ITHEA

International Journal

INFORMATION THEORIES & APPLICATIONS

2006 Volume 13 Number 1

International Journal
INFORMATION THEORIES & APPLICATIONS
ISSN 1310-0513
Volume 13 / 2006, Number 1

IJ ITA is official publisher of the scientific papers of the members of
the Association of Developers and Users of Intellectualized Systems (ADUIS).

IJ ITA welcomes scientific papers connected with any information theory or its application.
Original and non-standard ideas will be published with preferences.

IJ ITA rules for preparing the manuscripts are compulsory.
The **rules for the papers** for IJ ITA as well as the **subscription fees** are given on *www.foibg.com/ijita*.
The camera-ready copy of the paper should be received by e-mail: foi@nlcv.net.
Responsibility for papers published in IJ ITA belongs to authors.

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

International Journal "INFORMATION THEORIES & APPLICATIONS" Vol.13, Number 1, 2006
Printed in Bulgaria
Edited by the Institute of Information Theories and Applications FOI ITHEA, Bulgaria,
in collaboration with the V.M.Glushkov Institute of Cybernetics of NAS, Ukraine, and
the Institute of Mathematics and Informatics, BAS, Bulgaria.
Publisher: FOI-COMMERCE - Sofia, 1000, P.O.B. 775, Bulgaria. www.foibg.com, e-mail: foi@nlcv.net

ISSN 1310-0513

PREFACE

Verba volant, scripta manent !

The "**International Journal on Information Theory and Applications**" (**IJ ITA**) has been established in 1993 as independent scientific printed and electronic media. IJ ITA is edited by the *Institute of Information Theories and Applications FOI ITHEA* in collaboration with the leading researchers from the Institute of Cybernetics "V.M.Glushkov", NASU (Ukraine) and Institute of Mathematics and Informatics, BAS (Bulgaria).

During the years, IJ ITA became as well-known international journal. Till now, including this volume, more than 550 papers have been published. IJ ITA authors are widespread in 39 countries all over the world: *Armenia*, *Belarus*, Brazil, *Belgium*, *Bulgaria*, Canada, Czech Republic, Denmark, Egypt, Estonia, *Finland*, France, *Germany*, Greece, *Hungary*, Ireland, Israel, Italy, Japan, Kirghizia, Latvia, Lithuania, Malta, Mexico, Moldavia, Netherlands, *Poland*, Portugal, Romania, *Russia*, Scotland, Senegal, Serbia and Montenegro, *Spain*, *Sultanate of Oman*, Turkey, UK, *Ukraine*, and *USA*.

Volume 13/2006 of the IJ ITA contains 56 papers written by 98 authors from 13 countries (marked in italics above), selected from several international conferences, seminars and workshops organized or supported by the Journal.

At the first place, the main source for selection were the ITA 2005 Joint International Events on Information Theories and Applications, (June 20-30, 2005, Varna, Bulgaria):

- XI-th International Conference "Knowledge-Dialogue-Solution" (KDS 2005);
- Third International Conference "Information Research, Applications and Education" (i.TECH 2005);
- International Conference "Business Informatics" (Bi 2005);
- IV-th International Workshop on General Information Theory (GIT 2005);
- International INTAS-FET Strategic Workshop "Data Flow Systems: Algorithms and Complexity",
- Third International Workshop on Multimedia Semantics.

Several papers were selected from the pool of papers directly submitted to IJ ITA.

Main characteristic of ITA 2005 International Conferences was that the papers were combined into thematic sessions. Because of this, the selected papers are published in this volume following the thematic sessions' organisation.

Congratulations to Luis Fernando de Mingo López (Spain) and Milena Dobreva (Bulgaria) who were awarded by the International Prize "**ITHEA**" for the year 2005. The "ITHEA" Prize has been established in 1995. It is aimed to mark the achievements in the field of the information theories and applications.

More information about the IJ ITA rules for preparing and submitting the papers as well as how to take out a subscription to the Journal may be obtained from www.foibg.com/ijita.

Krassimir Markov

IJ ITA Founder and Editor in chief

International Prize "ITHEA"

Awarded Scientists till 2005:

1995	*Sandansky*	*K. Bankov, P. Barnev, G. Gargov, V. Gladun, R. Kirkova, S. Lazarov, S. Pironkov, V. Tomov*
1996	*Sofia*	*T. Hinova, K. Ivanova, I. Mitov, D. Shishkov, N. Vashchenko*
1997	*Yalta*	*Z. Rabinovich, V. Sgurev, A. Timofeev, A. Voloshin*
1998	*Sofia*	*V. Jotsov*
1999	*Sofia*	*L. Zainutdinova*
2000	*Varna*	*I. Arefiev, A. Palagin*
2001	*St.Peterburg*	*N. Ivanova, V. Koval*
2002	*Primorsko*	*A. Milani, M. Mintchev*
2003	*Varna*	*T. Gavrilova, A. Eskenazi, V. Lozovskiy, P. Stanchev*
2004	*Varna*	*B. Kokinov, T. Vamos*
2005	*Varna*	*L.F. de Mingo López, M. Dobreva*

IJ ITA major topics of interest include, but are not limited to:

INFORMATION THEORIES

Artificial Intelligence
- *Computer Intellectualisation*
- *Intelligent Networks and Agents*
- *Intelligent Technologies*
- *Knowledge Discovery and Engineering*
- *Knowledge Acquisition and Formation*
- *Distributed Artificial Intelligence*
- *Models of Plausible Reasoning*
- *AI Planning and Scheduling*

Bioinformatics
Business Informatics
Cognitive Science
Decision Making
Education Informatics
General Information Theory
Hyper Technologies
Information Models
Intellectualisation of Data Processing
Knowledge-based Society
Logical Inference
Natural language Processing
Neuroinformatics
Philosophy and Methodology of Informatics
Quality of the Programs
Software Engineering
Theory of Computation

APPLICATIONS

Business Information Systems
Communication Systems
Computer Art and Computer Music
Hyper Technologies
Intelligent Information Systems
Multimedia Systems
Programming Technologies
Program Systems with Artificial Intelligence
Pyramidal Information Systems
Very Large Information Spaces

USING SENSITIVITY AS A METHOD FOR RANKING THE TEST CASES CLASSIFIED BY BINARY DECISION TREES

Sabrina Noblesse, Koen Vanhoof

Abstract: Usually, data mining projects that are based on decision trees for classifying test cases will use the probabilities provided by these decision trees for ranking classified test cases. We have a need for a better method for ranking test cases that have already been classified by a binary decision tree because these probabilities are not always accurate and reliable enough. A reason for this is that the probability estimates computed by existing decision tree algorithms are always the same for all the different cases in a particular leaf of the decision tree. This is only one reason why the probability estimates given by decision tree algorithms can not be used as an accurate means of deciding if a test case has been correctly classified. Isabelle Alvarez has proposed a new method that could be used to rank the test cases that were classified by a binary decision tree [Alvarez, 2004]. In this paper we will give the results of a comparison of different ranking methods that are based on the probability estimate, the sensitivity of a particular case or both.

ACM Classification Keywords: I.2.6 Learning – induction, concept learning; I.5.2 Classifier design and evaluation

1. Introduction

Decision trees do not only classify a particular case, they will also give the probability or the estimate of the probability that this case eventually belongs to the predicted class. For computing the probability estimate of a case, decision tree algorithms use the raw training frequency of the leaf where the particular case belongs to. The raw training frequency, *P*, puts the number of positive training cases, *k*, in a leaf to the total number of training cases, *n*, in that leaf [Zadrozny and Elkan, 2001]. If the class of the leaf of a particular training case is the same as the real class of that case, it is called a positive training case. The formula, $P = k / n$, can be used to compute the raw training frequency. This way of computing probability estimates and the fact that decision tree algorithms want to maximize classification accuracy and minimize the size of the decision tree [Provost and Domingos, 2000] cause some problems concerning the use of probability estimates for ranking the test cases classified by decision trees. These problems are listed below.

- Probability estimates can have extreme values: decision tree algorithms try to make the leaves of the decision tree as homogeneous as possible. By doing so the observed frequencies of positive training cases will shift automatically to 0 and 1 [Zadrozny and Elkan, 2001; Provost and Domingos, 2000].
- Probability estimates are not statistically reliable: this applies especially when the number of training cases associated with a leaf is small [Zadrozny and Elkan, 2001]. Suppose our decision tree has a leaf that consists of cases that all belong to the same class. It's difficult to accept that other cases also belong to the same class if they will comply with the constraints imposed by the tests of the decision tree that are needed to come to this particular leaf especially when only a small number of training cases belongs to this leaf [Provost and Domingos, 2000].
- One probability estimate per leaf: a decision tree algorithm assigns the same probability estimate to every case that belongs to the same leaf. We know that every leaf of a decision tree corresponds with a certain decision space. Thus, regardless of the fact that the different cases of a leaf are situated on different places in this space, a decision tree algorithm will appoint to all of them the same probability [Margineantu and Dietterich, 2001].

A possible solution for these problems is the topic of this paper. If the attribute values of a dataset are numeric and a binary decision tree is used to classify the test cases, one can use a sensitivity algorithm to compute the sensitivity or distance between a case and the corresponding decision surface [Alvarez, 2004]. Eventually the computed distance or sensitivity then can be used to rank the test cases.

This paper will compare the results of different ranking methods that are based on either the probability estimate, the sensitivity of a particular case or both. This way it will be possible to find out if the sensitivity algorithm can be used for ranking the test cases classified by a binary decision tree.

2. Experimental Design

This experiment was conducted with the dataset of the data mining competition that took place in the run up to the 29th Annual Conference of the German Classification Society (GFKL 2005): From Data Analysis to Knowledge Engineering. This dataset was provided by the sponsor Deutsche Sparkassen und Giroverband (DSGV) and the goal of the competition was to predict a liquidity crisis based on a subset of 26 variables describing attributes of companies. When the companies were facing a liquidity crisis they should be classified as belonging to class 1 and if they don’t have a liquidity crisis they belong to class 0. Only 11% of the data set has a liquidity crisis. Because the values of these attributes were all numeric, this dataset could be used to test the sensitivity algorithm of Isabelle Alvarez and compare its results with the results of other ranking methods. In order to compare the different ranking methods we used all attributes to select the 2000 companies out of 10000 companies from the test dataset that were most likely to be correctly classified according to the used ranking method. The number of correctly and incorrectly classified cases of the test dataset will then be compared for the different ranking methods so we can draw some conclusions about the use of the sensitivity algorithm.

We used the Weka j48 algorithm [Witten and Frank, 2000] to grow a binary decision tree on the training dataset. From the resulting model the different decision surfaces could be derived. A decision surface can be seen as the boundary between regions with different class labels. Every leaf within the tree has its own decision surface determined by the tests a case has to comply with before belonging to a particular leaf. Because the emphasis of this study lays on the examination of the sensitivity algorithm as a ranking method and not on the development of a state of the art decision model/decision tree, we used the default values for the different thresholds and parameters that are necessary for developing a binary decision tree. Once Weka had developed the binary decision tree it could be applied to the test dataset and after predicting the class value for each case of the test dataset, we were able to apply the earlier mentioned sensitivity algorithm. Two kinds of standardization methods were applied on the attributes, the standard and the minmax method. Both methods are defined with information that could be easily inferred from the dataset itself.

$$y_i^{MinMax} = (x_i - Min_i) / (Max_i - Min_i)$$

$$\text{or} \quad y_i^{Standard} = (x_i - E_i) / S_i$$

with min, max, E,S respectively the minimum, the maximum, the estimated mean value and standard deviation of the corresponding attribute values. The sensitivity algorithm projects a given case onto the decision surface of every leaf that has a different class value than the class value predicted for this case. The algorithm then computes and ranks the distance between the case itself and its different projections. We assume that when the distance between a case and the decision surface of a leaf becomes smaller, the risk of a misclassified instance becomes more realistic. The final step of this experiment is to rank the classified instances by different ranking methods based on the probability estimates and/or sensitivity and compare the results.

3 Methods for Ranking Test Cases Classified by Binary Decision Trees and Experimental Results

In this section, we present the methods we used for ranking the test cases that were classified by a binary decision tree. More specifically we will look at some ranking methods and give in section 4 their impact on selection and ranking. The impact of selection is done by analyzing a) the number of correctly classified cases, b) the number of selected minority cases and c) the correct classification of minority cases. The impact on the ranking is done by analyzing the Roc curves. These ranking methods are based on the probability estimates, the sensitivity of a case or a combination of both. For comparing the different ranking methods we will look at the 2000 best cases. The ranking methods are based on:

- Transformed probability estimate
- Sensitivity
- Piecewise correction function

3.1 Transformed Confidence or Probability Estimates

Within the scope of the data mining competition it was important to correctly classify as much cases of class 1 as possible. These are the cases that either belong to the True Positive quadrant or to the False Negative quadrant. We assume that the cases that belong to the False Negative quadrant are the ones that were classified as class 0 but have a very low probability estimate. To be able to select also these cases we can try to convert the probability estimates of the cases that belong to class 0 to the probability that they would belong to class 1. This can be done with the following formula.

$$P^*0 = 1 - P0$$

With P0 the probability that the case belongs to class 0 and P*0 the transformed probability estimate. We can use this formula for transforming the probability estimates for cases belonging to class 0 because the classified cases can only belong to class 0 or class 1. This is why we can assume that if the case doesn’t belong to class 0, it can only belong to class 1. The probability estimates of the cases belonging to class 1 will not be transformed.

3.2 Sensitivity

The sensitivity of a case from the dataset corresponds with the smallest distance between the case and the corresponding decision surface with different class value. This surface is created by the decision tree. We assume that how smaller the distance, how larger the probability of an incorrect prediction or a wrong classification. For computing the sensitivity we have used the decision surfaces from the decision tree that was developed for the data mining competition. The distance between the classified cases and the corresponding decision surfaces was then computed. Like already stated we are using two different standardization methods for calculating the sensitivity, the Standard and the Minmax method. The Minmax method makes the attribute values laying between 0 and 1 and the standard method standardizes the different values of the attributes.

3.3 Piecewise Correction Function

In the following, we will describe a function that can be used for “correcting” the probability estimates made by a binary decision tree. Because the resulting ranking method has to take into account both the probability estimate and the sensitivity of the concerning/particular case, both factors need to be present in the equation of the correction function. After interpreting the definition of sensitivity and the initial goal of the sensitivity algorithm of Isabelle Alvarez [Alvarez, 2004] we believe that when the sensitivity or the distance from a case to the decision surface is small, the probability of an incorrect prediction or classification will be greater. When the distance is larger there would be, according to us, less chance on an incorrect prediction or classification by the developed decision tree. We assumed that there could be a different correction necessary for a sensitivity value that is below a certain threshold than for a probability estimate whose value is above a specific threshold. Keeping all this in mind we have chosen for a piecewise correction function.

The modified probability estimate will be presented by the symbol *W** and the initial probability estimate by the symbol *W*. We use the symbols *s* and s_0 to indicate the sensitivity and the threshold of the sensitivity. The symbols c_1 and c_2 denote the amount by which the original probability estimate will be altered. The following piecewise function could be used for correcting the original probability estimates.

$$W^* = W + c_1 \text{ if } s > s_0$$
$$W^* = W - c_2 \text{ if } s \leq s_0$$

Note that the modified estimate can be greater then one or less than zero and cannot be considered any more as a probability estimate. After the selection of a threshold s_0 we will examine for this threshold what the effects of different corrections of the probability estimates will be on the results of the ranking. The extent by which the original probability estimates will be corrected varies between 0 and 50 percent. The choice for a particular value for the parameters c_1, c_2 and s_0 will depend on the effect of the ranking of the training data set. A linear search has been applied to maximize the effects on the training data.

The Minmax metric and the Standard metric were treated separately when developing a correction function for the probability estimates given by the binary decision tree because they have for each case a different sensitivity and both metrics also have different maximum sensitivity values. The sensitivity that was computed with the Standard method lies in the interval [0 ; 1,375939965248] and in the interval [0 ; 0,0235140007] for the Minmax method.

4 Selection and Ranking Results

The evaluation of the results from the different ranking methods will be done by comparing the 2000 most likely cases of every applied ranking method with the solution dataset sent to us by the organizers of the data mining competition.

4.1 Selection

The individual selection results of each ranking method will be extracted from the corresponding confusion matrix. The results will be put in a tabular overview as can be seen in Table 1. A confusion matrix shows the number of cases that are correctly classified as class 0 and class 1 and the cases that are incorrectly classified as class 0 and class 1. Because we have only two possible classes in our experiment, the confusion matrix can be depicted as a 2x2-matrix. The classifications that end up in the True Positive and the True Negative quadrant are correctly made classifications, these are cases that were classified as class 0 and class 1 and also really belong to this classes. A case that lies in the False Positive quadrant is a case that is a class 0 case in reality but was classified as class 1. If a case was classified as class 0 but actually belongs to class 1, it will be placed in the False Negative quadrant.

Our quality criteria are

a) TP +TN : the total accuracy of classified cases,
b) TP +FN : the number of selected minority (positive) cases,
c) TP/ (FP+TP) : the accuracy of minority (positive) case classification.

The first criterion is important when both class values are of equal importance. So the bank can handle companies with our without a liquidity crisis in an appropriate way. This is called the accuracy criterion. The second criterion is important when the bank wants to reach all companies with a liquidity crisis (positive case) and it does not matter when companies without a liquidity crisis are also included. This is called the market share criterion. The third criterion is important when the bank wants to reach the companies with a liquidity crisis (positive case) but without reaching companies without a liquidity crisis. This is called the profit share criterion. Next table gives the results.

Table 1: Results of the selection of 2000 test cases using the different ranking methods

Ranking method	Criteria		
	Total Accuracy	Number of minority cases	Accuracy Minority
Transformed probability estimates	70.55	753	0.62
Sensitivity			
Standard metric	89.65	495	0.70
Minmax metric	86.50	501	0.68
Correction function			
Standard metric	79.3	753	0.63
Minmax metric	71.3	753	0.63

From this table we learn that indeed using sensitivity increases the accuracy of the classifier and therefore selects less minority cases (= most difficult to predict). When the goal of the company is to maximize profit or accuracy : the ranking method based on sensitivity is the best method. When the goal of the company is market share : the ranking method based on transformed probability is the best method. Applying a correction function delivers a compromise solution.

4.2 Ranking

To be able to compare the ranking performance of different ranking methods, a single number measure which reflects the ranking performance of the ranking method is needed. The area under the ROC curve (AUC) appears to be one of the best ways [Bradley A.P.,1997], [Ling, C. X., Huang, J. and Zhang, H., 2003]. This ROC space is a coordinate system where the rate of true positives is plotted on the Y-axis and the rate of false positives is plotted on the X-axis for every proportion of the ranked dataset. The true positive rate is defined as the fraction of positive cases classified correctly relative to the total number of positive examples. The false positive rate is defined as the fraction of negative cases classified erroneously relative to the number of all negative examples. From a visual perspective, one point in the ROC curve is better than another if it is located more to the north-west (TP is higher, FP is lower or both) on the ROC graph. Statistical analysis was applied to calculate upper and lower bounds (98% confidence level. The results are given in next table.

Table 2: Results of the Roc analysis using the different ranking methods

Ranking method	Criteria		
	Area under curve	Lower bound	Upper bound
Transformed probability estimates	0.731	0.709	0.754
Sensitivity			
Standard metric	0.887	0.865	0.908
Minmax metric	0.808	0.781	0.834
Correction function			
Standard metric	0.841	0.822	0.861
Minmax metric	0.747	0.725	0.769

These results confirm previous findings. With sensitivity we can better predict the minority cases and as a consequence in the ROC graph these cases are ranked first and a higher “Area under curve” will be obtained as can be seen by comparing the left and right part of figure 1.

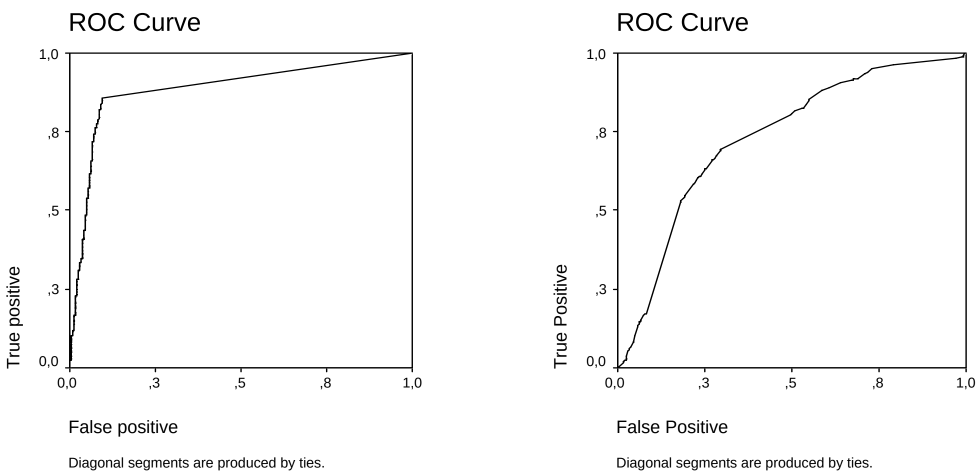


Figure 1: Roc Curve for standard metric (left) and for transformed probability (right)

It is clear that for the sensitivity ranking methods the difference is quite large and statistically significant. The confidence intervals do not overlap with the confidence intervals of transformed probability.

This figure clearly show that the use of probability estimates for ranking the test cases classified by decision trees can be improved by using sensitivity.

5 Conclusion

Based on the comparison of the ranking methods in the previous section, it has become clear that business goal (accuracy, profit or market share) and the way we rank the classified cases will determine the final results. The ranking methods based on sensitivity give the highest total accuracies, the best profit and give the highest Area under curve figures. If we use a ranking method that is based on the sensitivity then the used standardization method will also be of importance for the results. In our experiment the results are optimal when it is based on computations with the Standard metric. The mentioned problems concerning the use of probability estimates for ranking the test cases classified by decision trees are confirmed. When the task is selection of minority cases the position of the cut-off point (2000 out of 10 000 in our case) is also of crucial importance. It seems that decision trees perform well when the cut-off point is far away from the number of minority cases (1113).

Finally, we want to mention that the results of this experiment are valid for the data set under study and can not yet be generalized because further research will be necessary to decide if using the sensitivity of the cases in a dataset is a good basis for the correction of probability estimates of binary decision trees or for the improvement of ranking the results of binary decision trees.

Bibliography

[Alvarez, 2004] I.Alvarez. Sensitivity Analysis of the Result in Binary Decision Trees. In: Proceedings of the 15th European Conference on Machine Learning, Vol 3201, pp 51-62, Springer Verslag.

[Bradley A.P.,1997] The Use of the Area Under the ROC Curve in the Evaluation of Machine Learning Algorithms. Pattern Recognition 1997; 30 (7): 1145-1159.

[Ling, C. X., Huang, J. and Zhang, H., 2003] AUC: a statistically consistent and more discriminating measure than accuracy. In Proceedings of 18th International Conference on Artificial Intelligence (IJCAI-2003), 2003. p. 329–341.

[Margineantu and Dietterich, 2001] D.D.Margineantu and T.G. Diettterich. Improved Class Probability Estimates from Decision Tree Models. Departement of Computer Science, Oregon State University.

[Provost and Domingos, 2000] F. Provost and P. Domingos. Well-Trained PETs: Improving Probability Estimation Trees. Information Systems Department, Stern School of Business, New York University.

[Witten and Frank, 2000] I. Witten and E. Frank. Data Mining: Practical Machine Learning Tools and Techniques with Java Implementation. Morgan Kaufmann.

[Zadrozny and Elkan, 2001] B. Zadrozny en C. Elkan. Obtaining calibrated probability estimates from decision trees and naive Bayesian classifiers. Department of computer science and engineering, University of California.

Authors' Information

Prof. Dr. Koen Vanhoof – Universiteit Hasselt, Campus Diepenbeek, Departement BEDR/VEBI; Agoralaan gebouw D, B3590 Diepenbeek, Belgium; e-mail: koen.vanhoof@uhasselt.be

Sabrina Noblesse – Universiteit Hasselt, Campus Diepenbeek, Departement BEDR/VEBI; Agoralaan gebouw D, B3590 Diepenbeek, Belgium

THE STAPLE COMMODITIES OF THE KNOWLEDGE MARKET

Krassimir Markov, Krassimira Ivanova, Ilia Mitov

Abstract: In this paper, the "Information Market" is introduced as a payable information exchange and based on it information interaction. In addition, special kind of Information Markets - the Knowledge Markets are outlined. The main focus of the paper is concentrated on the investigation of the staple commodities of the knowledge markets. They are introduced as kind of information objects, called "knowledge information objects". The main theirs distinctive characteristic is that they contain information models, which concern sets of information models and interconnections between them.

Keywords: Information Market, Knowledge Market, Knowledge Information Objects, General Information Theory

ACM Classification Keywords: K.4 Computers and Society – K.4.0 General; K.4.4 Electronic Commerce

> "The speaker doesn't deliver his thought to the listener, but his sounds and performances provoke the thought of the listener. Between them performs a process like lighting the candle, where the flame of the first candle is not transmitted to another flame, but only cause it."
>
> *Pencho Slaveikov, Bulgarian poet,*
> *the beginning of the XX-th century*

Introduction

The main characteristic of the Information Markets is payable information exchange and based on it information interaction. Special kinds of Information Markets are the Knowledge Markets. The main goal of this paper is to continue the investigation of the Knowledge Markets started in [Ivanova et al, 2001], [Markov et al, 2002]. Now, our attention will be paid to the staple commodities of the Knowledge Markets. The usual talk is that at the Knowledge Market one can buy knowledge. But, from our point of view, this is not so correct.

The investigation presented in this paper is based on the *Theory of Information Interaction*, which is one of the main parts of the *General Information Theory* [Markov, 1984], [Markov, 1988], [Markov et al, 1993], [Markov et al, 2003].

Firstly, let remember some basic concepts of the Theory of Information Interaction.

At the first place, we need to remember the concept “INFOS”. Its genesis started from the understanding that the concept *"Information Subject"* is perceived as a single human characteristic. It is clear that in the nature there exist many creatures, which may be classified to this category, especially groups of persons and societies. To exclude the misunderstandings we decide to introduce new word to denote all possessors of the characteristics of the Information Subject. This word is **"***INFOS***"** [Markov et al, 2003], [Markov et al, 2004].

On given level of complexity of the entities, a new quality becomes **-** the possibility of self-reflection and internal activity appears. One special kind of activity is the secondary (information) one. The secondary activity need to be resolved by relevant possibilities of the entities from the environment. So, not every entity may be used for resolving the secondary activity. This way, the entity needs a special kind of (information) contacts and (information) interaction for resolving the information activity.

The entity, which has:

– *(primary) activity* for external interaction;

– possibility for ***reflection***, i.e. possibility for collecting the information;

– possibility for *self-reflection*, i.e. possibility for generating secondary (information) activity;

– *information expectation* i.e. available (secondary) information activity for internal or external contact for resolving it

is called *Infos*.

The resolving of the information activity is *the goal* of the Infos. This goal may be achieved by the establishment and providing (information) contacts and (information) interaction, which are remembered below.

Information Objects

When the Infos interact with the entities around in the environment, there exist at least two cases of reverberation:

– the contacts and interaction are casual and all reflections in the Infos as well as in the entities have casual origin;

– the contacts and interactions are determined by the information activity of the Infos.

In the both cases, the contacted entity may reflect any information model from Infos. The concept “information model” has been defined in [Markov et al, 2001]. In general, the information model is a set of reflections, which are structured by Infos and, from his point of view, represents any entity.

An entity, in which one or more information models are reflected, is called *"information object".*

The information objects can have different properties depending on:

– the kind of influence over the entities **-** by ordering in space and time, by partial or full modifying, etc.,

– the way of influence over the entities **-** by direct or by indirect influence of the Infos on the object,

– the way of development in time **-** static or dynamic,

etc.

Information Operations

The information is reflected relationship, i.e. it is a kind of reflection [Markov, 1988]. Therefore, the only way Infos to operate with information is to operate with the entity that contains it. Every influence on the entity may cause any internal changes in it and this way may change the information already reflected. Another type of influence is to change the location of entity or to provoke any contact between given entity and any other.

The influence over the information object is called "***information operation***" if it is determined by any Infos information activity.

The information operations may be of two main types:

– the Infos internal operations with the sub-entities that contain information,

– external operations with the information objects that contain information.

The internal operations with the sub-entities closely depend of the Infos' possibilities for self-reflection and internal interaction of its sub-entities.

The self-reflection (self-change) of the Infos leads to the creating of new relationships (and corresponding entities) in it. These are *subjectively* defined relationships, or shortly - *subjective relationships*. When they are reflected in the memory of the Infos they may initiate any new information model on a higher level. In such case, a relation between reflected relationships appears. The high-level information models may have not real relationships and real entities that correspond to them.

For instance, the possibility for creating the information models of similarity is a basis for realising such very high level operations as "comparing elements or substructures of the information models", "searching given substructure or element pattern in the part or in the whole structure of the information model", etc.

It is clear, the Infos is built by entities some of which may be also Infoses, but on the lowest levels. For instance the society and single human who belongs to it. So, the internal operations are determined by the concrete internal level and from the point of view of these low levels, they may be assumed as external operations. Because of this, we will concentrate out attention on the second type of operations.

The external operations with information objects may be differed in two main subtypes:

– basic information operations;

– service information operations.

There are two basic information operations which are called I-operations:

✓ I-reflection (reflecting the information object by the Infos, i.e. the origination of a relevant information model in the memory of the Infos);

✓ I-realisation (creating the information object by the Infos).

In the process of its activity, the Infos S reflects (perceives) information from the environment (entities $O_{I,}$ i=1,2...) by proper sub-entities (sensitive to video, acoustic, tactile, etc. influences) called "***receptors***" R_i (i=1,2...). Consequently, the Infos may receive some information models.

The Infos subjective reflection is called "*I-reflection*".

When necessary, the Infos can realise in its environment (entities O'_j, i=1,2...) some of the information models, which are in his memory, using some sub-entities called "***effectors***" E_j (j=1,2...). Consequently, new or modified already existing entities may reflect the information, relevant to these information models.

The Infos subjective realisation is called ***"I-realisation"***.

There are several operations, which can be realised with the information objects: transfer in space and time, destroying, copying, composition, decomposition, etc. Because of the activity of the Infoses, these operations are different from other events in reality. In this case, such Infos determined operations with information objects are called "service information operations".

For example, some of the very high-level service operations are based on the external influence on the information object to change any existing reflection:

✓ Including and removing an element in and from the object's structure;

✓ Copying or moving object's substructures from one place to another;

- ✓ Building new object's structure using as a basis one or several others;
- ✓ Composing or decomposing of object's elements or substructures;

Etc.

Information Processes

Let "O" is a set of real information objects i.e. $O = \{O_{ij} \mid i= 1,..,n; j=1,..,m\}$.

Let "I_s" is a set of information models in Infos S, i.e. $I_s = \{i_p \mid p=1,..,q\}$.

If the opposite is not stated, we will consider:

- every set of information objects is an information object,
- every set of information models is an information model.

Every information operation "t" can be treated as a function between two sets of information objects, which may be coincidental, i.e. $t: O_d \rightarrow O_r$.

I-realisation can be considered as a function $m: I_s \rightarrow O$.

I-reflection - in the opposite - as a function $r: O \rightarrow I_s$.

Let $t_1, t_2,..,t_n$ are information operations. The consequence of information operations P created by any composition, i.e.

$$P = t_1 \circ t_2 \circ \ldots \circ t_n$$

is called "*information process*".

It is possible that some of t_i, i=1,..,n may be I-realisation or I-reflection.

In particularly an information process can include only one operation.

Information Contact

If an information model from the Infos is reflected in another entity, there exist possibility, during the "a posterior" interactions of the given entity with another Infos, to transfer this reflection in it. This way an information model may be transferred from the Infos to another.

If the second Infos has already established information expectation, the incoming reflection will be perceptible for him. The information expectation will be resolved in some degree and the incoming information model and information in it will be received by the second Infos.

Let S_1 and S_2 are Infoses and O is an arbitrary entity.

The composition of two real contacts Θ_1 and Θ_2 :

$$S_1 \xrightarrow{\Theta_1} O \xrightarrow{\Theta_2} S_2$$

is called "*information contact*" between Infos S_1 and Infos S_2 iff during the contacts any information model from I_{S1} is reflected in the I_{S2} through the entity O.

The Infos S_1 is called "*information donor*", the Infos S_2 is called "*information recipient*", and, of course, the entity O is called "*information object*".

In this case, when the donor and the recipient are different Infoses the information contacts between them consist of the composition of at least two information operations - I-realisation and I-reflection. For the realisation of any direct information contact between two different Infoses is necessary the execution of the composition of these two "basic" operations. All the rest information operations are necessary for supporting the basic ones i.e. they are auxiliary (service) operations.

This way the elementary communicative action will be provided.

In general, every information process "c", having as a start domain the set I_{Sd} of information models of the Infos S_d and as a final domain the set I_{Sr} of information models of the Infos S_r, (I_{Sd} and I_{Sr} may be coincidental)I,

$$c: I_{Sd} \rightarrow I_{Sr}$$

is called ***"information contact"*** between S_d and S_r:

Note that for the realisation of one information contact at least one information object is necessary.

Information Interaction

The set "R" of all information contacts between two Infoses S_a and S_b

$$R= \{c_i \mid i=1,2..; c_i:I_{Sa}\rightarrow I_{Sb}\}$$

is called "*information interaction*".

When S_a and S_b are coincident, we call it Information interaction of the Infos with itself (through the space and time).

The set "B" of all information objects, used in the information interaction between given Infoses is called "*information base*".

Information Society

The "*Information Group*" (IG) is a set of Infoses, with common Information base of the information interactions between them.

In the small IG the service information operations may be provided by the every Infos. In the large IG this is impossible or not optimal. In such case, some Infoses became as "*information mediators*" between the others. They start to provide the service information operations. They realise "*Information Service*".

The "Information Society" is an IG with internal Information Service.

Information Market

Now we are ready to continue with introducing the basic ideas of the Information Markets.

Up to this moment, the discussion about essence of the information society has not resulted in uniform definition. Everyone from his point of view defines this stage of development of a society.

It is clear, at the stage of social growth, called “information society”, for existence of the separate individuals or social teams the information and information activity get decisive value. Certainly, at earlier stages of development of mankind, the information had the important value too. But never, in all known history, the other means for the existence have been so dominated by the information means as it is in the information society.

So, the direct conclusion is the understanding that *the information society differs from the other levels of the human been growth by the domination of the information interests above all others.*

From the origin, the human society has been the "information" one, but the levels of the information service differ in the different periods of the existence of the societies. So, it is possible to allocate the following levels:

- ✓ Primitive information society (people having knowledge, letters on stones etc.);
- ✓ Paper information society (books, libraries, post pigeons, usual mail etc.);
- ✓ Technology information society (telephone, telegraph, radio, TV, audio- and video-libraries etc.);
- ✓ High-Technology information society (automated systems of information service, local computer information networks etc.);

- ✓ Global information society (global systems for information service, opportunity for every body to use the information service with help of some global network etc.).

The information society does not assume compulsory usage of the information services by the part or all inhabitants of given territory. One very important feature thus is emphasized: for everyone will be necessary diverse and qualitative (from his point of view) information, but also everyone can not receive all necessary information. The enterprising experts will accumulate certain kinds of the information and will provide the existence through favourable to them information exchange with the members of the society. Thus, in one or other form, they will carry out payable information service (granting of information services for some income) [Ivanova et al, 2001]. This is the background of the Information Market.

