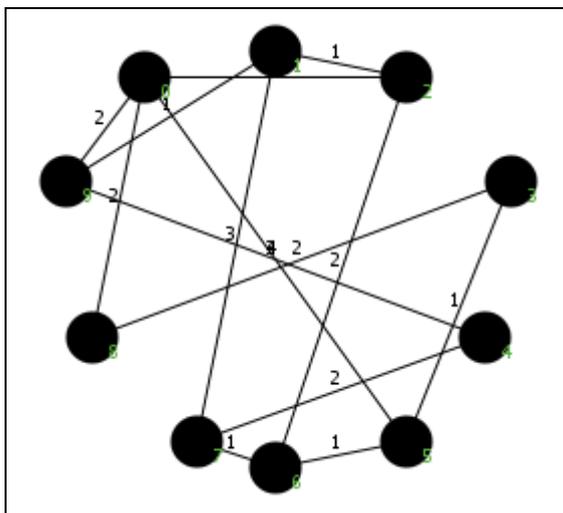


Рис. 1. Топология №7 сети для задачи коммивояжёра

Таблица 4. Подход к поиску оптимального пути с сохранением лучшего найденного пути

		Число муравьёв в узлах перед запуском алгоритма							
		1	2	3	4	5	10	30	50
Номер итерации	5	2	2	2	2	2	2	2	2
	50	2	2	2	2	2	2	2	2
	100	2	2	2	2	2	2	2	2
	250	2	2	2	2	2	2	2	2
	500	2	2	2	2	2	2	2	2
	1000	2	2	2	2	2	2	2	2
	1500	2	2	2	2	2	2	2	2

Эксперименты с использованием тактики поиска оптимального пути с приростом феромона на дугах, пропорциональным длине найденного замкнутого маршрута, состоящего из данных дуг

Таблица 5. Подход к поиску оптимального пути с ориентированием на наибольшую концентрацию феромона

		Число муравьёв в узлах перед запуском алгоритма							
		1	2	3	4	5	10	30	50
Номер итерации	5	0	0	0	0	2	2	2	2
	50	0	0	2	2	2	2	2	2
	100	0	2	2	2	2	2	2	2
	250	0	2	2	2	2	2	2	2
	500	0	2	2	2	2	2	2	2
	1000	0	2	2	2	2	2	2	2
	1500	0	2	2	2	2	2	2	2

Таблица 6. Подход к поиску оптимального пути с сохранением лучшего найденного пути

		Число муравьёв в узлах перед запуском алгоритма							
		1	2	3	4	5	10	30	50
Номер итерации	5	2	2	2	2	2	2	2	2
	50	2	2	2	2	2	2	2	2
	100	2	2	2	2	2	2	2	2
	250	2	2	2	2	2	2	2	2
	500	2	2	2	2	2	2	2	2
	1000	2	2	2	2	2	2	2	2
	1500	2	2	2	2	2	2	2	2

Результаты экспериментов с использованием вероятностной стратегии поиска оптимального маршрута

Эксперименты с использованием тактики поиска оптимального пути с приростом феромона на дугах, пропорциональным длине дуг

Таблица 7. Подход к поиску оптимального пути с ориентированием на наибольшую концентрацию феромона

		Число муравьёв в узлах перед запуском алгоритма							
		1	2	3	4	5	10	30	50
Номер итерации	5	0	2	0	0	2	0	1	0
	50	2	2	1	2	2	2	2	2
	100	1	2	1	2	2	2	2	2
	250	1	2	1	2	2	2	2	2
	500	1	2	1	2	2	2	2	2
	1000	1	2	1	2	2	2	2	2
	1500	1	2	1	2	2	2	2	2

Таблица 8. Подход к поиску оптимального пути с сохранением лучшего найденного пути

		Число муравьёв в узлах перед запуском алгоритма							
		1	2	3	4	5	10	30	50
Номер итерации	5	0	2	1	2	2	1	2	2
	50	2	2	2	2	2	2	2	2
	100	2	2	2	2	2	2	2	2
	250	2	2	2	2	2	2	2	2
	500	2	2	2	2	2	2	2	2
	1000	2	2	2	2	2	2	2	2
	1500	2	2	2	2	2	2	2	2

Эксперименты с использованием тактики поиска оптимального пути с приростом феромона на дугах, пропорциональным длине найденного замкнутого маршрута, состоящего из данных дуг

Таблица 9. Подход к поиску оптимального пути с ориентированием на наибольшую концентрацию феромона

		Число муравьёв в узлах перед запуском алгоритма							
		1	2	3	4	5	10	30	50
Номер итерации	5	0	0	0	0	0	0	0	1
	50	0	0	2	2	2	1	2	2
	100	0	0	2	2	2	1	2	2
	250	0	2	2	2	2	1	2	2
	500	0	2	2	2	2	1	2	2
	1000	0	2	2	2	2	1	2	2
	1500	0	2	2	2	2	1	2	2

Таблица 10. Подход к поиску оптимального пути с сохранением лучшего найденного пути

		Число муравьёв в узлах перед запуском алгоритма							
		1	2	3	4	5	10	30	50
Номер итерации	5	0	0	2	2	2	2	2	2
	50	2	2	2	2	2	2	2	2
	100	2	2	2	2	2	2	2	2
	250	2	2	2	2	2	2	2	2
	500	2	2	2	2	2	2	2	2
	1000	2	2	2	2	2	2	2	2
	1500	2	2	2	2	2	2	2	2

Анализ полученных результатов

На топологии, изображённой на Рис. 1, проводились экспериментальные исследования по применению разновидностей муравьиного алгоритма для решения задачи коммивояжёра. Для восстановления оптимального пути после использования алгоритмов использовалось два подхода: подход, сохраняющий лучший путь, полученный в течении работы алгоритма в памяти (при нахождении более короткого замкнутого маршрута, он помещался в память вместо предыдущего лучшего пути) и подход, основывающийся на запуске муравья из любого узла, который ориентируется только на наличие феромона в рёбрах (выбранные муравьём рёбра окрашиваются). Кроме того, в алгоритме исследовалось две стратегии выбора дуги для перехода муравья из текущей вершины графа: детерминированная (основывается на значении показателя принятия решения про переход) и случайная (в которой наличие феромона в смежных с исходной вершиной рёбрах задают вероятности перехода по ним, и дуга для перехода выбирается на основании этих вероятностей). Исследованию также подвергались и три тактики приращения феромона в рёбрах для детерминированной стратегии (константный прирост, прирост обратно пропорциональный длине дуги и прирост обратно пропорциональный длине замкнутого маршрута, который проходит через все вершины и который был пройден муравьём) и две для вероятностной стратегии (без константного прироста).

Из результатов экспериментов можно сделать *следующие выводы*:

Подход, основанный на сохранении лучшего пути в ходе работы алгоритма, оказался более эффективным.

Для детерминированной и вероятностной стратегий все тактики поиска, использующие подход, основанный на сохранении лучшего пути в ходе работы алгоритма, показали одинаковую высокую эффективность. Для подхода к восстановлению пути на основании прогона муравья, наиболее эффективной оказалась локальная тактика поиска для обеих стратегий.

Использование детерминированной стратегии поиска позволило найти оптимальный путь значительно быстрее по сравнению с вероятностной стратегией.

В целом, более эффективной оказалась детерминированная стратегия поиска. Это объясняется, на наш взгляд особенностями исследованных топологий: слабосвязанные и небольшой размерности. При переходе к большим размерностям сетей и большей степени связанности есть основания ожидать, что вероятностная стратегия окажется более предпочтительной.

