

**INFORMATION SCIENCE
&
COMPUTING**

International Book Series

Number 7

**Artificial Intelligence
and
Decision Making**

Supplement to
International Journal "Information Technologies and Knowledge" Volume 2 / 2008

**I T H E A
SOFIA, 2008**

Alexey Voloshin, Krassimir Markov, Krassimira Ivanova, Mykola Malyar, Iliia Mitov (ed.)

Artificial Intelligence and Decision Making

International Book Series "INFORMATION SCIENCE & COMPUTING", Number 7

Supplement to the International Journal "INFORMATION TECHNOLOGIES & KNOWLEDGE" Volume 2 / 2008

Institute of Information Theories and Applications FOI ITHEA

Sofia, Bulgaria, 2008

This issue contains a collection of papers in the field of Artificial Intelligence and Decision Making. Papers are selected from the International Conferences of the Joint International Events of Informatics "ITA 2008" - Varna, Bulgaria and Uzhgorod, Ukraine.

International Book Series "INFORMATION SCIENCE & COMPUTING", Number 7

Supplement to the International Journal "INFORMATION TECHNOLOGIES & KNOWLEDGE" Volume 2, 2008

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

- **V.M.Glushkov Institute of Cybernetics of NAS**, Ukraine,
- **T.Shevchenko National University of Kiev**, Ukraine
- **Uzhgorod National University**, Ukraine
- **Institute of Mathematics and Informatics, BAS**, Bulgaria,
- **Institute of Information Technologies, BAS**, Bulgaria.

Publisher: **Institute of Information Theories and Applications FOI ITHEA**, Sofia, 1000, P.O.B. 775, Bulgaria.

Издавател: **Институт по информационни теории и приложения ФОИ ИТЕА**, София, 1000, п.к. 775, България

www.ithea.org, www.foibg.com, e-mail: info@foibg.com

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

Printed in Sofia, Bulgaria and Uzhgorod, Ukraine

Copyright © 2008 All rights reserved

© 2008 Institute of Information Theories and Applications FOI ITHEA - Publisher

© 2008 Alexey Voloshin, Krassimir Markov, Krassimira Ivanova, Iliia Mitov – Editors

© 2008 For all authors in the issue.

ISSN 1313-0455 (printed)

ISSN 1313-048X (online)

ISSN 1313-0501 (CD/DVD)

PREFACE

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

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

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

IBS ISC rules for preparing the manuscripts are compulsory. The rules for the papers and books for IBS ISC are given on www.foibg.com/ibsisc. The camera-ready copy of the papers and books should be received by e-mail: info@foibg.com. Responsibility for papers and books published in IBS ISC belongs to authors.

The Number 7 of the IBS ISC contains collection of papers in the field of Artificial Intelligence and Decision Making. Papers are peer reviewed and are selected from the XIVth International Conference "Knowledge-Dialog-Solution", which was a part of the Joint International Events of Informatics "ITA 2008", Varna, Bulgaria and Uzhgorod, Ukraine.

ITA 2008 has been organized by

Institute of Information Theories and Applications FOI ITHEA
in collaboration with:

- ITHEA International Scientific Society
- International Journal "Information Theories and Applications"
- International Journal "Information Technologies and Knowledge"
- Association of Developers and Users of Intelligent Systems (Ukraine)
- Association for Development of the Information Society (Bulgaria)
- V.M.Glushkov Institute of Cybernetics of National Academy of Sciences of Ukraine
- Institute of Mathematics and Informatics, BAS (Bulgaria)
- Institute of Information Technologies, BAS (Bulgaria)
- Institute of Mathematics of SD RAN (Russia)
- Taras Shevchenko National University of Kiev (Ukraine)
- Universidad Politecnica de Madrid (Spain)
- BenGurion University (Israel)
- Rzeszow University of Technology (Poland)
- University of Calgary (Canada)
- University of Hasselt (Belgium)
- Kharkiv National University of Radio Electronics (Ukraine)
- Astrakhan State Technical University (Russia)
- Varna Free University "Chernorizets Hrabar" (Bulgaria)
- National Laboratory of Computer Virology, BAS (Bulgaria)
- Uzhgorod National University (Ukraine)
- Sofia University "Saint Kliment Ohridski" (Bulgaria)
- Technical University – Sofia (Bulgaria)
- New Bulgarian University (Bulgaria)

More information about ITA 2008 is given at the www.foibg.com.