The payable information exchange and services regulated by the corresponded laws and norms as well as by the government protection of the rights of the participants (members) of this kind of social interactions form the *Information Market*.

So, at the centre of discussion, we have discovered a simple true: *in the information society the payable information exchange and services will dominate above all other market activities.* In other words, the Information Market dominates over all other type of markets of the information society.

Knowledge Information Objects

V.P. Gladun correctly remarks that the concept “knowledge” does not have common meaning, especially after beginning of it’s using in the technical lexicon in 70-ies years of the last century. Usually, when we talk about the human knowledge we envisage all information one has in his mind. Another understanding sets the “knowledge” against the “data”. We talk about data when we are solving any problem or are making logical inference. Usually the concrete values of the given quantities are used as a data as well as the descriptions of the objects and interconnections between objects, situations, events, etc. During decision making or logical inference we operate with data involving some other information like descriptions of the solving methods, rules for inference of the corollaries, models of the actions from which the decision plan is formed, strategies for creating decision plans, and general characteristics of the objects, situations, and events. In accordance with this understanding, the “knowledge” is information about processes of decision making, logical inference, regularities, etc., which applying to the data creates any new information. [Gladun, 1994].

The usual understanding of the verb "to know" is: "to have in the mind as the result of experience or of being informed, or because one has learned"; "to have personal experience of smt.” etc. The concept "knowledge" usually is connected to concepts "understanding" and "familiarity gained by experience; range of information" [Hornby et al, 1987] or "organized body of information" [Hawkins, 1982].

In other words, the knowledge is a structured or organised body of information models, i.e. the knowledge is information model, which concerns a set of information models and interconnections between them.

In accordance with this the information objects, which contain such information models are called *“knowledge information objects”*.

This definition corresponds to everyday understanding of the concept “knowledge”. For instance, during the process of education the presented above operations I-realization and I-reflection correspond to creating and perceiving the *“knowledge information objects”*.

Knowledge Market

The growth of the societies shows that the knowledge information objects become important and necessary articles of trade. The open social environment and the market attitudes of the society lead to arising of the knowledge customers and knowledge sellers, which step-by-step form the "Knowledge Markets" [Markov et al, 2002].

As the other markets, the Knowledge Market is the organised aggregate of participants, who operate following common rules and principles. The knowledge market structure is formed by a combination of mutually-connected elements with simultaneously shared joint resources.

The staple commodities of the knowledge market are the knowledge information objects.

The knowledge information bases and tools for processing the knowledge information objects, such as tools for collecting, storing, distributing, etc., form the knowledge environment.

The network information technologies enable to construct uniform *global knowledge environment*.

It is very important, that it will be friendly for all knowledge market participants and open for all layers of the population without dependence from a nationality, social status, language of dialogue, place of residing. The decision of this task can become the important step of humanization of all world commonwealths.

In the global information society, on the basis of modern electronics, the construction of the global knowledge market, adapted to the purposes, tasks and individual needs of the knowledge market participants is quite feasible, but the achievement of this purpose is connected to the decision of a number of scientific, organizational and financial problems.

For more clear explanation let consider an example about the correspondence between concepts “information object” and “knowledge information object”. When an architect develops any constructive plan for future building, he creates a concrete “information object”. Of course, he will sell this plan. This is a transaction in the area of the Information Market. Another question is from where the architect has received the skills to prepare such plans. It is easy to answer – he has studied hardly for many years and received knowledge is the base for his business. So, we see that the textbooks are not concrete information for building concrete house, but they contain the information needed for creating such plans. The textbooks written by the lecturer in the architectural academy are special kind of “information objects” which contain special generalized information models. They are “knowledge information objects” and these textbooks have been sold to the students. It is clear; here we have a kind of transactions at the “Knowledge Market”.

At the end, we need to take into consideration the difference between responsibility of the architect and the lecturer. If the building collapses the first who will be responsible is the architect, but never the lecturer!

Conclusion

In this paper, we introduced the “Information Market” as a payable information exchange and based on it information interaction. In addition, special kind of Information Markets - the Knowledge Markets were outlined.

The identifying of the staple commodities of the knowledge markets was a step of the process of investigation of contemporary situation in the global knowledge environment.

The investigation of the staple commodities of the knowledge markets is very difficult but useful task. In this paper we introduced them as kind of information objects, called “knowledge information objects”. The main theirs distinctive characteristic is that they contain information models, which concerns sets of information models and interconnections between them.

This way, we have seen the usual talk that at the Knowledge Market one can buy knowledge is not so correct. But in everyday language in is accepted to say “knowledge” with the meaning of the “knowledge information object”. We need specially to say that there exists another meaning of knowledge, which points to the information models into the Infos memory. Usually we do not distinct these two meanings. When we say “receiving of knowledge” we assume the I-reflection operation; when we say “generating the knowledge” we assume the I-realization operation; and at the end, when we say simply “knowledge” it is context depended to understand what is the meaning – knowledge information models into the Infos memory or those which the knowledge information objects contain. However, if we remember that the Infoses are built by entities some of which may be also

Infoses, but on the lowest levels, the internal memory of given level of the organisation of the Infos may be considered as external set of information objects on the lower levels.

This has important role for future research of this social and information phenomenon.

Acknowledgements

This paper is partially supported by the project ITHEA-XXI.

Bibliography

[Gladun, 1994] Гладун В.П. Процессы формирования знаний. София, Педагог 6, 1994.

[Hawkins, 1982] Hawkins J.M., *The Oxford Paperback Dictionary*. Oxford University Press, 1982, ISBN 0-19-281209-2.

[Hornby et al, 1987] Hornby A.S., A.P. Cowie, A.C. Gimson, *Oxford Advanced Learner's Dictionary.* Oxford University Press. 1987, ISBN 0-19-431106-6.

[Ivanova et al, 2001] N. Ivanova, K. Ivanova, K. Markov, A. Danilov, K. Boikatchev. *The Open Education Environment on the Threshold of the Global Information Society*. IJ ITA, 2001, V.8, No.1 pp.3-12. (Presented at Int. Conf. KDS 2001 Sankt Petersburg, 2001, pp.272-280, in Russian, Presented at Int. Conf. ICT&P 2001, Sofia, pp.193-203)

[Ivanova et al, 2003] Кр.Иванова, Н.Иванова, А.Данилов, И.Митов, Кр.Марков. *Обучение взрослых на рынке профессиональных знаний*. Сборник доклади на Национална научна конференция "Информационни изследвания, приложения и обучение" (i.TECH-2003), Варна, България, 2003. Стр. 35-41.

[Markov, 1984] K.Markov. *A Multi-domain Access Method.* Proceedings of the International Conference on Computer Based Scientific Research. PLOVDIV, 1984, pp.558-563.

[Markov, 1988] Кр.Марков. *От миналото към бъдещето на определението на понятието "информация".* Сборник "Програмиране '88". БАН, Варна, Дружба, 1988, стр.150.

[Markov et al, 1993] K.Markov, K.Ivanova, I.Mitov. *Basic Concepts of* a *General Information Theory.* IJ ITA {International Journal "Information Theories & Applications"}; 1993; v.1, n.10; pp.3-10.

[Markov et al, 2000] Kr. Markov, Kr. Ivanova, I. Mitov, N. Ivanova, K. Bojkachev, A. Danilov. *Co-operative Distance and Long-Live Learning*. ITA-2000, Bulgaria, Varna, 2000.

[Markov et al, 2001] K. Markov, P. Mateev, K. Ivanova, I. Mitov, S. Poryazov. *The Information Model*. IJ ITA, 2001, V.8, No.2 pp.59-69 (Presented at Int. Conf. KDS 2001 Sankt Petersburg, 2001, pp.465-472)

[Markov et al, 2002] K. Markov, K. Ivanova, I. Mitov, N. Ivanova, A. Danilov, K. Boikatchev. *Basic Structure of the Knowledge Marke*t. IJ ITA, 2002, V.9, No.4, pp.123-134 (Presented at Int. Conf. ICT&P, 2002, Приморско)

[Markov et al, 2003] Kr.Markov, Kr.Ivanova, I.Mitov. *General Information Theory. Basic Formulations.* FOI-COMMERCE, Sofia, 2003. ISBN 954-16-0024-1.

[Markov et al, 2004] K. Markov, K. Ivanova, I. Mitov, E. Velikova-Bandova. *Formal Definition of the Concept "INFOS"*. Int. Journal "Information Theories and Applications", 2004, ISSN 1310-0513. V.11, N.1, pp.16-19. (Presented at i.TECH 2004, Varna, Bulgaria. pp. 71-74)

Authors' Information

Krassimir Markov – Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Institute of Information Theories and Applications FOI ITHEA, foi@nlcv.net

Krassimira Ivanova – Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, foi@nlcv.net

Ilia Mitov – Institute of Information Theories and Applications FOI ITHEA, foi@nlcv.net

BASIC INTERACTIONS BETWEEN MEMBERS OF THE KNOWLEDGE MARKET

Krassimira Ivanova, Natalia Ivanova,
Andrey Danilov, Ilia Mitov, Krassimir Markov

Abstract: The interconnections and information interactions between main members of the Knowledge Market are presented in the paper.

Keywords: Knowledge Market, Components of the Knowledge Market, Information Interaction

ACM Classification Keywords: K.4 Computers and Society – K.4.0 General; K.4.4 Electronic Commerce

Introduction

The growth of the global information society shows that the information and, especially – knowledge, becomes important and necessary article of trade. The open environment and the market attitudes of the society lead to arising of the knowledge customers and knowledge sellers, which step-by-step form the "Knowledge Markets". As the other markets the Knowledge Market is the organised aggregate of participants, which operates in the environment of common rules and principles [Markov et al, 2002].

The Structure of the Knowledge Market was presented in [Markov et al, 2002]. Let's remember its basic elements.

Usually a person or enterprise, called Employer (Er), hires Employees (Ee), who have exact skills and knowledge and transform them in real products or services during the work processes. This process is served by the Manpower Market. But the Employees, even owning a high education level, need additional knowledge to solve the new tasks of the Employers. In this moment they became customers of new knowledge, who arouse the necessity of the Knowledge Market, which should rapidly react to the customers' requests. In other words, the Manpower's Market causes the appearance of the Knowledge Market. These two members of KM form one side of the market – the knowledge customers.

The continuous changing of technological and social status of the society leads to appearance of new category – Consultants (C) – peoples/organisations, who have two main tasks:

- to promote new technologies to Employers in convenient way to implement them in practice;
- to determine the educational methods for training the staff for using the new technologies.

The educational process is carried out by the Lecturers (L), who transform new scientific knowledge into the pedagogical grounded lessons and exercises.

During the realising the concrete educational process the Lecturer is assisted by Tutor (T) who organises the educational process and supports the Employees to receive the new knowledge and to master theirs skills.

At the end of the educational process, a new participant of KM appears – Examiner (E) – who tests the result of the process and answers to the question "have the necessary knowledge and skills been received".

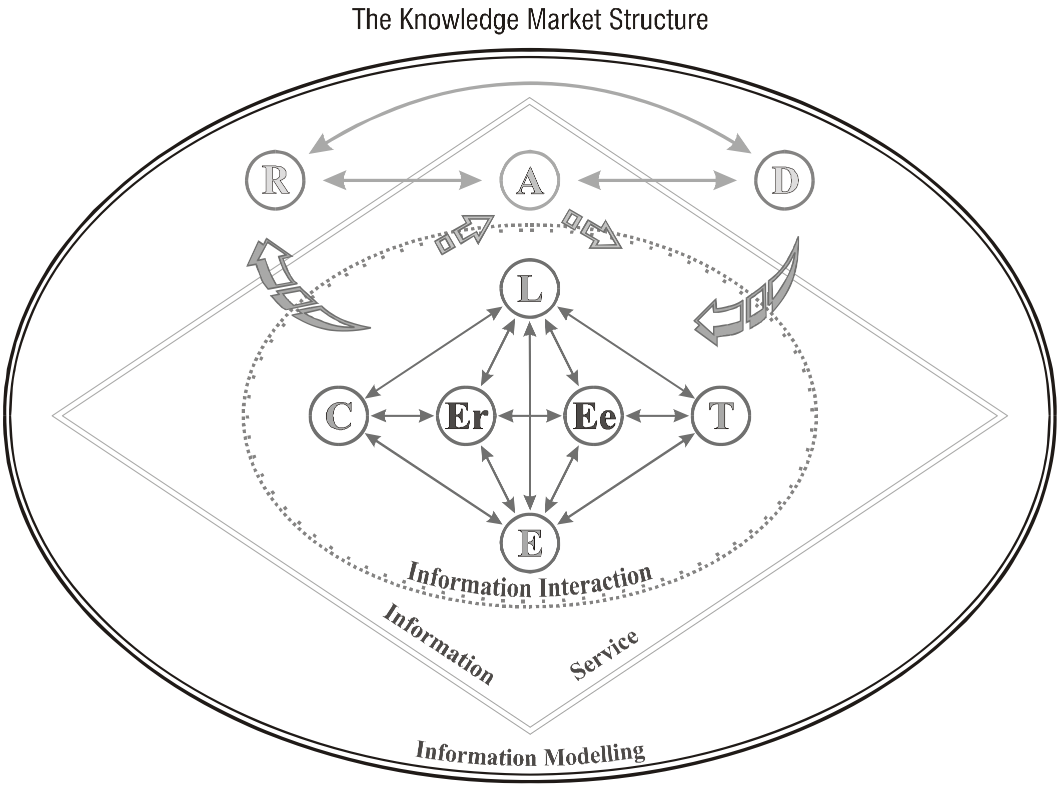
These six Components of the Knowledge Market, which contact each other via global information network, form the first level of the knowledge market, called the level of the "information interaction".

As far as these components are too much and are distributed in the world space, the organisation and co-ordination of theirs information interaction needs adequate "information service". It is provided by a new component called Administrator (A). Usually the Administrators are Internet and/or Intranet providers or organisations.

The rising activity of the knowledge market creates the need of developing modern tools for the information service in the frame of the global information network. This causes the appearance of the high knowledge market level, which allows the observing the processes, as well as developing and implementing new systems for information service. This is the "information modelling" level.

It consists of two important components – the Researchers (R) and the Developers (D).

On the figure below is presented the scheme of the basic structure of Knowledge Market.



Of course, the Knowledge Market as a kind of Market follows the rules and laws given by the social environment. The interrelation between government and social structure and Knowledge Market need to be studied in separated investigation.

In several papers we have already investigate different problems of the Knowledge Market.

Some of the results given below are received during these works. For five years we have seen that the Knowledge Market is very important scientific area and need to be investigated. The main received results are given in [Markov et al, 2000a], [Markov et al, 2000b], [Ivanova et al, 2001], [Boikatchev et al, 2001a], [Boikatchev et al, 2001b], [Markov et al, 2002], [Ivanova et al, 2003], [Markov et al, 2003].

In global information society the e-commerce becomes as fundamental way for financial support of the Knowledge Market. The advantages of e-commerce are obvious. In the same time there exist many risks for beginners at this kind of market. From this point of view the society need to provide many tasks for training the citizens to use properly opportunities of the new environment [Markov, 1999].

The Interconnections between Members

In this section, the interrelations between members of the Knowledge Market are outlined. For easy reading we will denote the Knowledge Market's members with abbreviated names as they were given in the structure above. The next convention concerns the style of description of the corresponded interconnections. We consider that in every such interconnection the both participants have equal rights and the order of writing is not important. For instance, the denotation "L – T" represents the interconnection between Lecturers and Tutors without any precedence of any of them.

<u>Er – Er</u>

The group of the Employers in given domain causes the growing of the Knowledge Market in this direction.

The products of some Employers develop the new technologies, which are used from other Employers. This causes the development of the Knowledge Market in these new directions.

The role of the Employer from the point of view of the Manpower market is to buy and hold more qualified and skilled workers for less money in competition with other Employers.

<u>Er – Ee</u>

The Employer buys the result of the work of the Employee, which is closely depended to the received knowledge and skills, in one hand, and the possibilities of the Employee to realise them in concrete results in everyday work, on other hand.

The Employee informs the Employer for already received knowledge and skills or his intention for future. He need to prove this knowledge by correspond qualification procedure.

The Employee has to realise the requested knowledge and skills during the working process, because the Employer pays for the real results of the work, but not for the possibilities.

<u>Er – C</u>

The Employer can take advises from the Consultant for optimisation or future expansion of his business.

The Employer may ask the Consultant about information for the other components of the Knowledge Market or for the Knowledge Market as whole.

Special Interest for the Employer may be the data about the degree of the knowledge and skills of the Examiners, because of the importance of theirs role for the employers business.

One of the main tasks of the Consultants is to advise the Employers who can be the appropriate Lecturers or Examiners for the manpower they need.

During the interaction with the Consultant the Employer defines his intentions for the future of his business. This way the Consultant receives information about the future growing of the Knowledge Market. The Consultant investigates this growing by the interaction with the two types of Employers:

- Manufacturers of new products and technologies, which production will income on the real market;
- Users of the existing products and technologies.

The first type of Employers grows up a new niche of the Knowledge Market – the knowledge about using, service, etc. of theirs new products, technologies or theories.

The second type influences over the increasing or decreasing of that niche. From other side, the investigation of this group can focus the attention to the new problem, which solving may generate a new niche in the Knowledge Market.

Er – L

The Employer may consult the Lecturer about the present' and future' needs and changes of the business, he manages. The initiative for such consulting process usually belongs to the Lecturer, or to the Consultant. The Employer becomes as the corrective of the futures Lecturers' work.

During the interaction with the Employers, the Lecturer receives theirs:

- requests for training and additional education of new skilled Employees;
- requests for consulting the specialists already appointed in the enterprises of the Employers;
- requirements for preparing adequate educational programs and courses for theirs Employees;
- In the same time, the Lecturer gives information to the Employers for the level of the skills received by the Employees during the implemented courses.

Er – T

The direct interactions between Employer and the Tutor are not so strong, i.e. the direct interaction between the Tutor and the Employer is not always present.

In one side, the Tutor organises the educational process in correspondence to the requirements of the Employers' business work environment. Usually the Consultant defines these conditions directly or through the Lecturer. The Employer defines the concrete duties, which usually are subset of the Consultant' recommendations.

Meanwhile the Tutor contacts directly with the Employer, if it's a question of consultation, check of the level of the skills and knowledge or re-education of already hired Employees.

The Tutor has to establish appropriate educational environment for training the Employees to satisfy given requirements.

In other side, the Tutor can receive the concrete information for the current Employees, which information may facilitate the learning process.

Er – E

The interaction between Employer and Examiner is aimed to define the specific characteristics, which the Employee has to satisfy. These characteristics are subset of the general requirements and rules for given business activity. Usually the Consultant in collaboration with the Employer generates the definition of the specific rules. It is possible, during the process of qualification of the Employee, the set of rules to be changed in direction to extend or to shrink in correspondence to:

- The real needs and possibilities of the Employer;
- The set of candidates for given workplace.

The Employer can ask the Examiner to examine some of the employees, which already work for the Employer in accordance with the current (or near-future) duties-rules.

The Examiner informs the Employer for the Employees results in the education course.

Er – A

It is clear that the concrete needs of the given Employer are subset of the general Employers' information model. In every concrete case as well as for every concrete event the Employer needs specialised administration support.

The Administrator interacts directly with the Employer for:

- co-ordinating and service the information interaction between him and other Employers as well as with all other Knowledge Market participants;

- advising him how to use the information service tools to establish interconnections;
- placing the tools and lines for distant information exchange at Employers' disposal;
- supporting the creation of the specialised internet and/or intranet sites and other communication possibilities for representing the Employers;
- Software and hardware support of Employers' activities.

Er – R&D

There are several sources for financing the scientific research and its implementation in the Knowledge Market.

The main sources are the Employers. They may order specialised research and developing of correspond tools, which may extend service of the Knowledge Market in the new directions.

Another source is the government or public financial support. In such case the corresponded authorised institutions became as specific Employers but the scientific results could not be implemented immediately. This possibility is very important because of the increase in the production of scientific "garbage".

As corollary we have two types of interaction between Researchers, respectively – Developers, and Employers:

- Interaction based on concrete request and corresponded scientific or practical projects;
- Interaction determined by future prognostic government, social or scientific goals, needs, expectations, etc.

Ee – Ee

The connections between Employees are mainly in two dimensions:

- collaboration with others during the educational process;
- competition with others in the Manpower market as well as during the educational process.

The collaboration is very important because of possibility to receive new knowledge from “colleagues” and the need to discuss in free style the problems arisen from the tasks given by the Lecturer and the Tutor.

Competition has significant role in growing the personal knowledge and skills.

Ee – C

The Employee receives information from the Consultant usually through the other components of the Knowledge Market and in rare case directly. This information concerns the present and future growth of the Manpower and Knowledge Markets and serves the orientation of the Employee in the social environment. This knowledge may help the Employee to make decision for his future professional and social growing.

In addition, the Consultant makes studies on the state of the Employees’ commonwealth and generates conclusions about the advance or reduction of any parts of Knowledge Market.

Ee – L

From point of view of the Employee the Lecturer is integrated set of education sources, from which he can receive general or specialized knowledge and skills.

The Lecturer transfers knowledge to the Employees, using modern methods and devices of delivering. He participates in development of the training program, study materials, study tasks and exercises that the Employees should be carried out during the training before beginning of the educational process.

During preparation of a course the Lecturer should give study materials, which can be written in appropriate formats and media and the references to the additional sources of information.

The Lecturer may be connected with each Employee directly and has access to the discussion forum and the Employees’ diaries where he can estimate the Employees’ knowledge and make the comments.

Ee – T

One very important point of the interaction between the Employee and the Lecturer is to determine the currently received knowledge and skills level of the Employee for choosing the appropriate approach and capacity of the material for education. This process needs to be supported by the Tutor.

The tutor directs and plans the education of the Employee. He also plans and organizes Employee's interaction with the Lecturer and informs the Employee about all the courses that he can get from the Lecturer.

The Employee demands from the Tutor maximal effective organization of the educational process for acceptable expenses.

Let's remark that the concept of acceptable expenses closely depends of financial status of the Employee. These expenses include not only concrete charges for education, but also the charges for existence of the Employee during the educational process.

Ee – E

The Employee must prove to the Examiner that his received knowledge and skills cover chosen qualification level. If the Employee does not cover all the requirements, the Examiner need to explain in appropriate form what the Employee have to learn in addition. These explanations need to be directed to the Tutor (respectively to the Lecturer) for future extension of the education of this Employee, his group or whole educational process.

The Examiner conducts the Employee's attestation on different levels of the education course. Within the interaction with the Employee, he gives him the terms and the conditions of the test or exam Employee's about take. The Examiner sets the tests and gets the Employee's answers.

At the end, the Examiner takes a decision for a final or level test and gives the Employee a certificate based on the tests.

Ee – A

The Employee interact with the Administrator for receiving the appropriate information service for every of given above information interactions.

The concrete needs of the given Employee are subset of the general Employee' information model.

The specific characteristics of the Employee such as physiological characteristic, language specific, professional background, national and religious affiliations, the level of skills for using the information service, etc. strongly influence to the type and degree of the automated information service, provided and conducted by the Administrator.

Ee – R&D

The interaction of the Employee with the Researches and Developers mainly is based on the using of the tools and approaches for information service of the education and self-education.

In this interaction the Employee usually play passive role.

As a rule, the Researchers carry out specialized research or general investigation and developing of corresponding tools for information service. These tools are provided for the specific characteristic of given group of Employees or, respectively, for common (future) needs of the Employees' group.

C – C

The interaction between Consultants may be structured at least on three levels:

- scientific level;
- business level;
- state (government) and social level.

At the scientific level, the Consultants interact to investigate new theoretical or practical domains and to generate new common knowledge. As a rule, this interaction is based on the scientific norms for collaboration and exchange of the new ideas. Some political or socioeconomic processes as well as the business restrictions in the given field may restrict this interaction.

At the business level the main goal of interaction between Consultants is to extend already existing knowledge for decision of concrete practical problems. It is important to remark that often the experts in the same domain do not collaborate because of the concurrency between them.

At the state (government) and social level the Consultants play subordinate role to the Government or Social Institutions, such as the parliament, ministry, syndicates, associations, foundations etc. The knowledge of the consultants plays advising role for the decision making of the governing organization or institution. In such way this knowledge became active in the Knowledge Market.

C – L

The interaction between Consultants and Lecturers has significance place in the Knowledge Market. The Consultants are the sources of the new knowledge and the Lecturers need to be continuously in touch with them.

In many cases the Lecturers collect parts of knowledge from different Consultants and integrate it in educational materials. This way the Consultants may affect to the educational processes in the Knowledge Market.

Beside of this, the Consultants may influence the Knowledge Market by theirs ability to investigate the results of the previous cycle of education and to draw conclusions for eventual corrections in the educational processes.

At the end, the Consultants are the persons or organisations that can and have to test and certify the Lecturers in given knowledge domain.

C – T

At glance, the Consultants and the Tutors not interact at all.

The interaction may be realised indirectly by other components of the Knowledge Market – mainly by the Lecturers.

Nevertheless, they exchange via Knowledge Environment much information for practical implementation of the Consultants recommendations for Employees training.

C – E

The main goal of the interaction between Consultants and Examiners is to clear the real needs of the Employers for high qualified and skilled workers. The Examiners need exact information what the Employees really have to know. The qualification procedure includes large scale of tests and other examination steps. For every one, the Consultant may be the source of the knowledge.

In this interaction the Consultants may influence by:

- conclusions for the results of the previous qualification cycles;
- certifying the Examiners in given knowledge domain;
- playing role of the Examiner.

The Examiner receives from the Consultant new test materials and education methods of conducting the attestation activities and informs the Tutor for the results of the use of the new attestation materials and methods.

C – A

The Consultant interact with the Administrator for receiving the appropriate information service for every of given above information interactions.

A specific characteristic of this interaction is the ability of the Consultant to interrupt any of the educational processes in the Knowledge Market via the Administrator's support. This interruption need to be done when from point of view of the Consultant:

- any new events, external for the educational process, appear and the structure of given steps of the educational process need to be changed;
- any deviations in the educational process have registered by the Consultant.

In this role the Consultant can be qualified as a regulator of the Knowledge Market. This is very important that the main way the Consultant can play this role is by the support of the Administrator.

C – R&D

The interaction of the Consultants with the Researchers and Developers usually is provided in two directions:

- the Consultants became as a source for information modelling of theirs work;
- the Consultants assist the creation and verify the implementation of the information models of the other components of the Knowledge Market. In this interaction the Consultant usually play active role.

Let's remark that the second function is very important not only of the work of the Researchers and Developers, but also for the stability and growing of the Knowledge Market.

The interaction of the Researchers and Developers with the Consultants is caused by:

- the needs of the scientific information modelling of the consultants' work as well as the activities of all connected to them Knowledge Market subjects;
- the implementation of the information models – in this case the Consultants need to play active role to assist the creation and to verify the realisations of scientific models.

The integration of the Researchers' and Developers' knowledge with the Consultants' one is very important for the making correct and appropriate solutions.

L – L

The main goal of the interaction between Lecturers is to deliver concrete material and pedagogical methods as well as to exchange theirs educational experience.

In other hand, the competition between Lecturers in one and the same subject can harm the Knowledge Market stability. In the same time, just the competition is the source of stimulus for the Lecturers advance and growth.

L – T

The Lecturer keeps in touch with the Tutor in mean to present him the information and educational materials needed for developing the curriculum during the education course as well as consults him in given application domain.

The Tutor is the Lecturer's main assistant – he has a permanent contact with the Lecturer and plans the interaction between him and the Employees. The Tutor gets the latest information about Lecturer's knowledge and education courses, and helps him to build the advertising strategy for the Knowledge Market. The Tutor also helps the Lecturer in using modern information technologies.

L – E

The Lecturer consults the Examiner about the plans for providing the tests and exams, as well as about the methodical recommendations for the contents of the tests and the educational methods for organizing and providing them.

The Examiner informs the Lecturer for the results of the attestation activities and suggests corrections to the current educational process.

L – A

Owning the basic wares of the knowledge market, namely the knowledge information objects, the Lecturer needs everyday information service for each one of the information interactions.

The Lecturer has to receive the actual knowledge and data and accumulate them, transforming in a resource for selling as a ware at the Knowledge Market.

The Lecturer always has to take into consideration all the specific characteristics of the customer of every ware (such as language specification, professional orientation, nationality and religion, information services' skill level etc.). So he has to own this kind of data as well. This type of information service of the Lecturer should be realised by the Administrator.

The interaction between the Administrator and the Lecturer consists in realization the hardware and software support as well as for all participants of the Knowledge Market, and in particular – service for realization the main function of the Lecturer – receiving new knowledge information objects and their transformation in the new articles of trade.

L – R&D

Special attention needs to be paid to the interaction of the Researchers and Developers with the Lecturers.

For some of the specific types of activity (like rapid and/or network education) the Lecturer can order special scientific research or a development accorded to the devices and the technologies, which can enlarge the service of the Knowledge Market in new direction. Such special scientific investigations and/or developments the Lecturer can order to the Researcher and the Developer, cooperating with the Employer. The Lecturer has to receive the information for the new scientific results and research-based technologies created from that participant in the Knowledge Market, and deliver it to the Employees.

The Lecturers play leading role for Employees' education and training. Because of this, the continual investigation of theirs possibilities and activities is very important.

The main goal of the interaction between Researchers and Developers and Lecturers is to enhance the effectiveness of the Lecturers work.

The achieving of this goal requires to build appropriate models of the Lecturer activities and to develop convenient subsidiary tools for receiving the new information and knowledge and creating educational courses and additional illustrative material.

T – T

The Tutors transfer to each other the educational plans, and compare the information needed for the coordination of the different courses of the whole education process.

In the same time, the competency of the Tutors is not enough to control the educational process. Because of this the collaboration between Tutors need to be provided through the Lecturers and under theirs management. Individual Tutors' initiative may cause embarrassments during the knowledge exchange.

T – E

The Examiner receives from the Tutor the plans for the expected tests and exams conducting as well as the test materials.

He also informs the Tutor for the results of the attestation activities.

T – A

The Administrator supplies the Tutor with all needed hardware and software support for realisation of his activities – mainly connection with other participants, especially with the Lecturers, Employees and Examiners, for organising the education cycle.

As far as the Tutor's role is often played by the automated systems, the Administrator may perform a part of the Tutor duties – especially in the cases of the self-education.

T – R&D

The interaction between the Tutors and the Researchers and Developers is usually realised through the Lecturers. The Tutor doesn't take decisions about using new methods or technologies without Lecturer's opinion. He can get the new information from the Researchers and Developers and transfer it to the Lecturer. Only the Lecturer can decide how to use it.

The accent of the investigation of the Tutor's activities lies on the increasing the effectiveness of the organization of educational process as hole. The major element of this organization is the coordination with the other Knowledge Market participants (mainly with the Employee, Lecturer and Examiner) and external structures for optimal planning the activities of the participants in time and in place.

E – E

The Examiners transfer to each other the results of using attestation materials and methods.

At the state (government) and social level there exist similar structures, which are aimed to examine the educational organizations and to give them rights to provide the education in accordance to government and social requests.

E – A

To keep the knowledge level needed for conducting the Employee's exams, as well as all the information interactions, the Examiner needs everyday information service. It should be realised by the Administrator. The Administrator supplies the Examiner with all needed hardware and software support for realisation of his activities – mainly connection with other participants of the Knowledge Market for preparing and processing the process of the educational testing.

E – R&D

Taking into account that the main activity of the Examiner is to estimate the available knowledge and skills of the learner in the beginning, intermediate and final stage of the educational process in order:

- to evaluate the possibilities of the exanimate person (Employee);
- to guide and correct the educational process, provided by the Lecturer and the Tutor;
- to estimate the adequacy of the received possibilities in given learning cycle to the requirements of the Employer, which usually come to Examiner via Consultant,

the major tasks of the Researcher and Developer are:

- to investigate pointed Examiner's activities;
- to build theirs appropriate information models;
- to develop correspond tools for information servicing of these activities.

A – A

As a rule the single Administrator could not support and serve all information activities in the global net. Because of this he needs of participation of other Administrators in his concrete information service activity.

The Administrators of the Knowledge Market are the part of the global administrators' society for the information service and support of the business and social activities.

A – R&D

Every Administrator became as a source of information for building the Knowledge Market Administrator abstract information model by the Researchers and Developers.

The Administrator is the integrating component that is always in connection with all participants in the Knowledge Market. Because of this, he is valued helper of the Researchers' and Developers' work for investigation and modelling the components and communication processes between them.

In other hand, the implementation of the results of the Researchers' and Developers' work closely depends on Administrator's activity.

R&D – R&D

As a rule the single Researcher could not support and serve all investigation activities in the global net. Because of this, he needs of participation of other Researchers in his concrete scientific work. The same is true for Developers. So, the Researchers and Developers of the Knowledge Market need to be part of the global society of Researchers and Developers.

Special interest for us is the interaction between Researchers and Developers.

As it is remarked above, the Developers are obligated to implement new scientific knowledge for advance of the Knowledge Market information service and to abide by the scientific rules and recommendations.

Contrariwise, the Researchers need to interact with the Developers for collecting knowledge for implementation of the scientific results and theirs possibilities as well as for the new and not investigated areas of the Knowledge Market information interaction and service. Such way the Researchers may extend theirs knowledge and generate new scientific results to be implemented in the practice. So, the cycle "practice – science – practice – etc." may be complete.

Conclusion

In this paper the interconnections and information interactions between main members of the Knowledge Market have been presented. The discussion has been based on the direct interconnections between Knowledge Markets' participants. It is clear that there exist many indirect relations and influences which need to be investigated.

The Knowledge Markets are social phenomenon and, because of this, one very important task, not included in this paper, is the investigation of the relations of Knowledge Markets and other social formations. At the first place, it is obligatory for the government organizations to regulate and control the functionality of the Knowledge Markets.

Till now there not exist special lows and other normative documents aimed to regulate the interactions at the Knowledge Markets. Our expectations are turned to future research just in this area.

At the end we need to point one crucial direction for future research work. The new kind of human interconnections at the Knowledge Markets requires new pedagogical approaches. We are at the beginning of new kind of knowledge exchange based on the market principles and regulated by market lows. Many beginners will be embarrassed by the unknown environment and this may cause great social problems. The payable electronic way for exchanging the knowledge information objects will throw aside the destitute groups of people as well as nations of the world.

Acknowledgements

This paper is partially supported by the project ITHEA-XXI.

Bibliography

[Markov, 1999] Кр. Марков. *Относно вредите от електронната търговия*. IECC'99: International e-Commerce Conference. ADIS & VIA EXPO, Bulgaria, Sofia, 1999.

[Markov et al, 2000a] Кр. Марков, Кр. Иванова, И. Митов. *Требования к автоматизации обучения на пороге информационного общества*. Новые информационные технологии в электротехническом образовании (НИТЭ-2000): Сборник научных трудов пятой международной научно-методической конференции. – Россия, Астрахань: Изд-во ЦНТЭП, 2000.

[Markov et al, 2000b] Kr. Markov, Kr. Ivanova, I. Mitov, N. Ivanova, K. Bojkachev, A. Danilov. *Co-operative Distance and Long-Live Learning*. ITA-2000, Bulgaria, Varna, 2000.