Использование подхода к нахождению оптимального пути с его сохранением в процессе работы алгоритма позволяет его найти за незначительное число итераций и имея небольшое количество муравьёв в вершинах (менее 5 итераций и менее 5 муравьёв в каждой вершине).

Заключение

В работе рассматриваются муравьиные алгоритмы оптимизации и дается их анализ. Описан предлагаемый муравьиный алгоритм оптимизации для задачи коммивояжера. Проведены его экспериментальные исследования, в ходе которых изменялась стратегия откладывания феромона: были рассмотрены и исследованы 3 стратегии: глобальная стратегия, локальная стратегия и постоянная величина феромона., а также изменялось число муравьев в узлах сети. Кроме того, применялись два типа стратегий выбора перехода: детерминированная и вероятностная. В результате проведенных экспериментов были получены следующие выводы.

1. Увеличение числа муравьев в узлах сети от одного до 2, 3 -5, позволяет улучшить характеристики алгоритма и достичь минимума при меньшем числе итераций. При дальнейшем увеличении числа муравьев до 10, 30, 50 не дает эффекта.

Подход, основанный на сохранение лучшего пути в ходе работы алгоритма, оказался более эффективным.

Тактика прироста феромона обратно пропорционально длине дуг (локальная тактика) оказалась более эффективной в сравнении с глобальной тактикой и константным приростом феромона..

2. Использование детерминированной стратегии поиска позволило найти оптимальный путь значительно быстрее по сравнению с вероятностной стратегией. Это объясняется, на наш взгляд особенностями исследованных топологий: слабосвязанные и небольшой размерности.

Bibliography

- [Deneubourg,1990] J-L. Deneubourg, S. Aron, S. Goss, and J-M. Pasteels. The Self-Organizing Exploratory Pattern of the Argentine Ant. *Journal of Insect Behavior*, 3:159–168, 1990
- [Dorigo, 1997] M. Dorigo and L.M. Gambardella. Ant Colony System: A Cooperative Learning Approach to the Traveling Salesman Problem. *IEEE Transactions on Evolutionary Computation*, 1(1):53–66, 1997.
- [Dorigo, 1999] M. Dorigo and G. Di Caro. Ant Colony Optimization: A New Meta-Heuristic. In *Proceedings of the IEEE Congress on Evolutionary Computation*, volume 2, page 1477, July 1999.
- [Dorigo, 2000] M. Dorigo, E. Bonabeau, and G. Theraulaz. Ant Algorithms and Stigmergy. *Future Generation Computer Systems*, 16(9):851–871, 2000.
- [Maniezzo, 1999] V. Maniezzo and A Carbonaro. Ant Colony Optimization: An Overview. In C. Ribeiro, editor, *Essays and Surveys in Metaheuristics*, pages 21–44. Kluwer, 1999.

Acknowledgement

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

Authors' Information

Юрий Зайченко– доктор технических наук, профессор. Институт прикладного системного анализа НТУУ «КПИ», 03056, Киев-56, Украина phone: 38044 -4068393, e-mail: baskervil@voliacable.com

МЕТОД ПОИСКА РЕШЕНИЙ В ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМАХ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ НА ОСНОВЕ ПРЕЦЕДЕНТОВ

Павел Варшавский, Роман Алехин

Аннотация: Рассматриваются методы поиска решений на основе прецедентов и аналогий в интеллектуальных системах поддержки принятия решений. Исследуется возможность представления прецедентов на основе онтологии предметной области. Предлагается метод поиска решения на основе прецедентов с использованием онтологического подхода и структурной аналогии.

Ключевые слова: интеллектуальные системы поддержки принятия решений, правдоподобные рассуждения, прецеденты, аналогия, онтологический подход.

ACM Classification Keywords: H.4.2 [Information systems applications]: Types of systems – Decision support; I.2.3 [Artificial intelligence]: Deduction and Theorem Proving – Uncertainty, "fuzzy," and probabilistic reasoning; I.2.6 [Artificial intelligence]: Learning – Analogies.

Введение

Одним из подходов к решению задачи моделирования человеческих рассуждений (рассуждений «здорового смысла») [Поспелов, 1989] в интеллектуальных системах (ИС) является использование аппарата нетрадиционных логик [Вагин и др., 2008] – индуктивных, абдуктивных, нечетких, а также методов рассуждений на основе аналогий и прецедентов.

Методы рассуждения на основе прецедентов (CBR – Case-Based Reasoning) и CBR-системы успешно используются в различных областях человеческой деятельности (медицина, техника, юриспруденция и др.), а также прецедентный подход активно применяется в динамических ИС, в системах экспертного диагностирования, в ИС поддержки принятия решений (ИСППР), системах машинного обучения, при решении задач прогнозирования, обобщения накопленного опыта, поиска решения в малоизученных предметных областях и др.

Рассуждения на основе прецедентов основываются на накоплении опыта и последующей адаптации решения известной задачи к решению новой. Прецедентный подход позволяет упростить процесс принятия решений в условиях временных ограничений и при наличии различного рода неопределенности в исходных данных и экспертных знаниях, а также в случае возникновения различных нестандартных (аномальных) ситуаций.

Прецедентный подход

Прецедент определяется как случай, имевший место ранее и служащий примером или оправданием для последующих случаев подобного рода.

Методы рассуждения на основе прецедентов включают в себя четыре основных этапа, образующие так называемый CBR-цикл (рис. 1) [Aamodt et al., 1994; Варшавский и др., 2006]:

- **Retrieve** - извлечение наиболее соответствующего (подобного) прецедента (или прецедентов) для сложившейся ситуации из библиотеки прецедентов (БП);

- **Reuse** - повторное использование извлеченного прецедента для попытки решения текущей проблемы (задачи);
- **Revise** - пересмотр и адаптация в случае необходимости полученного решения в соответствии с текущей проблемой (задачей);
- **Retain** - сохранение вновь принятого решения как части нового прецедента.

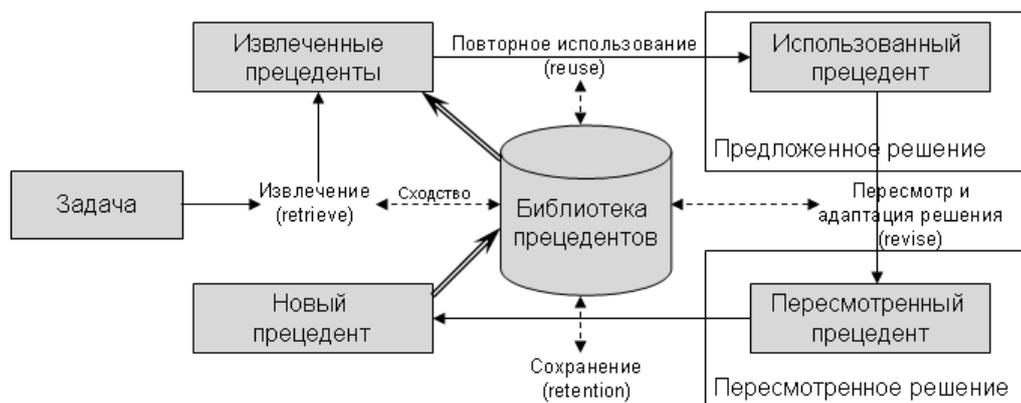


Рис. 1. CBR-цикл

Последние два этапа в CBR-цикле могут исключаться и выполняться экспертом или лицом, принимающим решения (ЛПР). Это связано с необходимостью при формировании БП использовать только достоверную информацию или информацию, подтвержденную экспертом. Таким образом, можно минимизировать количество прецедентов в БП CBR-системы и повысить степень их достоверности.