The XIVth International Conference "Knowledge-Dialogue-Solution" (KDS-2008) took place in June, 21-26, 2008 in Varna, Bulgaria (First part) and in September 28 - October 04, 2008 in Uzhgorod, Ukraine (Second part). This volume includes the papers presented at the second part of the KDS-2008.

KDS-2008 continues the series of international scientific meetings, which were initiated more than sixteen years ago. It is organized owing to initiative of ADUIS - Association of Developers and Users of Intelligent Systems (Ukraine), Institute of Information Theories and Applications FOI ITHEA, (Bulgaria), and IJ ITA - International Journal on Information Theories and Applications, which have long-term experience of collaboration. Now we can affirm that the international conferences "Knowledge-Dialogue-Solution" in a great degree contributed to preservation and development of the scientific potential in the East Europe.

KDS-2008 is dedicated to 15th Anniversary of the International Journal on Information Theories and Applications.

The KDS conferences are traditionally devoted to discussion of current research and applications regarding three basic directions of intelligent systems development: knowledge processing, natural language interface, and decision making. The basic approach, which characterizes presented investigations, consists in the preferential use of logical and linguistic models. This is one of the main approaches uniting investigations in Artificial Intelligence. The main topics of KDS conferences include but are not limited to: Neural Nets, Mathematics of Computing, Decision Tools and Techniques, Decision Support, Formal Models, Expert Systems, Ontologies, Knowledge Acquisition, Natural Language Processing, Pattern Recognition, Distributed Information Processing, Philosophy and Methodology of Informatics.

The official languages of the Conference are Russian and English.

KDS 2008 Program Committee

	Victor Gladun (Ukraine)	Krassimir Markov (Bulgaria)			
	Alexey Voloshin (Ukraine)	Mikola Malyar (Ukraine)			
Adil	Timofeev (Russia)	Jury	Zaichenko (Ukraine)	Vasil	Sgurev (Bulgaria)
Alexander	Eremeev (Russia)	Krassimira	Ivanova (Bulgaria)	Vitaliy	Lozovskiy (Ukraine)
Alexander	Kleshchev (Russia)	Nadezhda	Kiselyova (Russia)	Vitaliy	Velichko (Ukraine)
Alexander	Palagin (Ukraine)	Natalia	Filatova (Russia)	Vladimir	Donskoy (Ukraine)
Alexander	Reznik (Ukraine)	Natalia	Kussul	Vladimir	Jotsov (Bulgaria)
Anatoly	Shevchenko (Ukraine)	Nikolay	Zagoruiko (Russia)	Vladimir	Khoroshevsky (Russia)
Arkady	Zakrevskij (Belarus)	Oleg	Kuznetsov (Russia)	Vladimir	Lovitskii (GB)
Boris	Fedunov (Russia)	Olga	Nevzorova (Russia)	Vladimir	Pasechnik (Ukraine)
Constantine	Gaindric (Moldova)	Rumyana	Kirkova (Bulgaria)	Vladimir	Ryazanov (Russia)
Ekaterina	Solovyova (Ukraine)	Stanislav	Wrycza (Poland)	Vladislav	Shelepov (Ukraine)
Galina	Rybina (Russia)	Tatyana	Gavrilova (Russia)	Volodimir	Donchenko (Ukraine)
Genady	Osipov (Russia)	Vadim	Stefanuk (Russia)	Xenia	Naidenova (Russia)
Iliia	Mitov (Bulgaria)	Vadim	Vagin (Russia)	Zinoviy	Rabinovich (Ukraine)
		Valery	Tarasov (Russia)		

The papers in this volume are organized in four sections: Intelligent Technologies; Decision Making; Algorithmic and Mathematical Foundations of the Artificial Intelligence; Philosophy and Methodology of Informatics.

The great success of ITHEA International Journals, International Book Series and International Conferences belongs to the whole of the ITHEA International Scientific Society. We appreciate the contribution of the members of the KDS-2008 Program Committee. We express our thanks to all authors, editors, collaborators and sponsors.