[Ivanova et al, 2001] N. Ivanova, K. Ivanova, K. Markov, A. Danilov, K. Boikatchev. *The Open Education Environment on the Threshold of the Global Information Society*. IJ ITA, 2001, V.8, No.1 pp.3-12. (Presented at Int. Conf. KDS 2001 Sankt Petersburg, 2001, pp.272-280, in russian, Presented at Int. Conf. ICT&P 2001, Sofia, pp.193-203)

[Boikatchev et al, 2001a] K. Boikatchev, N. Ivanova, A. Danilov, K. Markov, K. Ivanova. *Authoring Tools for Courseware Designing*. IJ ITA, 2001, V.8, No.3 pp.115-121. (Presented at Int. Conf. KDS 2001 Sankt Petersburg, 2001, pp.32-37)

[Boikatchev et al, 2001b]K. Boikatchev, N. Ivanova, A. Danilov, K. Markov, K. Ivanova. *Teacher's and Tutor's Role in Long Life Distance Learnin*g. IJ ITA, 2001, V.8, No.4 pp.171-175. (Presented at Int. Conf. KDS 2001. Sankt Petersburg, 2001, pp.38-41)

[Markov et al, 2002] K. Markov, K. Ivanova, I. Mitov, N. Ivanova, A. Danilov, K. Boikatchev. *Basic Structure of the Knowledge Marke*t. IJ ITA, 2002, V.9, No.4, pp.123-134 (Presented at Int. Conf. ICT&P, 2002, Primorsko)

[Ivanova et al, 2003] Кр.Иванова, Н.Иванова, А.Данилов, И.Митов, Кр.Марков. *Обучение взрослых на рынке профессиональных знаний*. Сборник доклади на Национална научна конференция "Информационни изследвания, приложения и обучение" (i.TECH-2003), Варна, България, 2003. Стр. 35-41.

[Markov et al, 2003] Кр. Марков, А.Данилов, Кр.Иванова, Н.Иванова, И.Митов. *Массовое профессиональное обучение в условиях рынка знаний*. Пленарен доклад. Сборник доклади на: VI-та Международна научно-методическа конференция «Новые информационные технологии в электротехническом образовании», НИТЭ-2003, Астрахан, Русия, октомври 2003. стр. 9-18.

Authors' Information

Krassimira Ivanova – Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, foi@nlcv.net

Natalia Ivanova – Petersburg State Transport University, Russia, natali@main.ru

Andrey Danilov – St. Petersburg State University of Aerospace Instrumentation, Russia, danilov@main.ru

Ilia Mitov – Institute of Information Theories and Applications FOI ITHEA, Bulgaria, foi@nlcv.net

Krassimir Markov – Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Institute of Information Theories and Applications FOI ITHEA, foi@nlcv.net

ABOUT PROBLEMS OF DECISION MAKING IN SOCIAL AND ECONOMIC SYSTEMS

Oleksiy Voloshyn

Abstract: The reasons of a restricted applicability of the models of decision making in social and economic systems. 3 basic principles of growth of their adequacy are proposed: "localization" of solutions, direct account of influencing of the individual on process of decision making ("subjectivity of objectivity") and reduction of influencing of the individual psychosomatic characteristics of the subject (" objectivity of subjectivity ") are offered. The principles are illustrated on mathematical models of decision making in ecologically- economic and social systems.

Keywords: decision making, ecological, economic and social systems, sequential analysis of variants, indistinct analysis, methods of expert estimation, collective decision making, and decision-making support system.

ACM Classification Keywords: K.3.2 Computer and Information Science Education.

Introduction

The second half of XX century gave birth to the many mathematical models of ecological, economical and social process [1]. However the success in application of them is incomparable with practical achievements in application of the models of the «lifeless nature». Economical, political man-caused and the ecological crises are typical to separate countries, regions, political and economical systems, and mankind. Penetrating into the depths of the Universe, uncovering secrets of a micro-world, modern civilization is not able to prevent military conflicts, famine, and terrorist acts. It is enough to tell, that for 5500 years of a modern history the mankind lived without wars only 292 years, and from 6.4 billion of dwelling people are starve more then 800 million one, though food is made on 20% more indispensable. There are unsolved ecological problems, which one threatens to the survival of mankind.

There is a natural question — why the use of the modern achievements in the decision-making theory in practice of modeling of economical and social processes is very limited? The answer is obvious, because they have very low "adequacy" to a substantial world and very high scale of result uncertainty. Moreover, frequently this result is "meaningless" (i.e. basically is not testable). It is referred to normative mathematical models of economic and social processes, and positive one. The normative models are answer of a problem "what it is necessary to do to achieve desirable?" and positive models are answer of a problem " why we have that we have?" and, at the best case, "what will be?"

Within the framework of existing paradigm of the development and analysis of ecological, economical and social processes for overcoming the indicated defects are necessary for taking into account following aspects. At first, to take into account as a lot of the factors as possible, which one influences on a process of decision-making. This fact lead to "damnation of dimension". Secondly, it is necessary to take into account an inaccuracy and illegibility of parameters means of model (quite often and it full absence). These inaccuracies are connected as with "by objective uncertainty" [2] (which one is proper of the organization of our world), and with "by subjective uncertainty " (which one is characteristic to the human nature as a whole). The standard approaches of development of algorithms, which one guarantee convergence (in classic comprehension) to precise solution, create more illusion of their solution, than decide the delivered problems.

The problem consists not in finding to any desired degree of solutions precision of a problem (especially if the initial data is not exact!), but its "localization" [3]. It is necessary to determine intervals of solution component changes. These intervals depend on accuracy of data calculation. In an ideal it is desirable to achieve maximum localization of solution - minimum (in some metrics) intervals of solution uncertainty. Such approach in natural sciences, for a long time already became natural (we shall recollect "a principle of uncertainty " of the Heisenberg [4]). At simulation of social and economic processes this approach has not found broad application. And it provided that the subject is their active component! All features of subjective perception of a choice of decision-making in the "classic" approach include in to an axiomatic. In this case problem of the contributor is lead to testing of their "reality" [5].

In such aspect is a priori accepted, that "reality" of a hypothesis means its "reality" for each subject at any moment of time (or in a continuous enough time interval). From a beginning 70 years of past century are successfully used in acceptance of political-military solutions principles of the theory of group thinking [6]. In modern mathematical economy these principles have not found application. In 90 years of XX century the necessity of subjective features usage for decision making in economy has lead to appearance "situational economy ". Originating in 60 years of past century of the "indistinct analysis" [7] also is confirmation of necessity of direct usage of the subject at construction of social and economic models. In this case "functions of an accessory" meaning of indistinct set can be interpreted as "subjective probability".

Thus, during the mathematical informational model construction of social and economic processes there are two problems. At first "subjectivity of objectivity that is"direct consider of influencing of the subject on decision-making process. In second "objectivity of subjectivity" that is a reduction of influence of the individual characteristics of the subject and accounting of subject's "objective" psychosomatic features of reality knowledge.

"Subjectivity of objectivity", lead to necessity of usage at construction and research of models of the indistinct analysis theory, group thinking theory, development of algorithms solution locating. "Objectivity of subjectivity " in models can be carried out on the basis of social and psychological disciplines. For example, to take into account subject attitude to the one of the socionic types, tendency to risk, independence degrees etc. [8]. In turn to increase the objectivity it is reasonable to use the description of the subject in an unclear form. Ideally "objectivity" of the subject should be imposed on "subjective" description of an object for the increase of the objectivity degree.

The offered principles are illustrated below on idealized and applied models of decision-making problems in social and economic systems.

Models of the Analysis of Static Balance of Large Dimensional Ecological and Economic Models

The static balance Leontieff — Ford model [9] is considered in these settings:

$$\begin{cases} x_i^1 = \sum\limits_{p \in I} a_{ip}^{11} x_p^1 + \sum\limits_{q \in J} a_{iq}^{12} x_q^2 + y_i^1, \\ x_j^2 = \sum\limits_{p \in I} a_{jp}^{21} x_p^1 + \sum\limits_{q \in J} a_{jq}^{22} x_q^2 - y_j^2, \\ i \in I,\ j \in J; \quad x^1 > 0, x^2 \geq 0; \end{cases} \tag{1}$$

where $x^{1T} = \left(x_i^1\right)_{i \in I}$ — a vector of production volumes, $x^{2T} = \left(x_j^2\right)_{j \in J}$ — a vector of contaminants volumes destroying, I = {1,2, …, n}, J = {1,2, …, m} — are sets of indexes of variables representing «economical» and «ecological» model component, T — a sign of transposition; $A_{11} = \left\| a_{ji}^{11} \right\|_{i \in I}^{j \in I}$ — a square matrix of normative factors of the product cost j by production of a product unit i, $A_{12} = \left\| a_{ij}^{12} \right\|_{j \in J}^{i \in I}$ — a rectangular matrix of normative factors of the production cost i at killing unit of contaminant j, $A_{21} = \left\| a_{ji}^{21} \right\|_{i\in I}^{j\in J}$ — a rectangular matrix of normative factors of ejection of contaminant j by production of product unit i $A_{22} = \left\| a_{ji}^{22} \right\|_{i\in J}^{j\in J}$, — a square matrix of normative factors of contaminant ejection j at killing unit of contaminant i, $a_{ij} \in [0;1]$, $y^{1T} = \left(y_i^1\right)_{i\in I}$ — a vector of volumes of final product, $y^{2T} = \left(y_j^2\right)_{j\in J}$ — a vector of volumes of contaminants, which one are not deleted, $y_i^1 > 0, i \in I$, $y_j^2 \ge 0$. $j \in J$

$X = \prod_{i\in I} X_i^1 \times \prod_{j\in J} X_j^2$ — a hyper parallelepiped of problem solutions, which takes into account the contents of limitations of economic and ecological components:

$$X_i^1 = \begin{cases} \left[d_{i(H)}^1, d_{i(B)}^1\right], & i \in I^{(H)}, \\ \left\{d_{i(H)}^1, d_{i(H)}^1 + 1, \dots, d_{i(B)}^1\right\}, & i \in I^{(Ц)} = I \setminus I^{(H)}, \end{cases} \tag{2}$$

$$X_j^2 = \begin{cases} \left[d_{j(H)}^2, d_{j(B)}^2\right], & j \in J^{(H)}, \\ \left\{d_{j(H)}^2, d_{j(H)}^2 + 1, \dots, d_{j(B)}^2\right\}, & j \in J^{(Ц)} = J \setminus J^{(H)}, \end{cases} \tag{3}$$

where $d_{i(H)}^1$, $d_{i(B)}^1$ and $d_{j(H)}^2$, $d_{j(B)}^2$ — are variables border (upper and lower accordingly) x_i^1 and x_j^2 accordingly, $d_{i(B)}^1 \ge d_{i(H)}^1 \ge 0$, $d_{j(B)}^2 \ge d_{j(H)}^2 \ge 0$ (in case of an integer formulation (1) $d_{i(H)}^1$, $d_{i(B)}^1$ and $d_{j(H)}^2$, $d_{j(B)}^2$ are considered integer), the indexes (H) and (Ц) divide set of indexes on sets of indexes accordingly of continuous and integer variables. If the initial borders do not given, $d_{i(B)}^1$ and $d_{j(B)}^2$ are considered as zero points, and $d_{i(H)}^1$, $d_{j(H)}^2$ are selected large enough (proceeding from economic reasons).

In a basis of methods, which one are offered for overcoming the difficulties, indicated in the introduction, which one arise at the analysis of models (1-3) large dimensions, the scheme of a sequential analysis of variants lay [10-15]. The basis of algorithm of problem solving (1-3) is compounded by a WB procedure [16] approximating of set D of solutions by a hyper parallelepiped X^* by such, that $X \supseteq X^* \supseteq D$.

The analytical estimation of performance of a WB procedure looks like the following. Let ε_1 : $\varepsilon_1 > 0$, ε_2 : $\varepsilon_2 > 0$, - percentage errors of magnitudes nearing $Q^{1(k)}$ and $Q^{2(k)}$, $k = 0, 1, \dots,$ to boundaries

$$Q^{1*} = \lim_{k\to\infty}\left(y_{\min}^1 \sum_{s=0}^{k} \lambda_1^s\right) = y_{\min}^1 \lim_{k\to\infty} \sum_{s=0}^{k} \lambda_1^s = \frac{y_{\min}^1}{1-\lambda_1} \quad , \quad Q^{2*} = \lim_{k\to\infty}\left(y_{\min}^2 \sum_{s=0}^{k-1} \lambda_2^s\right) = y_{\min}^2 \lim_{k\to\infty} \sum_{s=0}^{k-1} \lambda_2^s = \frac{y_{\min}^2}{1-\lambda_2} .$$

Then $k_1 \le \dfrac{\log_2 \varepsilon_1}{\log_2 \lambda_1} + 1$, $k_2 \le \dfrac{\log_2 \varepsilon_2}{\log_2 \lambda_2} + 1$, where $\lambda_1 = \min\limits_{i\in I} \dfrac{\sum\limits_{p\in I,\, p\ne i} a_{ip}^{11}}{1 - a_{ii}^{11}}$, $\lambda_2 = \min\limits_{j\in J} \dfrac{\sum\limits_{q\in J,\, q\ne j} a_{jq}^{22}}{1 - a_{jj}^{22}}$.

Let $L = \left((n+m)^2 + 3(n+m)\right)\log_2 a$ - length of an input of a problem (1), $a = \max\left|\log_2 a_{ij}\right| + 1$, then a percentage error of a solution $\max\{\varepsilon_1, \varepsilon_2\} = \varepsilon \ge 2^{-L}$.

The amount of elementary operations M WB procedures has the order $M = O(L)$, and the computational complexity is given in to [17]: $N = O(Mk) = O\left(-\frac{L^2}{\log_2 \lambda}\right)$, where $k \le \frac{\log_2 \varepsilon}{\log_2 \lambda} + 1$, $\lambda = \min\{\lambda_1, \lambda_2\}$. At non-existence of original limitations: $d^{1(0)}_{i(H)}$, $d^{1(0)}_{i(B)}$, $i \in I$, $d^{2(0)}_{j(H)}$, $d^{2(0)}_{j(B)}$, $j \in J$, for a WB procedure it is possible to put: $d^{1(0)}_{i(H)} = \frac{y^1_{\min}}{1-\lambda_1}$, $i \in I$, $d^{2(0)}_{j(H)} = \frac{y^2_{\min}}{1-\lambda_2}$, $j \in J$, $d^{1(0)}_{i(B)} = \frac{y^1_{\max}}{1-\bar{\lambda}_1}$, $i \in I$, $d^{2(0)}_{j(B)} = \frac{y^2_{\max}}{1-\bar{\lambda}_2}$, $j \in J$,

where $y^1_{\max} = \max\limits_{i \in I} y^1_i$, $\bar{\lambda}_1 = \max\limits_{i \in I} \frac{\sum_{p \in I,\, p \ne i} a^{11}_{ip}}{1 - a^{11}_{ii}}$, $y^2_{\max} = \max\limits_{j \in J} \{A_{21} y^1 - y^2\}$, $\bar{\lambda}_2 = \max\limits_{j \in J} \frac{\sum_{q \in J,\, q \ne j} a^{22}_{jq}}{1 - a^{22}_{jj}}$.

More detailed results of computational experiment on a WB procedure are given into [18].

The productivity of a computational procedure is instituted by time spent for calculus, and fidelity of result. At a percentage error to within the sixth sign and dimension of a problem m+n=1000 the time for a solution of a continuous and integer problem compounds about 30 min. As at magnification of dimension of a problem the specific weight of nonzero members is moderated, introduces practical concern update of a WB procedure on a case of sparse matrixes [18]. At a reduced percentage error and dimension about 10000 at matrixes entry on 1 % the time for a continuous integer problem compounds about 10 min.

Fuzzy and Multicriteria Models

"Subjectivity of objectivity" guesses fissile involvement of the subject in decision making, account in the patterns of social and economic processes subjective reluctance. Let's esteem two basic methods of realisation of this principle - illegibility (for example - the static Leontieff pattern [1]) and multicriterion (for example - a collective decision making problem [21]).

Let's esteem the Leontieff pattern x = Ax + y, x≥0. . The bulk of final consumption, as a rule, is set by the way of hyper parallelepiped Y = $\Pi[y_j^-, y_j^+]$, where y_j^- - - lower “norm” of consumption j-th of a product, y_j^+ - upper. Moreover, is logical (as a rule, it is done) on an interval $I_j = [y_j^-, y_j^+]$ the function of an accessory μ_{Ij} by the way of expert recommendations for bulk of consumption j-th of a product is set. Similarly for the pattern Leontieff-Ford is logical to set lower and upper “norms” of bulks of the not erased contaminants, that is to set a hyper parallelepiped $Y^2 = \Pi[y_j^-, y_j^+]$. Then also solution of the patterns $x \in X$ is logical to discover, allowing wishes of the experts, by the way functions of an accessory μ_X,

For the Leontieff pattern (for the pattern Leontieff-Ford the constructing similar) are necessary for finding $\{x \in \{x: x = Ax + y, x \in X, y \in Y\}$ at given functions of an accessory μ_X, μ_Y. Let's record indistinct set by the way: X~ = $\bigcup\limits_{\alpha \in [0,1]} \alpha X_\alpha$, were $X_\alpha = \{x \mid \mu_X(x) \ge \alpha\}$.

On definition the function of an accessory of indistinct set α A is set as following: $\mu_{\alpha A}(a) = \begin{cases} \alpha, a \in A \\ 0, a \notin A \end{cases}$.

Similarly Y~ = $\bigcup\limits_{\alpha \in [0,1]} \alpha Y_\alpha$. The basic conjecture, which one is superimposed on a solution of the indistinct pattern Leontieff, consists in following: $x \in X_\alpha$ in only case when, when $x - Ax \in Y_\alpha$

Let's esteem piecewise linear functions of an accessory, then:

$$X_\alpha = \prod_{k=1}^{n} \left[x_k^-(\alpha), x_k^+(\alpha)\right] = \prod_{k=1}^{n} I_\alpha(x_k),\ X^{\pm}_\alpha(\alpha) = x^{\pm}_\alpha(0)(1-\alpha) + x^{\pm}_\alpha(1) \cdot \alpha,\ \mu_X(x) = \prod_{k=1}^{n} \mu_i(x_k).$$

Let's designate: $AX = \{Ax \mid x \in X\}$, $X + Y = \{x + y \mid x \in X, y \in Y\}$.

Allowing, that $(AX)_\alpha = AX_\alpha$ and the additive property of a WB procedure, boundaries calculus of a hyper parallelepiped for each value a is carried on separately by the formula:

$$AX = \prod_j \left[\sum_{j,a_{ij}>0} a_{ij}x_j^- + \sum_{j,a_{ij}<0} a_{ij}x_j^+ ; \sum_{j,a_{ij}>0} a_{ij}x_j^+ + \sum_{j,a_{ij}<0} a_{ij}x_j^- \right].$$

At applying the depicted above approach to a problem with indistinct data we have:

1 Solutions of an indistinct problem is lead to a solution p (on an amount α - - levels) customary problems Leontieff, which one are decided separately;

2 As solution of an indistinct problem we shall consider indistinct set with an accessory function to making up by level lines for each of a these *p* components of problems solution;

3 The algorithm of a sequential analysis of alternatives does not degrade arguments of a solution: „equivocation» of a solution of an indistinct problem will be «not greater » uncertainty of original data.

The proposed approach is applicable and in case of an indistinct matrix of normative coefficients.

Most general formulation of a problem of acceptance of a collective solution having a numerous application in economics, policies and other fields of human activity, bound with analysis and resolution of conflicts, is lead to a following mathematical model:

$$U_i(x) \rightarrow \max, \ i \in I, \ x \in \Pi X_i \qquad (4)$$

Where: U_i - utility function i-th of the agent, X_i- set of its policies, the set of policies x is called as a situation.

In classic posing of a gamble problem of strategy are selected by the gambler (agents) simultaneously and separately. Generally players can agree about queue of courses, about a share choice of strategy the etc. Most popular principle of optimal behaviour ("principle of an optimality ") in a problem (4) is considers "the Nash equilibrium" [19]. In this principle the individual deflections of the gamblers from strategies, which are included in this situation, do not augment a scoring of the declined gambler, if the remaining gamblers hold on to the strategies, fixed in this situation. The considerable situations of equal balance as the challenger for optimal behaviour are natural enough. However situations of equal balance can have a series of properties handicapping their operational use. First of all is non-uniqueness, and the miscellaneous situations of equal balance are preferential to the miscellaneous gamblers. In [19] two yardsticks of a choice of single equal balance - prevalence on a scoring and prevalence on risk are selected. When prevalence on a scoring and prevalence on risk have the different directions, the writers of [19] return priority to prevalence on a scoring. However in practice is reasonable to allow for psychosomatic features of the agents. In this case this is "tendency to risk" [25]. Is generally logical to esteem a multicriteria choice with allowance for two given criteria. With the purpose of “subjectivity” of the model (4) are expedient for esteeming not scalar utility functions Ui, but vector functions [20,21], and to apply convolutions depending on psychosomatic features of a making decision person (MDP). In absolute majority of cases of value of utility functions there is a result of an expert estimation. For this reason practical concern consideration of fuzzy formulations (4) [22] is interesting.

Methods of an Expert Estimation

In the given partition we shall esteem a possibility of sharing of principles of localisation and subjectivity of the pattern for a problem of an expert estimation [23].

Was held studies about deriving the consistent information from the expert about numerical values of weighting coefficients of criteria. It demonstrates that the expert or (MDP) can adequately define weighting coefficients, if the amount of object arguments does not exceed "magic" number 5-9 [24]. If these objects are characterised by a great number of arguments, it is necessary to apply indirect methods. In these methods the ratios of preference sequentially are updated on the basis of accepted before solutions (intervals of relative relevance of plants "are localised").

Let A - set of objects a^j, $j \in J$. Each of objects a_j is characterised by a set of arguments

$a^j = (a^j_i)$, $i \in I$. It is necessary a_j to deliver object in conformity vector estimation in R^n, which defined by a set of criterions, on which one MDP values objects.

Let's esteem two objects a^1 and a^2 from set of effective objects And. The object $a^1$is considered as "best", than the object a^2, if the sum of fluidised deflections of arguments from their best values for object a^1 is less, than for a^2, i.e.

$$\sum_i \rho_i \omega\left(a_i^1\right) < \sum_i \rho_i \omega\left(a_i^2\right) \tag{5}$$

Where ρ- normalised vector of relative relevance of objects arguments for the assertion MDP about a ratio of preference between objects; $\omega\left(a_i^j\right)$ - some monotonic transforming, defining extent of deflection from a best value of argument and conversing all value of arguments to a dimensionless aspect in a spacing [0,1].

On the basis of a method of localisation of solutions [3] the procedures of calculus of intervals are proposed [$\rho_i^н$, $\rho_i^в$], conserving a ratio (5). The software is designed, which one in real-time mode allows to decide problems with 50 objects.

Decision–making Support Systems

Methods of social and economical processes progressing estimation which one are "a past prolongation", allows to receive the forecast, as a rule, with a very high scale of ambiguity. As the knowledge, operating of the subjects of these processes is "suspended" past expertise in these methods. Desire as much as possible to approximate subjective perceptions to an actuality is lead to following. First of all it led to necessity of the expert information usage in the given point of time. Secondly, the complication of modelled processes does not allow using immediate knowledge of the expert in broad fields. Therefore there is a problem of creation of "flexible" systems of decision making. Such systems should be customised on a particular data domain. They require narrow professional, "local" knowledge. This knowledge, naturally, will be ill structured and fuzzy. "Objection" of such shapes of knowledge is possible by the account of psychosomatic features of the expert and his previous expertise.

It is generally recognised that one of decision-making, most adequately modelling process, by the person is the method of a decision-tree [25]. However its applying is hampered by a "damnation of dimension", which one arises at its usage. Mining of a special processing techniques of a decision-tree therefore is indispensable [8].

On the basis of the above-stated concepts the development support system of creation applied decision-making support systems (DMSS) in different fields is designed. The constructing of applied DMSS is resulted to separation by the experts of problems and sub problem (tops of a tree) and links between them (arcs of a tree). Further experts are assigned the weight coefficients of transferring probability between tops. The indistinct estimations of the experts with the help of logical variables are enabled. These variables are described by values of a membership function (vectors of true numbers from 0 up to 1). Each expert sets three estimations: optimistic, realistic and pessimistic. Scalarization of these estimations is carried out with allowance for psychological type of the expert. The type is assigned on a foundation of the psychological tests, which one is input in a system. On the basis of the psychological tests the coefficients of "truthfulness", "independence", "caution" etc. [8] are assigned also.

The tree is under construction on the basis of collective estimations of the experts with applying of a pairwise comparison method. The algebraic processing techniques of the expert information are applied to constructing a resultant tree. As spacing interval between ranging the metric of the Hamming and measure of straddling of ranks of objects is applied. The resultant tree is assigned as a Kemeni- Snell median:

$$Arg\,\min_{A}\sum_{i=1}^{n} d(A, A^{i}) \text{ Or as the compromise: } Arg\,\min_{A}\max_{i}\ d(A, A^{i}) .$$

Where A^{i} - matrix of i-th expert, in which one a member a_{ij} =1 in only case when, when i-th the top is more preferential to the j-th expert, a_{ji} = -1, for equivalent objects a_{ij} =0, a_{ii} =0. .

In case of the definition of advantage in the indistinct shape the members of a matrix are set through functions of an accessory.

For determination of optimal paths in a tree the algorithms of a sequential analysis of alternatives are t suggested [10-15]. These algorithms allow to process trees with hundreds tops.

The tables set the solution-tree. Each table is a separate level of a tree. Each table line is separate top at this level. Each element of line is a probability, with which one the transferring from the given top in top of a lower layer is possible. These probabilities are set by functions of accessories representing of real numbers vector from 0 up to 1 of any lengths. The table is filled by interrogation of the experts. The existing functions allow to add columns, line, to set the dictionary. This dictionary allows verbal estimations of the expert to pose in conformity of probabilities, by the definition of definite levels, to save the tables in the file, to read out the tables from the file.

Matrixes assigned by the expert way. This is the result of tops variants comparison. They can be included in a tree. On the basis of analysis of matrixes the tops are assigned. They are included in a tree. As on the basis of matrixes analysis the probabilities are assigned, with which one the transferring in them from top-level tops is possible. If a tree of a solution decomposed on some sub trees, and these sub trees have identical leafs, probabilities of these leafs in each of them in the beginning are evaluated, and then evaluated the probabilities for all tree as a whole.

Up to the moment there are created a series of application systems: forecasting of exchange, account of a gross national product, diagnostics of cardiovascular diseases, forecasting of index inflation and so on.

Bibliography

1. Леонтьев В.В. Межотраслевая экономика. - М.: Экономика, 1997. - 479 с.
2. Нариньяни А.С. Неточность как Не - фактор. Попытка доформального анализа. – Москва – Новосибирск, 1994. Препринт РосНИИ ИИ, №2. - 34с.
3. Волошин А.Ф. Метод локализации области оптимума в задачах математического программирования //Доклады АН СССР. - 1987. - Т. 293, № 3. - С. 549–553.
4. Орир Дж. Физика. - М.:Мир,1981,том 2. - 288с.
5. Nikolson W. Microeconomic theory. -The DRYDEN PRESS, 1998. -821 p.
6. Janis I.L. Groupthink//Psychology Today, 1971. –P.43-46.
7. Орловский С.Г. Проблемы принятия решений при нечеткой исходной информации. - М.:Наука,1981.
8. Voloshin O.F., Panchenko M.V. The System of Quality Prediction on the Basis of a Fuzzy Data and Psychografphy of the Experts// “Information Theories & Applications”, 2003, Vol.10, №3.-P. 261-265.
9. Леонтьев В.В., Форд Д. Межотраслевой анализ воздействия структуры экономики на окружающую среду // Экономика и математические методы. - 1972. - т. VIII, вып. 3. - С. 370–400.
10. Михалевич В.С. Последовательные алгоритмы оптимизации и их применение // Кибернетика. - 1965. - № 1. - С. 45–55; -№2. - С.85-89.
11. Волкович В.Л., Волошин А.Ф. Об одной схеме последовательного анализа и отсеивания вариантов // Кибернетика. - 1978. - № 4. - С. 98–105.

12. Михалевич В.С., Волкович В.Л., Волошин А.Ф. Метод последовательного анализа в задачах линейного программирования большого размера // Кибернетика. - 1981. - № 4. - С. 114–120.
13. Волошин О.Ф., Мащенко С.О., Охрименко М.Г. Алгоритм послідовного аналізу варіантів для розв'язання балансових моделей // Доповіді АН УРСР. - 1988. - Сер. А, № 9. - С. 67–70.
14. Волкович В.Л., Волошин А.Ф. и др. Методы и алгоритмы автоматизированного проектирования сложных систем управления. - К.: Наукова думка, 1984. - 216 с.
15. Волкович В.Л., Волошин А.Ф. и др. Модели и методы оптимизации надежности сложных систем. - К.: Наукова думка, 1993. - 312с.
16. Волошин О.Ф., Чорней Н.Б. Алгоритм послідовного аналізу варіантів для розв'язання міжгалузевої моделі Леонтьєва-Форда // Вісник Київського університету. - 1999. - Сер. фіз. - мат. н., № 1.
17. Пападимитриу Х.,Стайглиц К. (Christos H. Papadimitriou, Kenneth Steiglitz) Комбинаторная оптимизация. - М.:Мир,1985. - 510 с.
18. Волошин А.Ф. Методы анализа статических балансовых эколого-экономических моделей большой размерности//Киевский национальный университет им.Т.Шевченко, «Научные записки», 2004, Том. VII.-С.43-55.
19. Харшаньи Дж., Зельтен Р. (John C. Harsanyi, Reinhard Selten) Общая теория выбора равновесия в играх.-Санкт-Петербург: «Экономическая школа», 2001.-406с.
20. Мащенко С.О. Равновесие Нэша в многокритериальной игре//Вестник Киевского университета,2001,№3.-С.214-222.
21. Волохова О.В. Задача коллективного принятия решений с векторными функциями полезности//Труды школы-семинара «Теория принятия решений», Ужгород, 2004.-С.14.
22. Мащенко С.О. Равновесие Нэша в нечетких играх//Вестник Киевского уни-та,2004,№2.-С.204-212.
23. Волошин А.Ф., Гнатиенко Г.Н. Метод косвенного определения интервалов весовых коэффициентов параметров объектов//Проблемы управления и информатики, 2003, №2.-С.34-41.
24. Тоценко В.Г. Методы и системы поддержки принятия решений.-Киев:Наукова думка, 2002.-382с.
25. Ларичев О.И. Теория и методы принятия решений.-М.:Логос,2000.-296с.

Author's Information

Oleksiy F. Voloshyn – T. Shevchenko Kiev National University, Faculty of Cybernetics, professor. Kiev, Ukraine. e-mail: ovoloshin@unicyb.kiev.ua

OPTIMUM TRAJECTORY OF DYNAMIC INPUT-OUTPUT BALANCE MODEL FOR OPEN ECONOMY

Igor Lyashenko, Olena Lyashenko

Abstract: When export and import is connected with output of basic production, and criterion functional represents a final state of economy, the generalization of classical qualitative results of the main-line theory on a case of dynamic input-output balance optimization model for open economy is given.

Keywords: dynamic input-output balance, optimization problem, main-line theory, open economy, trajectory of the balanced growth.

ACM Classification Keywords: J.1 Administrative Data Processing: Business, Financial, Government

Introduction

The basic difficulties at practical application of economic dynamics diversified models are connected with the large dimension of linear programming optimization problems. In this connection there was an interest of experts to qualitative methods of optimum trajectories research. On this way the interesting results concerning creation of the so-called main-line theory were received. One of the most beautiful results in the theory of extending economy is the theorem by Dorfman, Samuelson and Solow [Dorfman, 1958] and it assert that the effective trajectory of economic growth has the long-term tendency to come nearer to Neumann way of the steady balanced growth. After the publication of the book by Dorfman, Samuelson and Solow [Dorfman, 1958] the theorems on main-line were established by Hics, Morishima and Mac-Kensey - for model of Neumann-Leontiev; by Radner and Nikaido - for model of Neumann-Gail with strictly convex set of productions; by Mac-Kensey - for generalized Leontiev model including the capital boons.

Basic Concepts and Corollaries of the Main-line Theory

It is possible to show the basic concepts and corollaries of the main-line theory on an example of optimization problem for Neumann model [Ashmanov, 1984; Ponomarenko, 1995]:

$$\begin{gathered} c_T x_T \to \max, \\ Ax_t \le Bx_{t-1}, \quad x_t \ge 0, \quad t = 1,2,...,T, \end{gathered} \tag{1}$$

where $A \ge 0, \quad B \ge 0$ - non-negative rectangular $n \times m$ matrixes of expenses and release accordingly, Bx_0 - given vector, $c_T > 0$ - given positive vector, x_t - vector of intensities of technological process in time interval t.

The stationary trajectory of intensities for Neumann model (*A, B*) is determined by rate of growth $\alpha = \overline{\lambda}^{-1}$ and Neumann beam $\overline{x}$ and looks like $x_t = \overline{\lambda}^{-t}\overline{x}$, where $\overline{\lambda} > 0$, $\overline{x} > 0$ - unique decision of inequalities system to within scalar multiplier

$$A\overline{x} \le \overline{\lambda} B\overline{x} \tag{2}$$

It is significant that the main-line $\overline{x}$ appears not sensitive to changing of coefficients of criterion functional $c_T > 0$, the problem (1) is reduced to the following Neumann problem consequently

$$\lambda \to \min, \quad Ax \le \lambda Bx, \quad x \ge 0 \tag{3}$$

The basic result concerning the minimal eigenvalue $\overline{\lambda} < 1$ of Neumann model (maximal rate of growth) is formulated as the following theorem.

<u>Theorem 1.</u> *Let non-negative matrixes $A \ge 0, \quad B \ge 0$ are such that the matrix of production B has no zero lines, and the matrix of expenses A has no zero columns. Then the indecomposable productive Neumann model (3) has unique rate of growth $\overline{\lambda} < 1$ and main-line $\overline{x} > 0$.*

Here indecomposability of Neumann model is understood as impossibility by simultaneous rearrangement of lines and columns in matrixes A and B to reduce them to the form

$$A = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}, \quad B = \begin{pmatrix} B_{11} & B_{12} \\ 0 & B_{22} \end{pmatrix},$$

where the rectangular matrixes-blocks A11 and B11 have equal dimension, 0 - zero matrix of dimension more than 1. Indecomposability of Neumann model is equivalent to the condition: Frobenius number of model (3) is prime number, and Frobenius vector is strictly positive.

The *productivity* of Neumann model is understood as existence of the solution of inequalities system

$$Bx - Ax \geq c, \quad x \geq 0$$

at any $c \in R_+^n$. The productivity of Neumann model is equivalent to the condition: the Frobenius number of model (3) is less than 1.

General Scheme of π-model.

One of the most known scheme of dynamic input-output balance of closed economy is so-called general scheme of π-model (detailed scheme is developed by Yu.P.Ivanilov and A.A.Petrov [Ivanilov, 1971]):

$$\begin{gathered} Ax_t + D\eta_t + L_t c \leq x_t, \\ x_t \leq \xi_{t-1}, \quad \xi_t \leq \xi_{t-1} + \eta_t, \\ lx_t \leq L_t, \\ (x_t, \xi_t, \eta_t, L_t) \geq 0, \quad t = 1,2,\ldots,T, \end{gathered} \tag{4}$$

where the index t is number of time interval (year), $x = (x_1, x_2, \ldots, x_n)^T$ - aggregate stock of goods, the technological expenses of every of n branches are described by Leontiev matrix *A*, $\xi = (\xi_1, \xi_2, \ldots, \xi_n)^T$ - greatest possible total output (capacities of branches), $\eta = (\eta_1, \eta_2, \ldots, \eta_n)^T$ - desirable increment of capacities, the material input on increment of the basic capacities of all n branches are described by matrix *D*, $l = (l_1, l_2, \ldots, l_n)^T$ - expense of manpower for unit of total output, L - total employment, $c = (c_1, c_2, \ldots, c_n)^T$ - vector of consumption on one worker (his natural wages), ξ_0 - basic capacities at the initial moment *t*=0.