В общем случае модель представления прецедента включает описание ситуации и решение для данной ситуации [Еремеев и др., 2012]: $CASE = (Situation, Solution, Result)$, где *Situation* – ситуация, описывающая данный прецедент, а *Solution* – решение (например, диагноз и рекомендации ЛПР), *Result* – результат применения решения, который может включать список выполненных действий, дополнительные комментарии и ссылки на другие прецеденты, а также в некоторых случаях может приводиться обоснование выбора данного решения и возможные альтернативы. Различия способов представления прецедентов заключаются в разных способах описания указанных компонент.

Прецеденты могут быть представлены в виде записей в БД, концептуальных графов, семантической сети, древовидных структур, предикатов, фреймов, рисунков и мультимедийной информации.

В большинстве случаев для представления прецедентов достаточно простого параметрического представления в виде набора параметров с конкретными значениями и решения [Варшавский, 2008]: $CASE = (x_1, x_2, \dots, x_n, R)$, где $x_1 \dots x_n$ – параметры ситуации, описывающей данный прецедент ($x_1 \in X_1, x_2 \in X_2, \dots, x_n \in X_n$), n – количество параметров, R – решение (диагноз и рекомендации ЛПР), а X_1, \dots, X_n – области допустимых значений соответствующих параметров.

В некоторых случаях такого представления бывает недостаточно, так как имеются ограничения, связанные с выразительными возможностями параметрической модели представления прецедентов. При параметрическом представлении трудно обеспечить учет зависимости между параметрами прецедента (например, временные зависимости или причинно-следственные).

Одним из возможных способов решения этой проблемы является представление прецедентов с использованием онтологии предметной области [Варшавский и др., 2012].

Среди специалистов, занимающихся проблемами компьютерной лингвистики, наиболее устоявшимся (классическим) считается определение онтологии, данное Губертом: «онтология – это спецификация концептуализации».

Определение онтологии как формального представления предметной области (ПО), построенного на базе концептуализации, предполагает выделение ее трех взаимосвязанных компонентов: таксономии терминов, описаний смысла терминов, а также правил их использования и обработки. Таким образом, модель онтологии O задает тройка [Гаврилова и др., 2000]: $O = (X, R, \Phi)$, где X – конечное множество концептов (понятий, терминов) ПО, которую представляет онтология; R – конечное множество отношений между концептами; Φ – конечное множество функций интерпретации, заданных на концептах и (или) отношениях.

Выбор онтологии для представления прецедентов обусловлен рядом важных достоинств, отличающих ее от других моделей представления знаний. Использование онтологии для представления прецедентов позволяет задать сложную структуру прецедента, включающую данные разных типов, и обеспечить естественность представления структурированных знаний и достаточно простое их обновление в относительно однородной среде. Последнее свойство особенно важно для ИСППР, ориентированных на открытые и динамические ПО.

Онтология содержит знания по ПО, которые используются для поддержки СВР-цикла, а также онтология задает структуру прецедента и обеспечивает его хранение.

Знания о ПО и модель прецедентов описываются в виде иерархии концептов онтологии, а каждый прецедент из БП в виде иерархии экземпляров концептов, связанных отношениями языка описания онтологий для Semantic Web (OWL) [Алехин, 2011].

Метод аналогии и прецедентов на основе теории структурного отображения

Теория структурного отображения (SMT – structure mapping theory) позволяет формализовать некоторый набор неявных ограничений, которыми пользуется человек, оперируя такими понятиями, как аналогия и подобие [Falkenhainer et al, 1989]. Согласно SMT предполагается, что аналогия является отображением знаний одной области (базы) в другую область (цель), базирующимся на системе отношений, которые имеются между объектами базовой области и объектами целевой области, а также, что человек (ЛПР) предпочитает оперировать некоторой целостной системой взаимосвязанных глубинных отношений, а не простым набором поверхностных и слабо связанных фактов.

Процесс вывода на основе аналогий согласно SMT включает следующие этапы:

1. **Определение потенциальных аналогов.** Имея целевую ситуацию (цель), определить другую ситуацию (базу), которая является аналогичной или подобной целевой ситуации;
2. **Отображение и вывод.** Построить отображение, состоящее из соответствий между базой и целью. Это отображение может включать дополнительные знания (факты) о базе, которые могут быть перенесены в цель.
3. **Оценка "качества" соответствия.** Оценить полученное соответствие, используя такие структурные критерии, как число подобий и различий, степень структурного соответствия, количество и тип новых знаний, полученных по аналогии из кандидатов заключений.

Рассмотрим механизм структурного отображения (SME – structure mapping engine), базирующийся на SMT [Варшавский и др., 2005]. Этот механизм предназначен для моделирования поиска решений на основе аналогий и прецедентов, и позволяет сформировать наиболее общие соответствия (*Gmaps*) для

структурированных представлений базовой и целевой ПО, а также обеспечивает оценку полученных соответствий.

Входными данными для алгоритма SME являются структурные представления базовой и целевой ПО.

Алгоритм SME реализует следующие четыре этапа:

1. Построение локальных соответствий (гипотез соответствия *MHs*). Определить соответствия между элементами в базовой и целевой ПО с помощью правил:

- если два отношения имеют одинаковое имя, тогда создается гипотеза соответствия;
- для каждой гипотезы соответствия между отношениями проверяется их соответствующие аргументы: если их количество и тип совпадает, тогда создается гипотеза соответствия между ними.

Далее определить оценку правдоподобия локальных соответствий, используя следующие правила и коэффициенты, задаваемые экспертом (CF_1 – величина, на которую необходимо увеличить оценку правдоподобия *MH*, если имена базового и целевого элементов совпадают; CF_2 – величина, на которую необходимо увеличить оценку правдоподобия *MH*, если хотя бы базовый элемент имеет родительское отношение – отношение более высокого уровня):

- увеличить оценку правдоподобия для соответствия на CF_1 , если базовый и целевой элементы имеют одинаковые имена;
- увеличить оценку правдоподобия для соответствия на CF_2 , если хотя бы у базового элемента имеется родительское отношение.

2. Построение глобальных соответствий (*Gmaps*). Сформировать системы соответствий, которые используют совместимые пары объектов (1:1), называемые *Emaps*.
3. Построение кандидатов заключения (*Inferences*). С каждым *Gmap* связать множество (возможно пустое) кандидатов заключения *Inferences* - факты, которые присутствуют в базовой ПО, но не присутствуют изначально в соответствующей целевой ПО;
4. Оценка глобальных соответствий *Gmaps* (*SES*). Получить оценку для *Gmaps*, которая зависит от оценок правдоподобия локальных соответствий *MHs*.

Таким образом, в результате выполнения алгоритма формируются наиболее общие соответствия *Gmap*, включающие следующие составляющие:

- соответствия – множество парных соответствий между базовой и целевой ПО;
- кандидаты заключения – множество новых фактов, которые предположительно могут содержаться в целевой области;
- оценка глобальных соответствий – числовая оценка качества *Gmap*.

Метод поиска решений в ИСППР на основе прецедентов

Существует целый ряд методов поиска решений на основе прецедентов (извлечения прецедентов) и их модификаций [Варшавский и др., 2009], например: метод ближайшего соседа, метод извлечения прецедентов на основе деревьев решений, метод извлечения прецедентов на основе знаний, метод извлечения с учетом применимости прецедента.

Предлагается осуществлять извлечение и определение сходства прецедента *S* и текущей ситуации *Q* в два этапа:

- определение сходства прецедента с текущей ситуацией на основе онтологии предметной области и формирование парных соответствий с помощью алгоритма на базе SMT;

- определение сходства прецедента и текущей ситуации по методу ближайшего соседа с учетом полученных парных соответствий.