TABLE OF CONTENTS

<i>Preface</i>	3
<i>Table of Contents</i>	5
<i>Index of Authors</i>	8
Intelligent Technologies	
Hierarchical Three-level Ontology for Text Processing	
Victor Gladun, Vitalii Velychko, Leonid Svyatogor	11
Trained Neural Network Characterizing Variables for Predicting Organic Retention by Nanofiltration Membranes	
Arcadio Sotto, Ana Martinez, Angel Castellanos	19
Гиперплоскости в «множествах и расстояниях соответствия»: кластеризация	
Николай Кириченко, Владимир Донченко	25
Предварительная обработка Кирлиан изображений в диагностической системе	
Владимир Калмыков, Виталий Вишнеvский, Татьяна Романенко	37
Method of Finding Hamilton Routes in Transport Network	
Irina Garashchenko, Anatoliy Panishev	43
Analysis of Dermatoglyphic Signs for Definition Psychic Functional State of Human's Organism	
Anatoly Bykh, Elena Visotska, Vladimir Kobzev, Yury Mischeryakov, Nataly Bilous, Irena Kulya, Alexander Kobzev, Nataly Makhaikina, Irena Makartseva, Katarine Dyakonenko.....	49
Automatic ECG Analysis for Preliminary and Detailed Diagnostics Based on Scale-space Representation	
Natalie Belous, Gleb Kobzar.....	53
Отношения ассоциации для анализа предметных областей учебных курсов	
Вера Любченко, Алексей Ярмула	61
Методы интеграции учебных материалов для дистанционного обучения	
Игорь Шубин, Ярослав Святки, Ирина Белоус	65
Decision Making	
Experience of Development and Application of Expert Systems for Economic Macroparameters Forecasting and Operation	
Alexey Voloshyn, Victoria Satyr	73
Нелинейная схема компромиссов в многокритериальных задачах	
Альберт Воронин, Юрий Зиятдинов	79

Анализ свойств модели Леонтьева при нечётко заданных параметрах критериальной функции методом базисных матриц	
Владимир Кудин, Григорий Кудин, Алексей Волошин	86
О задаче нахождения строгой результирующей ранжировки в виде медианы Кемени-Снелла	
Павел Антосяк, Алексей Волошин	91
Применение нечеткой логики для задач коллективного выбора	
Николай Маляр	99
Сравнительный анализ методов оценки риска банкротства предприятий Украины	
Юрий Зайченко, Светлана Рогоза, Владимир Столбунов	103
Оператор мутации в эволюционной технологии решения задачи составления расписаний	
Елена Сипко	111
Технология нечеткого прогнозирования характеристик сложных объектов и систем	
Виталий Снитюк, Сергей Говорухин	117
Интегрированный моделирующий комплекс поддержки принятия решений в комплексной корпоративной системе управления магистральной газотранспортной сетью	
Виктор Борисенко, Юрий Пономарев, Татьяна Борисенко	123
Применение метода параметрической оптимизации для задач фокусировки пучков ионов в линейных ускорителях	
Федор Гаращенко, Игорь Харченко	131
Использование алгоритма ускоренного вероятностного моделирования в схеме табу-поиска	
Леонид Гуляницкий, Александр Турчин	137
Algorithm of Decision Support in Emergency	
Oleksandr Kuzomin, Valentina Davidenko, Ilyya Klymov	143
Modeling of Effective Process of Network Maintaining Based on Statistical Data	
Oleksandr Kuzomin, Nataliya Goloviy, Yaser Dayoub	147
Emergency Prevention and Situation Monitoring in Carpathian Region	
Oleksandr Kuzomin, Dmytro Malytskyy, Asanbek Toroev	153
Индивидуально-оптимальные равновесия в играх двух лиц	
Сергей Мащенко	157
О классификации приближенных методов комбинаторной оптимизации	
Сергей Сиренко	164
Algorithmic and Mathematical Foundations of the Artificial Intelligence	
Алгоритмы решения систем линейных диофантовых уравнений в дискретных областях	
Сергей Крытый	173