For model (4) terminal criterion is used more often in the form

$$c_T x_T \rightarrow \max , \tag{5}$$

where $c_T > 0$, that directly contacts with maximal rate of economy growth.

The basic result concerning existence of main-line for the problem (4) - (5) is formulated as the following theorem [Ashmanov, 1984].

<u>Theorem 2.</u> *Let ξ_0 >0, matrix $R = (c_i l_j)_1^n$, matrix $A+R$ is indecomposable and productive, matrix $Q(\lambda) = \lambda(A + R) + (1 - \lambda)D$ is primitive. Then the vector $(\overline{x}, \overline{\xi}, \overline{\eta}, \overline{L})$, where $\lambda = \overline{\lambda} < 1$ and $\overline{x} > 0$ are Frobenius number and right eigenvector of matrix $Q(\overline{\lambda})$ accordingly , is a main-line for model (4-5).*

Expansion of π-model.

The problem on expansion of the scheme (4) of model on a case of open economy is unsolved today, when export and import run up to such large volumes, that the refusal of them results in a situation of impossibility of economic development. The next problem is a problem on existence of a main-line open economy development, as it is shown for a case of closed economy. This paper is devoted to the solution of these two problems.

For open economy having export and import in large volumes, it is offered to allocate exporting branches (first group of branches) and importing branches (second group of branches), and on vectors of export e1 (t) and import i2 (t) to impose industrial restrictions on its volumes. The model is offered:

$$
\begin{gathered}
c_1^T x_1(T) + c_2^T x_2(T) \to \max, \\
A_{11}x_1(t) + A_{12}x_2(t) + D_{11}\eta_1(t) + D_{12}\eta_2(t) + c_1 L(t) + e_1(t) \le x_1(t), \\
A_{21}x_1(t) + A_{22}x_2(t) + D_{21}\eta_1(t) + D_{22}\eta_2(t) + c_2 L(t) - i_2(t) \le x_2(t), \\
x_1(t) \le \xi_1(t-1), \quad x_2(t) \le \xi_2(t-1), \\
\xi_1(t) \le \xi_1(t-1) + \eta_1(t), \quad \xi_2(t) \le \xi_2(t-1) + \eta_2(t), \\
l_1 x_1(t) + l_2 x_2(t) \le L(t), \\
e_1(t) \ge F_1 x_1(t), \quad i_2(t) \le H_2 x_1(t), \\
x_1(t) \ge 0, \quad x_2(t) \ge 0, \quad \xi_1(t) \ge 0, \quad \xi_2(t) \ge 0, \quad \eta_1(t) \ge 0, \quad \eta_2(t) \ge 0, \\
e_1(t) \ge 0, \quad i_2(t) \ge 0, \quad L(t) \ge 0, \quad t = 1,2,\ldots,T.
\end{gathered} \tag{6}
$$

In model (6) stocks of production of exporting group of branches x1 (t) in time interval of t provide direct industrial expenses A11x1 (t) and A12x2 (t), consumption L (t) c1, creation of capacities increments of both groups of branches $D_{11}\eta_1(t)$ and $D_{12}\eta_2(t)$, and export e1 (t) also. At the same time stock of production of importing group of branches x2 (t) in time interval t can ensure direct industrial expenses A21x1 (t) and A22x2 (t), consumption L (t) c2, creation of capacities increments $D_{21}\eta_1(t)$ and $D_{22}\eta_2(t)$, but with the help of import i2 (t) use. In model (6) F1 means a non-negative coefficients matrix of the minimal production export of the first group of branches, H2 - non-negative matrix of coefficients of the maximal import for maintenance of industrial needs of the first group of branches. In particular, the F1 can be diagonal matrix with diagonal elements which are smaller than one.

Model (6) is dynamic model, as a result of its functioning we receive at the initial data $\xi(0)$ a sequence of vectors $X(t) = \left(x_1(t), x_2(t), \xi_1(t), \xi_2(t), \eta_1(t), \eta_2(t), e_1(t), i_2(t), L(t)\right)$, t=1,2, ..., T, satisfying to all restrictions of the model. Such sequence represents a trajectory. At the end of the researched period (at the time moment T) the state of the model is characterized by a vector X (T) (so-called terminal state of model).

State of Equilibrium

Let's research a state of equilibrium in model (6). The appropriate stationary trajectory of intensities is determined by rate of growth $\alpha = \overline{\lambda}^{-1}$ > 1, by a Neumann beam $X = \left(x_1, x_2, \xi_1, \xi_2, \eta_1, \eta_2, e_1, i_2, L\right)$ and looks like

$$
\begin{gathered}
x_1(t) = \lambda^{-t} x_1, \quad x_2(t) = \lambda^{-t} x_2, \quad \xi_1(t) = \lambda^{-t}\xi_1, \quad \xi_2(t) = \lambda^{-t}\xi_2, \\
\eta_1(t) = \lambda^{-t}\eta_1, \quad \eta_2(t) = \lambda^{-t}\eta_2, \quad e_1(t) = \lambda^{-t} e_1, \quad i_2(t) = \lambda^{-t} i_2, \quad L(t) = \lambda^{-t} L.
\end{gathered} \tag{7}
$$

If correlations (7) to substitute to (6), for the state of equilibrium $\left(\lambda, x_1, x_2, \xi_1, \xi_2, \eta_1, \eta_2, e_1, i_2, L\right)$ at large T we receive optimization problem:

$$
\begin{gathered}
\lambda \to \min, \\
x_1 \geq A_{11}x_1 + A_{12}x_2 + D_{11}\eta_1 + D_{12}\eta_2 + Lc_1 + e_1, \\
x_2 \geq A_{21}x_1 + A_{22}x_2 + D_{21}\eta_1 + D_{22}\eta_2 + Lc_2 - i_2, \\
x_1 \leq \lambda\xi_1, \quad x_2 \leq \lambda\xi_2, \\
(1-\lambda)\xi_1 \leq \eta_1, \quad (1-\lambda)\xi_2 \leq \eta_2, \\
l_1x_1 + l_2x_2 \leq L, \\
e_1 \geq F_1x_1, \quad i_2 \leq H_2x_1, \\
x_1 \geq 0, \quad x_2 \geq 0, \quad \xi_1 \geq 0, \quad \xi_2 \geq 0, \quad \eta_1 \geq 0, \quad \eta_2 \geq 0, \\
e_1 \geq 0, \quad i_2 \geq 0, \quad L \geq 0.
\end{gathered} \tag{8}
$$

Let's consider matrixes

$$R_{11} = \left(c_i^1 l_j^1\right), \quad R_{12} = \left(c_i^1 l_j^2\right), \quad R_{21} = \left(c_i^2 l_j^1\right), \quad R_{22} = \left(c_i^2 l_j^2\right).$$

Since at $0 < \lambda < 1$ we have

$$x_1 \leq \lambda\xi_1 \leq \frac{\lambda}{1-\lambda}\eta_1, \quad x_2 \leq \lambda\xi_2 \leq \frac{\lambda}{1-\lambda}\eta_2,$$

and consequently

$$\eta_1 \geq \frac{1-\lambda}{\lambda}x_1, \quad \eta_2 \geq \frac{1-\lambda}{\lambda}x_2,$$

then subject to $D_{11} \geq 0, \quad D_{12} \geq 0, \quad D_{21} \geq 0, \quad D_{22} \geq 0$ we receive

$$D_{11}\eta_1 \geq \frac{1-\lambda}{\lambda}D_{11}x_1, \quad D_{12}\eta_2 \geq \frac{1-\lambda}{\lambda}D_{12}x_2, \quad D_{21}\eta_1 \geq \frac{1-\lambda}{\lambda}D_{21}x_1, \quad D_{22}\eta_2 \geq \frac{1-\lambda}{\lambda}D_{22}x_2.$$

Further, since

$$\left(l_1x_1\right)c_1 = R_{11}x_1, \quad \left(l_2x_2\right)c_1 = R_{12}x_2, \quad \left(l_1x_1\right)c_2 = R_{21}x_1, \quad \left(l_2x_2\right)c_2 = R_{22}x_2,$$

then subject to non-negativity of matrixes F1 and H2, we come to inequalities

$$
\begin{aligned}
& x_1 \geq A_{11}x_1 + A_{12}x_2 + D_{11}\eta_1 + D_{12}\eta_2 + Lc_1 + e_1 \geq \\
& \geq \left(A_{11} + R_{11} + \frac{1-\lambda}{\lambda}D_{11} + F_1\right)x_1 + \left(A_{12} + R_{12} + \frac{1-\lambda}{\lambda}D_{12}\right)x_2, \\
& x_2 \geq A_{21}x_1 + A_{22}x_2 + D_{21}\eta_1 + D_{22}\eta_2 + Lc_2 - i_2 \geq \\
& \geq \left(A_{21} + R_{21} + \frac{1-\lambda}{\lambda}D_{21}\right)x_1 + \left(A_{22} + R_{22} + \frac{1-\lambda}{\lambda}D_{22}\right)x_2 - H_2x_1.
\end{aligned} \tag{9}
$$

After transferring the term H2x1 on the left and multiplication of both parts of these inequalities on $\lambda > 0$ we receive

$$\lambda(E+H)x \geq \left[\lambda(A+R+F) + (1-\lambda)D\right]x,$$

where $A=\begin{pmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{pmatrix},\quad R=\begin{pmatrix}R_{11} & R_{12}\\ R_{21} & R_{22}\end{pmatrix},\quad F=\begin{pmatrix}F_1 & 0\\ 0 & 0\end{pmatrix},\quad H=\begin{pmatrix}0 & 0\\ H_2 & 0\end{pmatrix},\quad D=\begin{pmatrix}D_{11} & D_{12}\\ D_{21} & D_{22}\end{pmatrix}$ are non-negative square matrixes, E is diagonal identity matrix.

Thus, the problem on maximization of growth rate of open economy (6) is reduced by us to such generalized Neumann model

$$\lambda \to \min,\quad Q(\lambda)x \le \lambda Bx,\quad x \ge 0, \tag{10}$$

where

$$Q(\lambda) = \lambda(A+R+F) + (1-\lambda)D,\qquad B = E + H. \tag{11}$$

Let's begin a finding of equilibrium state of model (6). The equilibrium trajectory of intensities is determined by rate of growth $\alpha = \lambda^{-1}$ and Neumann beam $X = (x_1, x_2, \xi_1, \xi_2, \eta_1, \eta_2, e_1, i_2, L)$. To ensure existence of the non-trivial solution of inequalities system (8), it is necessary to impose some restrictions on parameters of model.

Let's consider, that the matrix A is non-negative and indecomposable, $l > 0,\quad c \ge 0,\quad c \ne 0,$ and matrix A+R+F is productive, i.e. its Frobenius number is less than 1. This restriction means, that the existing technology (A, I, F) allows every worker "to support" itself, carrying out production and the foreign trade operations.

Besides, we consider, that $D \ge 0$ and if $\eta \ge 0,\quad D\eta = 0$, $\eta = 0$. The given assumption means, that any increment of the basic capacities requires material inputs. In other words, we consider, that in the matrix D there are'nt present zero columns.

Let's construct for (10) the dual problem

$$pQ(\lambda) \ge \lambda pB,\quad p \ge 0,$$

where $p = (p_1, p_2)$ is vector-line of dual estimations. As we are interested in a case $x > 0$, it is possible only then, when

$$pQ(\lambda) = \lambda pB. \tag{12}$$

System of linear algebraic equations (12) has non-trivial solution $p \ne 0$ only under condition of

$$\det\big(Q(\lambda) - \lambda B\big) = 0,$$

i.e.

$$\det\big[(1-\lambda)D - \lambda(E - A - R - F + H)\big] = 0. \tag{13}$$

Since the matrix $A + R + F$ is considered productive, then [Ponomarenko, 1995] there is a non-negative matrix $(E - A - R - F)^{-1} \ge 0$.

Let also matrix H is such, that $H \le A + R + F$ (it can be carried out, as $A \ge 0,\quad R > 0,\quad F \ge 0$). Then the matrix $\overline{A} = A + R + F - H \ge 0$ remains productive, i.e. there is a non-negative matrix $\left(E - \overline{A}\right)^{-1} = (E - A - R - F + H)^{-1} \ge 0$. The last statement appears the most important requirement.

Equation (13) is possible to rewrite as

$$\det\left[(E - A - R - F + H)^{-1} D - \mu E\right] = 0,$$

where $\mu = \dfrac{\lambda}{1-\lambda} = \dfrac{1}{1-\lambda} - 1 > 0$ at $0 < \lambda < 1$. To the least value λ corresponds the greatest value μ.

Matrix $(E - A - R - F + H)^{-1} \geq 0$. Then in accordance to the Perron-Frobenius theorem [Ponomarenko, 1995] there is a Frobenius number $\overline{\mu} > 0$ and appropriate Frobenius vector $\overline{z} \geq 0$ such, that

$$(E - A - R - F + H)^{-1} D\overline{z} = \mu\overline{z}\ .$$

Let's note thus, that

$$(E - A - R - F + H)^{-1} D \leq (E - A - R - F)^{-1} D,$$

and therefore $\overline{\mu} \leq \mu^*$, where μ^* - Frobenius number of a matrix $(E - A - R - F)^{-1} D$.

Let's return now to problem (10), which rewrite as

$$\mu \to \min, \quad (E - A - R - F + H)^{-1} x \leq \mu x, \quad x \geq 0.$$

The solution of this problem is reached at $\mu = \overline{\mu}, \quad x = \overline{z}$. Thus rate of growth $\overline{\lambda}^{-1} = 1 + \dfrac{1}{\overline{\mu}} \geq 1 + \dfrac{1}{\mu^*} > 1$, and also structure of output $\overline{x} \geq 0$.

The Basic Result

The basic result of this paper is formulated as such theorem.

<u>Theorem 3.</u> If the matrix $A + R$ is productive, matrix $H \leq A + R + F$, and matrix D has no zero columns, then in model (6) there is a condition of equilibrium with rate of growth $\overline{\lambda}^{-1} = 1 + \dfrac{1}{\overline{\mu}}$, to which corresponds a unique Neumann beam $(x_1, x_2, \xi_1, \xi_2, \eta_1, \eta_2, y_2, L)$, and:

1) $\overline{\mu}$ - Frobenius number, $\overline{x}$ - right Frobenius vector of matrix $(E - A - R - F + H)^{-1} D$;

2) $\overline{\xi} = \overline{\lambda}^{-1}\overline{x}, \quad \overline{\eta} = \dfrac{1-\overline{\lambda}}{\overline{\lambda}}\overline{x}, \quad \overline{l}_1 = F_1\overline{x}_1, \quad \overline{i}_2 = H_2\overline{x}_1, \quad \overline{L} = l\overline{x}.$

Bibliography

[Dorfman, 1958] Dorfman R., Samuelson P.A., Solow R.M. Linear Programming and Economic Analysis, New York: McGraw-Hill, 1958.

[Ashmanov, 1984] Ashmanov S.À. Introduction to Mathematical Economy. - M.: Science, 1984.- 296 p. (in Russian)

[Ponomarenko, 1995] Ponomarenko O.I., Perestyuk M.O., Burym V.M. Basis of Mathematical Economy. - K.: Informtekhnika. 1995.- 320 p. (in Ukrainian)

[Ivanilov, 1971] Ivanilov Yu.P., Petrov A.A. Dynamic Model of Expansion and Reorganization of Production (model). - In book: Cybernetics - on a Service to Communism, vol.6, M.: Energy, 1971, p. 23-50. (in Russian)

Authors' Information

Igor Lyashenko – Kiev National Taras Shevchenko University , Vasilkivska Str., 42, apt. 44, Kiev -22, 03022, Ukraine; e-mail: lyashenko@unicyb.kiev.ua

Olena Lyashenko – Kiev National Taras Shevchenko University , Vasilkivska Str., 42, apt. 44, Kiev -22, 03022, Ukraine; e-mail: lyashenko@unicyb.kiev.ua

FUZZY SET THEORY APPROACH AS THE BASIS OF ANALYSIS OF FINANCIAL FLOWS IN THE ECONOMICAL SECURITY SYSTEM

Alexander Kuzemin, Vyacheslav Lyashenko

Abstract: The problems of formalization of the process of matching different management subjects' functioning characteristics obtained on the financial flows analysis basis is considered. Formal generalizations for gaining economical security system knowledge bases elements are presented. One of feedback directions establishment between knowledge base of the system of economical security and financial flows database analysis is substantiated.

Keywords: financial flows, economical security, knowledge database, fuzzy sets.

ACM Classification Keywords: K.4.4 Electronic Commerce - Security; I.5.1 Pattern recognition Models – Fuzzy set

Introduction

Objective laws of economical development assume constant interaction of a managing subject (economical agent) with different manifestations of the external medium. At the same time variations of the external medium stipulate rebuilding of the economical agent internal structure. It is connected with the fact that the given process promotes preservation of the management subject stability, overcoming of crisis-forming external factors; those as a whole can be regarded as a result of purposeful response of the corresponding system of economical security. At the same time coordination and efficiency of the economical security system can be increased at the cost of introduction of financial flows analysis subsystem both of the managing subject itself and the flows connected with economics functioning as a whole. It can be explained by the fact that the direction and velocity of financial flows' movements in many respects reflects transient nature of economical processes and changes taking place in the external medium of managing subject functioning. In this case, the range of the indicated problems becomes apparent in the period of transformations which in fact permanently embrace all without exceptions institutional formations both in the developed and developing countries due to evaluation process of economical relations' development. Such a situation emerges because just at the period of transformation changes the possibility of the rupture between the flows of the real and financial sectors of economics increases.

That is why the problems of various management subjects financial and goods flows control are constantly in view of researches.

At the same time, the effectiveness of financial flows' analysis depends in many instances on the existing systems of their estimate and the system of conclusion concerning any decision-making. This is already directly associated with the stage of formalization and analysis of different characteristics of financial flows in the system of any management subject economical security.

Substantiation of Investigation Aims and Problems

Traditionally designation of investigations' direction is solved in the context of definite problems of the managing subject taking into account the temporal factor of the economical situation development. Nevertheless, one of the widespread tooling of the set problem is the use of the methods and approaches of simulation modeling [1], this makes it possible to link them with visualization and semantic interpretation of movement of the flows under investigation. In this case recognizing that definite characteristics of financial flows' analysis are interrelated as a rule to the temporal factor, the use of the apparatus of the analysis theory and time sequence prediction [2,3] turns out to be rather important.

But in the first and in the second cases the problem of matching the data obtained in the course of financial flows analysis which reflect dynamics of different components of economical activity for adequate decision-making as a rule remains beyond the field of vision.

Yet, a managing subject constantly comes up against the problem on some decision-making in the process of his functioning, this makes his system of economical security an integral part of the general management system. The reason is that a definite managing subject is developing through widening of his vital resource space on the basis of solving conflicts with internal and external media which are an integral part in any decision-making and which require a distinct agreement. That is why the problem of agreement of different characteristics and indices of the managing subjects functioning processes obtained on the financial flows analysis basis for adequate decision-making in the system of their economical security is presented as the main problem of investigation.

The Problems of Financial Flows' Analysis in the Social-economical Systems of the Transition Period. The Decision-making Formalization Apparatus Choice.

But before proceeding directly to the preset problem treatment it is necessary to note that for the modern managing subject the problem of the economical security system creation, development and control of this security assumes solution of the complicated and multi-aspect problems complex. At the same time one should take into account that the economical security of development of some economical agent is defined not only through widening a set of elements of its resource space but the degree of protection against internal and external threats. Their activation can significantly decrease or even destroy this space. That is why the choice of economical security system functioning criteria, creation of data bases and knowledge bases serving the foundation of adequate economical security system creation, matching different characteristics of the processes for objective and appropriate decision-making are among the problems of paramount importance in the given direction.

In this case, it is also important to realize that there are objective reasons for the limited application of the decision-making models in the social-economical systems [4].

At the same time, the preset problem solution complexity is enhanced with the presence of different estimates of the same processes taking place during the transformation period of economical relations development. Several of the problems connected with estimation of tendencies in development of the funds market can be shown as the most striking example. This funds market on one side is rather little predicted for the countries with c transient economy, and on the other side it is very important component for decision-making in the system of economical security as it reflects the processes of redistribution of free monetary and financial resources.

Fig.1 shows the profitableness distribution density of some fund indices of the Ukrainian securities' market (KAC-20 (a), SOKRAT (b), Sofia Bondar Priadka Ukraine (c), TEKT-KKOC Price Bonds (d), PFTS (e), KINDEX (f) calculated for the period which covers a year and a half from 01.06.03 based on the sites of the corresponding companies).

As can be seen from Fig.1 the analysis and proper conclusion is hampered with a distinct uncertainty of fund indices' dynamics. Significant dissimilarity of statistical characteristics of indices under consideration both from the standard meanings and from each other can be considered as its manifestation.

The problem of estimation of financial flows mutual influence on the side of different economical agents on the activity of each other is also no less important in the context of the considered problems. This also manifests itself both on the micro- and macro level [5].

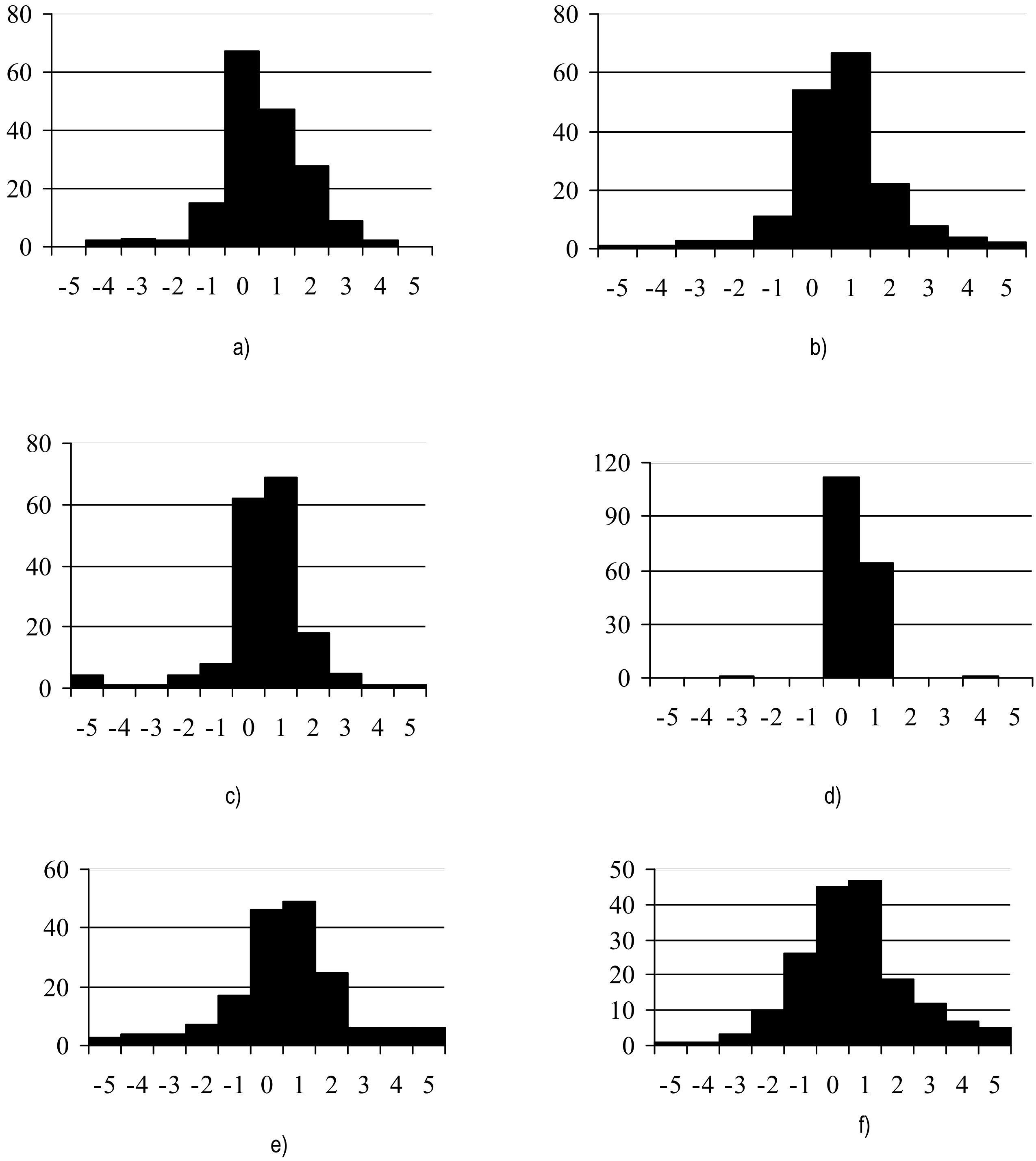


Fig.1. Profitableness distribution density of some Ukrainian fund indices

Thus, in the process of analysis of different financial flows we obtain the database, which represents in the general case some dissimilar database and which should be transformed into the knowledge base of full value later on. In this case considering the structure of the given process (Fig.2) it can be characterized in the most general terms as the procedure of obtaining information images which in the following suffer a definite grouping and analysis.

That is to say, from the formal point of view the structure of the knowledge base formation process for decision-

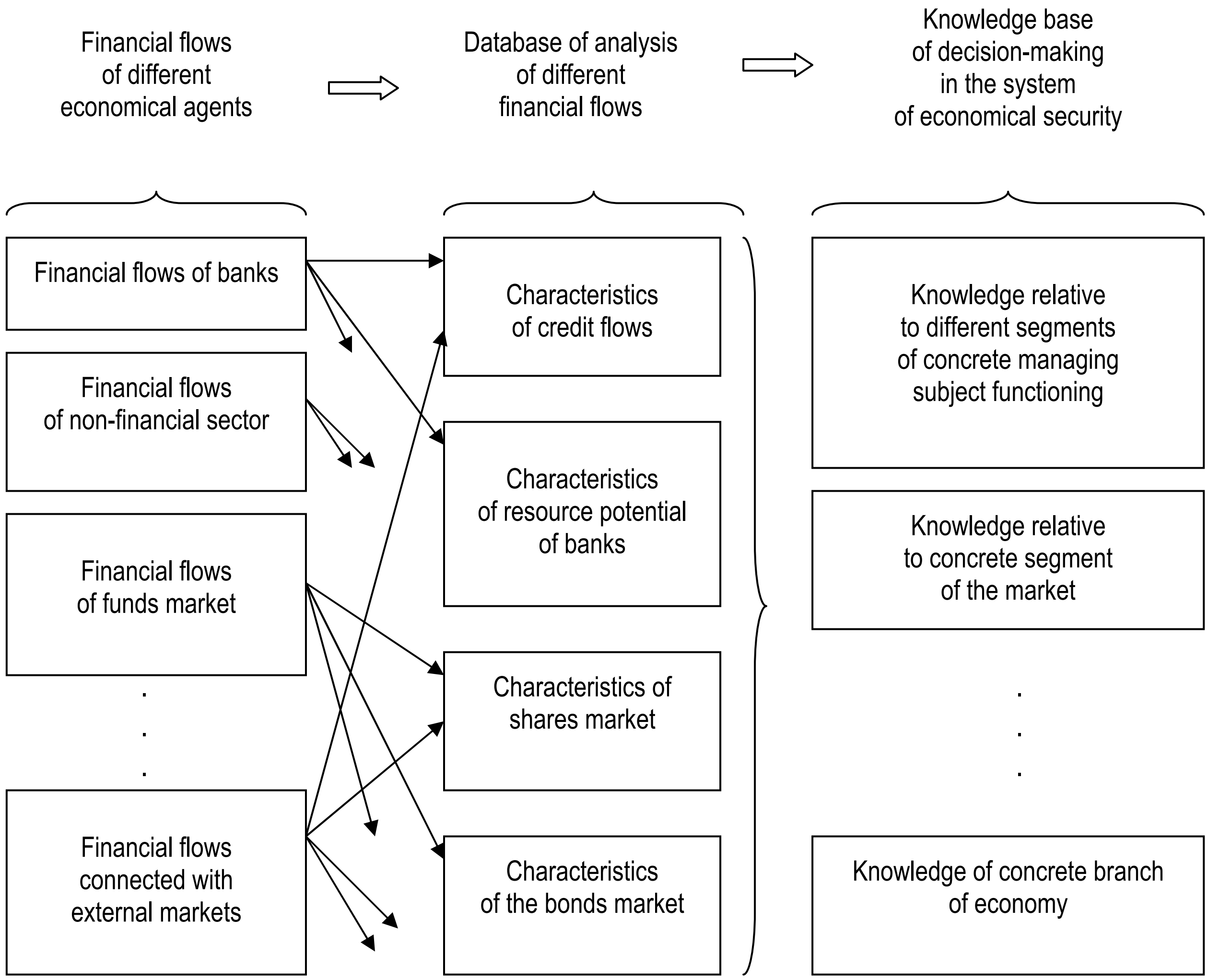


Fig.2. Generalized structure of the knowledge base formation process for decision-making in the system of economical security on the basis of different financial flows analysis

making in the system of economical security based on different financial flows analysis can be presented in the form of image recognition theory procedures. However, taking into account the fact that all economical and financial processes carry the randomness treats and inertia properties of the objects under investigations (though they give a high degree of conditionality of the future behavior of the previous ones) do not guarantee their complete execution. This fact determines to consider the theory of fuzzy set as an additional mathematical apparatus.

The utility of such a union of the indicated theories is connected, first of all, to the fact that to establish links between elements of the database and knowledge base, to clarify which of them should be or shouldn't be considered linked with some relation, one is forced to solve the problems of interaction between enterprises which sometimes do not correlate well. At the same time, it is necessary to take into account the difference in transitions for gaining knowledge between different characteristics of financial flows for the same economical processes. In other words, the case in point is only a degree of belonging of the concrete characteristics of financial flows required for gaining one or other knowledge. Then, the higher is degree of the alternative belonging to fuzzy set the greater is degree of attaining this aim when choosing the given alternative as a solution.

Matching of Different Characteristics of Financial Flows in the Economical Security System

Along with the discussed above one should note that the standard operations with fuzzy sets could be used as the simplest procedures for knowledge base elements' derivation. As an example the inference concerning the use of a definite set of the fund indices combinations existing on the market for gaining the criterion of the enterprise assets maximal diversification into different securities can be considered. Let us assume that we have belonging functions $\mu_F(P)$ which represent profitableness of investments (P) based on the analysis of three different fund indices of some their fuzzy set (F) (Fig. 3),

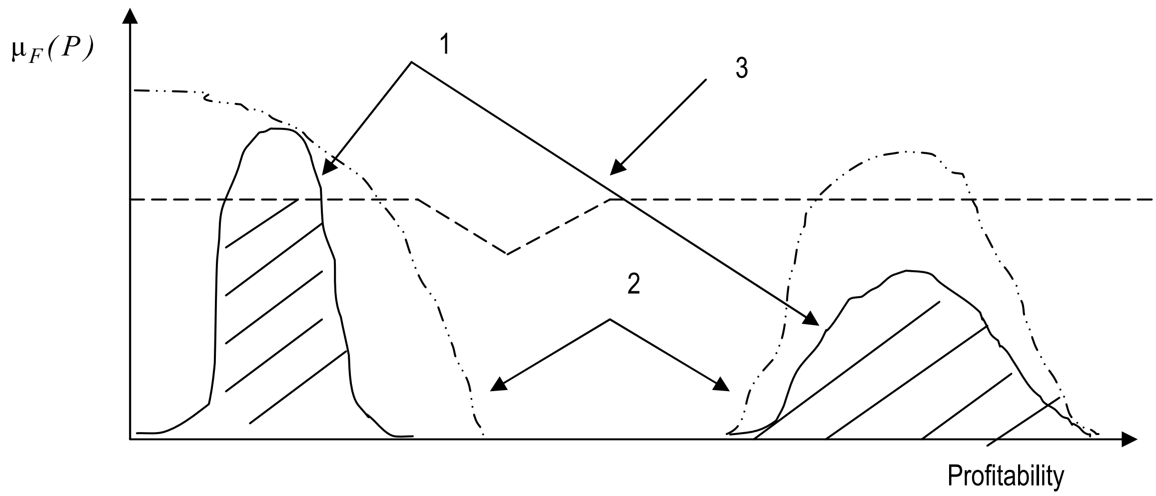


Fig.3. Fuzzy setting of profitability function on the basis of different fund indices analysis and knowledge base element formation results

where

curve 1 – represents insufficiently high and stable profitability degree by investments into securities according to the values of the first of the fund indices under investigation;

curve 2 – represents sufficiently high and not very stable profitability degree by investments into securities according to the values of the second of the fund indices under investigation;

curve 3 – represents sufficiently mean and sufficiently stable profitability degree by investments into securities according to the values of the third of the fund indices under investigation.

Then to attain sufficiently stable and high profitability by the investments into securities it is necessary to consider the fuzzy crossing operation for the obtained belonging function by every of the fund indices under investigation (the shaded region in Fig.3).

Nevertheless, for more adequate solution of the problems emerging when choosing different managing subjects functioning processes' characteristics being formed on the basis of the existing financial flows analysis it is expedient to consider their generalized matching to obtain the knowledge base elements. Such solution can be obtained starting from the fact that the corresponding crossing of the considered sets which represent the concrete characteristics of the managing subjects functioning processes should be the maximal one and the symmetric difference approaches zero.

In the formalized form this can be written as follows: so the crossing maximization condition assumes the following form:

$$\mu_{A1\cap A2\cap\ldots\cap An}(x) = min(\mu_{A1}(x),\mu_{A2}(x),\ldots,\mu_{An}(x)),$$

and the condition for the symmetrical difference has the form:

$$\mu_{A1\oplus A2\oplus\ldots\oplus An}(x) = max(\mu_{A1}(x),\mu_{A2}(x),\ldots,\mu_{An}(x)) - min(\mu_{A1}(x),\mu_{A2}(x),\ldots,\mu_{An}(x)),$$

where $\mu_{A1\cap A2\cap\ldots\cap An}(x)$ and $\mu_{A1\oplus A2\oplus\ldots\oplus An}(x)$ – belonging function of some knowledge base element which is defined from the concrete belonging functions $\mu_{An}(x)$ data base X. Characterizing corresponding fuzzy sets of different analyzed financial flows' indices.

In this case, the condition of the fuzzy sets' matched join for knowledge base elements' formation is also a good tool:

$$\frac{\mu_{A1\cap A2\cap\ldots\cap An}(x)}{\mu_{A1\oplus A2\oplus\ldots\oplus An}(x)} = \frac{min(\mu_{A1}(x),\mu_{A2}(x),\ldots,\mu_{An}(x))}{max(\mu_{A1}(x),\ldots,\mu_{An}(x)) - min(\mu_{A1}(x),\ldots,\mu_{An}(x))},$$

as can be seen from the above formulae the symmetrical difference approach to zero can point to maximization of joining the considered fuzzy subsets and, as a result, characterize in some way the degree of matching of the obtained corresponding characteristics of financial flows when deriving the concrete element of the knowledge base.

It is precisely the given approach was used as a basis for building the developed economical security system; their main components are:

- subsystems for analysis of external factors acting on the financial flows movement on the basis of the indices system characterizing the position of a country in the external financial markets;
- subsystems for analysis of the degree and direction of action on financial flows on the side of other countries and organization;
- subsystems for analysis of financial flows movement on the basis of economy bank sector indices taking into account researches of the possibility of emerging crises situations with a mutual movement of flows of the real and financial sectors of economy, influence of the external bank capital;
- subsystems for analysis of efficiency of monetary and financial resources redistribution through the securities market.