На первом этапе сравниваются по структуре описания ситуации прецедента и текущей ситуации. Цель данного этапа – определить возможные парные соответствия между прецедентом и текущей ситуацией и оценить их сходство.

Множество парных соответствий формируется в результате выполнения процедуры *Pairs*:

Входные параметры: Q – текущая ситуация, C – прецедент, O – онтология ПО, CF_1 и CF_2 – величины из алгоритма SME.

Выходные параметры: F – множество парных соответствий, LS – множество оценок правдоподобия.

Промежуточные параметры: i, j – параметры циклов, $Ftemp$ – параметр для временного хранения множества парных соответствий, $Frec$ – результат рекурсивного вызова процедуры.

Шаг 1. $i = 0, Frec = \emptyset, Ftemp = \emptyset$.

Шаг 2. Если нет непроверенных концептов, связанных с Q , то $F = F \cup Ftemp$ и перейти к шагу 6. Иначе $i = i + 1$, выбираем непроверенный концепт Q_i связанный с Q и переходим к шагу 3.

Шаг 3. Поиск концепта C_j связанного с C и удовлетворяющего условию – имя концепта C_j совпадает с Q_i . Если такой C_j найден, то увеличиваем оценку правдоподобия LS_j на CF_1 и формируем соответствие $F_j = \langle Q_i, C_j \rangle$, выполняем рекурсивный вызов процедуры *Pairs* ($Frec = Pairs(Q_i, C_j, O)$), производим добавление полученных парных соответствий во временное множество $Ftemp$ ($Ftemp = Frec \times F_j$, где \times – декартово произведение множеств) и переходим к шагу 2, иначе переходим к шагу 4.

Шаг 4. Поиск концептов C_j связанных с C и удовлетворяющих условию – концепт C_j подобен Q_i (подобие между концептами онтологии будем определять руководствуясь следующим принципом: концепты A и B подобны, если родительские концепты A совпадают с родительскими концептами B). Переходим к шагу 5.

Шаг 5. Пока есть C_j , удовлетворяющие условию шага 4, $F = F \cup Ftemp$. Увеличиваем оценку правдоподобия LS_j на CF_2 и формируем соответствие $F_j = \langle Q_i, C_j \rangle$ и выполняем вызов процедуры *Pairs* ($Frec = Pairs(Q_i, C_j, O)$), добавляем парные соответствия $Ftemp = Frec \times F_j$. Если нет C_j , удовлетворяющих условию шага 4 переходим к шагу 2.

Шаг 6. Выход.

На втором этапе для оценки близости текущей ситуации и прецедента используется метод ближайшего соседа. Для каждого парного соответствия в выбранной метрике определяются расстояние d_{CQ} между текущей ситуацией и прецедентом. Для определения значения степени сходства $Sim(C, Q)$ необходимо найти максимальное расстояние d_{MAX} в выбранной метрике, используя границы диапазонов параметров ($x_i^{нач}$ и $x_i^{кон}$, $i = 1, \dots, n$).

В результате применения данной процедуры получим множество прецедентов, каждому из которых сопоставлены две оценки сходства с текущей ситуацией, которые могут быть выражены в процентах:

- оценка на основе онтологии предметной области: $Sstruct = \sum_{i=1}^k LS_i / SES_{MAX}$, где k – количество соответствий, LS_i – оценка правдоподобия для i соответствия, SES_{MAX} – оценка для случая, когда каждый элемент в базовой области имеет родительское отношение и в качестве базовой области выбирается целевая;
- оценка по методу ближайшего соседа: $Sim(C, Q) = 1 - d_{CQ} / d_{MAX}$, где d_{CQ} – расстояние между текущей ситуацией и прецедентом, d_{MAX} – максимальное расстояние в выбранной метрике.

Исходя из этих данных, ЛПП может выбрать наиболее подходящий прецедент и получить решение для текущей ситуации.

Реализация прототипа CBR-системы

Предлагаемый подход реализован в прототипе CBR-системы (рис. 2). Программная реализация прототипа CBR-системы выполнена с использованием языка C# [Шилдт, 2013] и среды программирования MS Visual Studio 2010, а также с использованием редактора онтологий Protégé (<http://protege.stanford.edu>).

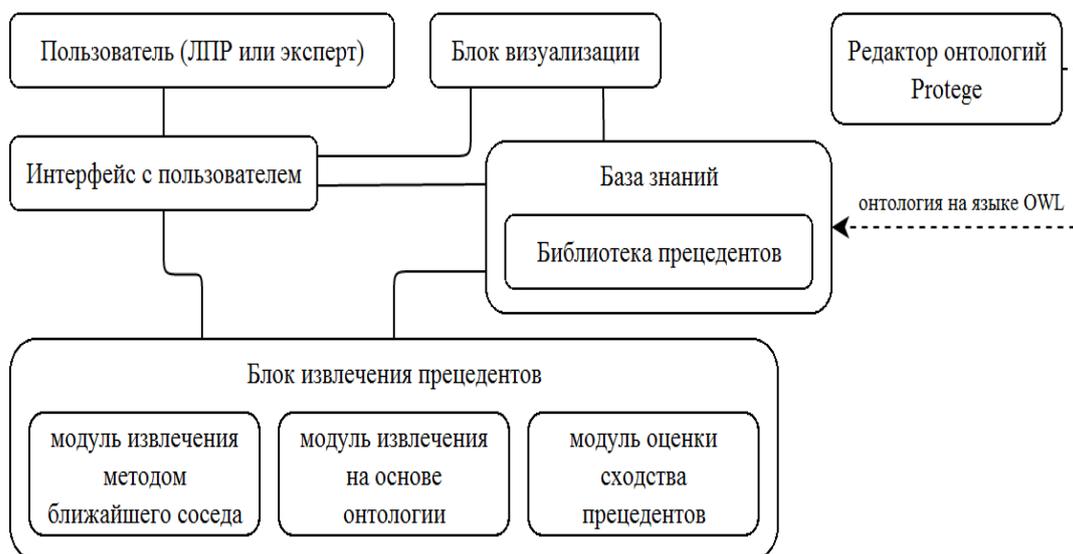


Рис. 2. Архитектура прототипа CBR-системы

Работа реализованного прототипа CBR-системы рассмотрена на примере решения задачи экспертного диагностирования с использованием онтологии из открытого репозитория онтологий по биологии и медицине NCBO BioPortal (<http://bioportal.bioontology.org>), которая была дополнена моделью прецедента (рис. 3), спроектированной на основе набора данных из хранилища UCI Machine Learning Repository (<http://archive.ics.uci.edu/ml/>).

В данном примере использована онтология с записями пациентов (<http://bioportal.bioontology.org/ontologies/1059>), включающая 236 концептов и 315 отношений и набор медицинских данных, описывающих состояние пациента после операции (<http://archive.ics.uci.ml/datasets/Post-Operative+Patient>).

В данном случае поставленную задачу экспертного диагностирования можно свести к задаче классификации. То есть необходимо отнести текущую ситуацию к одному из известных прецедентов.

На вход прототипа CBR-системы подается ситуация (рис. 4). Прототип CBR-системы позволяет вычислить оценки сходства текущей ситуации и прецедентов из БП. ЛПР имеет возможность выбрать наиболее подходящий прецедент исходя из двух оценок сходства: по структуре (на основе онтологии предметной области и SMT); по методу ближайшего соседа.

Заключение

В работе рассмотрены общие понятия и особенности CBR-технологии, процесс поиска решения на основе прецедентов. Исследована возможность интеграции прецедентного и онтологического подхода, в частности, возможность представления прецедентов с помощью онтологии предметной области. Предложен метод поиска решения на основе прецедентов с использованием онтологического подхода и структурной аналогии на базе SMT. Разработана архитектура и выполнена программная реализация

прототипа CBR-системы в среде MS Visual Studio 2010 на языке C#. Рассмотрен пример использования разработанной системы для решения задачи экспертной диагностики на основе онтологии из репозитория NCBO BioPortal и реальных наборов данных из хранилища UCI Machine Learning Repository.