Многокритериальные задачи на комбинаторном множестве полиразмещений: структурные свойства решений	
Людмила Колечкина	180
Векторные задачи на комбинаторном множестве полиразмещений: условия оптимальности и подход к решению	
Наталия Семенова	187
Поиск общих принципов формализации в теоретических и прикладных исследованиях	
Петр Василик, Александр Провотар.....	196
Using SAT for Combinational Implementation Checking	
Liudmila Cheremisinova, Dmitry Novikov	203
Philosophy and Methodology of Informatics	
Business Informatics	
Krassimir Markov, Krassimira Ivanova, Iliia Mitov, Valeria Mitova.....	213
Basic Principles of Organization of the Medium and Thinking Process of the Human in its Conceptual Presentation	
Yuriy Belov, Zinoviy Rabinovich	219
Обязательно ли ощущения являются изоморфными «образами» мира: анализ с нейрофизиологических позиций некоторых аспектов теории отражения	
Геннадий Воронков.....	228
Гиперслучайные марковские модели	
Игорь Горбань	233
Структуры, неопределённость: математическое моделирование	
Владимир Донченко.....	243

INDEX OF AUTHORS

Natalie Belous	49, 53	Павел Антосяк	91
Yuriy Belov	219	Ирина Белоус	65
Anatoly Bykh	49	Виктор Борисенко	123
Angel Castellanos	19	Татьяна Борисенко	123
Liudmila Cheremisinova	203	Петр Василик	196
Valentina Davidenko	143	Виталий Вишневецкий	37
Yaser Dayoub	147	Алексей Волошин	86, 91
Katarine Dyakonenko	49	Альберт Воронин	79
Irina Garashchenko	43	Геннадий Воронков	228
Victor Gladun	11	Федор Гаращенко	131
Nataliya Goloviy	147	Сергей Говорухин	117
Krassimira Ivanova	213	Игорь Горбань	233
Illya Klymov	143	Леонид Гуляницкий	137
Gleb Kobzar	53	Владимир Донченко	25, 243
Alexander Kobzev	49	Юрий Зайченко	103
Vladimir Kobzev	49	Юрий Зиатдинов	79
Irena Kulya	49	Владимир Калмыков	37
Oleksandr Kuzomin	143, 147, 153	Николай Кириченко	25
Irena Makartseva	49	Людмила Колечкина	180
Nataly Makhaikina	49	Сергей Крытый	173
Dmytro Malytskyy	153	Владимир Кудин	86
Krassimir Markov	213	Григорий Кудин	86
Ana Martinez	19	Вера Любченко	61
Yury Mischeryakov	49	Николай Маляр	99
Ilia Mitov	213	Сергей Мащенко	157
Valeria Mitova	213	Юрий Пономарев	123
Dmitry Novikov	203	Александр Проватар	196
Anatoliy Panishev	43	Светлана Рогоза	103
Zinoviy Rabinovich	219	Татьяна Романенко	37
Victoria Satyr	73	Ярослав Святки	65
Arcadio Sotto	19	Наталия Семенова	187
Leonid Svyatogor	11	Елена Сипко	111
Asanbek Toroev	153	Сергей Сиренко	164
Vitalii Velychko	11	Виталий Снитюк	117
Elena Visotska	49	Владимир Столбунов	103
Alexey Voloshyn	73	Александр Турчин	137
		Игорь Харченко	131
		Игорь Шубин	65
		Алексей Ярмула	61

Intelligent Technologies

HIERARCHICAL THREE-LEVEL ONTOLOGY FOR TEXT PROCESSING

Victor Gladun, Vitalii Velychko, Leonid Svyatogor

Abstract: *The principal feature of ontology, which is developed for a text processing, is wider knowledge representation of an external world due to introduction of three-level hierarchy. It allows to improve semantic interpretation of natural language texts.*

Keywords: *ontology, text processing, thematic text analysis.*

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

Introduction

One of the practical purposes of an artificial intelligence methods use is *Text Processing* - the semantic analysis of naturally-language texts both general thematic orientations and concerning to various domains [1, 2, 3, 4]. Language thesauruses and linguistic ontologies are developed for this purpose and set the certain conceptually-expressed system of knowledge. Distinction between thesauruses and linguistic ontologies consists in volume knowledge representation and methods of classification (structurization) the conceptual environment. Among significant achievements in the field of lexicographical representation of natural languages it is possible to specify: Roget P.M. and Dornzaif F. thesaurus [5], the Ideographic dictionary of the Russian language [6], the Thesaurus of the Russian language *RuThes* [1], resource *WordNet* [7]. However, ontology is more adequate to the conceptual texts analysis, because by means of ontology connections between objects, concepts and its properties can be most full presented by all of them a variety.