Conclusion

Here it should be noted that the considered approaches to matching of different managing subjects functioning characteristics obtained on the basis of the financial flows analysis mustn't be regarded as the unique ones. For example, the problem of account of the possible range of financial flows' indices quantity variation when deriving different knowledge elements is rather significant. As a whole, it can result in correction of the corresponding belonging functions. Nevertheless, despite the given remark it should be noted that its elimination is possible through widening functional subsystems of the general economical security system. In this case, it should be spoken about correction of the knowledge base; establishment of the feedback with database, which will control the process of gaining, needed indices of financial flows' analysis. At the same time, it is also expedient to use the theories of fuzzy sets as formal apparatus for such problem solution.

Bibliography

1. Gorbunov A. Financial flows' management. – M.: Globus, 2003, 224p.
2. Murray C.J., Nelson C.R. The Uncertain Trend in U.S. GDP // Journal of Monetary Economics. 2000, 46. pp.79–95.
3. Dutt S.D., Ghosh D. An Empirical Examination of Exchange Market Efficiency // Applied Economic Letters. 1999, 6. №2. pp.89–91.
4. Voloshin A.F. On the problems of decision-making in the social-economical systems // XI-th International Conference «Knowledge-Dialogue-Solution» – Varna, 2005. V.1. pp.205–212.
5. Kuzemin A., Liashenko V., Bulavina E., Torojev A. Analysis of movement of financial flows of economical agents as the basis for designing the system of economical security (general conception) // Third International Conference «Information Research, Applications, and Education». June 27-30. Varna, Bulgaria. – Sofia: FOI-COMMERCE, 2005. pp.204–209.

Authors' Information

Alexander Ya. Kuzemin – Prof. of Information Department, Kharkov National University of Radio Electronics, Head of IMD, Ukraine, kuzy@kture.kharkov.ua

Vyacheslav V. Lyashenko – senior scientific employee Kharkov National University of Radio Electronics, Ukraine

DYNAMICAL SYSTEMS IN DESCRIPTION OF NONLINEAR RECURSIVE REGRESSION TRANSFORMERS

Mykola Kirichenko, Volodymyr Donchenko, Denys Serbaev

Abstract: The task of approximation-forecasting for a function, represented by empirical data was investigated. Certain class of the functions as forecasting tools: so called RFT-transformers, – was proposed. Least Square Method and superposition are the principal composing means for the function generating. Besides, the special classes of beam dynamics with delay were introduced and investigated to get classical results regarding gradients. These results were applied to optimize the RFT-transformers. The effectiveness of the forecast was demonstrated on the empirical data from the Forex market.

Keywords: empirical functions, learning samples, beam dynamics with delay, recursive nonlinear regressive transformer, Generalized Inverse, Least Square Method.

ACM Classification Keywords: G.1.2 Approximation, G.1.3 Numerical Linear Algebra, G.1.6 Optimization

Introduction

Approximation of the function represented by its values is classical direction of researches for mathematics both in the deterministic statement, and in statistical variant (see, [4,5], and also [1-3]). Classical results in this area are full enough represented in the specified works. They show importance of superposition-recurrence as means of generation of a class of approximating functions.

Natural way of use of approximation in applied researches is the forecast of values of researched function. Means of forecasting have special importance in the modern unified systems of automation of management of firm: in so-called Business Intelligence systems and, in particular, in their structural elements as DSS.

The Method of optimal constructing of forecasting means, based on application of classical LSM in combination with superposition-recursion, linear transformations of coordinates and component-wise nonlinear transformations in the context of so called RFT-transformers is offered in present article. The variant of building of RFT-transformers together with algorithm of synthesis of basic element was discussed in [6]. Special classes of systems with delay both for single trajectories and for their beam is introduced and investigated in the article, since optimization is connected to representation of researched objects by control systems with delay.

Recursive Regression Nonlinear Transformer: RFT-transformer

As it was already marked, the idea of recursive regression nonlinear transformer (RFT-transformer) is offered in [6], in variant, which can be named reverse recursion. Such transformer: with reverse recursion [6], and a forward one, which will be offered and considered below, - are constructing by recursive application of the certain standard element, which will be designated by abbreviation ERRT (Elementary Recursive Regression Transformer). Process of construction of the RFT-transformer consists in connection of the next ERRT to already constructed during execution of the previous steps transformer according to one of three possible types of connection. Types of connection which will be designated as “parinput”, “paroutput” and “seq”, realize natural variants of use of an input signal: parallel or sequential over input, - and parallel over output. An input of the again attached ERRT, in variant of reverse recursion [6], is the input of whole RFT, and in the variant of a forward recursion, which will offered below, the input of whole system is an input of the already constructed RFT.

The basic structural element of the RFT-transformer is ERRT - an element [6], which is determined as mapping from R^{n-1} in R^m of a kind:

$$y= A_+\Psi_u\left(C\begin{pmatrix}x\\1\end{pmatrix}\right), \tag{1}$$

which approximates the dependence represented by training sample $(x_1^{(0)}, y_1^0),...,(x_M^{(0)}, y_M^{(0)})$, $x_i^{(0)} \in R^{n-1}$, $y_i^{(0)} \in R^m$, $i=\overline{1,M}$,

where:

- C–(n×n) - matrix, which performs affine transformation of the vector x$\in R^{n-1}$ - an input of the system; it is considered to be given at the stage of synthesis of ERRT;
- Ψu – nonlinear mapping from R^n in R^n , which consists in component-wise application of scalar functions of scalar argument $u_i \in \Im$, $i=\overline{1,n}$ from the given final set $\Im$ of allowable transformations, including identical transformation: must be selected to minimize residual between input and output on training sample during synthesis of ERRT;
- A₊– solution A with minimal trace norm of the matrix equation

$$AX_{\Psi_u C} = Y, \tag{2}$$

in which matrix $X_{\Psi_u C}$ formed from vector-columns $\Psi_u(C\begin{pmatrix}x_i^{(0)}\\1\end{pmatrix}) = \Psi_u(z_i^{(0)})$, and Y – from columns $y_i^{(0)}, i=\overline{1,M}$.

In effect, ERRT represents empirical regression for linear regression y on Ψu(C$\begin{pmatrix}x\\1\end{pmatrix}$), constructed with method of the least squares, with previous affine transformation of system of coordinates for vector regressor *x* and following nonlinear transformation of each received coordinate separately.

Remark 1. Further we shall assume that functions of component-wise transformations from $\Im$ would have a necessary degree of smoothness where it is necessary.

Task of synthesis of ERRT by an optimal selection of nonlinear transformations of coordinates on the given training sample was introduced and solved in already quoted above work [6]. The solution of a task of synthesis is based on methods of the analysis and synthesis of the pseudoinverse matrices, developed in [7-9]. Particularly, reversion of Grevil's formula [10] was received in these works that recurrently allows recalculating pseudoinverse matrices at replacement of a line or a column of the appropriate matrix.

Generalized Recursive Regression Transformers with Forward Recursion

Recursion in construction of the RFT-transformer in variant of forward recursion offered below will be considered in the generalized variant in which several ERRTs is used in recurrent connection to already available RFT-structure. Total quantity of recurrent references we shall designate through N, and quantity of ERRTs used on a step m – through k_m , $m = \overline{1, N}$. The common number of ERRTs, used for construction of whole transformer will be designated through T: $T = \sum_{m=1}^{N} k_m$.

Variants of the generalized forward recursion which depend from type of connection of attached ERRT: parinput, paroutput and seq, - are represented on figures 1-3. Where $\hat{y}$ designates an output of already available RFT-structure or an output of the same structure after connection of the next ERRT from the total number k_m of such elements, attached on a step m of the recursion: $m = \overline{1, N}$. Each figure is accompanied by the system of equations which determine transformation of a signal on the next step of recursion.

- Type parinput:

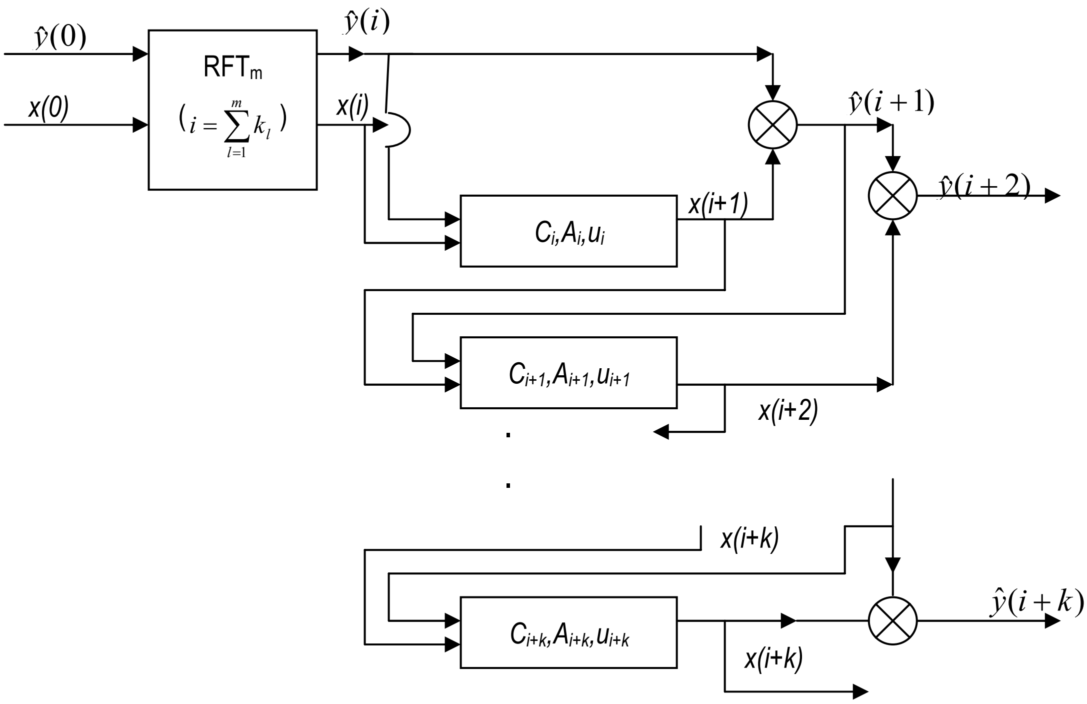


Figure 1. Scheme of connection of parinput type – forward recursion.

In parinput type of connection already constructed structure approximates an output of training sample by its input, and set of ERRTs - resulting residual of such approximation depending from an input of training sample. Transformation of the information at this type of connection is described by the following system:

$$\begin{aligned} & x(i+j) = A_{i+j-1}\Psi_{u_{i+j-1}}(C_{i+j-1}x(i)), \\ & \hat{y}(i+j) = \hat{y}(i+j-1) + A_{i+j-1}\Psi_{u_{i+j-1}}(C_{i+j-1}\cdot x(i)). \\ & i = \sum_{l=1}^{m} k_l,\ j = \overline{1, k_{m+1}} \end{aligned} \tag{3}$$

- Type paroutput:

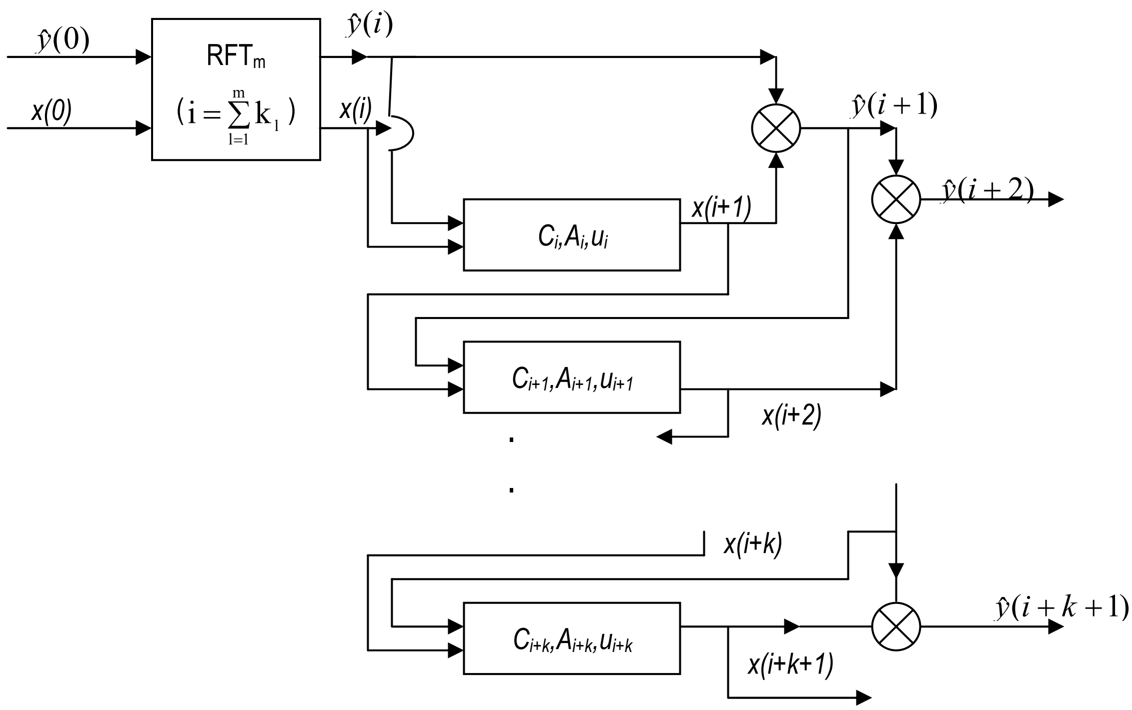


Figure 2. Scheme of connection of paroutput type – forward recursion.

The system describing transformation inside the transformer and communication between an input and an output, at this type of connection looks like:

$$\begin{aligned} & x(i+j) = A_{i+j-1}\Psi_{u_{i+j-1}}(C_{i+j-1}x(i+j-1)), \\ & \hat{y}(i+j) = \hat{y}(i+j-1) + A_{i+j-1}\Psi_{u_{i+j-1}}(C_{i+j-1}\cdot x(i+j-1)). \\ & i = \sum_{l=1}^{m} k_l,\ j = \overline{1, k_{m+1}} \end{aligned} \tag{4}$$

- Type seq:

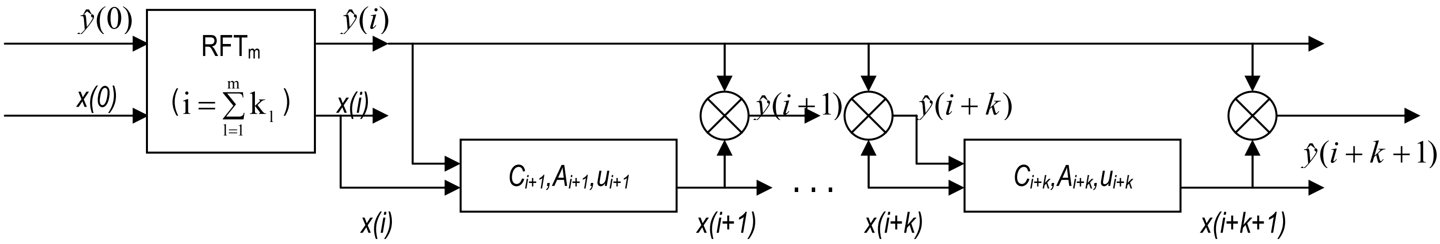


Figure 3. Scheme of connection of seq type – forward recursion.

The equations describing transformation of the information on the next step of recursion look as follows:

$$x(i+j) = A_{i+j-1}\Psi_{u_{i+j-1}}(C_{i+j-1}x(i+j-1)),$$
$$\hat{y}(i+j) = \hat{y}(i) + A_{i+j-1}\Psi_{u_{i+j-1}}(C_{i+j-1}\cdot x(i+j-1)).\tag{5}$$
$$i=\sum_{l=1}^{m} k_l,\ j=\overline{1,k_{m+1}}$$

In this scheme of connection RFT_{m-1} approximates an output part of training sample by input part, and set of ERRTs approximates residual which depends from an output of the next ERRTs.

Entry conditions for all types of connections are described by equations:

$x(0) = x$ – an input of whole RFT-transformer, (6)

$\hat{y}(0) = 0,\ \hat{y}(1) = x(1)$ for all types of connections. (7)

According to (6) in a training mode the inputs of RFT-transformer are $x_{i1}^{(0)}$: $x_i^{(0)} \in R^{n-1}$, $y_i^{(0)} \in R^m$, $i=\overline{1,M}$, and outputs are – $y_i^{(0)} \in R^m$, $i=\overline{1,M}$.

Connections of recursive construction of the RFT-transformer are determined so, that standard functional of the least squares method is minimized during its construction, i.e.

$$\sum_{i=1}^{M} \| y_i^{(0)} - RFT(x_i^{(0)}) \|^2 .\tag{8}$$

Equations (3) - (7) for N steps of recursion with common number T of used ERRTs, $T=\sum_{m=1}^{N} k_m$, k_m – quantity of ERRTs, used on a step with number $m=\overline{1,N}$, and also efficiency functional (8) - represent mathematical model of RFT-transformer.

Discrete control systems with delay

Equations (3)-(8) represent the system of the recurrent equations being certain generalization of a classical control system with discrete time (for details see [12-16]) and first of all in referring to delay.

2.1. The simple and combined control systems with delay

Definition. Simple, accordingly - combined, - nonlinear control system with delay on a time interval $\overline{0,N}$ is a control system whose trajectories are defined by system of recurrent equations (9), accordingly - (10), entry conditions (11) and efficiency functional (12) and represented below:

$$x(j+1) = f(x(j-s(j)), u(j), j) \ , \tag{9}$$

$$x(j+1) = f(x(j), x(j-s(j)), u(j), j) \tag{10}$$

$$j = \overline{0, N-1}, \qquad x(0) = x^{(0)} , \tag{11}$$

$$I(U) = \Phi(x(\mathrm{N})) , \tag{12}$$

where function s(j): s(0)=0, $s(j) \in \{2,..., j\}$, $j = \overline{0, N-1}$ – is known.

Evidently, systems with delay for a beam of trajectories are determined. Entry conditions of trajectories of a beam we shall designate $(x(0))_i = x_i^{(0)}$, $i = \overline{0, M}$, M– quantity of trajectories of a beam. Trajectories of a beam for both types of systems with delay we shall designate by the appropriate indexation: $x_i(j)$, $j = \overline{0, N-1}$, $i = \overline{0, M}$.

Let's define efficiency functional for a beam of dynamics which we shall consider dependent only from final states of trajectories of a beam, with equation:

$$I_p(U) = \sum_{i=1}^{M} \Phi^{(i)}(x_i(N)) . \tag{13}$$

Remark 2. Evidently, efficiency functional, as well as for classical control systems, may be defined on all trajectory or trajectories. However, systems with delay at which efficiency functional depends only from final states of a trajectory will be considered in context of RFT-transformers.

Phase trajectories of simple systems with delay are defined by a set of functions f(z, u, j), $j = \overline{0, N-1}$ with one argument z, responsible for a phase variable, and for combined one - a set f(z,v,u,j), $j = \overline{0, N-1}$ with two variables: z, v, which respond for phase variables. Gradients on a phase variables will be designated accordingly $grad_z f$, $grad_v f$, in this last case.

The problem of optimization for both types of such systems with delay is being solved, as well as in a classical case, with construction of the conjugate systems and functions of Hamilton. The assumptions of the smoothness providing correct construction of conjugate systems and functions of Hamilton, and also their use for gradients calculations on controls completely coincide with classical and further will be considered automatically executed.

Optimization in simple control systems with delay

Definition. The conjugate system of a simple control system with delay (9), (11), (12) we shall name the following recurrent equation concerning p(k), $k = \overline{N,0}$:

$$p(k) = \sum_{j \in \{j: j - s(j) = k, j > k\}} grad_{x(k)} \{ p^T(j+1) f(x(k), u(j), j) \} \quad , \qquad k = \overline{N-1,0} \tag{14}$$

with the initial condition

$$p(N) = -grad_{x(N)}\Phi(x(N)) \tag{15}$$

Accordingly, in the case of a beam of trajectories the conjugate systems are defined for each trajectory by equations:

$$p^{(i)}(N) = -grad_{x_i(N)}\Phi^{(i)}(x_i(N)), \tag{16}$$

$$p^{(i)}(k) = \sum_{j\in\{j:j-s(j)=k,\,j>k\}} grad_{x_i(k)}\{p^{(i)^T}(j+1)f(x_i(k),u(j),j)\}$$
$$k=\overline{N-1,0},\, i=\overline{1,M}. \tag{17}$$

Function of Hamilton for simple system with delay is defined by a classical equation:

$$H(p(k+1),x(k-s(k)),u(k),k) = p^T(k+1)f(x(k-s(k)),u(k),k),\, k=\overline{N-1,0}. \tag{18}$$

For a beam of trajectories of a simple control system with delay the set of functions of Hamilton $H^{(i)}$, $i=\overline{1,M}$ for each of trajectories $x_i(k)$, $k=\overline{N-1,0}$.

Theorem 1. The gradient on control from the efficiency functional which depends only from final states of trajectories of a beam, for a simple control system with delay is defined by equations:

$$grad_{u(k)}I(U) = -\sum_{i=1}^{M} grad_{u(k)}\left\{p^{(i)^T}(k+1)f(x^{(i)}(k-s(k)),u(k),k)\right\} = -grad_{u(k)}\sum_{i=1}^{M} H_v^{(i)}(x_i(k),u(k),k),$$
$$k=\overline{0,N-1} \tag{19}$$

The evidence. The evidence will be carried out precisely the same as in a classical case: first - for one trajectory, and then by use of additivity of efficiency functional on trajectories of system.

Optimization in the combined control systems with delay

Definition. The conjugate system for the combined control system with delay is the system determined by recurrent equations:

$$p(k) = grad_z\{p^T(j+1)f(x(k),x(k-s(k)),u(j),j) \; +$$
$$+ \sum_{j\in\{j:j-s(j)=k,\geq kj\}} grad_v\{p^T(j+1)f(x(k),x(j),u(j),j)\},$$
$$k=\overline{N-1,0},\, i=\overline{1,M} \tag{20}$$

with the initial condition

$$p(N) = -grad_{x(N)}\Phi(x(N)). \tag{21}$$

Accordingly, function of Hamilton H(p(k+1),x(k),x(k–s(k),u(k),k)) of the combined system is defined by a equation:

$$H(p(k+1),x(k),x(k-s(k),u(k),k)) = p^T(k+1)f(x(k),x(k-s(k)),u(k),k). \tag{22}$$

As before, the upper index (i), $i=\overline{1,M}$: $p^{(i)}(k)$, $H^{(i)}$, $k=\overline{N-1,0}$, will define objects for trajectories of a beam.

Theorem 2. Gradients on the appropriate controls from the efficiency functional which depends only from final values of trajectories of the combined control system with delay, are defined by gradients from the appropriate functions of Hamilton:

$$grad_{u(k)} I_p(U) = \text{- grad}_{\text{u(k)}} \text{H(p(k+1), x(k), x(k - s(k)), u(k)), k))}\ . \qquad (23)$$

And, hence, for a beam of dynamics the appropriate gradient is defined by equation:

$$grad_{u(k)} I_p(U) = -\sum_{\text{i}=1}^{\text{M}} \text{grad}_{\text{u(k)}} \text{H}^{(\text{i})}(\text{p}^{(\text{i})}(\text{k}+1), \text{x}_\text{i}(\text{k}), \text{x}_\text{i}(\text{k - s(k), u(k), k)})\ . \qquad (24)$$

The evidence. The evidence will be carried out in a standard way for use of the conjugate systems and functions of Hamilton.

RFT-transformers and control systems with delays

As it was already marked RFT-transformer may be represented by a control system with delay. More precisely, the following theorem is fair.

Theorem 3. Regression RFTN-transformer with the direct N-times recursive reference to $k_m, m=\overline{1,N}$ ERRT elements on each of the steps of the recursion is represented by the nonlinear combined beam of dynamics with delay on an interval $\overline{0,T}$ $, T=\sum_{l=1}^{N} k_l$:

with a phase variable

$$z(t)=\begin{pmatrix} z_1(t) \\ z_2(t) \end{pmatrix},\ z_i(t)\in R^m, i=1,2\ , t=\overline{0,T}\ ; \qquad (25)$$

- by the system of the recurrent equations which determines trajectories of a beam:

$$z^{(i)}(t+1)= f(z^{(i)}(t), z^{(i)}(t-k(t)), C_t, t) = \begin{pmatrix} f_1(z^{(i)}(t), z^{(i)}(t-k(t)), C_t, t) \\ f_2(z^{(i)}(t), z^{(i)}(t-k(t)), C_t, t) \end{pmatrix}, \qquad (26)$$

  with $f_1, f_2 \in R^m,\ t=\overline{1,T-1}$, $i=\overline{1,M}$, dependent on topology of the RFT-transformer and determined by equations (29)-(31);

- initial conditions:

$$z^{(i)}(0)=\begin{pmatrix} x_i^{(0)} \\ 0 \end{pmatrix}, i=\overline{1,M}\ , \qquad (27)$$

  where $z^{(i)}(0)$ $, i=\overline{1,M}$ initial conditions of trajectories of a beam, and $x_i^{(0)}$ $, i=\overline{1,M}$ – elements of an input of training sample;

- efficiency functional I(C), $C=(C_1,\dots,C_T)$, which depends on matrices $C_1,\dots,C_T$ as on controls:

$$I(C)=\sum_{k=1}^{M} \| y_k^{(0)} - z_2(T) \|^2\ , \qquad (28)$$

where $y_k^{(0)}$ $, k=\overline{1,M}$ components of an output of training sample.

The proof can be found in [8].

The proved statement enables using of methods of optimization of the theory of control for optimization of already constructed RFTN-transformer on residual size depending on matrices $C_0, C_1, ..., C_{T-1}$. This statement also is a subject of the following theorem.

Theorem 4. By presence of continuous derivatives up to the second order inclusive of functions of family $\Im$ RFT - transformer may be optimized by gradient methods with gradients of the efficiency functional on matrices C, as on parameters.

The proof. According to the theorem 3 RFT-transformer may be represented by the combined control system with delay, and according to the theorem 2 gradients on the appropriate controls - parameters of the RFT-transformer are defined by equation (24). Importance of the given theorem lies in that it gives exhaustive interpretation of Back Propagàtion algorithm, outlining at the same time borders of the specified method

RFT-transformers in various variants: linear, nonlinear, and also with gradient optimization - were used for the forecast of values on days in such high volatility market described by modeling uncertainty, as the currency market. Such use demands the decision of some questions, concerned formation of training sample, choice of an interval of forecasting and some other which remain behind frameworks of discussion. We shall note only, that the standard set of characteristics was predicted: close rate, the maximal and minimal rates for the day. It is necessary to note, that the prediction well catches turns of a exchange rate and together with other technical parameters forecasts on the basis of suitable RFT - transformer may be effectively used in the appropriate market

Bibliography

1. Ивахненко А.Г. Самоорганизующиеся системы распознавания и автоматического управления. – К.: Техника. – 1969. – 395 с.
2. Колмогоров А.Н. О представлении непрерывных функций нескольких переменных суперпозициями непрерывных функций меньшего числа переменных.// Докл. АН СССР. – 1956. – Т. 108. – № 2. – С. 179- 182.
3. Арнольд В.И. О функциях трех переменных.// Докл. АН СССР. – 1957. – Т. 114. – № 4. – С. 953- 956.
4. Кириченко Н. Ф., Крак Ю. В., Полищук А.А. Псевдообратные и проекционные матрицы в задачах синтеза функциональных преобразователей.// Кибернетика и системный анализ –2004.–№3.
5. Кириченко Н.Ф., Лепеха Н.П. Применение псевдообратных и проекционных матриц к исследованию задач управления, наблюдения и идентификации.// Кибернетика и системный анализ. – 2002. – № 4. – С. 107-124.
6. Бублик Б.Н., Кириченко Н.Ф. Основы теории управления. – К.: Высшая школа, 1975.–328 с.
7. Бублик Б.Н., Гаращенко Ф.Г., Кириченко Н.Ф. Структурно-параметрическая оптимизация и устойчивость динамических пучков .– К.: Научная мысль, 1985.–304 с.
8. Нелинейные рекурсивные регрессионные преобразователи: динамические системы и оптимизация / Кириченко Н.Ф., Донченко В.С., Сербаєв Д.П. .// Кибернетика и системный анализ.–2005.–№ .– 4.

Authors' Information

Mykola Kirichenko - professor, Institute of Cybernetics of NASU, e-mail: kir@dept115.icyb.kiev.ua

Volodymyr Donchenko - professor, Kiev National Taras Shevchenko University, Department of System Analysis and Decision Making Theory, e-mail: vsdon@unicyb.kiev.ua

Denis Serbaev - post graduate course student, Kiev National Taras Shevchenko University, Department of System Analysis and Decision Making Theory, e-mail: sdp@unicyb.kiev.ua

DATA FLOW ANALYSIS AND THE LINEAR PROGRAMMING MODEL[1]

Levon Aslanyan

Abstract: The general discussion of the data flow algorithmic models, and the linear programming problem with the variating by data flow criterion function coefficients are presented. The general problem is widely known in different names - data streams, incremental and online algorithms, etc. The more studied algorithmic models include mathematical statistics and clustering, histograms and wavelets, sorting, set cover, and others. Linear programming model is an addition to this list. Large theoretical knowledge exists in this as the simplex algorithm and as interior point methods but the flow analysis requires another interpretation of optimal plans and plan transition with variate coefficients. An approximate model is devised which predicts the boundary stability point for the current optimal plan. This is valuable preparatory information of applications, moreover when a parallel computational facility is supposed.

Keywords: data flow algorithm, linear programming, approximation

ACM Classification Keywords: G.1.6 Numerical analysis: Optimization

1. Introduction

Data flow is a concept, traditionally appearing in the sensor based monitoring systems. Advanced global networks brought a number of novel applied disciplines intensively dealing with data flows. The network monitoring itself and optimal management of telecommunication systems, search engines with consequent data analysis, the network measuring instruments and network monitoring for security, etc. are the novel examples of data flow models. These deal with continuous data flows and unusual, non-finite and non-stored data set. In this case, the queries (the data analysis requests) are long-term and continuous processes in contrast to usual one-time queries. The traditional databases and data processing algorithms are poorly adjusted for the hard and continuous queries in data flows. This generates the necessity of new studies for serving continuous, multilayered, depending on time and subjected to indefinite behaviour of data flows [MM 2003]. Concerning the mentioned problem area, systems and algorithms are devised for different needs: real time systems, automation control systems, modelling processes, etc., but they are episodes in point of view of the formulated general problem. Traditional trade offs of such systems include one-pass and multi-pass algorithms, deterministic and randomized algorithms, and exact and approximate algorithms. Off-line algorithms solve a problem with full knowledge of the complete problem data. Online algorithms construct partial solutions with partial knowledge of the problem data, and update their solutions every time some new information is provided. In other words, they must handle a sequence of closely related and interleaved sub-problems, satisfying each sub-problem without knowledge of the future sub-problems. Standard examples of online problems include scheduling the motion of elevators, finding routes in networks and allocating cache memory. The usual way of measuring the quality of an online algorithm is to compare it to the optimal solution of the corresponding off-line problem where all information is available at the beginning. An online algorithm that always delivers results that are only a constant factor away from the corresponding optimum off-line solution, is called a competitive algorithm.

The “incremental update” algorithmic model of data analysis [AJ 2001] modifies the solution of a problem that has been changed, rather than re-solving the entire problem. For example, partial change of conditions of a time-table problem must be force to only partial reconstruction of the table. It is obvious that it is possible to construct a

[1] The research is supported partly by INTAS: 04-77-7173 project, http://www.intas.be

theoretical problem, where any particular change brings to the full reconstruction of the problem. It is also clear that there are numerous problems, which are not so critical to the local transformations. It is an option to try to solve the given data flow problem by the mentioned incremental algorithms, moreover, in the specific conditions it is the only possible way for solving the problem including the data flows analysis.

Measuring the “variability”, “sortedness” and similar properties of data streams could be useful in some applications; for example, in determining the choice of a compression or sort algorithm for the underlying data streams. [MM 2003] have studied the bit level changes in video sequences and [AJ 2001] - the problem of estimating the number of inversions (the key element of the Shell type sorting algorithms) in a permutation of length n to within a factor $1 \pm \varepsilon$, where the permutation is presented in a data stream model. [MM 2003] proves the decreasing bit level changes of image pixels in video sequences and in [AJ 2001] - an algorithm obtained requiring the space $O(\log n \log \log n)$.

Sketching tools are usual for many data oriented applications. These include approximations of statistical parameters, histograms, wavelets, and other similar general descriptors. The simplest calculations for data streams serve the base statistical means like the averages and variations [AC 2003]. Other data flow descriptors also appear in publications: frequency moments [AM 1996], histograms [GG 2002], etc.

The paper below discusses an important applied model for the flow environments. We consider the linear programming mathematical problem, parameters of which are formed by data flows. In a moment it is assumed that the optimal plan is found and the coordinates of the target function are variable by the flow. In this case, there is an emerging question: which mechanisms are well suited to follow the coefficients variations in creating the configuration of the next resulting optimal plan. It is clear that the small changes of coefficients lead to simple changes of the current optimal plan, probably not requiring the total analysis of the problem by the complete flow information.

2. Linear Programming in Data Flows

Let's consider the linear programming problem in its canonical form:

$$\min\ c'x$$
$$Ax = b, \quad x \geq 0,$$

where $c \in R^n,\ b \in R^m,\ A$ is an $m \times n$ full rank real matrix, and $m < n$. Without a loss of generality we may also suppose that $b_i \geq 0$, $i = \overline{1,m}$. Particular examples of linear programming problem are given through the definition of coefficient values: a_{ij}, c_j, b_i, for $i = \overline{1,m};\ j = \overline{1,n}$. Let's imagine a specific situation arising from application that the mentioned coefficients are changing in time. Such problems appear, for example, in data flow analysis.

Le us consider a data flow $B(t,n)$, which is finite but a very large-sized sequence of values $b_1, b_2, \cdots, b_n$, where $b_t, t = \overline{1,n}$ are certain structures. The data flows processing algorithms may use comparatively small storages than the input size. A limitation window is given in certain cases for processing separate data fragments. The time-dependent values of parameters, forming the applied problem model are formed as a result of algorithmic analysis. Unlike other natural and similar definitions, the variation of parameters is unpredictable here, as it has not probabilistic distribution and is not described by one or another property. Instead, it is considered that the variation takes place very slowly, because of the accumulation concept. In its turn, the applied problem demands to have ready answers to the certain questions in each time stamp.

There are two strategies: (1) solving a problem for each stage by all actual information which is practically impossible because of the large data sizes; and (2) structuring hereditary systems when the new data analysis is relatively easily integrated with the results of the previous data analysis.

We are going to consider the linear programming model in the mentioned conditions. The arbitrary variation of the coefficients is not allowed, instead, slow variations are considered so that the variation is fully monitored and it changes the solutions very slowly. Of course, it is possible to formalize this fully. At the same time, it is possible to consider partial behaviour of parameters variations, providing simple scenes of the algorithmic developments.