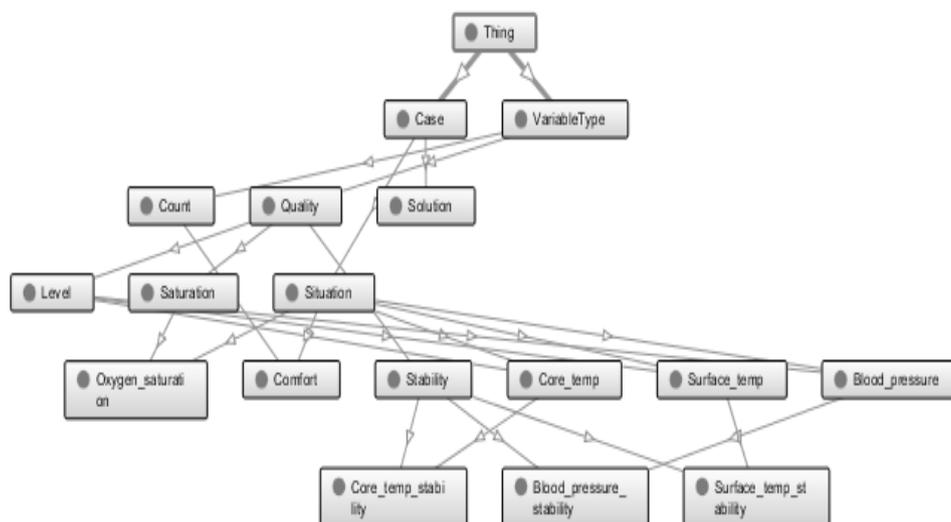


Рис. 3. Описание прецедента

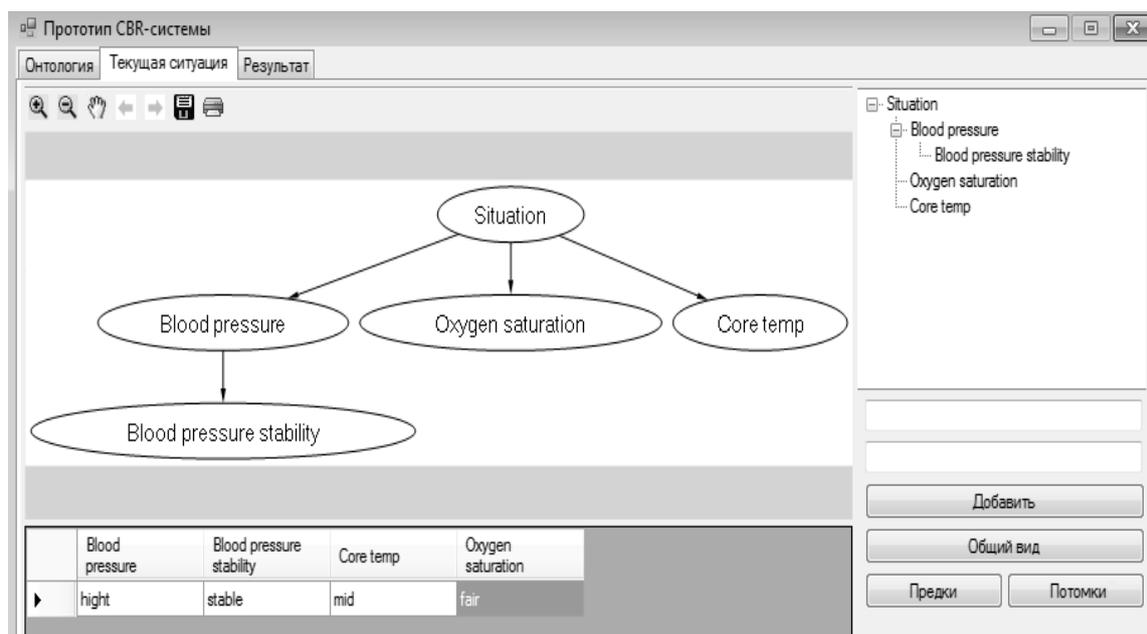


Рис. 4. Описание текущей ситуации

Благодарности

Работа поддержана Российским Фондом Фундаментальных Исследований (проекты 11-01-00140, 12-07-00508).

Список литературы

- [Aamodt et al., 1994] Aamodt A., Plaza E. Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches // Artificial Intelligence Communications. IOS Press. 1994. – Vol. 7, № 1. – P. 39-59.
- [Falkenhainer et al., 1989] Falkenhainer B., Forbus K., Gentner D. The Structure-Mapping Engine: Algorithm and examples // Artificial Intelligence, 41, 1989, – P. 1–63.
- [Алехин, 2011] Алехин Р.В. Использование языка OWL для формирования онтологий предметной области // Радиоэлектроника, электротехника и энергетика: XVII Междунар. науч.-техн. конф. студентов и аспирантов: Тезисы докладов в 3 т. Т. 1. М.: Издательский дом МЭИ, 2011. – С. 353–354.
- [Вагин и др., 2008] Вагин В.Н., Головина Е.Ю., Загорянская А.А., Фомина М.В. Достоверный и правдоподобный вывод в интеллектуальных системах // Под ред. В.Н. Вагина, Д.А. Поспелова. –М.: ФИЗМАТЛИТ, 2004. – 704 с.
- [Варшавский и др., 2005] Варшавский П.Р., Еремеев А.П. Поиск решения на основе структурной аналогии для интеллектуальных систем поддержки принятия решений // Известия РАН. Теория и системы управления. – № 1. – 2005. – С. 97–109
- [Варшавский и др., 2006] Варшавский П.Р., Еремеев А.П. Методы правдоподобных рассуждений на основе аналогий и прецедентов для интеллектуальных систем поддержки принятия решений // Новости искусственного интеллекта. 2006. №3. – С. 39-62.
- [Варшавский и др., 2009] Варшавский П.Р., Еремеев А.П. Моделирование рассуждений на основе прецедентов в интеллектуальных системах поддержки принятия решений // Искусственный интеллект и принятие решений. 2009. №2. – С. 45–57.
- [Варшавский и др., 2012] Варшавский П.Р., Алехин Р.В., Зо Лин Кхаинг Применение онтологического подхода для реализации поиска решения на основе прецедентов в интеллектуальных системах поддержки принятия решений // Труды 13-ой национальной конференции по ИИ с международным участием (КИИ-2012). Т. 3. –Белгород: Издательство БГТУ, 2012. – С. 72-79.
- [Варшавский, 2008] Варшавский П.Р. Механизмы правдоподобных рассуждений на основе прецедентов (накопленного опыта) для систем экспертной диагностики // Труды 11-ой национальной конференции по ИИ с международным участием (КИИ-2008). Т. 2. –М: ЛЕНАНД, 2008. – С. 321–329.
- [Гаврилова и др., 2000] Гаврилова Т.А., Хорошевский В.Ф. Базы знаний интеллектуальных систем. –СПб.: Питер, 2000. – 384 с.
- [Еремеев и др., 2012] Еремеев А.П., Варшавский П.Р., Куриленко И.Е. Моделирование временных зависимостей в интеллектуальных системах поддержки принятия решений на основе прецедентов // International Journal «Information technologies and knowledge», Vol. 6, № 3, 2012, С. 227-239.
- [Поспелов, 1989] Поспелов Д.А. Моделирование рассуждений. Опыт анализа мыслительных актов. –М.: Радио и связь. 1989.
- [Шилдт, 2013] Шилдт Г. С# 4.0: полное руководство. –М.: Вильямс, 2013.