As is known, ontology in a general view represents the sets of the domain terms, relations between terms and the domain interpretation functions on the terms and the relations. It is possible to point out the next advantages of ontologies: a) deep interaction between both described objects and phenomena and the contextual environment; b) the economical storage of the information demanding storing domain terms and relations, instead of stages memorizing; c) universal character of ontology, which allows within its structure to solve both problem – knowledge synthesis and analysis; d) "flexibility" of the knowledge structure which is adjusted on diverse domains.

Ontologies are developing, more often, for concrete domains ("narrow ontologies"), at that process begins with the analysis of a text collection [1, 2]. However, in a context of the given work the principal interest have linguistic ontologies ("wide ontologies"), which cover the knowledge continuum in different spheres of human activity, so as *Mikrokosmos* [8], *SUMO* [9] and *Knowledge Representation* by J. Sowa [10].

Two approaches to creation of ontologies. Topicality of an compromise

Now two opposite approaches to construction of the linguistic ontologies were formed. First of them – "*search ontologies*" – is purposeful on problems of automatic text processing in various domains. For example, Sociopolitical thesaurus is focused on the social and political life and used in such applications of automatic text processing as conceptual indexing, automatic text categorization [1]. The ontology synthesis technology is based on the analysis of the representative text collection (sometimes – tens thousands sources). Sizeable difficulties are in the fact that is impossible "to fish out" the *context knowledge* out of professional text collections, what is quite necessary to texts understanding. Therefore, it requires substantial experts' efforts for completion of the ontology by complementary concepts and to organisation semantic relationships between domain and external

world. Moreover, in many cases the semantic communications do not lend its to formalization. As a result, it is impossible to avoid of essential intervention of expert. The final structure becomes too bulky in use and difficult for tuning.

On the other hand, there is an opposite approach. As a starting point of construction "abstract ontologies" consider *Universe* [10], *Essence* [9], *All* [8]. The knowledge representation is set by the branched out hierarchical structure. Concepts ontology reaches the maximal generalization and abstraction, and on this height, more often is remaining. By virtue of it, application of universal ontology to concrete texts analysis seems to be very problematic. Besides that, the formal relations used in abstract ontologies far not always describe the properties of real world in its reflection by man. Therefore, some authors introduce into consideration "flexible" relations: "conceptual dependence relations" [11], "role relations" [1], "symmetric and asymmetrical associations" [2] or "associative relations" [12]. Using of similar relations to thematic analysis NL text seems very constructive.

Thus, in a context of the lead analysis, the problem consists in separation between professional-focused and abstract ontologies, i.e. – between a particular description and abstract presentation of knowledge. Resources, which are available in a free access, do not allow to realize thematic text analysis in Russian. The way out from this opposition consist, as it seems, in synthesis "wide" and "narrow" ontologies. It is necessary to create an integrated structure, where it can be distributed and balanced described both the general, meaningful knowledge about nature and society, and concrete domain resources. Such structure can reflect a hierarchical picture of the whole world.

The purpose of the given work consists in offering the concept of integrated hierarchical (three-level) representation model of environment, which: a) in the compressed kind and with a different degree of generalization (or detailed elaboration) reflects actual knowledge about structure of an external world; b) is focused on text processing both the general subjects and separate areas of knowledge and c) allows to integrate a professional knowledge into a conceptual network without reorganization of the upper and middle ontology levels.

This conception develops the *semantic (thematic) analysis method* of NL text processing by creation of the document synopsis. Procedure of making a synopsis is based on disclosing of the given theme by means of sequence of keys words, which are automatically generated with the program "KONSPEKT" [13]. However, defect of this method was in strongly simplified one-level model of the real world representation, which was not taken into account hierarchy and depth of external world.

Substantiation of the approach to construction of hierarchical ontology

For construction of multilevel ontology, the methods and mathematical models that contain models of ontology, knowledge and domain are used [4]. Offered here three-level associations ontology is intended for the decision of more specific problem - the thematic texts analysis. It defines a number of preconditions and features.

1. Gnosiological conceptions in ontologies. Paradigm acceptance

There are many approaching methods to problem of universe representation in philosophy, natural sciences and linguistics. The authors of different universal ontologies describe a world with such general categories as: *Essence = Material, Abstract* [3, 9]; *All = Object, Event, Property* [3, 8]; *Universum* (is divided on seven components) [3, 10]. That is not exclude and others, exotic variants: for example, classification of *All* into *Goodness* and *Evilness* may be successful... The question of substantiation usually is not considered.