Let's suppose that the coefficients c_j of linear criterion function $Z = \sum_{j=1}^{n} c_j x_j$ of the linear programming problem are varying by flow $B(t,n)$. Assume that t_0 is the moment where the complete analysis exists, i.e. we know about the existence of optimization at that moment and the optimal vertex and plan, if the latter exists. This vertex obeys the property of stability of optimality for certain variations of coefficients c_j [GL 1980]. The stability area is described by a set of simple inequalities and it is clear that it is an issue to consider the border of this area. The theoretical analysis of optimality of vertex set elements of the area of base restrictions is well known as the simplex method [V 1998]. The simplex method looks for sequence chains of vertex transitions, which converge to an optimal plan. Complementary, in our case, we study all the possible ways of optimality transitions driven by the changes of coefficients c_j.

The novelty is that we device the concept of equivalency of vertices groups of the feasible polyhedron vertices set and prove that the transition from one optimal vertex to another takes place through these groups. So the continuous change of target function coefficients generates the continuous change of optimality vertices.

From practical point of view – a path prediction process is possible to apply to the situation with varying coefficients. Prediction of intersection of the trajectory extrapolation of coefficient changes to the boundary of stability area of the current optimal plan helps to determine the vertex equivalency cluster and so - the further possible transitions and by these – the most likely arriving optimums when coefficients keep the track of their current modifications.

Going through the transitions some vertices might neighbour with comparatively large equivalency groups of vertices and then the total number of those vertices can become large. Theoretically, in terms of flows, having the current optimization vertex, it is necessary to prepare neighbouring equivalent vertices by calculating them by the current and predicted coefficients c_j. The weakness of the direct application of the given approach is in drastic increase in the number of calculations for the vertex sets involving the predictions and equivalencies. The considered below natural approach gives primary significance to the vertices which correspond to the linear approximations of the given variations.

Let's denote the optimal vertex at the moment t_0 by $\tilde{x}^{t_0}$ and let $\tilde{c}^{t_0}$ is the corresponding vector of coefficients. Let's follow the transition of $\tilde{c}^{t_0}$ to the $\tilde{c}^{t}$. It is clear that this transition is more or less arbitrary and it is controlled by the flow. It is important if during the transition the vector $\tilde{c}^{t}$ of coefficients approaches to the boarder of stability of current optimal plan $\tilde{x}^{t_0}$, - or not. To see this we have to monitor the changes of $\tilde{c}^{t}$. Alternatively, it is possible to approximate the transition, activating the possible future “optimal plans”. For example, spline approximations or a more simple approximation by the main values and standard deviations might be applied. The most simple is the linear approximation model, which we consider below. As an extrapolation, it leads to the intersection with the stability boundary (shell) of the vertex $\tilde{x}^{t_0}$ at the most probability point. In case of sufficient computational resources, it is also possible to consider some neighbourhood of that point, and it is important that in contrast to the above mentioned theoretical model, this applied approach gives an opportunity to work with the limited quantity of the possible candidate vertices. Depending on the considered problem, an algorithmic system is able to choose a corresponding extrapolation scheme, which deals with different numbers of neighbouring vertices. The approximation of $\tilde{c}^{t}$ by the flow averages and dispersions requires their calculation, which is a simple flow problem (it is shown in [MM 2003]). Supposing that this question is clarified, let's consider the problem behaviour in the case of linear approximations.

3. Linear Approximation

In the case mentioned, variation of the optimization criteria function coefficients is supposed to be changed by an expression $c_j(\lambda) = c_j^{t_0} + \lambda(c_j^t - c_j^{t_0})$, where λ varies in certain limits. The interval $[0,1]$ for λ is internal and characterizes the variation from $\tilde{c}^{t_0}$ to $\tilde{c}^t$, and the values $\lambda > 1$ are extrapolating the further behaviour of coefficients in a linear model. Let's denote $c_j^{\Delta} = c_j^t - c_j^{t_0}$.

So we are given the linear function

(1) $$Z = \sum_{j=1}^{n} (c_j^{t_0} + \lambda c_j^{\Delta}) x_j$$

and the system of linear requirements, given by

(2) $$\sum_{i=1}^{n} a_{ij} x_j = b_i, \ i = 1,2,\cdots,m,$$
$$x_j \geq 0, \qquad j = 1,2,\cdots,n.$$

It is necessary to accompany the changes of λ, finding out in the interval $1 < \lambda$ the minimal value at which the change of the optimal plan takes place for the first time. Assume that the vector $\tilde{x}^t = (x_1^t, x_2^t, \cdots, x_n^t)$ which satisfies the system (2) introduces the corresponding new optimization basis.

According to the assumptions, we have optimal solution when $\lambda = 0$. Assume that the solution basis consists of the first m vectors of $\bar{a}_1, \cdots, \bar{a}_n$. In accord to the simplex algorithm and its optimization condition, all the "estimations" in this case must obey to the following condition: $z_j - c_j^{t_0} \leq 0, j = 1,2,\cdots,n$.

As $c_j(\lambda) = c_j^{t_0} + \lambda c_j^{\Delta}$ then that general optimization condition becomes:

$$z_j - c_j(\lambda) = (c_j^{t_0} + \lambda c_j^{\Delta}) x_j^0 - (c_j^{t_0} + \lambda c_j^{\Delta}) \leq 0, \ j = 1,2,\cdots,n.$$

Let's group the expression in the following way:

$$c_j^{t_0}(x_j^0 - 1) + \lambda c_j^{\Delta}(x_j^0 - 1) \leq 0, \ j = 1,2,\cdots,n,$$

and let introduce the notations: $\alpha_j = c_j^{t_0}(x_j^0 - 1)$ and $\beta_j = c_j^{\Delta}(x_j^0 - 1)$. The constants α_j and β_j are defined by the initial configuration: optimization criterion function coefficients and the corresponding solution basis, criterion current coefficients with the supposition that optimization did not change during that period.

For $\lambda = 0$ we have the optimization vertex $\tilde{x}^0$, and therefore, we get the following limitations: $\alpha_j = c_j^{t_0}(x_j^0 - 1) \leq 0$. The optimization vertex change does not take place when $0 \leq \lambda \leq 1$, so we get also: $\alpha_j + \lambda\beta_j = c_j^{t_0}(x_j^0 - 1) + \lambda c_j^{\Delta}(x_j^0 - 1) \leq 0$.

In particular, when $\lambda = 1$ we get $\alpha_j + \beta_j \leq 0$. The extreme condition will be written in the following general form: $\alpha_j + \lambda\beta_j \leq 0, j = 1,2,\cdots,n$. Let's find the minimal value of λ at which at least one of this inequalities violates for the first time.

Let's separate negative and positive cases of β_j. The restrictions on λ will accept the following forms:

$\lambda \geq -\alpha_j / \beta_j$ for all $\beta_j < 0$, and

$\lambda \leq -\alpha_j / \beta_j$ for all $\beta_j > 0$.

Let's introduce one more notation:

$$\bar{\lambda} = \{\min\left(-\alpha_j/\beta_j\right) \ if \ a\,\beta_j > 0, and +\infty, \ when\, all\, \beta_j \leq 0.$$

The optimal solution for $\lambda = 0$ coincides with optimal solutions for all λ which obeys the condition $0 \leq \lambda \leq \bar{\lambda}$. It is ensued that $\bar{\lambda}$ is the possible transition configuration. If $\bar{\lambda} = +\infty$ then there is no change of optimal plan. If $\bar{\lambda}$

is limited then it is necessary to consider two cases: the first one (a/) is the point $\bar{\lambda}$ with the possible equivalent optimal plans and possible continuations in this case, and the second one (b/): if there is a new optimal plan and if the problem has no solution at $\lambda > \bar{\lambda}$.

a/ Assume that $\bar{\lambda}$ is finite, i.e. $\bar{\lambda} = -\alpha_k / \beta_k$ for the corresponding value of parameter k. It means that $z_k - c_k(\lambda) = 0$ from which follows that the optimization plan is not single. Actually, let's insert the k-th vector into the basis and according to the simplex method let's exclude one of the vectors from the previous basis. We will get a new optimal plan the criterion value of which will stay unchanged. It follows even more – that, by all null estimations and by all basis modifications we can get many optimization equivalent vertexes and all elements of their linear closure also have the same discussing optimization value.

b/ In this case, we consider the values $\lambda > \bar{\lambda}$ and the $\bar{\lambda}$ is finite. If the coefficients of above mentioned k-th vector all not positive, i.e. $\tau_{ik} \leq 0$, by optimization basis, then according to the simplex method, the criterion function becomes unlimited. This takes place any time when according to the increasing character of the criterion function we get the vector which is going to be involved into the basis $z_k - c_k(\lambda) > 0$, but it becomes clear that the vector has no positive $\tau_{ik} > 0$ coordinate because of we could exclude it from the basis. In this case, it is impossible to choose such a coefficient $\theta > 0$ that any $x_i - \theta\tau_{ik} = 0$ when $i = \{1, \cdots, m\}$. Therefore, we get the optimization plan with $m+1$ positive components; the set of $\bar{a}_1, \bar{a}_2, \cdots, \bar{a}_m, \bar{a}_k$ vectors are linearly depending and this corresponds to the non-angle vertex. Therefore, linear criterion function could not get to its minimal value. This means that hyper-plane which is defined by linear function could not become supporting hyper-plane of permissible polyhedron at any shift in the direction of gradient.

If a $\tau_{ik} > 0$ then the vector $\bar{a}_k$ is included into the basis and another vector $\bar{a}_l$ is excluded from it. As the new basis is constructed by the simplex method then it corresponds to a new optimal solution, and at those inequalities

(3) $$\alpha'_j + \lambda\beta'_j \leq 0, \; j = 1, 2, \cdots, n$$

are compatible.

Let's show that any $\lambda < \bar{\lambda}$ does not satisfy the system (3) of inequalities. Really, for the vector $\bar{a}_l$, excluded from the basis we will get the following:

(4) $$\alpha'_l = -\alpha_k / \tau_{lk}; \qquad \beta'_l = -\beta_k / \tau_{lk},$$

where $\tau_{lk} > 0$. Suppose that (3) takes place for any $\lambda < \bar{\lambda}$ then $\alpha'_j + \lambda\beta'_j \leq 0$, or according to (4) $-\alpha_k - \lambda\beta_k \leq 0$. As $\beta_k > 0$ then, from the latter inequality follows that $\lambda \geq -\alpha_k / \beta_k = \bar{\lambda}$.

4. Conclusion

The paper is devoted to the discussion of applied algorithms for data flows. The linear programming problems and the simplex algorithm of their solution were considered. This research is not about the simplex algorithm developments, but is about the approaches processed in this sphere that also help when according to the problem assumption the coefficients of criterion function variate in the result of the data flows analysis. We got that it is possible to introduce and develop the concepts and tools related to the simplex algorithm by approaches, which solve flow linear optimization problems. The core result is the construction of the extrapolation mechanism that applies linear extrapolation by predicting the stationary data. The concept of equivalency of optimal vertices is introduced, which helps to accompany the variation process preparing the possible optimization vertexes in advance.

This is important from the viewpoint of linear programming systems and optimization in applied data flow systems.

Bibliography

[AJ 2001], M. Ajtai, T. Jayram, R. Kumar, and D. Sivakumar. Counting inversions in a data stream. manuscript, 2001.

[AC 2003], Aslanyan L., Castellanos J., Mingo F., Sahakyan H., Ryazanov V., Algorithms for Data Flows, International Journal "Information Theories and Applications", ISSN 1310-0513, 2003, Volume 10, Number 3, pp. 279-282.

[AM 1996], N. Alon, Y. Matias, and M. Szegedy. The space complexity of approximating the frequency moments. In Proc. of the 1996 Annual ACM Symp. on Theory of Computing, pages 20-29, 1996.

[GL 1980], E.N. Gordeev and V.K. Leontiev, Stability in problems of narrow places, JCM&MP, No. 4, Moscow, 1980.

[MM 2003], Manoukyan T and Mirzoyan V., Image Sequences and Bit-Plane Differences. "Computer Science & Information Technologies Conference", Yerevan, September 22-26, pp. 284-286 (2003).

[GG 2002], A. Gilbert, S. Guha, P. Indyk, Y. Kotidis, S. Muthukrishnan, and M. Strauss. Fast, small-space algorithms for approximate histogram maintenance. In Proc. of the 2002 Annual ACM Symp. on Theory of Computing, 2002.

[V 1998], R.J. Vanderbei, Linear Programming; Foundations and Extensions, Kluwer Academic Publishers, Boston/London/Dordrecht, 1998.

Author's Information

Levon Aslanyan – Institute for Informatics and Automation Problems, NAS Armenia, P.Sevak St. 1, Yerevan-14, Armenia; e-mail: lasl@sci.am

MATRICIAL MODEL FOR THE STUDY OF LOWER BOUNDS

Jose Joaquin Erviti, Adriana Toni

Abstract: Let V be an array. The range query problem concerns the design of data structures for implementing the following operations. The operation update(j,x) has the effect $v_j \leftarrow v_j + x$, *and the query operation retrieve(i,j) returns the partial sum* $v_i + \ldots + v_j$. *These tasks are to be performed on-line. We define an algebraic model – based on the use of matrices – for the study of the problem. In this paper we establish as well a lower bound for the sum of the average complexity of both kinds of operations, and demonstrate that this lower bound is near optimal – in terms of asymptotic complexity.*

Keywords: zero-one matrices, lower bounds, matrix equations

ACM Classification Keywords: F.2.1 Numerical Algorithms and Problems

1 Introduction

Let V=(v_1 .. v_n) be an array of length n storing values from an arbitrary commutative semigroup S. We define the operations:

- retrieve(j,k): returns v_j+..+v_k $\quad \forall 1 \le j \le k \le n$
- update(j,x) : v_j := vj +x $\quad \forall 1 \le j \le n, \quad x \in S$ (1)

We refer to n as the size of the range query problem. We see that the complexity of executing an update(j,x) operation is constant meanwhile the worst complexity of a retrieve(i,j) operation is linear on n.

If we define the operations in a different way:

- retrieve(j,k) : returns v_k-v_{j-1} $\forall 1 \le j \le k \le n$ (consider v_0=0)
- update(j,x) : [forall $k \ge j$ do v_k:=v_k+x] $\forall 1 \le j \le n$, $x \in S$ (2)

then the complexity of a retrieve(i,j) operation is constant meanwhile the worst case complexity of an update(j,x) operation is linear on n.

In both cases we are solving the same computational problem, that is to say: any sequence of update and retrieve operations over V produces the same results, no matter if the operations are implemented as in (1) or (2). With results we mean the outputs produced by the retrieve operations.

We can use different data structures involving a different number of variables storing values in the semigroup – others than the n variables organized as an array V - and provide the corresponding algorithms to implement the operations update and retrieve, and still be solving the same computational problem.

We work inside the semigroup model of computation: the array V can store values from an arbitrary commutative semigroup, and the implementations of the update and retrieve operations must perform correctly irrespective of the particular choice of the semigroup. In particular, the implementations are not permitted to utilize the substraction operation.

What kind of different data structures and programs are to be considered? We introduce right now the formal statement of the problem provided in [7].

We consider the class of data structures that involve program variables z_1, z_2 .. which store values in S. Stored in a variable z_i is the value $\sum_{j \in Y_i} v_j$ where Y_i is a specified subset of {1,2..n}. The query retrieve(j,k) is to be implemented with the program return $\sum_{i \in R_{jk}} z_i$ where R_{jk} is a specified set of integers. The update(j,x) operation is to be implemented with the program $z_l \leftarrow z_l$+x $\forall l \in U_j$, where $U_j = \{i \,/\, j \in Y_i\}$.

It is shown that a data structure in this class is correct if and only if

$$|U_l \cap R_{jk}| = \begin{cases} 1 & j \le l \le k \\ 0 & otherwise \end{cases}$$

In other words, we consider all the solutions consisting in a set of variables z_1, z_2 … which store values in S, and such that a retrieve operation consists in adding up a subset of these variables, and an update(j,x) operation consists in incrementing by x a subset of these variables.

The complexity of the operation is defined as : complexity of update(j,x)=$|U_j|$, and complexity of retrieve(j,k) =|Rjk|.

From now on we will denote:

$$p = \frac{\sum_{j=1}^{n} U_j}{n} \quad \text{and} \quad t = \frac{\sum_{1 \le j \le k \le n} |R_{jk}|}{\binom{n}{2}}$$

that is to say, p represents the average complexity of the update operations – n is the number of such operations – and t represents the average complexity of the retrieve operations – in this case the number of operations is $\binom{n}{2}$.

The following results are discussed in this paper:

- In section 2 we define an algebraic model – based in the use of matrices – for the study of the problem.
- In section 3 we obtain a lower bound for p+t. We use a previously known result about the optimal lower bound for the average complexity of operations update and retrieve concerning the partial sums problem, which is a particular case of the range query problem in which the first argument of operation is fixed and equal 1.
- In section 4 we provide the corresponding amount p+t for some previously defined data structures solving our problem. This had not been calculated before, and it serves to verify that our lower bound is near optimal.

2 Matricial Model of Computation

An algebraic model of computation has been defined by M.L.Fredman (see [3]) for the study of the partial sums problem – a particular case of the range query problem in which the first argument of any operation is fixed and equal 1. It consists in defining the operations update and retrieve through pairs of zero-one matrices R, U verifying that their product is matrix T, where T is defined as

$$T=(t_{ij})_{i,j=1..n} \qquad \text{with} \quad t_{ij}=\begin{cases}1 & i \geq j\\ 0 & i \prec j\end{cases}$$

We establish in this section an algebraic model for the study of the more general problem known as the range query problem.

Our model includes all programs verifying:

- A set of variables $Z=\{z_1,z_2\ldots z_m\}$ which stores values in S is maintained, organized as an array.
- The operation retrieve(i,j) is executed adding up a subset of these variables.
- The operation update(j,x) is executed incrementing a subset of these variables by amounts which depend linearly on x.

Thus the model consists on triples <Z,R,U> where Z is the array, $R=(r_{ij})$ is a zero-one matrix of dimension $\frac{n(n+1)}{2}\times m$, and $U=(u_{ij})$ is a zero-one matrix of dimension $m\times n$ (m, the number of required program variables, may change although it has to be greater or equal n).

Associated with a triple there are the following programs to implement the update and retrieve operations.

<u>Definition 1</u>

If <Z,R,U> is a triple for the range query problem of size n, then the operations update and retrieve must be implemented through the following programs:

- *update(j,x) :* *for l:=1 to m do* $[z_l \leftarrow z_l + u_{lj}x]$
- *retrieve(i,j) :* *output* $\sum_{l=1}^{m} r_{kl}z_l$, *where* $k=\sum_{s=0}^{i-2}(n-s)+(j-i+1)$

Lemma 2 below establishes a condition on R,U which implies the correction of the programs given in Definition 1.

<u>Lemma 2</u>

Let H be the $\frac{n(n+1)}{2}\times n$ *dimensional matrix defined by:*

$$H_{ij}=\begin{cases}1 & l \le j \le l+(i-w_{l-1}-1)\\ 0 & otherwise\end{cases}$$

where

$w_k = \sum_{l=0}^{k-1}(n-l)$, *k=0…n*

$i \in \{(w_{l-1}+1)\ldots w_l\}$, *l=1…n*

Then the programs in Definition 1 are correct if and only if RxU=H.

Proof

We find inspiration in the proof used in [3] (see Lemma 1 of [3]) to prove the correctness of the algorithms implementing the update and retrieve operations concerning the partial sums problem.

We have to consider the effect of the consecutive execution of operations [retrieve(i,j) ,update(r,x),retrieve(i,j)] for any i,j,r $\in \{1 \ldots n\}$, x$\in$S.

Let q_1,q_2 be the output produced as a result of executing the two operations retrieve(i,j), the first and the second respectively.

From the definition of the operations given at the beginning of the Introduction section (see (1) in that section), it is obvious that the programs defined in Definition 1 are correct if and only if

$$q_2 - q_1 = \begin{cases} x & i \le r \le j \\ 0 & otherwise \end{cases}$$

But $q_1 = \sum_{l=1}^{m} r_{kl} z_l$, $q_2 = \sum_{l=1}^{m} r_{kl}(z_l + u_{lr}.x)$ with $k = \sum_{s=0}^{i-2}(n-s) + (j-i+1) = w_{i-1} + (j-i+1),$

and so $q_2 - q_1 = \sum_{l=1}^{m}(r_{kl}u_{lr})x.$

By the definition of matrix H given in Lemma 2, and given that

$$\left[k = \sum_{s=0}^{i-2}(n-s) + (j-i+1)\right] \Rightarrow k \in \{(w_{i-1}+1)\ldots w_i\},$$

we have

$$H_{kr} = \begin{cases} 1 & i \le r \le i + (k - w_{i-1} - 1) \\ 0 & otherwise \end{cases}$$

But $i + (k - w_{i-1} - 1) = i + (j - i + 1) - 1 = j$, and the result may be deduced immediately.

Remark 3

Let us observe that if T^n *is the nxn dimensional matrix*

$T=(t_{ij})_{i,j=1..n}$ *with* $t_{ij}=\begin{cases} 1 & i \ge j \\ 0 & i \prec j \end{cases}$

then we have that for all $k = 0 \ldots (n-1)$

$$H_{[w_k+1..w_{k+1}],[k+1..n]} = T^{n-k}$$
$$H_{[w_k+1..w_{k+1}],[1..k]} = 0$$

where $w_i = \sum_{s=0}^{i-1}(n-s)$ *for i=0..n,* $H_{[i..j],[r..s]} \equiv$ *the block of matrix H integrated by the rows belonging to the interval [i..j] and the columns belonging to the interval [r..s].*

So, for any size n of the range query problem, we may describe the corresponding matrix H^n through matrix T as

$$H = \begin{pmatrix} T^n \\ \overrightarrow{T^{n-1}} \\ \vdots \\ \overrightarrow{T^{n-(n-1)}} \end{pmatrix}$$

where

$$\overrightarrow{T^{n-i}} = \left(\begin{matrix} 0 & \dots & \dots & 0 \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ 0 & \dots & \dots & 0 \end{matrix} \middle| T^{n-i} \right) \qquad \text{i=1..(n-1)}$$

Example 4

Let us see the matrix H corresponding to the range query problems of size n=2 and n=4, which are named as H^2 and H^4 respectively. We have set the natural division in blocks off.

$$H^2 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ \hline 0 & 1 \end{pmatrix}, \qquad H^4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ \hline 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ \hline 0 & 0 & 0 & 1 \end{pmatrix}$$

Lemma 5 below establishes a result concerning the number of program variables $z_1..z_m$ required to implement any solution for the range query problem.

Lemma 5

Let R, U be zero-one matrices of dimensions nxm and mxn respectively such that RxU=H^n .

Then we have that $n \le m$.

Proof

First of all we prove that for any A, B zero-one matrices of dimensions nxm and mxn respectively such that AxB=T^n, we have that $n \le m$.

Let us proceed with this proof. We denote r, u, t the linear mappings associated with matrices A, B, T^n respectively. Then

$$R^n \xrightarrow{u} R^m \xrightarrow{r} R^n$$

From $T = A \circ B$, it follows $n = \dim(\mathrm{Im}(t)) = \dim(r(u(R^n))) \le \dim(\mathrm{Im}(r)) \le m$. So, necessarily $n \le m$.

Now let $R \times U = H^n$. Given that the product of the first n rows of R by matrix U is matrix T^n, we conclude trivially our result.

The following result connects the solutions for the range query problem inside our matricial model with the ones within the setting introduced in [7] that may be found also at the beginning of the Introduction section in this paper.

Remark 6

Given a triple <Z,R,U> which represents a solution for the range query problem of size n within our matricial model, where Z=($z_1..z_m$), R=(r_{ij})$_{(n(n+1)/2)xm}$, U=(u_{ij})$_{mxn}$, and given $\{Y_1 \ldots Y_m\}$, $\{R_{ij} \;/\; i, j = 1 \ldots n\}$ defined as in Section 1, we have that

$$r_{ij} = \begin{cases} 1 & j \in R_{kl} \quad i = (l-k+1) + \sum_{s=0}^{k-2}(n-s) \\ 0 & otherwise \end{cases}$$

$$u_{ij} = \begin{cases} 1 & j \in Y_i \\ 0 & otherwise \end{cases}$$

Definition 7 below establishes the complexity of the operations within our model.

Defintion 7

Given the triple <Z,R,U> which represents a solution for the range query problem of size n within our matricial model, where Z=($z_1..z_m$), R=(r_{ij})$_{(n(n+1)/2)xm}$ and U=(u_{ij})$_{mxn,}$, we define

- *Complexity of the operation retrieve(i,j):*

$$|\{r_{kl} \;/\; r_{kl} \neq 0 \wedge (1 \leq l \leq m)\}| \quad \text{where} \quad k = (j-i+1) + \sum_{s=0}^{i-2}(n-s)$$

- *Complexity of the operation update(j,x):*

$$|\{u_{lj} \;/\; u_{lj} \neq 0 \wedge (1 \leq l \leq m)\}|$$

3 Lower Bound for the Sum of the Average Complexities

In this section we obtain a lower bound for the sum of the average complexity of the operations update and the average complexity of the operations retrieve which improves previously obtained lower bounds relating to this sum. We work inside our matricial model.

Let R, U be zero-one matrices representing a solution for the range query problem of size n, that is to say, such that RxU=H^n.

Assuming that m is the number of columns of matrix R and thus the number of rows of U, we have that the average complexity of the update operations is

$$p = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} u_{ij}}{n}$$

and the average complexity of the retrieve operation is

$$t = \frac{\sum_{i=1}^{n(n+1)/2} \sum_{j=1}^{m} r_{ij}}{n(n+1)/2}$$

In [7] has been established that for any data structure we have that $p + t = \Omega(\log n)$ (see Theorem 1 of [7]), and more exactly, the result says that

$$(2n^2/9)\log_e n + O(n^2) \le 2n^2 p + 2\frac{n(n+1)}{2}t$$

(from this inequality a lower bound for p+t may be obtained with little calculation).

We provide a different method which let us obtain easily a lower bound for $2n^2 p + n(n+1)t$ by reusing the knowledge of an optimal lower bound for the average complexity of the operations update and retrieve (all of them being considered together in this case) with regard to the partial sums problem of size n.

It has been proved (see [6]) that for $2^k \le n < 2^{k+1}$, we have that if AxB=T^n, being A and B nxm and mxn dimensional matrices respectively, then

$$\sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij} + \sum_{i=1}^{m}\sum_{j=1}^{n} b_{ij} \ge n[\log_2 n] + [\log_2 n] + (m - 2^k) + 2$$

We proceed to obtain a lower bound for $2n^2 p + n(n+1)t$. It must be observed that we will reach a lower bound I which the constant factor affecting the term which leads the expression is greater that the one in the lower bound obtained in [7]: ours is $n^2 \log_2 n$, meanwhile the one in [7] is $n^2\left(\frac{2}{9}\log_e n\right)$. In particular, $\log_2 n > 5\left(\frac{2}{9}\log_e n\right)$.

<u>Theorem 8</u>

Let R, U be zero-one matrices such that RxU=H^n.

Then we have that

$$2n^2 p + n(n+1)t \ge n^2(\log_2 n - \frac{1}{2}\log_2 e) + 2n\log_2 n + 2n + \frac{10}{4}\log_2 e$$

where

$$p = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} u_{ij}}{n} \quad , \quad t = \frac{\sum_{i=1}^{n(n+1)/2}\sum_{j=1}^{m} r_{ij}}{n(n+1)/2}$$

and m represents the number of columns in matriz R and tus the number of rows in U.

<u>Proof</u>

We have that

$$R \times U = H^n = \begin{pmatrix} \overrightarrow{T^n} \\ \overrightarrow{T^{n-1}} \\ \vdots \\ T^{n-(n-1)} \end{pmatrix}$$

where

$$\overrightarrow{T^{n-i}} = \begin{pmatrix} 0 & \dots & \dots & 0 & \\ \vdots & & & \vdots & \\ \vdots & & & \vdots & \Big| T^{n-i} \\ 0 & \dots & \dots & 0 & \end{pmatrix} \qquad \text{i=1..(n-1)}$$

Let $R_{[i\dots j],[1\dots m]}$ be the block consisting of the rows of R belonging the interval [i..j], for $1 \le i \le j \le n(n+1)/2.$

We consider matrix R as divided in n blocks, $R_1..R_n$, where R_1 consists of the first n rows of R, R_2 of the following (n-1) rows…and being the last block R_n formed exclusively by the last row of R. That is to say, if we call $z_i = \sum_{j=0}^{i-1}(n-j), \quad i = 0\dots n,$ then $\quad R_i = R_{z_{i-1}+1\dots z_i, 1\dots m}.$

Using this notation we have that $R_i \times U = T^{n-i+1}$.

We apply the lower bound that we already know for the partial sums problem to every pair R_i , U with $i = 1\dots n$ (see [6]). Let us observe that $|R| + n|U| = \binom{n}{2}t + n^2 p$ where |R|, |U| denotes the number of non-zero entries in matrix R and U respectively .

We have

$$|R_i| + |U| = \sum_{l=z_{i-1}+1}^{z_i}\sum_{s=1}^{m} r_{ls} + \sum_{l=1}^{m}\sum_{s=1}^{n} u_{ls} \ge (n-i+1)log_2(n-i+1) + log_2(n-i+1) + (m-2^k) + 2$$

So, and without taking account of the factor $(m-2^k)$,

$$\begin{aligned} |R| + n|U| &= \sum_{i=1}^{n(n+1)/2}\sum_{j=1}^{m} r_{ij} + n\sum_{i=1}^{m}\sum_{j=1}^{n} u_{ij} \ge \\ &\ge [(n+1)\log_2 n + 2] + [n\log_2(n-1) + 2] + \dots + [(2+1)\log_2 2 + 2] + 2 = \\ &= [(n+1)\log_2 n + n\log_2(n-1) + \dots + (2+1)\log_2 2] + 2n \end{aligned}$$

Let us call $\quad \delta(n) = (n+1)\log_2 n + \dots + (2+1)\log_2 2 \quad$ and then we have

$$|R| + n|U| \ge \delta(n) + 2n$$

We observe that

$$\sum_{i=1}^{n(n+1)/2}\sum_{j=1}^{m} r_{ij} + n\sum_{i=1}^{m}\sum_{j=1}^{n} u_{ij} = \frac{n(n+1)}{2}t + n^2 p,$$

and we are trying to establish a lower bound for $2n^2 p + n(n+1)t.$

Let $f(x) = (x+1)\log_2 x.$ Then

$$\int_1^n f(x)dx \le \delta(n) \le \int_2^{n+1} f(x)dx$$

Considering $\log_2 x = \log_2 e \log_e x,$ we integrate the expression

$$\log_2 e \int (x+1)\log_e x dx$$

so if we call

$$u = \log_e x \Rightarrow du = dx / x$$
$$dv = (x+1)dx \Rightarrow v = (x+1)^2 / 2$$

then

$$\int (x+1)\log_e x dx = \frac{(x+1)^2}{2}\log_e x - \frac{1}{4}x^2 - x - \frac{1}{2}\log_e x$$

(the constant factor $\log_2 e$ will be taken into account later)

So

$$\int_1^n f(x)dx = \frac{(n+1)^2}{2}\log_e n - \frac{1}{4}n^2 + \frac{5}{4} - n - \frac{1}{2}\log_e n$$

Then

$$n^2 p + \frac{n(n+1)}{2}t \geq \delta(n) + 2n \geq$$
$$\geq 2n + \log_2 e\left[\frac{(n+1)^2}{2}\log_e n - \frac{1}{4}n^2 + \frac{5}{4} - n - \frac{1}{2}\log_e n\right] \geq$$
$$\geq \frac{n^2}{2}\log_2 n + n\log_2 n + n - \frac{1}{4}n^2\log_2 e + \frac{5}{4}\log_2 e$$

Multiplying by 2 both sides of the inequality we obtain the result stated in this theorem, that is

$$2n^2 p + n(n+1)t \geq n^2(\log_2 n - \frac{1}{2}\log_2 e) + 2n\log_2 n + 2n + \frac{10}{4}\log_2 e\,.$$

Remark 9

Let us observe that from this result a lower bound of $\Omega(\log n)$ *for* $p+t$ *can be deduced: dividing both sides of the final inequality by* $2n^2$ *we obtain* $p + \left(\frac{1}{2} + \frac{1}{2n}\right)t \geq \log_4 n + (\log_2 n)/n + 1/n - (\log_2 e)/4 + (10\log_2 e)/8n^2$, *and so, given that* $\frac{1}{2} + \frac{1}{2n} \leq 1 \quad \forall n \geq 1,$ *we conclude that* $p + t \in \Omega(\log n.)$

4 Complexity of Certain Data Structures

In this section we establish the value of $p+t$ for some previously defined data structures solving our problem – more precisely , we provide the amount $\binom{n}{2}t + n^2 p$. In particular, trees have been defined in [7] (see Theorem 2 in that paper) verifying that given a fixed integer k, the worst case complexity of an update operation is k , and the worst case complexity of a retrieve is $O(n^{\frac{1}{k-1}})$. The definition of the trees satisfying these conditions is straightforward, based on the use of an $n^{1/(k-1)}$-ary tree of height k with n leaves. An update(j,x) operation has to be executed incrementing by x the value stored in the j-th leaf from left to right and in all the nodes belonging to the path from this leaf to the root, and a retrieve(j,x) operation has to be executed adding up the values in the minimum set of nodes verifying that the union of its successors includes exactly the leaves from i to j, and the intersection of its successors is disjoint pairwise - the idea is that each internal node stores the sum of the values in its sons.

Let $T_{r^{h-1}}$ denote the r-ary tree of deep h with $n=r^{h-1}$ leaves. We claim

Theorem 10

Given $T_{r^{h-1}}$ *, we have that*

$$\binom{n}{2}t+n^2p=(r-1)(h-1)\left(n+\frac{n^2}{2}\right)-\frac{n(n-1)(r+1)}{3}+1+n^2h$$

Proof

The proof involves heavy calculation and will not be included in this paper.

If we compare the result stated in Theorem 10 with our lower bound for $\binom{n}{2}t+n^2p$ we obtain the following corollary.

Corollary 11

Given $T_{r^{h-1}}$ *, in order to minimize the amount* $\binom{n}{2}t+n^2p$ *we must choose* $r=2$*,* $h=\log_2 n+1$*, being* $n=r^{h-1}$*.*

In this case we have that $\binom{n}{2}t+n^2p = 3n^2\log_4 n-\Theta(n^2)$ *. In Theorem 8 we proved that* $\binom{n}{2}t+n^2p\geq n^2\log_4 n-\Theta(n^2)$*, so this lower bound is near optimal. Let us observe that this means that the* $\Omega(\log n)$ *lower bound for* $p+t$ *is optimal in terms of asymptotic complexity.*

5 Bibliography

[1] W.A. Burkhard, M.L. Fredman, D.J.Kleitman, *Inherent complexity trade-offs for range query problems*, Theoretical Computer science, North Holland Publishing Company 16, (1981) pp.279--290.

[2] M.L. Fredman, *Lower Bounds on the Complexity of some Optimal Data Structures*, Siam J. Comput., Vol.10, No.1 (1981) pp.1--10.

[3] M.L. Fredman, *The Complexity of Maintaining an Array and Computing its Partial Sums*, J.ACM, Vol.29, No.1 (1982) pp.250--260.

[4] M.L. Fredman, H.Hampapuram, *Optimal Bi-Weighted Binary Trees and the Complexity of Maintaining Partial Sums*, Siam J. Comput., Vol.28, No.1 (1998) pp.1--9.

[5] M.L. Fredman, M.E. Saks, *The Cell Probe Complexity of Dynamic Data Structures*, Proceedings of 21 st ACM Symposium on Theory of Computing, (1989) pp.345--354.