Сведения об авторах

Варшавский Павел Романович – к.т.н., доцент кафедры Прикладной математики НИУ «МЭИ», 111250, Россия, Москва, Красноказарменная ул., 14; e-mail: VarshavskyPR@mpei.ru

Область научных интересов: искусственный интеллект, принятие решений, методы правдоподобных рассуждений

Алехин Роман Викторович – магистр, кафедра Прикладной математики НИУ «МЭИ», 111250, Россия, Москва, Красноказарменная ул., 14; e-mail: r.alekhin@gmail.com

Область научных интересов: искусственный интеллект, принятие решений, методы правдоподобных рассуждений, модели представления знаний

МОДЕЛИРОВАНИЕ СОГЛАСОВАННОСТИ ЦЕЛЕЙ В ИЕРАРХИЧЕСКИХ СТРУКТУРАХ ПРИНЯТИЯ РЕШЕНИЙ

Н.Н. Маляр

Аннотация: В работе рассмотрен подход к моделированию многокритериальных задач принятия решений с иерархической структурой. Приводится способ согласования целей при помощи модели задачи многокритериального выбора с целевой точкой, координаты которой являются удовлетворяющими значениями оценок по критериям.

Ключевые слова: иерархические структуры, многокритериальный рациональный выбор, точка удовлетворения, функция принадлежности, нечеткое множество.

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

Введение

Как показывает анализ технической литературы, сложным объектом обычно считается субстанция (система), которая удовлетворяет трем основным системным концепциям:

1. Структурность - система определяется как целостность взаимосвязанных элементов; отношения между элементами придают системе дополнительное качество, получившее название эмерджентность (эффект целостности, интегральный эффект);
2. Иерархичность - система любой сложности обязательно входит в систему более высокого уровня, а каждый из ее элементов может рассматриваться, в свою очередь, в качестве системы (подсистемы);
3. Функциональность - система характеризуется параметрами входа - выхода и параметрами состояния.

Иерархия – это определенная структура, основанная на предложении, что ее элементы могут группироваться в несвязные множества. Эта совокупность подмножеств должна обладать тем свойством, что пересечение подмножеств является пустым множеством, а их объединение дает исходное множество. При рассмотрении проблем функционирования или развития сложных объектов часто удается представить их связи между системно организующими факторами в виде многоуровневых иерархических структур. Значительные трудности в управлении сложной системой возникают в связи с многоуровневым иерархическим характером объектов (подсистем) и критериев их оптимизации. В основном это происходит из-за слабой разработанности методов принятия решений в многоуровневых иерархических системах в условиях неопределенности. При изучении сложных кибернетических систем исследователь сталкиваемся с иерархией процессов принятия решений.

Проблема принятия решений является одной из основных в современной теории и практике управления. Известный американский специалист по управлению Герберт Саймон назвал принятие решений «сутью управленческой деятельности». Иерархия является распространенным типом структуры системных объектов. Особенно характерна она для систем управления социально-экономическими процессами. Везде, где приходится сталкиваться с иерархией, обнаруживается одна важная особенность: целостность оказывается "разложимой" на элементы, каждый из которых, в свою очередь, ведет себя как целостность. Возникновение иерархической структуры было обусловлено всевозрастающей сложностью технологии

управляемых объектов, создающей большие трудности для централизованного управления. Поэтому появилась необходимость разделения всего процесса принятия решений на такое число уровней, чтобы решение задачи оптимизации на каждом из них было относительно не сложным. Но с возникновением многоуровневых иерархических систем управления появилась и новая задача согласования и координации решений, принимаемых на всех уровнях управления. В системах управления нередко соседствует сразу несколько иерархических структур, между которыми возникает сложное взаимодействие. Примером может служить экономическая система народного хозяйства, где предприятия, с одной стороны, расчленены и управляются по отраслевому принципу, а с другой – образуют иерархию региональных систем. В результате возникает сложная проблема связи отраслевого и территориального управления, актуальность которой в настоящее время имеет тенденцию к возрастанию.

Принципы иерархичности

На сегодняшний день существует большое количество проблемных ситуаций в сложно структурированных областях знаний. Поэтому представление их в виде иерархичных структур является наиболее удобным. В качестве модели используется дерево целей, в котором общие цели разделяются на цели частного характера. Любая цель непосредственно описывается критерием, по которому и/или проверяется ее реализуемость (достижимость).

Иерархия рассматриваемой проблемной ситуации строится методом структурной декомпозиции, то есть формирование структуры осуществляется «сверху вниз», и должна соответствовать общеизвестным принципам построения иерархических структур [Михалевич, 1982; Моисеев, 1981]:

1. Определяется «глобальная» цель управления (функционирования, развития) рассматриваемой проблемной ситуации, которая располагается на верхнем уровне иерархии, называется «центром» и имеет порядковый номер 0.
2. Путем структурной декомпозиции ситуации определяются подцели глобальной цели, а также критерии реализации глобальной цели и подцелей нижестоящего уровня таким образом, что они должны раскрывать смысл критериев вышестоящего уровня.
3. Между критериями, находящимися на одном уровне иерархии, должна отсутствовать взаимозависимость.
4. Критерии данного уровня непосредственно зависят только от критериев соседнего нижестоящего уровня иерархии.
5. Полнота раскрытия критерия. Смысл критерия является раскрытым, если на следующем нижестоящем уровне иерархии все влияющее на него критерии известны (существуют).

В теории многоуровневых систем двухуровневая система принятия решений представляет специфический интерес. Во-первых, это простейший тип систем, в котором проявляются все наиболее существенные характеристики многоуровневой системы. Во-вторых, более сложные многоуровневые системы могут быть построены из двухуровневых подсистем, как из блоков.

Постановка задачи

Рассмотрим двухуровневую иерархическую структуру принятия решений. Пусть некий субъект, впрямь именуемый центром, стремится достичь определенных целей, реализация которых описывается множеством критериев $K^0 = \{K_1^0, K_2^0, \dots, K_m^0\}$. Центр воздействует на отдельные звенья P_1, P_2, \dots, P_s , которые назовем агентами. Они обладают собственными целями, реализация которых описывается множествами $K^l = \{K_1^l, K_2^l, \dots, K_{m_l}^l\}, l = \overline{1, s}$. Задано дискретное множество альтернативных

вариантов $A = \{a_1, a_2, \dots, a_n\}$, над которым центром и агентами необходимо осуществить задачу выбора. Примером такой задачи может служить выбор производственной программы для фирмы - производителя. В данном случае центром является дирекция фирмы, а агентами – производственный отдел, отдел сбыта, финансовый отдел, отдел кадров.

Принятие решения центром требует дополнительных информационных затрат для согласования своих целей с целями агентов. Предлагается подход, который позволяет увязывать выбор агентов с выбором центра. Предположим, что центру и каждому агенту известны множество альтернатив и критерии, которые реализуют их цели и служат для оценки полезности множества альтернатив. Таким образом, центром и каждым агентом решается задача многокритериального выбора.

В общем случае задачу многокритериального иерархичного выбора можно сформулировать в следующем виде. Заданы множество альтернатив решения некоторой проблемы $A = \{a_1, a_2, \dots, a_n\}$ и множество критериев $K = \{k_1, k_2, \dots, k_m\}$ для оценки полезности альтернатив. Необходимо ранжировать альтернативы относительно интегрированной функции полезности.

Предположим, что каждой альтернативе a_j определены оценки x_{ij} по всему множеству критериев K .

То есть x_{ij} - оценка j -й альтернативы по i -му критерию. Оценки могут задаваться в количественной (например, в привычной шкале оценок в системе образования) или в качественной форме (например, лингвистическими переменными). Результаты оценивания представляются в виде матрицы решений.

Модель принятия решений такого вида - это модель задачи выбора на языке матриц решений оценок альтернатив $X = (x_{ij}), i = \overline{1, m}, j = \overline{1, n}$.

Далее для каждого вектора оценок $\overline{x}_j = (x_{1j}, x_{2j}, \dots, x_{mj})$ вычисляется, в зависимости от используемой модели многокритериального выбора решения, функция полезности U_j . Наилучшей объявляется та альтернатива, для которой функция полезности имеет максимальное значение – в случае, например, максимизации прибыли, или минимальное значение, например, при минимизации убытков. При вычислении значений полезности могут приниматься во внимание веса критериев w_j .

Процедура многокритериального выбора

Как показывает практика, поведению людей отвечают как оптимизационные шаги, так и шаги получения удовлетворяющих результатов [Саймон, 1999].

Не ограничивая общности, предположим, что все альтернативы множества $A = \{a_1, \dots, a_n\}$ принадлежат множеству Парето и критерии эффективности измеряются в разных единицах измерения, а наилучшей считается альтернатива, для которой оценки достигают своего максимального или минимального значения. Пусть, без ограничения общности, все оценки – положительные числа,

Введем в рассмотрение фиктивную альтернативу T , которую назовем “точкой удовлетворения” или “точкой ограниченной рациональности” [Маляр, 2005]. Координаты этой точки представляют собой оценки по соответствующим критериям, значениями которых лицо, принимающее решение, может быть удовлетворенным.

Понятие удовлетворенности является нечетким понятием и для его описания применяется теория нечетких множеств. В современной практике различают два типа нечетких множеств [Батыршин, 2007]:

1. Нечеткие множества, которые определены на некоторой числовой шкале. Например, на интервале действительных чисел. В таких случаях нечеткие множества - это нечеткие величины. Примерами нечетких величин являются нечеткие числа и нечеткие интервалы.
2. Нечеткие множества определены на нечисловом множестве. Например, на множестве целей и альтернатив, экспертных оценок, бинарных отношений между объектами. Нечеткие множества такого типа - это множества нечетких объектов, которые записываются в виде $B = \{x, \mu(x)\}$, где x – множество объектов, $\mu(x)$ - функция принадлежности объекта множеству.

На самом деле функция принадлежности определяет порядок элементов. Для более сложных категорий, которые определяются на декартовом произведении линейных шкал, функция принадлежности может быть получена за счет свертывания информации простых категорий.

Проблемой оценки удовлетворенности является то, что сама удовлетворенность представляет собой нечеткое, размытое понятие, на значение которого сильное влияние оказывают суждения, восприятие и эмоции человека. В связи с этим предлагается использовать при измерении удовлетворенности лингвистические переменные, т.е. такие переменные, значениями которых являются не числа, а слова или предложения на естественном или формальном языке [Кофман, 1982].

Для значений лингвистических переменных, представляющих собой нечеткое подмножество, строится функция принадлежности, т.е. такая функция, которая каждому элементу из универсального множества всех возможных оценок ставит в соответствие число из интервала от 0 до 1, которое характеризует степень принадлежности данного элемента рассматриваемому нечеткому подмножеству.

Далее, каждый агент обязан задать свою «точку удовлетворения» и описать нечеткое множество относительно этой точки.

Построим матрицу нечетких оценок $Y = (y_{ij})$, используя: матрицу решений $X = (x_{ij})$, точку «удовлетворения» $T = (t_1, \dots, t_m)$ и значения лингвистической переменной «альтернатива». Например, опишем нечеткие множества альтернатив «близких», «худших» и «лучших» по отношению к «точке удовлетворения».

Нечеткое множество $B_1 = \{\text{альтернативы, близкие относительно точки «удовлетворения»}\}$:

$$y_{ij} = 1 - \frac{|t_i - x_{ij}|}{\max\{t_i - \min_j x_{ij}; \max_j x_{ij} - t_i\}}; \quad i = \overline{1, m}; j = \overline{1, n}.$$

Нечеткое множество $B_2 = \{\text{альтернативы, лучшие относительно точки «удовлетворения»}\}$:

$$y_{ij} = \begin{cases} \frac{|t_i - x_{ij}|}{\max_j x_{ij} - t_i}, & x_{ij} \geq t_i; \\ 0, & x_{ij} \leq t_i. \end{cases} \quad i = \overline{1, m}; j = \overline{1, n}.$$

Нечеткое множество $B_3 = \{\text{альтернативы, худшие относительно точки «удовлетворения»}\}$:

$$y_{ij} = \begin{cases} \frac{|t_i - x_{ij}|}{t_i - \min_j x_{ij}}, & x_{ij} \leq t_i; \\ 0, & x_{ij} \geq t_i. \end{cases} \quad i = \overline{1, m}; j = \overline{1, n}.$$

Следующим шагом будет построение функции полезности U_j , как свертки столбцов матрицы Y с

весовыми коэффициентами критериев. Результатом работы этой процедуры будет профиль $U = \{U_1, U_2, \dots, U_n\}$, заданный функцией принадлежности данному нечеткому множеству. Таким образом, каждый агент задает свой профиль ранжирования альтернатив.

Схема решения задачи согласования целей

Опишем схему согласованности целей между центром и агентами. Сначала решается задача многокритериального выбора для каждого агента. Затем решается задача многокритериального выбора центром. На первом этапе для центра построим матрицу нечетких оценок $Y = (y_{ij})$, используя: матрицу решений $X = (x_{ij})$, точку "удовлетворения" $T = (t_1, \dots, t_n)$ и значения лингвистической переменной «альтернатива». Далее добавим в данную матрицу Y s строк, которые соответствуют профилям каждого агента, и построим профиль предпочтений главной цели центра. Профиль предпочтений центра может быть построен разными способами. Например, в виде свертки (описанной выше), композиционного правила [Маляр, 2010], критериев коллективного ранжирования и т. п.

Выводы

Данный подход позволяет представить согласованность целей в иерархических структурах при помощи модели задачи выбора в терминах размытых множеств. На каждом уровне иерархии центр и агенты строят свое размытое множество, которое зависит от множества критериев эффективности, точки «удовлетворения», его определения и способа построения функции принадлежности.

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

- [Батыршин, 2007] Нечеткие гибридные системы: Теория и практика / Батыршин И.З., Недосекин А.О., Стецко А.А. и др. ; под ред. Ярушкиной Н.Г. – М.: Физматлит, 2007. – 208 с.
- [Кофман, 1982] Кофман А. Введение в теорию нечетких множеств. – М.: Радио и связь, 1982. - 432 с.
- [Маляр, 2005] Маляр М.М. Описання задач вибору на мові розмитих множин/ Маляр М.М. // Вісник Київського університету. Вип.4. Серія фіз.-мат. Науки, Київ, 2005. – С. 197-201.
- [Маляр, 2010] Маляр Н.Н. Нечеткая модель удовлетворительного решения задачи выбора / Н.Н. Маляр // Information Models of Knowledge. – ITHEA – Kiev, Ukraine – Sofia, Bulgaria, 2010. – С.220-225.
- [Михалевич, 1982] Михалевич В.С., Волкович В.Л. Вычислительные методы исследования и проектирования сложных систем. – М.: Наука. Главная редакция физико-математической литературы, 1982. -286 с.
- [Моисеев, 1981] Моисеев Н.Н. Математические задачи системного анализа.– М.: Наука. Главная редакция физико-математической литературы, 1981. – 488 с.
- [Саймон, 1999] Герберт А. Саймон. Теория принятия решений в экономической теории и науке о поведении // Вехи экономической мысли. Теория потребительского поведения и спроса. Т.1./ Под ред. В.М. Гальперина. СПб.: Экономическая школа, 1999г.
-