[6] A. Toni, *Lower Bounds on Zero-one Matrices*, Linear Algebra and its Applications, 376 (2004) 275--282.

[7] D.J. Volper, M.L. Fredman, *Query Time Versus Redundancy Trade-offs for Range Queries*, Journal of Computer and System Sciences 23, (1981) pp.355--365.

[8] A.C. Yao, *On the Complexity of Maintaining Partial Sums*, Siam J. Comput., Vol.14, No.2 (1985).

Authors' Information

Jose Joaquin Erviti – jerviti@fi.upm.es

Adriana Toni – atoni@fi.upm.es

Facultad de Informática, Universidad Politécnica de Madrid, Spain

HI!MVC: HIERARCHICAL MVC FRAMEWORK FOR WEB APPLICATIONS

Micael Gallego-Carrillo, Ivan Garcia-Alcaide, Soto Montalvo-Herranz

Abstract: This paper presents Hi!MVC, a framework for developing high interactive web applications with a MVC Architecture. Nowadays, to manage, extend and correct web applications can be difficult due to the navigational paradigm they are based on. Hi!MVC framework helps to make these tasks easier.

This framework allows building a web based interface, generating each page from the objects that represent its state. Every class to be showed in the interface is associated with two entities: its html representation (view) and its interactions in the view manager (controller). The whole html page is generated by composition of views according to the composition relationship of objects. Interactions between user and application are managed by the controller associated to the view which shows interaction elements (links or forms). Hi!MVC allows building web interface in a hierarchical and distributed way.

There are other frameworks and APIs offering MVC architectures to web applications, but we think that they are not applying exactly the same concepts. While they keep on basing their architectures on the navigational paradigm we are offering a new point of view based on an innovator hierarchical model.

First, we present the main ideas of our proposal. Next, we expose how to implement it using different Java technologies. Finally, we make a first approach to our hierarchical MVC model. We also compare shortly our proposal with the previously cited technologies.

Keywords: Web Applications Engineering, Model, View, Controller, MVC, framework, J2EE.

ACM Classification Keywords: H.3.5 Online Information Services: Web-based services; H.5.3 Group and Organization Interfaces: Web-based interaction

Introduction

When the web was created, its main goal was to provide to the scientific community the chance of sharing information easily. At the beginning, files supporting this information were completely static and it could only be changed modifying the content of the files manually. Nowadays, we find a great evolution at this respect and we can see really complex services offered by web sites. This way, we are able of managing complete applications based on web. This fact is more than interesting for giving the possibility of using an application almost everywhere through a network without installing complicated clients but a web browser.

This is the main reason for enterprises and other organizations to base their development interfaces in web, but this may not be so easy. Patterns and many other design tools look forward to find a way to make software products extensible, manageable, reusable and easy to support. MVC architecture appeared to reach this point for window based interfaces. Now, it seems to be apprehensible to think if it would be possible to take MVC advantages to web based interfaces.

In this paper, we get deeper into a previously proposal [Gallego-Carrillo, 2005] that presented the basic guidelines of Hi!MVC framework. First, we describe the main concepts and organization of our proposed framework. At this point, it can be found that this description would be applicable to any other development tool than Java. Next, we describe in detail the framework implementation. Following, we focus on the hierarchical application of MVC. Finally, related works and conclusions are exposed.

Description of the Hi!MVC architecture

Here, it is presented the base and main concepts of the framework. It shows how to develop web interfaces for applications using this MVC architecture [Krasner, 1998]. This architecture reduces the coupling between classes and increases their cohesion. This fact gets the code to be more independent so it can be easily reused to save time and effort in further developments. MVC has been widely implemented in graphics user interfaces to separate the entity responsible of showing information (view), the one responsible of storing it (model) and the one responsible of receiving user events (controller). The standard UI technology used in java, Swing [Swing, 2005], also uses MVC.

As expected, the design we are writing about has three clearly separated parts, each one of them assume a different responsibility corresponding to MVC model. They are the *model*, the *view* and the *controller*. As follows, we describe them and its *event system*.

Model

This component is the responsible for data maintenance. It has to keep it available for the application in a consistent and safe environment, not allowing any external intervention to affect in any negative aspect. Every single code that implements the model has to be reusable by any other kind of interface without modifying it.

To implement this, we need a set of classes. In order to take advantage of many other technologies, these classes have to preserve JavaBeans specifications [JavaBeans, 2005].

View

View is the responsible for the application interface. In this case, it is the HTML and JavaScript code to send it to the client (or any other format like WML). It has to show the information managed by the application, so it can interact with the user. From here, it has to be given the chance of sending information to the application through events. View has to provide mechanisms like links and buttons to carry out the appropriate requests.

The way to implement this part is through JSP technology [JSP, 2005]. Every JavaBean of the model may have at least one JSP file to be presented. From these files it is used different technologies to access the data to be managed like expression language and tag libraries [Taglibs, 2005].

Controller

It is the responsible of managing the interactions with the user. Whenever the application receives an interaction or data produced by the user, the controller has to decide what to do next. This is the part that manages the global flow control of the application.

From a web browser, there is only one possibility of sending information to the server. This is done sending a request to a web resource. When requesting, the user can attach more information (for example from a form) than the reference to the new resource. There should be one controller by each JSP file. The controller will be a Java common class with methods that receive data about what happened, interpret it and execute its code consequently to continue with the application.

The event system

In order to follow the MVC architecture, the user interacts with the application raising events. An event is the state change notification of view or model. We are only considering in this paper, those events produced by the view. But the problem is that web pages have not the possibility of raising events directly. What the server is able to receive is just the requests from the client, so events are based on them.

When the user clicks a link, it is produced a request to the server. We call to the reception of this request, *real event*, because it is produced at the same moment that the user clicks.

On the other hand, when the user fills the set of components contained on a form, their information is sent all together into the same request when submit button is pressed. The set of changes in the data of the components from what it was shown is what we call *deferred events pack*. We say they are deferred because they are not sent at the same time they are produced. So, it is important to remark that the order or the time in which they are produced can not be considered.

Implementing MVC with J2EE

In this point it is explained how to implement our MVC model and what technologies we are using for it. In Figure 1 we show a scheme of this and in Figure 2 a web application request sequence diagram.

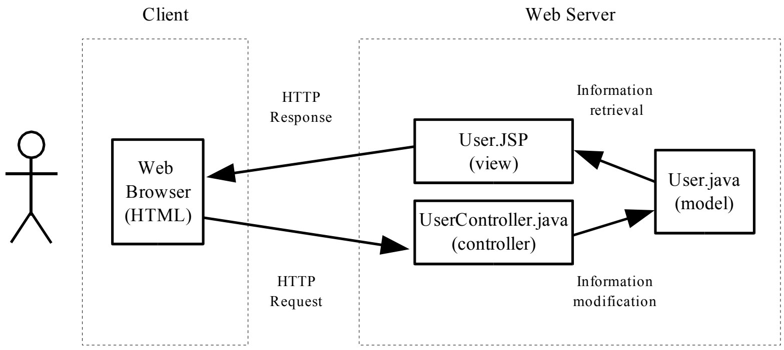


Figure 1. Scheme of design. It shows the collaboration between different layers

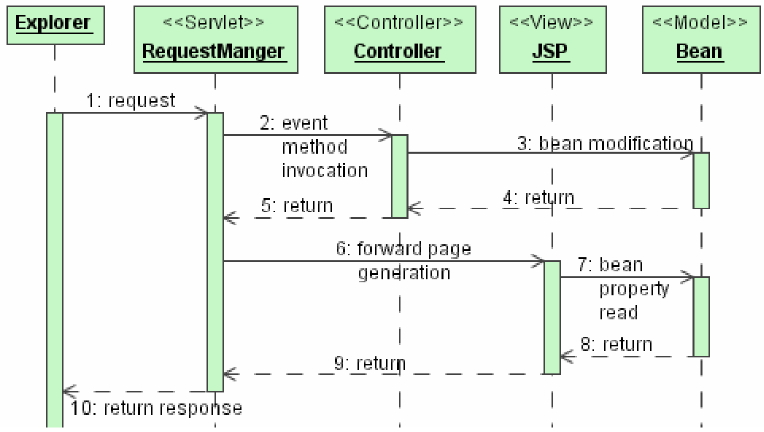


Figure 2. Web application request sequence diagram

Implementing the Model

In the model, we need java classes that support the data to be managed by the application. Fulfilling JavaBeans specification will later help us to be able of using technologies that have JavaBeans as requirement. Information supported by these classes is contained in their attributes, and the type of them can be another class of the same model giving rise a composition hierarchy or an association relation. Hence, basically these classes will have just some attributes and methods to access them. In this component may appear classes related to persistent storage technologies, like databases. In Figure 3 we show a sample JavaBean to implement a bills.

Bill.java

```
public class Bill {
	private String customerName;
	private String desc;
	private int price;

	public String getCustomerName() { return this.customerName; }
	public String getDescription() { return this.desc;  }
	public int getPrice() { return this.price; }
```

```
        public void setCustomerName(String cn) { this.customerName = cn; }
        public void setDescription(String desc) { this.desc = desc; }
        public void price(int prince) { this.price = price; }
}
```

Figure 3. A sample JavaBean to implement bills

Implementing the View

View implementation is mainly based on JSP files that access to model to show it. As classes in the model are JavaBeans, it will be more comfortable to apply Expression Language and User Defined Tag Libraries. One JSP file is always associated to a class in the model and its responsibility is to present the whole or a part of the information containing on it. One class of the model may be associated to one or more JSP files. It will be convenient to associate more than one JSP file, if it is needed, to represent the information with different formats or parts of the associated class.

Commonly, web applications are developed from a *navigational point of view*, namely, there are an independent pages set associated by links included on them. User interacts with the application navigating through these links. In our proposal we consider the web application as the representation in HTML, WML, etc., of an object. So, interactions with the application are made by changing its state and representing it again. We call this *object representation point of view*. This way there is not a home page to navigate to the rest of application. Page generated by JSP has to give the appropriate links and forms to the client so the object state can be changed. Instead of letting developer to write them directly (what would induce to keep navigational point of view) links and forms are generated dynamically and automatically by User Defined Tags. It is needed to pay special attention to the code generated for this, because it will be the responsible of notifying what events are produced and what controller has to listen to them. Further details about this code will be explained later in this point.

In Figure 4 we show two samples of JSPs corresponding to Bill class shows in figure 3. Bill.jsp is used to display bill's information in HTML. Bill.new.jsp is used to modify the state of a bill.

Bill.jsp

```
<%@ taglib uri="http://java.sun.com/jsp/jstl/core" prefix="c" %>
<%@ taglib uri="http://www.lpsi.eui.upm.es/es/upm/lpsi/himvc" prefix="himvc" %>
<%@ page contentType="text/html;charset=ISO-8859-1" language="java" %>
Bill for <strong>${this.customerName}</strong> <br />
Description: ${this.description} <br />
Total price: ${this.price} <br />
```

Bill.Modify.jsp

```
<%@ taglib uri="http://java.sun.com/jsp/jstl/core" prefix="c" %>
<%@ taglib uri="http://www.lpsi.eui.upm.es/es/upm/lpsi/himvc" prefix="himvc" %>
<%@ page contentType="text/html;charset=ISO-8859-1" language="java" %>
<himvc:form id="modify">
   Customer name:
      <himvc:text textId="customerName" text="${this.customerName}" />
   Description:
      <himvc:text textId="description" text="${this.description}" />
   Price:
      <himvc:text textId="price" text="${this.price}" />
</himvc:form>
```

Figure 4. Bill and Bill.Modify views implemented with JSPs corresponding to Bill class

Implementing the Controller

There will be one Java common class by each JSP file that represent a class in the model so, the controller will be a set of classes that receives information from the requests and modify the state of the model according to it. Because it is not necessary to keep the state of these classes, all their methods will be static. Every controller class has one method associated to each link or form that its JSP file is able to generate in order to attend it. Every method will receive one parameter with the object of the model that the JSP used to represent it. If there is additional information on the request like components of a form or parameters of a link, this information will be received by the method. In Figure 5 we can see the BillModifyController class belongs to Bill.Modify view. There is not a controller for Bill.jsp because it haven't any link or form.

BillModifyController.java

```
public class BillModifyController {
   public static void modifyCustomerName(Bill bill, String text){
      bill.setCustomerName(text);
   }
   public static void modifyDescription(Bill bill, String text){
      bill.setDescription(text);
   }
   public static void modifyPrice(Bill bill, String text){
      bill.setPrice(Integer.parseInt(text));
   }
}
```

Figure 5. Bill.Modify controller

It is needed a Servlet that manages requests and determine the correct controller to take control. This Servlet is called *RequestManager* and it is included in the framework. It expects to be able of determine the controller through the URL of the request previously generated by the User Defined Tag. On the other hand, it also expects to receive the form components data to forward them to the corresponding controller. This way the controller knows what events had been produced. So, it can change the model according to this.

Hierarchical MVC architecture

To build the model with object oriented programming, it is often used composition hierarchies between classes. To develop the user interface, it would be very useful to use the same composition hierarchy. To get it, each class in the model has one or more associated JSP files in the view that will take the responsibility of presenting it. If the class to be presented by its associated JPS file has attributes of other classes then, the JSP file may delegate the representation of them on its associated JSP files regarding the composition hierarchy. This is showed in Figure 6. We provide User Defined Tags to make easier the delegation in page generation.

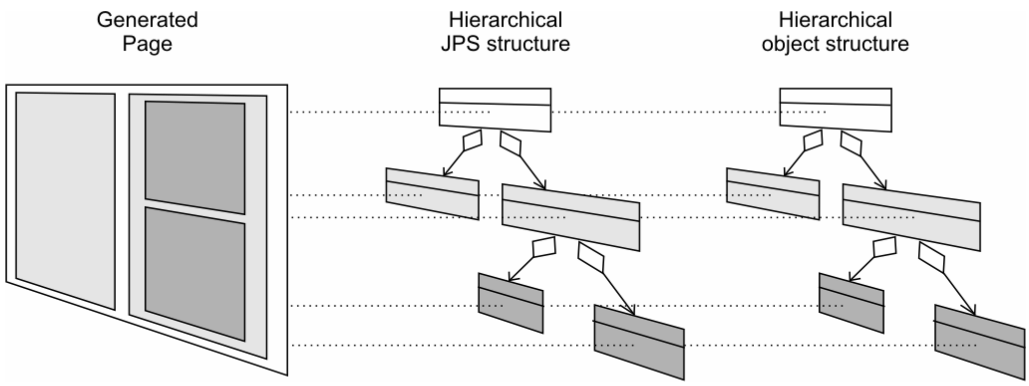


Figure 6. Hierarchical MVC architecture

To show a composition relation between classes in Figure 7 we can see BillsManager model class. This class hold a list of bills and allows to manage the held bills. Figure 8 shows a view for BillsManager. Finally, in Figure 9 we can see the BillsManager controller.

BillsManager.java

```
public class BillsManager {
   List<Bill> bills = new ArrayList<Bill>();
   int changingBill = -1;
   public List<Bill> getBills(){ return bills; }
   public int getChangingBill(){ return changingBill; }
   public void setChangingBill(int cb){ this.changingBill = cb; }
}
```

Figure 7. BillsManager model class

BillsManager.jsp

```
<%@ taglib uri="http://java.sun.com/jsp/jstl/core" prefix="c" %>
<%@ taglib uri="http://www.lpsi.eui.upm.es/es/upm/lpsi/himvc" prefix="himvc" %>

<%@ page contentType="text/html;charset=ISO-8859-1" language="java" %>

<himvc:form id="manager">
  <table>
    <c:forEach var="i" begin="0" end="${this.bills.size()}">
      <td><tr>
        <c:choose>
          <c:when test="${this.changingBill == i}">
              <himvc:include object="${this.bills.get(i)}" type="Modify">
            <br/><himvc:submit id="submit" text="Aceptar"/>
          </c:when>
          <c:otherwise>
              <himvc:include object="${this.bills.get(i)}">
            <br/><himvc:link id="modify" text="Modify" params="${i}"/>
          </c:otherwise>
        </c:choose>
      </td></tr>
    </c:forEach>
  </table>
</himvc:form>
```

Figure 8. BillsManager view

BillsManagerController.java

```
public static class BillsManagerController {
   public static void managerSubmit(BillsManager bm){
      bm.setChangingBill(-1);
   }
   public static void managerModify(BillsManager bm, int index){
      bm.setChangingBill(index);
   }
}
```

Figure 9. BillsManager controller

Related Works

Nowadays, there are several technologies that apply MVC architecture for developing web applications. The most representative in J2EE are Struts [Struts, 2005] and Java Server Faces [JSF, 2005]. Both are very similar in the way they apply MVC architecture.

JSF applies it at page level, so there is one JSP file per page. The model is represented by one object per page as well. The controller is implemented by the own technology and it is configured through defined tags in the JSP file. The controller updates JavaBeans properties with the values found in the interface components generated by the JSP file.

Struts also uses JavaBeans in the model and it implements the view with JSP files that generate the output from those JavaBeans. The controller manages the requests received by the web application. It associates the actions to logic names used to build the links and forms in the JSP file. Both technologies apply navigational paradigm and manage links through configuration files.

Main difference between these technologies and our proposal is that they have the navigation and page concepts as their base. In our approach, the page concept is not applied; instead of, we build a view from an object in the web application. The user does not interact with the application navigating through links but interacting with objects which representation is refreshed every time it changes. Due to this approach, there is not one JSP file per page but when an object needs to be presented then, it will use its own JSP file. If the state of the object is based in other objects, it can delegate the presentation to them including other JSP files.

Conclusion

In high interactive web applications it is not possible to apply in a natural way the navigational paradigm because the user needs to see the state of the application on every moment. The architecture proposed here presents a set of classes that shows their content and modifications at a given time, so not navigational point of view is used in this approach. This fact helps a lot to increase the manageability of the project at developing time. On the other hand, this approach takes good care of encapsulation and reusability benefits offered by classes saving time in the develop process.

In the future, incorporating this architecture into an Integrated Development Environment will enable faster developing giving to the programmer the possibility of generating big parts of his code automatically.

Moreover, building the user interface hierarchically, so it can be used the same delegation as in object oriented technology, makes that repercussions in changing code of the application are very limited.

Next to this work we will continue developing tools, code generators and libraries while going deeper into these concepts to form a complete framework that will help in development of web applications with a high interaction.

Acknowledgement

We would like to thank the following people for their help and their support: Fernando Arroyo Montoro, José Ernesto Jiménez Merino and Carmen Luengo Velasco.

Bibliography

[Gallego-Carrillo, 2005] Gallego-Carrillo M., García-Alcaide,I., Montalvo-Herranz, S. Applying Hierarchical MVC Architecture to High Interactive Web Applications. In: Proceedings of the Third International Conference I.TECH, 2005, pp. 110-115.

[J2EE, 2005] Java 2 Enterprise Edition. http://java.sun.com/j2ee/

[Krasner, 1998] Krasner, G. and Pope, S., 1988. A Description of the Model-View-Controller User Interface Paradigm in the Smalltalk-80 system. Journal of Object Oriented Programming, Vol. 1, No. 3, pp 26-49.

[Swing, 2005] Swing. http://java.sun.com/products/jfc/

[JavaBeans, 2005] JavaBeans. http://java.sun.com/products/JavaBeans/

[JSP, 2005] JavaServer Pages. http://java.sun.com/j2ee/JSP/

[Taglibs, 2005] Tag libraries. http://java.sun.com/products/JSP/taglibraries/

[Struts, 2005] Struts. http://jakarta.apache.org/struts/

[JSF, 2005] JavaServer Faces. http://java.sun.com/j2ee/javaserverfaces/

Authors' Information

Micael Gallego-Carrillo – ESCET, Universidad Rey Juan Carlos, C/Tulipan s/n, 28933 – Móstoles (Madrid), Spain ; e-mail : micael.gallego@urjc.es

Ivan Garcia-Alcaide – LPSI, Universidad Politécnica de Madrid, Campus Sur, Carretera de Valencia Km. 7, 28031 Madrid, Spain; e-mail : igarcia@eui.upm.es

Soto Montalvo-Herranz – ESCET, Universidad Rey Juan Carlos, C/Tulipan s/n, 28933 – Móstoles (Madrid), Spain; e-mail: soto.montalvo@urjc.es

A SENSITIVE METRIC OF CLASS COHESION

Luis Fernandez, Rosalia Pena

Abstract: Metrics estimate the quality of different aspects of software. In particular, cohesion indicates how well the parts of a system hold together. A metric to evaluate class cohesion is important in object-oriented programming because it gives an indication of a good design of classes.

There are several proposals of metrics for class cohesion but they have several problems (for instance, low discrimination). In this paper, a new metric to evaluate class cohesion is proposed, called SCOM, which has several relevant features. It has an intuitive and analytical formulation, what is necessary to apply it to large-size software systems. It is normalized to produce values in the range [0..1], thus yielding meaningful values. It is also more sensitive than those previously reported in the literature. The attributes and methods used to evaluate SCOM are unambiguously stated. SCOM has an analytical threshold, which is a very useful but rare feature in software metrics. We assess the metric with several sample cases, showing that it gives more sensitive values than other well know cohesion metrics.

Keywords: Object-Oriented Programming, Metrics/Measurement, Quality analysis and Evaluation.

ACM Classification Keywords: D.1.5 Object-oriented Programming; D.2.8 Metrics

Introduction

The capacity to measure a process facilitates its improvement. Software metrics have become essential in software engineering for quality assessment and improvement. Class cohesion is a measure of consistency in the functionality of object-oriented programs. High cohesion implies separation of responsibilities, components' independence and less complexity. Therefore, it augments understandability, effectiveness and adaptability. Actually, these are major factors of the great interest in using object-oriented programming in software engineering.

"Current cohesion metrics are too coarse criteria that should be complemented with finer-grained factors (..). Then it will be easier to assess the trade off involved in any design activity, which would make it possible to see whether a system is evolving in the right direction [Mens, 2002].

One software metric must be simple, precise, general and computable in order to be applicable to large-size software systems. Automated metric producing sensitive values of class cohesion can be of great value to most designers, developers, managers, and of course to beginners, in identifying the class cohesion in an object-oriented design.

This paper presents, as follows, a metric of class cohesion that has several advantages over previous cohesion metric proposals. The next section presents a brief summary of the state or art about proposed metrics of class cohesion. In the third section, we formulate the new metric. The existence of thresholds is studied in fourth section, where two representative values are analytically determined. This section also provides guidelines to improve a class definition when its cohesion value is below the threshold. In fifth section, the method and attributes to be considered on the evaluation of a class are discussed. The sixth section assesses qualitative and quantitatively the desirable features of our cohesion metrics. Particularly interesting is a comparative analysis with previous cohesion metrics, based on a sample of classes with different cohesion features. Finally, we summarize our conclusions.

Related Work

In a cohesive class, all the methods collaborate to provide the class services and to manage the object state; in others words methods work intensively with the attributes of the class. A metric for class cohesion was first proposed by Chidamber and Kemerer in 1991 and then revised in 1994 [Chidamber, 1994]. Then, it has been repeatedly reinterpreted and improved [Li, 1993][Li, 1995][Hitz, 1996]. Properly speaking, these proposals do not evaluate cohesion but lack of cohesion (hence it name, LCOM) in terms of the number of pairs of the class methods that use disjoints sets of attributes.

There are two problems with Chidamber and Kemerer's expression [Chidamber, 1994]:

i. the metric has no upper limit, so it is not easy to grasp the meaning of the computed value, and
ii. there are a large number of dissimilar examples, all of them giving the same value (LCOM=0) but not necessarily indicating the same cohesion.

In spite of its pitfalls, LCOM still is the most widely used metric for cohesion evaluation [Bansiya, 1999]. Li et al. [Li, 1995] estimate the number of non-cohesive classes that should be redesigned by computing the number of subsets of methods that use disjoint subsets of attributes. We call this value the number of clusters of the class. Etzkorn et al. [Etzkorn, 1998] study which formulation provides the best interpretation of lack of class cohesion.

Henderson-Sellers [Henderson-Sellers, 1996] proposed a completely new expression for cohesion evaluation, called LCOM*:

$$LCOM^* = \frac{\left[\frac{1}{a}\sum_{j=1}^{a}\mu(A_j)\right] - m}{1-m} \tag{1}$$

where a and m are the number of attributes and methods of the class, respectively, and $\mu(A_j)$ is the number of methods that access the datum A_j ($1 \le j \le a$). Notice that this expression is normalized to the range [0..1]. He claims that this metric has the advantage of being easier to compute than previous ones.

A problem with LCOM* is that it is not clear how it can give any account of class cohesion or inter-method cohesion (relation between methods). In effect, the summatory in the numerator just counts how many times all the attributes are accessed, independently of which method accesses each datum. This problem is illustrated in Table 1, with a comparison of several metrics, including ours. The table includes a representation of classes in the second column. As well, classes are represented by another table in which rows can be found the attributes and in the columns the methods accessed by them. So, a filled $cell_{i,j}$ means that the i-th method uses the j-th attribute. For example, class C has 6 attributes and 6 methods, and its method M_4 uses attributes A_1, A_2, A_3 and A_4.

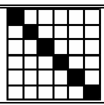
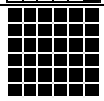
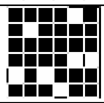
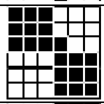
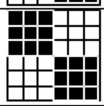
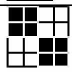
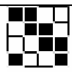
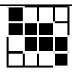
In particular, rows F, G and H show that whereas class F seems to be a good candidate to be split into two classes, it is not clear whether a more cohesive design can be found for classes G or H. However, the three classes are evaluated by LCOM* with the same value, it seems to be clear that it does not give a good account of class cohesion at least with these two examples.

Metrics measuring lack of cohesion, as proposed in [Chidamber, 1994][Li, 1993][Li, 1995][Hitz, 1996], evaluate the number of disjoint pairs of methods. From the opposite point of view, cohesion is related to the overlapping of these sets. It would be desirable to measure not only whether two methods are disjoint, but also, how big the intersection set is. Let us consider that two methods are more cohesive if they co-use more attributes. Consequently, cohesion is directly proportional to the cardinality of intersection for all the possible pairs of methods. [Park, 1998] suggested an evaluation of Class Cohesion (CC) as the sum of the connection intensity between all the possible pairs of methods. Let us denote I_k for the subset of attributes used by a given method M_k. The connection intensity of a pair of methods is defined as:

$$C_{i,j} = \frac{Card(I_i \cap I_j)}{a} \tag{2}$$

where *a* is the total number of instance attributes in the class. CC is more intuitive than LCOM*, but it also has deficiencies, as is shown in Table 1, where it also computes the same value for F, G, H classes.

Table 1. Values of different cohesion metrics on 8 sample classes.

CLASS		*LACK of COHESION*			*COHESION*	
Name	*Methods/Attribute Table*	*Chidamber (1994)*	*Li (1995)*	*Henderson (1996)*	*Park (1998)*	*SCOM*
A		15	6	1	0	0
B		0	1	0	1	1
C		0	1	0.23	0.62	0.82
D		0	1	0.57	0.23	0.29
E		3	2	0.60	0.20	0.20
F		2	2	0.67	0.17	0.17
G		0	1	0.67	0.17	0.25
H		0	1	0.67	0.17	0.36

Sensitive Class Cohesion Metric (SCOM)

In this section, we introduce a new cohesion metric that we call SCOM (standing for Sensitive Class Cohesion Metric). At this stage, let $\{A_1,\ldots,A_a\}$ and $\{M_1\ldots.M_m\}$ be the set of attributes and methods of the class, respectively, to be considered.

Firstly, let us consider connection intensity of a pair of methods (see formula 2 above). It would be more accurate if it were divided by the maximum possible value, rather than by the total number of attributes of the class, as proposed by Park in formula 2. As the largest possible cardinality of the intersection of two sets is the minimum of their cardinalities, the connection intensity formula becomes:

$$C_{i,j} = \begin{cases} 0, & if\ I_i \cap I_j = \phi \\ \dfrac{card(I_i \cap I_j)}{\min(card(I_i), card(I_j))} & otherwise \end{cases}$$

Connection intensity of a pair of methods must be given more weight when such a pair involves more attributes. Of course, the largest contribution is found when every attribute in the class is used by any of the two methods. The weight factor is:

$$\alpha_{i,j} = \frac{card(I_i \cup I_j)}{a}$$

Finally, the formula of our cohesion metric results:

$$SCOM = \frac{2}{m(m-1)} \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} C_{i,j} * \alpha_{i,j} \qquad (3)$$

where the coefficient of the summatory is the inverse of the total number of method pairs: $\binom{m}{2} = \frac{m(m-1)}{2}$.

As a consequence, the metric is normalized to the range [0..1], where the two extreme values are:

i. Value zero: there is no cohesion at all, i.e. every method deals with an independent set of attributes.

ii. Value one: full cohesion, i.e. every method uses all the attributes of the class.

Thresholds

The threshold or “alarm value” minimum (respectively, maximum) of a metric is a value that indicates a design problem for a software entity whose evaluation lies below (or over) it. The threshold calls for the developer’s attention to focus on a particular module or chunk for further evaluation [Lorenz, 1994].

As we mentioned in the previous section, the SCOM cohesion metric is ranged between zero (representing no cohesion at all) and one (representing full cohesion). Although an empirical threshold must be determined with real projects, we consider that a theoretical study about singular points gives information about how to interpret the values the metric delivers. In the next two subsections, we present the analytical expressions for the minimal value of a class that has one cluster and the maximum value of a class that has at least two clusters. The third subsection deals with the influence of these values on the SCOM threshold and its applicability for class evaluation.

Minimal cohesion value for one cluster.

Let us consider a class with *m* methods and *a* attributes defined so that it is impossible to find two clusters.

We start with *m=2* and then we include one by one new methods to the class satisfying the former condition. The lower contribution to cohesion is given by a method with just one attribute. The connection intensity $C_{i,j}$ of a pair of methods with only one attribute has the denominator equal to one (see formula 3). Counting the number of non-null terms, we induce the total number of pairs of the class with *m* methods and *a* attributes.

Table 2 shows the minimum number of pairs of non-null methods for the values of *a=3* and *a=4* and consecutive values of *m* starting from *m=2*. These minimums are expressed in terms of *a*'s and *m*'s. From this table, the general expression establishing the number of terms that contributes to cohesion in a significant manner can be deduced by induction.

Table 2. Number of pairs non-null methods.

	a=3		a=4	
m	S	S(a,m)	S	S(a,m)
2	1	a·0+1	1	a(0)+1
3	2	a·0+2	2	a(0)+2
4	3	a·(0+1)	3	a(0)+3
5	5	a(0+1)+2	4	a(0+1)
6	7	a(0+1)+4	6	a(0+1)+2
7	9	a(0+1+2)	8	a(0+1)+4
8	12	a(0+1+2)+3	10	a(0+1)+6
9	15	a(0+1+2)+6	12	a(0+1+2)
10	18	a(0+1+2+3)	15	a(0+1+2)+3
…	…	…	…	…

From these data, it can be induced that the number of terms that contributes in a significant manner to cohesion, named *S*, has a linear behavior. Hence, *S* can be expressed by $S = a{\cdot}t + r$, where:

$$t = \frac{1}{2} INT(\frac{m-1}{a}) \cdot \left[INT(\frac{m-1}{a}) + 1 \right]$$

and

$$r = \left[1 + INT(\frac{m-1}{a}) \right] \cdot Mod(\frac{m-1}{a})$$

hence,

$$S = \frac{a}{2} INT(\frac{m-1}{a}) \cdot \left[INT(\frac{m-1}{a}) + 1 \right] + \left[1 + INT(\frac{m-1}{a}) \right] \cdot Mod(\frac{m-1}{a})$$

and extracting common factors:

$$S(m,a) = \frac{1}{2} \left[1 + \mathrm{int}(\frac{m-1}{a}) \right] \left[\mathrm{mod}(\frac{m-1}{a}) + m - 1 \right]$$

The minimum weight factor αi,j occurs when both methods deal with the same attribute being 1/a, but some coefficients will actually be higher. Therefore:

$$SCOM\ \min > \frac{2}{m(m-1)a} S(m,a) = SCOM \min K \tag{4}$$

We call $SCOM_{minK}$ to the minimal known value. A class with SCOM < $SCOM_{minK}$ does not satisfy the induction premise. Consequently, we can claim that it has at least two clusters and it must be subdivided into smaller, more cohesive classes.

Maximum cohesion value for two clusters.

From the opposite point of view, it is possible to evaluate the maximum value of SCOM in presence of two clusters. In this situation there are at least *m-1* null terms in (3); in other words, there are at most $(m-1)\cdot(m-2)/2$ non-null terms. The biggest $C_{i,j}$ is valued one. The biggest weight factor occurs when one cluster has *a-1* attributes and the other has just one attribute and all the methods in each cluster involve all the attributes in such a cluster. In this situation, $\alpha_{i,j}=(a-1)/a$.

Finally, the maximum value of SCOM for a class with m methods arranged in two clusters is:

$$SCOM2\max = \frac{(a-1)(m-2)}{ma} \tag{5}$$

Threshold applicability.

As we explained above, the equation (4) provides the minimum value below which certainly there are two subsets using non-overlapping attributes. This class can be automatically split into two (or more) smaller and simpler classes.

The metrics for class size by Lorenz and Kidd [Lorenz, 1994] suggest a threshold of 20 instance methods and 4 class methods, being 12 the number average of methods per class in a project. They also proposed 3 as a threshold both for instance attributes and class attributes in a class, and 0, 1 the class attributes average in a project. With these particular values (a=*3* and m=*12*), the minimal cohesion is $SCOM_{minK}$=0.13.

Moreover, SCOM2max provides the analytical value that guarantees the existence of one cluster. For a=*3* and m=*12*, $SCOM_{2max}$=0.56. Figure 1 places both values on the cohesion range.

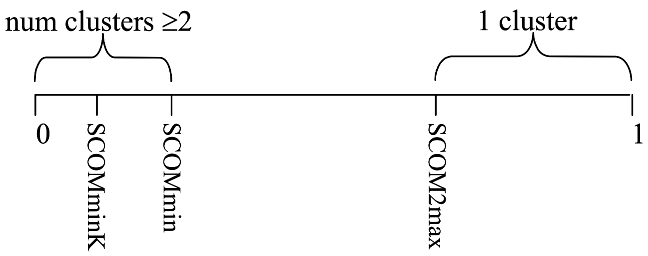


Figure 1. Significant values on cohesion range.

The presence of two clusters clearly indicates that the class may be split. Not all the classes with more than one cluster can be detected by $SCOM_{minK}$. Li formulation could help but, even in presence of just one cluster, the class design can be low cohesive and SCOM is giving account of such a situation.

The threshold considered in software evaluation should be between $SCOM_{minK}$ and 1. It looks feasible to demand guarantee of just one cluster, then SCOM will actually be between $SCOM_{2max}$ and 1. The adequate value, inside these limits, should be established by studying real projects.

Quoting Lorenz and Kidd [Lorenz], “metrics are guidelines of the quality of the design and not absolute laws of nature”. Thus, when the alarm is raised, the corresponding software entity is suspicious to have design problems, so it must be re-analyzed. However, thresholds are affected by many factors (as pointed out in the next section), so the warning may sometimes be disregarded.

Criteria of Application

The features of object-oriented programming do not always make obvious what methods and attributes to consider for estimating cohesion. Moreover, different authors use different criteria and even they say nothing about what to do with them. This un-definition or not precise definition in the way of evaluating cohesion makes very difficult to interpret the obtained values from the metrics. We discuss here the most relevant aspects.

i. Transitive closure of attributes. Whenever M_i invokes method M_j, M_i transitively uses each attribute being used by M_j. Consequently, these attributes must be taken into account for the metric evaluation. The same reason is argued to consider objects that are exported to clients which map a subsequent client request (method invocation sending the argument “this”), as it happens in the double dispatching and visitor patterns. Attributes involved in subsequent request must be counted for the exporter method.
ii. Class and instance methods and attributes. Class cohesion involves both instance and class methods and attributes. It is common that class attributes are less involved in methods than instance attributes, which would decrease inadequately the class cohesion. These attributes could be removed to evaluate cohesion of a single class. However, the rate between class and instance attributes use to be too small in the average to affect evaluation substantially, so they can either be removed or considered.
iii. Constructor and destructor methods. Etzkorn *et al*. [Etzkorn, 1998] suggested to exclude constructor methods but to keep the destructor method for LCOM calculations. The argument was that these particular functions often access most or all of the variables of a class. As LCOM looks for methods with non-overlapping attributes, a method using all the attributes lowers its capability of discrimination. However, we think that both constructor and destructor methods must be considered. For instance, a constructor that initializes all the attributes (as probably happens in the Car class of a concessionaire) contributes to a larger cohesion value than a constructor that just initializes some attributes (as will likely happen in a Student class, where the name will be required from the very beginning, but the student final evaluation will not).
iv. Access methods. The *get* and *put* private methods make easier future maintenance. They typically cope with just one attribute. However, the resulting cohesion value should not be negatively affected by a good programming style. Therefore, we suggest excluding them.
v. Methods overload. It is common in object-oriented programming to have a given method called with a concrete value. Some programming languages (such as C++) allow parameter initialization in absence (for example, an interval class being displaced to the coordinate origin). When a language lacks this feature (for instance, in Java or Smalltalk), the method is overloaded by defining a new method without this parameter declaration, where the new method’s implementation just consists in the invocation of the former with a constant value. Both methods cope with the same functionality and in this particular case, only the general one must be considered for cohesion evaluation.
vi. Inheritance in cohesion evaluation. The question of whether inherited variables must be considered for cohesion evaluation was posed by Li *et al*. [Li, 1995]. They only considered local attributes. However, in [Etzkorn, 1998] argues in favor of the consideration of inherited attributes.
The state of one instance is defined by its own and parent attributes, so both of them must be considered. Thus, we propose to consider the parent’s methods and attributes in the parent class cohesion evaluation, and the parent’s and its own attributes, but just its own methods being involved on subclass cohesion evaluation. In some types of inheritance, the subclass methods seldom use the inherited attributes; in such a situation, the metric value decreases.
vii. Methods dealing with no attributes. We find this situation in abstract classes and sometimes in static methods. On the one hand, abstract methods must not be considered for cohesion evaluation of the abstract class. They must be considered for the subclass where they are actually implemented. On the other hand, static methods not dealing with attributes (as happens in the factory pattern) are not considered for cohesion. It is expected that instance methods manage at least one attribute; otherwise, a different metric from cohesion must be used and it must raise the appropriate alarm. Therefore, methods, which deal with no attributes, must not be considered.
viii. Attributes not used by any method. This is an undesirable situation in absence of inheritance. Actually, some compilers (e.g. Fortran90 or Java) warn about it. A different metric will deal with it. In presence of inheritance, the parent class sometimes references an attribute to be managed by its subclasses. These attributes must not be considered for superclass evaluation, but they must be considered for subclass evaluation.

Metric Assessment

In this section, we assess our SCOM metric. Firstly, we make an informal assessment. Then, we make an empirical comparison with other metrics.

Qualitative analysis of cohesion metrics.

The desirable properties (or metametrics) of a software metric have been proposed elsewhere [Henderson-Seller, 1996][Xenos, 2000]. We assess SCOM informally with respect to such properties:

i. Simplicity (the definition of the metric is easily understood). SCOM has a simple definition.

ii. Independence (the metric does not depend on its interpretation). SCOM is independent because it has a clear definition based on syntactic information (attributes and methods).

iii. Accuracy (the metric is related to the characteristic to be measured). SCOM is accurate since it is based on cohesion features, namely dependencies between attributes and methods.

iv. Automation (effort required to automate the metric). SCOM has an analytical definition that depends on the attributes used by the methods in a class and their transitive closure. This information can be automatically extracted from the class implementation.

v. Value of implementation (the metric may be measured in early stages or it evaluates the success of the implementation). SCOM measures the class implementation.

vi. Sensitivity (the metric provides different values for non-equivalent elements). SCOM is sensitive for different cases, as it is shown in the next subsection.

vii. Monotonicity (the value computed for a unity is equal or greater/less than the sum of the values of its components). SCOM is decreasingly monotonous.

viii. Measurement scale (rate/absolute/nominal/ordinal, being more useful rate and absolute scales than nominal and ordinal ones). SCOM gives a normalized rate scale.

ix. Boundary marks (upper and/or lower bounds). SCOM has bounds on both sides of the scale since it ranges between 0 and 1.

x. Prescription (the metric describes how to improve the entity measured). SCOM establishes how to improve the class as we describe below.

There are two typical cases where some restructuring actions can be performed on a class. Firstly, when two (or more) non-overlapping clusters are detected in a class, it must be split into two (or more) classes. The methods using a subset of non-overlapping attributes must be grouped, taking care to preserve the appropriate granularity (size) of the class.

Secondly, when the threshold alarm is raised but isolated clusters are not detected, the class is a candidate to be re-analyzed. The programmer must inquire whether any attribute can be moved to another class, look for any undiscovered class, or check whether all the class responsibilities are necessary (or any can be moved to a different class). By doing this clean up, some methods may go away or may be shortened. As a consequence, the comprehensibility of the application and the quality of the class design are improved, while code length actually decreases.

Quantitative analysis of cohesion metrics.

We have applied different cohesion metrics, including SCOM, to several sample classes in order to empirically comparing their results. Table 1 illustrated such results for eight sample classes.

Notice that while Chidamber, Li-Henry and Henderson evaluate lack of cohesion, Park and SCOM evaluate cohesion, so their extreme values have the opposite interpretation.

Class A has no cohesion at all while class B has the highest cohesion, so they obtain the limit values for metrics with bounds (Henderson, Park and SCOM).

Classes B, C and D represent very different situations. Chidamber and Li-Henry metrics show low sensibility, rating the same value (zero) for these classes, whereas the other metrics give an account of their differences.

Classes E and F are very similar: they share the property of having two clusters, being obvious that they could be split into two. Chidamber's metric yields different values for these classes, showing that the meaning of non-zero values in this formulation are difficult to understand because it does not have an upper bound. Henderson's metric also exhibits low discrimination because the lack of cohesion measure obtained is not as high as we could expect.

Classes F, G and H also represent different situations. Henderson and Park cohesion metrics yield the same value, showing lower discrimination capability than expected. However, SCOM gives account of their differences, yielding the lowest value for the obviously worst design (F).

SCOM is more sensitive than Park's formulation, as can be seen comparing the cases D and E. Moreover, Park's formulation seems to be too restrictive for the cohesion's upper extreme, as it is shown in class C. This class has a good cohesive appearance, so a value higher than 0.62 is intuitively expected for a magnitude ranging between 0 and 1. SCOM behaves better, yielding the value 0.82.

Remember that our metric also allows computing the analytical values $SCOM_{minK}$ and $SCOM_{2max}$ that give information about the cohesion for a particular class, without any previous human analysis of its table structure. For a class with 4 attributes and 4 methods, $SCOM_{minK}$=0.13 and $SCOM_{2max}$=0.38. It can be seen that classes F, G and H are low cohesive; in particular, F should be split. For a class with 6 attributes and 6 methods, $SCOM_{2max}$=0.56. The classes D and E have values far below $SCOM_{2max}$ (0.29 and 0.20, respectively), indicating design problems. The value 0.82 obtained for class C suggests a good cohesive design, as expected analyzing its methods/attributes table.

Conclusion

This paper proposes a metric to evaluate class cohesion, that we have called SCOM, having has several relevant features. It has an intuitive and analytical formulation. It is normalized to produce values in the range [0..1]. It is also more sensitive than those previously reported in the literature. It has an analytical threshold, which is a very useful but rare feature in software metrics. The attributes and methods used for SCOM computation are also unambiguously stated.

The metric is simple, precise, general and amenable to be automated, which are important properties to be applicable to large-size software systems. By following the prescription provided by the metric, the understanding and design quality of the application can be improved.

Bibliography

[Mens, 2002] T.Mens and S.Demeyer. Future trends in software evolution metrics. In: Proc. IWPSE2001, ACM, 2002, pp. 83-86.

[Chidamber, 1994] S.R.Chidamber, and C.F.Kemerer. A Metrics Suite for Object Oriented Design. In: IEEE Transactions on Software Engineering, 1994, 20(6), pp. 476-493.

[Li, 1993] W.Li, and S.M.Henry. Maintenance metrics for the object oriented paradigm. In: Proc. 1st International Software Metrics Symposium, Baltimore, MD: IEEE Computer Society, 1993. pp. 52-60.

[Li, 1995] W.Li, S.M.Henry et al. Measuring object-oriented design. In: Journal of Object-Oriented Programming, July/August, 1995, pp.48-55.

[Hitz, 1996] M.Hitz, and B.Montazeri. Chidamber & Kemerer's metrics suite: a measurement theory perspective. In: IEEE Transactions on Software Engineering, 1996, 22(4), pp. 267-271.

[Bansiya, 1999] J.Bansiya, L.Etzkorn, C.Davis, and W.Li. A Class Cohesion Metric For Object-Oriented Designs. In: JOOP, 1999, 11(8), pp. 47-52.

[Etzkorn, 1998] L.Etzkorn, C.Davis and W.Li. A practical look at the lack of cohesion in methods metric. In: JOOP, 1998, 11(5), pp. 27-34.

[Henderson-Sellers, 1996] B.Henderson-Sellers. Object-Oriented Metrics: Measures of Complexity. In: New Jersey, Prentice-Hall, 1996, pp. 142-147.

[Park, 1998] S.Park, E.S.Hong et al. Metrics measuring cohesion and coupling in object-oriented programs. In: Journal of KISS, 1998, 25(12), pp. 1779-87.

[Lorenz, 1994] M.Lorenz, and J.Kidd. Object-Oriented Software Metrics: A Practical Guide. In: New Jersey, Prentice Hall, 1994, p. 73.

[Xenos, 2000] M.Xenos, D.Stavrinoudis, K.Zikouli, and D.Christodoulakis. Object-Oriented Metrics: A Survey. In: Proc. FESMA, Madrid, 2000, pp. 1-10.

Authors' Information

Luis Fernandez – UPM, Universidad Politécnica de Madrid; Ctra Valencia km 7, Madrid-28071, España; e-mail: setillo@eui.upm.es

Rosalia Pena – UA, Universidad de Alcalá; Ctra Madrid/Barcelona km 33, Alcalá-28871, España; e-mail: rpr@uah.es

HANDLING THE SUBCLASSING ANOMALY WITH OBJECT TEAMS

Jeff Furlong, Atanas Radenski

Abstract: Java software or libraries can evolve via subclassing. Unfortunately, subclassing may not properly support code adaptation when there are dependencies between classes. More precisely, subclassing in collections of related classes may require reimplementation of otherwise valid classes. This problem is defined as the subclassing anomaly, which is an issue when software evolution or code reuse is a goal of the programmer who is using existing classes. Object Teams offers an implicit fix to this problem and is largely compatible with the existing JVMs. In this paper, we evaluate how well Object Teams succeeds in providing a solution for a complex, real world project. Our results indicate that while Object Teams is a suitable solution for simple examples, it does not meet the requirements for large scale projects. The reasons why Object Teams fails in certain usages may prove useful to those who create linguistic modifications in languages or those who seek new methods for code adaptation.

Keywords: Languages; Code reuse; Subclassing

ACM Classification Keywords: D.3.2 [Programming Languages]: Language Classifications – Extensible languages, object-oriented languages; D.3.3 [Programming Languages]: Language Constructs and Features – Classes and objects, data types and structures, inheritance, polymorphism

1 Introduction

As the requirements for an object-oriented application evolve, so do the applications themselves. For example, an application or even a programming language may need to be enhanced with new functionalities or new linguistic features. Consequently, the set of Java classes that define the application or the set of classes that define a compiler may need to be adequately adapted for such changes [10].

Subclassing is the principle object-oriented programming language feature that provides code adaptation. Patterns may provide an alternative solution, but we are interested in a direct linguistic primitive. Subclassing allows for the derivation of new classes from existing ones through extension and method overriding. A subclass can inherit variables and methods from a parent class, can extend the parent class with newly declared variables and methods, and can override inherited methods with newly declared ones [10].

When a class that needs to be updated belongs to a collection of classes but is independent from all other classes in the collection, the functionality of that class can be easily updated through subclassing and method overriding. Subclassing is a straightforward code adaptation mechanism in the case of independent classes [10].

However, when working with dependent classes, subclassing may utilize outdated and unintended parent classes, a phenomenon that has been termed the *subclassing anomaly*. Hence, there is a serious concern because the benefits of inheritance are immediately lost in the presence of the subclassing anomaly [10].

Class overriding is a linguistic mechanism to overcome the subclassing anomaly. When new classes are created via subclassing, the correct parent class is overridden and new methods or variables can be introduced. Thus, the subclassing anomaly can be eliminated by using this technique.

Object Teams, an extension to Java, implements a limited version of class overriding and is a proposed method to overcome the subclassing anomaly. It is also an active project today, and is gaining popularity in the Java community. Additionally, Object Teams sums up a good representation of many efforts to enhance Java's code adaptation capabilities with linguistic techniques.

For our evaluation, we use Object Teams version 0.6.1c, which was the latest available release at the time of testing. Our tests were performed during late 2004. Unless otherwise noted, we will indicate Object Teams version 0.6.1c as simply Object Teams.

In this paper, our goal is to evaluate how well Object Teams provides a solution to the subclassing anomaly. We perform this by implementing a complex, real world case study instead of trivial sample code. In section 2, we discuss the subclassing anomaly in more depth and relate it to Object Teams. We present our case study and the evaluation results in section 3. Lastly, in section 4, we conclude our findings, provide future work, and indicate some related works.

2 Subclassing with Object Teams

Applications can be compiled with classes that are either polymorphic or monomorphic references [9]. The subclassing anomaly is limited to occur in monomorphic references. There are three distinct contexts, which include constructor invocations, subclass definitions, and static member accesses. Because the monomorphic reference often points to an old class, new applications must be recompiled to point to an enhanced subclass, which, in turn, may refer to an original parent class. Figure 1 depicts this concept.

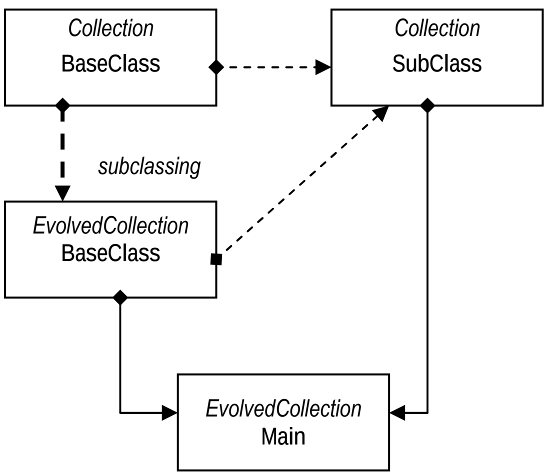


Fig. 1. A subclass definition example.

The outer Java class *Collection* contains *BaseClass*, which is subclassed to *SubClass*. Attempting to reuse code, a programmer subclasses the *Collection* outer class and expects the *Main* class to use newly updated methods in the *Collection.Subclass* class. However, only the methods in *Collection.BaseClass* and *EvolvedCollection.BaseClass* are used. In class overriding, we create a link between the new *BaseClass* and the existing *SubClass*.

There are several options when attempting to overcome the subclassing anomaly. A programmer could reimplement the exact same source code in the new project, if the source is available. This creates obvious code bloating and wastes development time. Alternatively, a project specific fix could be created through the use of additional base classes, interfaces, and message forwarding. However, this option requires some knowledge of how future updates would be implemented, and is often self-defeating. Alternatively, a project could be designed with Object Teams.

Object Teams is an extension of Java [2] with a few new linguistic mechanisms, a new compiler, and an optional new class loader (JMangler). This extension provides much more flexibility without the need to learn a new language because Object Teams is so similar to Java. The new team object extension allows for adaptation and refinement without the expense of new language syntax. Additionally, the compiled class files are native to the JVM, so users need not install a new runtime environment.

In fact, programmers could swap out the Java compiler with the Object Teams compiler. Everything that can be done with the Java compiler can be done in the Object Teams compiler. However, there are three main notable components of the Object Teams language. The first is family polymorphism [2], which redefines the way in which Java handles inheritance. Collections of classes are represented as teams with inner classes. With regard to the subclassing anomaly, some previously problematic inner classes are created as new virtual classes via implicit inheritance. It is this change that allows the subclassing anomaly to be fixed. Object Teams also makes use of translation polymorphism [2], which addresses inheritance with different containment objects. Lastly, many aspect-oriented features are available in Object Teams, such as the ability to remap method parameters and change return types. Such capabilities are not of interest to our goals with the subclassing anomaly.

```
public team class Collection {
  public class BaseClass {
    public BaseClass() {
      method();
    }
    public void method() {
      System.out.println("Collection.BaseClass.method()");
    }
  }
  public class SubClass extends BaseClass {
  }
}

public team class EvolvedCollection extends Collection {
  public class BaseClass {
    public void method() {
      System.out.println("EvolvedCollection.BaseClass.method()");
    }
  }
  public static void main (String args []) {
    final EvolvedCollection evolvedCollection = EvolvedCollection();
    evolvedCollection.BaseClass baseClass = evolvedCollection.new BaseClass();
    evolvedCollection.SubClass subClass = evolvedCollection.new SubClass();
  }
}
```

Fig. 2. A subclass definition: The Object Teams code is very similar to the Java equivalent. For *EvolvedCollection*, the first output is `"EvolvedCollection.BaseClass.method()"` and the second statement correctly outputs the same updated message; although, the Java equivalent would produce the old message. Hence, the subclassing anomaly is indicated and implicitly overcome.

In Figure 2, an example of a subclass definition version of the anomaly is shown. An outer class *Collection* contains inner classes *BaseClass* and *SubClass*. However, a programmer decides to evolve the original functionality, and wants to change the method within *BaseClass*. In standard Java, the new method is only called when line 22 is executed, not again at line 23. Instead, the original non-overriden method is called. This result is likely to be opposite of the desire of the programmer.

In order to fix this problem while using Object Teams, no additional design strategies are needed. Users can simply develop code as is normally done in Java. The cure is handled implicitly by the Object Teams compiler, specifically by creating virtual classes where they are needed. Minor syntactic changes, such as the keyword `team` denote slight differences with Object Teams code; but, there are no great differences with Java equivalent code. What is different is the presence of the anomaly in the standard Java code and the absence of the anomaly in the Object Teams code.

3 Case Study

The context of the subclassing anomaly is not likely to occur in every updated Java design. However, when software evolution and code reuse are motivations of programmers, problems may quickly be realized. When evolving projects with a simple GUI or theme enhancement, subclassing could be used to change what colors methods paint to a display. More complicated projects may not be linear, and several classes may need to be updated in part so that they are compatible with other original classes. As applications evolve, so do languages and hence compilers.

Our real world test is a case study of an evolving compiler. Because a compiler exhibits a complex dependency relationship between an AST, parser, analyzer, and code generator, it is a very suitable choice to test Object Teams' capabilities. Further, it has been well studied and understood by many. We have implemented the Triangle (educational/testing) language compiler [12] and runtime environment in Object Teams. We can successfully compile the Java equivalent code, but our goal is to evolve the language in some fashion, such as by adding new data types or capabilities. After performing these updates, we find several obstacles that prevent the subclassing anomaly from being solved.

3.1 Object Teams and the subclassing anomaly

Unlike the above simple example where Object Teams worked well, Object Teams has several weaknesses when applied to realistic usages. When typecasting or using the `instanceof` operator in Object Teams, users will quickly find such features to be unsupported. Additionally, a real world project will have complex relationships, and Object Teams fails to always determine a project level dependency analysis. As an alternative to inner classes, users may wish to use packages. However, in some cases with packages, name clashes may cause errors within the Object Teams compiler. Further, extended subclasses may be limited in usage context, as not every situation has been tested by the Object Teams Developers.

Typecasting and instanceof. Attempting to typecast or use `instanceof` with inner classes will produce obvious problems. In the Triangle compiler code, a *Contextual Analyzer* implements a visitor pattern and visits the parsed abstract syntax tree. The *Contextual Analyzer* has a class *Checker*, which determines the types of expressions, among other things. Figure 3 indicates how typecasting and the `instanceof` keyword might be used. An expression in the AST may be visited and then returned, so it can be typecast to determine the type of the expression, such as `int`, `char`, etc. Since *TypeDenoter* is an inner class of *AST*, we can typecast with the form *myAST.TypeDenoter*. However, in Object Teams, there is no support to typecast with inner classes. Similarly, using the `instanceof` operator is not supported. These features are simply issues that have not been fully developed in Object Teams. For small projects, it is possible to simply abandon inner classes and use outer team classes to typecast. However, this modification dismisses the notion of encapsulation and also produces even larger problems in large projects, including the Triangle compiler. Unfortunately, this means that the subclassing anomaly must be solved with other methods, if possible.

```
public Object visitBinaryExpression(myAST.BinaryExpression ast, Object o) {
  myAST.TypeDenoter e1Type = (myAST.TypeDenoter) ast.E1.visit(this, null);
  myAST.TypeDenoter e2Type = (myAST.TypeDenoter) ast.E2.visit(this, null);
  myAST.Declaration binding = (myAST.Declaration) ast.O.visit(this, null);
  if (binding instanceof myAST.BinaryOperatorDeclaration)
    myAST.BinaryOperatorDeclaration bbinding =
                                    (myAST.BinaryOperatorDeclaration) binding;
    ...
  return ast.type;
}
```

Fig. 3. Typecasting and instanceof: Attempting to typecast with inner classes fails because it is not currently supported in Object Teams.

Project level dependency analysis. The Object Teams compiler does not support a project wide dependency analysis on all classes. The consequences of this issue can be indicated in sample code. Figure 4 shows a simple example of how instances of an object, team *AST*, may be used between different outer team classes.

```
public team class SyntacticAnalyzer {
   public AbstractSyntaxTrees myAST = new AbstractSyntaxTrees();
   public myAST.BinaryExpression binaryExpAST = myAST.new BinaryExpression();
   ...
}
_____________________________________________________________________________
public team class ContextualAnalyzer {
   public AbstractSyntaxTrees myAST = SyntacticAnalzyer.myAST;
   myAST.Expression expressionAST = SyntacticAnalyzer.binaryExpAST;
   ...
}
```

Fig. 4. Externalized Roles: References must be to the same instance of an object. In the code, there is one instance of *AbstractSyntaxTrees*, which is used to create two other objects, regardless of what class they are located in. However, the Object Teams compiler does not always know that one object is shared through classes.

In our object-oriented compiler case study, methods do the work of returning objects and assigning them to appropriate locations. However, Object Teams requires that if a role (inner class) is to be used outside of its immediate team class, then the role that it is being assigned to must not only have the same team type, but also must be the same team instance. Further, the team object (instance of team class) that it is part of must be denoted as `final`. Type safety could not be guaranteed otherwise, especially if a user changes the type of the team. In order for the compiler to know that each *myAST* is of the same instance of *AST*, it must determine the dependency requirements of the whole project. Such a feature is not yet supported in Object Teams, which makes performing such assignments difficult. A programmer can compile dependent projects knowing which classes to compile in successive order. But, when projects have multiple instances of the same object being referenced from several classes, Object Teams fails to understand that each object is really the same instance, despite help from the programmer. A full project dependency analysis is imperative for this use, something that IBM's Jikes compiler [3] supports so that classes can be built autonomously. Without this feature in Object Teams, the only option to fix this issue is to model more inner classes as outer team classes. This workaround creates the same problems mentioned above and is not suitable for large scale projects.

Packages. As alternatives to inner classes, packages may be used to create a program hierarchy. Each package is represented by a team class and each inner class is defined by a separate file. Using this approach, we can successfully compile the Triangle language in an appropriate package design. However, when we attempt to evolve this package to, say an *EvolvedTriangle* package, we notice some problems. Under some circumstances of the constructor invocation version of the subclassing anomaly, there are compile time errors. These issues refer to name clashes, stack overflow errors, and internal bugs under Object Teams. However, the same source code, converted to the Java equivalent form, does compile without problems on the regular Java compiler. Unfortunately, executing such Java code quickly proves that the subclassing anomaly is an inherent problem. Hence, there is a small gap of what can be ported from Java to Object Teams.

Usage context. In our case study, we found a class that used throwable objects to catch exceptions. The Triangle compiler has a *Parser* class, which reads the input source files and determines if the format is correct. Figure 5 shows a simplified version of the *Parser* class. If we want to override the *parseDeclaration()* method, necessary if we need to add a new type of declaration to the Triangle language, we can do so by the approach followed in *ExtendedSyntacticAnalyzer*. However, the Object Teams compiler does not correctly determine the correct context of *SyntaxError*. Another bug or lack of design implementation prevents this ability to be used in Object Teams. Although it is possible to compile the original collection, any attempt to override it will produce this error, which prevents us from further testing of the subclassing anomaly under Object Teams. If the equivalent code is tested in regular Java, there are no problems with compilation or usage. This issue further hinders our attempts to produce evolved software.

Using only Object Teams, it is possible to convert all of the Triangle Java source code files to team classes. However, inner classes cannot be used often in our case study, because of the described problems. Therefore, outer team classes may be used as replacements. If a programmer knows all of the limitations and necessary adjustments for Object Teams, it is not a massive task to move original Java code to Object Teams code. However, our real goal is to add extensibility to the Triangle compiler, which requires a new collection or package to be created. Otherwise, we have just remodeled an old piece of software without making any changes. We at least need to make a base to allow for easy future updates. Since there is no method to allow for such evolution, either through inner classes or packages, it is not currently possible to overcome the subclassing anomaly in a real world complex project. Not enough of the compiler has been thoroughly tested or developed; otherwise it ought to be possible to overcome the anomaly on a large scale.

```
public team class SyntacticAnalyzer {
  public class Parser {
    public void parseDeclaration() throws SyntaxError { ... }
  }
  public class SyntaxError extends Exception { ... }
}
_______________________________________________________________________
public team class ExtendedSyntacticAnalyzer extends SyntacticAnalyzer {
  public class Parser {
    public void parseDeclaration() throws SyntaxError { ... }
  }
}
```

Fig. 5. Extended classes that throw objects:
Unfortunately, this appears to be an untested and unsupported case in Object Teams.

3.2. Performance evaluation

As part of our goal, we can evaluate the overhead in Object Teams in the cases where it succeeds in providing code adaptation. We can run performance tests on the converted version of the Triangle source, now implemented in Object Teams, as wells as smaller evolution tests. Such results will indicate what penalties there are when working with such a tool that claims to abstract implementation details from the programmer. After a series of detailed tests on Java implementations and their equivalent Object Teams implementations, we can indicate exactly this notion. The most important of these is the run time performance. Object Teams uses and creates more class files, so method lookups and class loading require more time. Also, we measured the extra run time requirements of using overridden classes.

Test Case	Java	Object Teams
Runtime for overridden methods	8 % more	11+ % more
Triangle implementation, compiled classes	115	341 (60 % more bytes)
Compile time	Average	128 % more

Table 1. Java vs. Object Teams Performance:
The most important performance result, runtime, indicates that Object Teams overcomes the anomaly at a small price.

As projects grow large, it can be expected that run time will grow. About eleven percent more runtime is required with Object Teams than for Java equivalent code. The number of classes and hence number of files also increases. The only bound on extra file size would be restricted to just over twice the bytes of the original project, since the worst case is overriding all existing code and adding the Object Teams core code. Overridden classes are achieved by adding virtual classes via implicit inheritance. Thus, original source is virtually copied to new extensions, creating larger class files. The amount of time it takes to compile projects using the Object Teams compiler is not of great concern, since programmers can accept extra time once here, not every time a user invokes an overridden class. Our indications report that the extra time required with Object Teams scales with the size of the project; in our real world example, just more than twice the amount of time is necessary. As long as the run time performance is similar with both techniques, the compilation time can be neglected. The most important performance issues indicate that Object Teams is currently a sound choice for the anomaly, unlike reflection, which requires significantly more run time. If Object Teams will be fixed with solutions to the above problems, these performance tests must be reiterated to determine if greater functionality arrives at performance costs.

3 Conclusion

We have indicated the context of the subclassing anomaly and the problems faced when attempting to implement a solution to a real world complex example. Although Object Teams successfully overcomes the subclassing anomaly in small scale, simple examples, it fails on real world cases. Missing essential features, such as typecasting and instanceof, provide insight to the downfall of Object Teams. A project level dependency analysis that fails and package behavior that is unexpected confirm that not all features are supported in Object Teams. As certain usages produce errors, it is clear that there needs to be additional tests performed on Object Teams by the developers. Our implementations may not have pointed out every issue, as we are only reporting on what is relevant to the subclassing anomaly. Other concerns such as cyclic redundancies and any bugs with the aspect-oriented features remain undocumented. However, at the least, we hope that we have assisted the Object Teams developers by pinpointing some ambiguous areas.

It should be noted that the current version of Object Teams, version 0.6.1c at the time of this writing, is not said to be a final release, but a work in progress. Object Teams developers seek to release a new version, which perhaps will address some of these problems. At such a time, the problems detailed in this paper should be reattempted. If the issues can be corrected without additional performance penalties or additional drawbacks, Object Teams stands to be a possible choice for overcoming the subclassing anomaly. However, until such a time when there is a near perfect solution, we remain without a technique to overcome all versions of the subclassing anomaly on any scale.

There are several other techniques that attempt to overcome the subclassing anomaly. Keris [13] is essentially a flavor of Java, and is strikingly similar to Object Teams, but at the least does not support implicit inheritance or class overriding. MultiJava [1] allows for open classes, but may be difficult to implement and lacks community support for such software. Extendibility is addressed in the generic visitor pattern [8], but any results here are affected by the run time degradation because of the usage of reflection. OpenJava [11] again produces results with reflection, and also is influenced by the run time loss; it, too, is inadequately supported. The most exciting recent addition is Jx, which attempts to do almost exactly what Object Teams does [6]. The developers have also followed a testing phase that includes nearly an identical plan with the evaluation presented in this paper. Unfortunately, the problems with Jx are often exactly the same as with Object Teams, and Jx is currently not available for public use. None of the available frameworks are the perfect solution because of the difficulty in applying them, the amount of time needed to implement evolved software, the unsupported features, or various other reasons. At first glance, Object Teams appears to excel all of these other alternatives, but it is now apparent that it is not perfect for significant projects. If Object Teams did perform as claimed, it would clearly be a suitable choice for the anomaly and for programmers because of its ease of use, nearly identical syntax as Java, small performance penalties, and compatibility with the existing JVMs. Thus, the overall efficiency of Object Teams far exceeds any other option.

Acknowledgements

This material is based in part upon work supported by the National Science Foundation under Grant Number CCF - 0243284. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Special thanks go to Stephan Herrmann, who provided much needed support for Object Teams.

Bibliography

[1] Clifton, C., Leavens, G., Chambers, C., and Millstein, T., 2000. MultiJava: Modular Open Classes and Symmetric Multiple Dispatch for Java. OOPSLA'00, Minneapolis, Minnesota, October 2000, ACM Press, New York, 130-145. Available from <http://www.cs.iastate.edu/~cclifton/multijava-data/papers/TR00-06.pdf>

[2] Herrmann, S., 2004. Confinement and Representation Encapsulation in Object Teams. Technical Report June 2004, Technical University Berlin. Available from <http://www.objectteams.org/publications/TR2004-06.pdf>.

[3] Jikes Homepage. <http://www-124.ibm.com/developerworks/oss/jikes/>.

[4] Joy, B., et al. 2000. Java Language Specification, Addison-Wesley Pub Co, New York, New York.

[5] Laffra, C., 2003. Java Bytecode Manipulation with JikesBT.

[6] Nystrom N., S. Chong, and A. Myers. Scalable extensibility via nested inheritance. Proceedings of Object Oriented Programming, Systems, Languages, and Applications, OOPLSA 04, October 24-8, 2004, Vancouver, British Columbia, ACM, 99-115.

[7] Object Teams Language Definition. <http://www.objectteams.org/def/index.html>.

[8] Palsberg J., and Jay, C. B., 1997. The essence of the visitor pattern. Technical Report 05, University of Technology, Sydney, Australia.

[9] Radenski, A., 2004. Anomaly-Free Component Adaptation with Class Overriding, Journal of Systems and Software 71 (2004), 37-48.

[10] Radenski, A., 2003. The Subclassing Anomaly in Compiler Evolution. International Journal on Information Theories and Applications, Institute of Information Theories and Applications, Sofia, Vol. 10, No 4, 2003, 394-399.

[11] Tatsubori, M., 1999. An Extension Mechanism for the Java Language. Master's dissertation, University of Tsukuba, Tsukuba City, Japan. Available from <http://www.csg.is.titech.ac.jp/openjava/papers/mich_thesis99.pdf>.

[12] Watt, D. and Brown, D., 2000. Programming Language Processors in Java: Compilers and Interpreters, Prentice Hall, New York, New York.

[13] Zenger, M., 2002. Evolving software with extensible modules. In: International Workshop on Unanticipated Software Evolution, Málaga, Spain. Available from <http://zenger.gmxhome.de/papers/use02.pdf>.

Authors' Information

Jeff Furlong – Department of Computer Science, Chapman University, 1 University Dr, Orange, CA 92866, USA; furlo101@chapman.edu

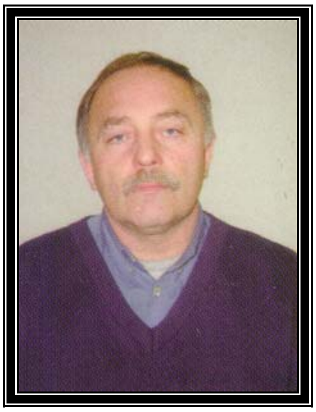
Atanas Radenski – Professor of Computer Science, Chapman University, 1 University Dr, Orange, CA 92866, USA ; radenski@computer.org; http://www.chapman.edu/~radenski

IN MEMORIAM

Russian and world computer science lost a prominent colleague:

Andrey Danilov

24th June 1949 – 7th December 2005



On the 7th December 2005 the eminent scientist in the field of computer sciences, research associate in the Department of Control and Informatics in Technical Systems in St. Petersburg State University of Aerospace Instrumentation, expert in the learning center of the Federal State Enterprise «Admiralty Shipyard», consultant in the Committee on Information, Press and Telecommunication of the Government of the Leningrad Region, member of many international projects for development and using of information technologies in the field of education and training of the staff, doctor of technical sciences, Andrey Dmitrievich Danilov pass away.

The scientific community has lost talented scientist, a grand person and friend.

Andrey Dmitrievich Danilov has published more than 60 works in the field of controlling of technical and information systems, knowledge technologies and distance learning.

A lot of these publications have been published on the pages of the International Journal «Information Theories and Applications»; he was one of the most active participants of the conferences, which has been organised by the Journal.

The Members of the Institute on Information Theories and Applications FOI ITHEA and Editorial Board of the International Journal «Information Theories and Applications» feel deep regret of bereavement of our colleague and friend Andrey Dmitrievich Danilov.

TABLE OF CONTENTS OF VOLUME 13, NUMBER 1