Сведения об авторе

Маляр Николай Николаевич - декан математического факультета, заведующий кафедрой кибернетики и прикладной математики Ужгородского национального университета, кандидат технических наук, доцент, Украина, Ужгород, ул. Подгорная, 46; e-mail: malvayarmm@gmail.com

TABLE OF CONTENT OF IJ IMA VOL.2, NO.1, 2013

<i>A Joint Global and Local Tone Mapping Algorithm for Displaying Wide Dynamic Range Images</i> Alain Horé, Chika A. Ofili, and Orly Yadid-Pecht, Fellow, IEEE	3
<i>Component Modeling: on Connections of Detailed Petri Model and Component Model of Parallel Distributed System</i> Elena Lukyanova.....	15
<i>Model for Astronomical Dating of the Chronicle of Hydatius</i> Jordan Tabov	23
<i>Connectivity Control in AD Hoc Systems: a Graph Grammar Approach</i> Alexander Mikov, Alexander Borisov.....	37
<i>Citation-Paper Rank Distributions and Associated Scientometric Indicators – a Survey</i> Vladimir Atanassov, Ekaterina Detcheva	46
<i>Sub-Optimal Nonparametric Hypotheses Discriminating with Guaranteed Decision</i> Fedor Tsitovich, Ivan Tsitovich.....	62
<i>Linguistic and Program Tools for Debugging and Testing of Simulation Models of Computer Networks</i> Elena Zamyatina, Alexander Mikov, Roman Mikheev.....	70
<i>The Inverse Maslov Method and ANT Tactics for Exhaustive Search Decreasing</i> Tatiana Kosovskaya, Nina Petukhova.....	81
<i>Application of Some Cybernetic Models in Building Individual Educational Trajectory</i> Borislav Lazarov.....	90

TABLE OF CONTENT OF IJ IMA VOL.2, NO.2, 2013

<i>A Comparison of Some Approaches to the Recognition Problems in Case of Two Classes</i> Yurii I. Zhuravlev, Yuriy Laptin, Alexander Vinogradov, Aleksey Likhovid	103
<i>Adaptive Fuzzy Probabilistic Clustering of Incomplete Data</i> Yevgeniy Bodyanskiy, Alina Shafronenko, Valentyna Volkova	110
<i>Crop Classification in Ukraine Using Satellite Optical and SAR Images</i> Nataliia Kussul, Sergii Skakun, Andrii Shelestov, Oleksii Kravchenko, Olga Kussul.....	118
<i>Use of Information Value in AVO-Polynomial Method Training</i> Alexander Dokukin	123
<i>Short Graph-Scheme of a Successful System Idea</i> Nikolay Kosovskiy	127
<i>Analysis of Features and Possibilities of Bank Functioning Efficiency Based on the Method of Stochastic Frontiers</i> Oleksandr Kuzomin, Vyacheslav Lyashenko	132
<i>Peculiarities of Linked Data Processing in Semantic Applications</i> Sergey Shcherbak, Ilona Galushka, Sergey Soloshich, Valeriy Zavgorodniy	139
<i>Key Frame Partition Matching for Video Summarization</i> Olena Mikhnova, Nataliia Vlasenko	145

<i>О приближенном решении задач восстановления зависимостей с помощью алгоритмов распознавания</i> Владимир Рязанов, Антон Щичко	153
<i>Логико-лингвистическая модель извлечения фактов из слабоструктурированной текстовой информации</i> Нина Хайрова, Наталья Шаронова	167
<i>Выделение текста на сложном цветном фоне</i> Роман Телятников, Иван Шумский, Ариф Мамедов, Анатолий Протосавицкий, Екатерина Матусевич, Екатерина Степанькова	176
<i>О вариативности некоторых буквенных частот в Супраслском сборнике</i> Й. Табов, Св. Христова	188
<i>Модели интеллектуальной адаптивной поддержки навигации в компьютерных обучающих системах</i> Игорь Шубин, Владимир Чернов, Владимир Гриценко, Ирина Кириченко.....	194

TABLE OF CONTENT OF IJ IMA VOL.2, NO.3, 2013

<i>Classification of Incomplete Data</i> Vasily V.Ryazanov, Vladimir V.Ryazanov	203
<i>On a Logical Regularities Based Method of Definite Quality Object Synthesis</i> Alexander Dokukin	212
<i>Modeling of an Intellectual Problem Solver by Transformation of Semantic Models</i> Elena Shalfeeva	217
<i>Modeling of the Organizational and Expert Activities of the Scientists at the Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, in 2005-2010</i> J. Tabov, S. Hristova	223
<i>Methods and Models for Selection of Rational Solutions in Decision-Making Systems</i> Oleksandr Kuzomin, Mikhail Shtukin.....	228
<i>Storing Dictionaries and Thesauruses Using NL-Addressing</i> Krassimira B. Ivanova, Koen Vanhoof, Krassimir Markov, Vitalii Velychko	239
<i>Задача выбора топологической структуры сети доступа</i> Галина Гайворонская, Антон Бондаренко	252
<i>Качество управления интеллектуальными услугами в сетях последующего поколения</i> Сергей Шестопапов.....	262
<i>Повышение структурной живучести телекоммуникационной сети</i> Нина Князева	275
<i>Разработка модели пользователя информационных сетей как плохоформализованного объекта</i> Анастасия Смирнова.....	285
<i>Програмный комплекс анализа конъюнктуры оптового рынка электроэнергии и мощности</i> Галина Старкова	292

TABLE OF CONTENT OF IJ IMA VOL.2, NO.4, 2013

<i>Интеграция интеллектуальных систем навигации и управления движения мехатронных роботов</i>	
Р.М. Юсупов , А.В. Тимофеев	303
<i>Выбор вычислительного кластера в сокращенном пространстве критериев</i>	
Алексей Петровский, Василий Лобанов, Алла Заболеева-Зотова, Татьяна Шитова	313
<i>Методы моделирования временных зависимостей в интеллектуальных системах с использованием темпоральных прецедентов</i>	
Александр Еремеев, Иван Куриленко.....	324
<i>Темпоральные сети Петри и их применение в интеллектуальных системах поддержки принятия решений реального времени</i>	
Александр Еремеев, Юрий Королев.....	336
<i>Метод описания изменения количества пользователей телекоммуникационной сети</i>	
Галина Гайворонская, Олег Домаскин.....	345
<i>Управление качеством услуг в телекоммуникационных сетях</i>	
Анастасия Кальченко	349
<i>Problem of Restoring the FunctionS-SIGNALs by Finite Set of Data with Errors</i>	
Veda Kasyanyuk, Iryna Volchyna.....	361
<i>Исследование Муравьиных алгоритмов оптимизации в задаче коммивояжера</i>	
Юрий Зайченко, Николай Мурга.....	370
<i>Метод поиска решений в интеллектуальных системах поддержки принятия решений на основе прецедентов</i>	
Павел Варшавский, Роман Алехин.....	385
<i>Моделирование согласованности целей в иерархических структурах принятия решений</i>	
Н.Н. Маляр.....	393
<i>Table of content of IJ IMA Vol.2, No.1, 2013.....</i>	398
<i>Table of content of IJ IMA Vol.2, No.2, 2013.....</i>	398
<i>Table of content of IJ IMA Vol.2, No.3, 2013.....</i>	399
<i>Table of content of IJ IMA Vol.2, No.4, 2013.....</i>	400