However, more practical and pragmatic methods recommend the scientific methodology of the system analysis. It operates the following types of resources: *Substance; Energy; Information; Man; Organisation; Space; Time* [14].

These categories, in our opinion, possess a necessary diversity and they are objects of researches in physical and social sciences. This world outlook may be taken as a base of description of the external world.

However, the most modern is the materialistic idea proposed by academician V.I. Vernadsky. In accordance with it, all, what is just known, may be divided on two fundamental categories: *Inert substance* and *Alive substance*. The *Alive substance* is realized in *Biosphere* and *Noosphere* - sphere of Human activity. Both categories characterize the *Matter* fundamentally. This materialistic paradigm is put in a basis of offered ontology. It works out in detail in partition "Choice of categories for upper level of ontology".

2. A choice of the general structure

In correspondence with hierarchical picture of the knowledge about the world, we are distinguishing three levels in ontology construction. On the upper level are the general categories of universe; here the strict taxonomy is possible. Upper level summarizes the concepts of general knowledge and reflects the ontological basis.

The middle level is disclosing the base terms in more detail - by concepts having lesser level of commonality. Concepts, used on middle level, reflect the universal, popular and well-known terms existing in nature, society and environment; there are presented relations between them also.

At last, on the lower level is concentrated knowledge, which characterizes concrete situations and describe some environment. At this level is presented knowledge of the problem-focused area. Therefore, on this level two intersected blocks exist: one of them contains the concepts and words of common usage, which exist in general, and interdisciplinary texts; other block serves the domains. Due to this two-blocks structure the ontology may grow at the expense of new domains. In principle, the domain block may be empty - the working capacity of ontology is remained fully.

Corresponding to hierarchical structure of ontology we will use the term **HiO**.

3. A choice of connections

At upper level **HiO** the formal connection of type «the whole - a part» is applied. At middle and upper levels, except the formal connections, following specific types of connections are entered into consideration:

- a) **A** reveals through **B**; **B** explains **A**;
- b) **A** is characterized by property **E**;
- c) **A** is *associatively* connected with **C** .

Widely used in **HiO** the term "associative connection" is not formal. It is necessary to reflect individual semantic correlation of two (any) concepts if it takes place. Associations have a situational and dynamical character. At an *information level* they can be able to open unexpected properties (or laws) of some object (or phenomena). At a *functional level*, they fix some dependence. At a *cognitive level* associative connection of two concepts means that one image has excited another. At a *logic level* association are predicative and implicative, but in most cases – not transitive.

Constructive properties of associations ontology

Based on a practical orientation of the text processing, we shall specify some functional **HiO** features.

1. **Completeness and taxonomy** on upper level of the ontology signify that the chosen categories in aggregate are representing the Matter in the exhaustive way. Outside these categories, there should be no manifestation of the reality. Categories are subordinated to strict hierarchy and classification that excludes logic ambiguity of concepts.

2. **Natural-scientific lexicon**. The ontology categories and concepts should be common, simple and clear. They lexically expressed by those concepts and terms which were established in sphere of the general

knowledge, in natural and social sciences, in the socio-cultural environment. The upper level **HiO** can be supplemented with a special terms.

3. **Connectivity on association.** Connections between concepts inside of levels and between ones include both formal and informal (associative) connections. Associative connections are actively used at middle and upper levels of **HiO** to fuller description of theme.

4. **Antagonism reflection.** Concepts, reflecting properties, which have (inside of the given measure) its contrast, can be designated by pair's words (antonyms).

Synthesis of hierarchical associations ontology

On the base of the formulated preconditions there is clear the following prospect of actions. It is necessary to choose categories, concepts and connections between them and distribute on the upper, middle and lower levels of hierarchical associations ontology. Thus on upper level the simplified strict model of the world is presented. On the middle level, it is necessary to disclose the categories of upper level, using wide concepts of interdisciplinary dialogue. On the lower level, the conceptual basis of middle level is to be described in detail. In addition, here professional domains knowledge is localized. Received three-levels network ontology should be later connected to linguistic database and integrated in the working system of NL texts analysis. On the Fig. 1 general block diagram of three-level hierarchical associations ontology is presented.

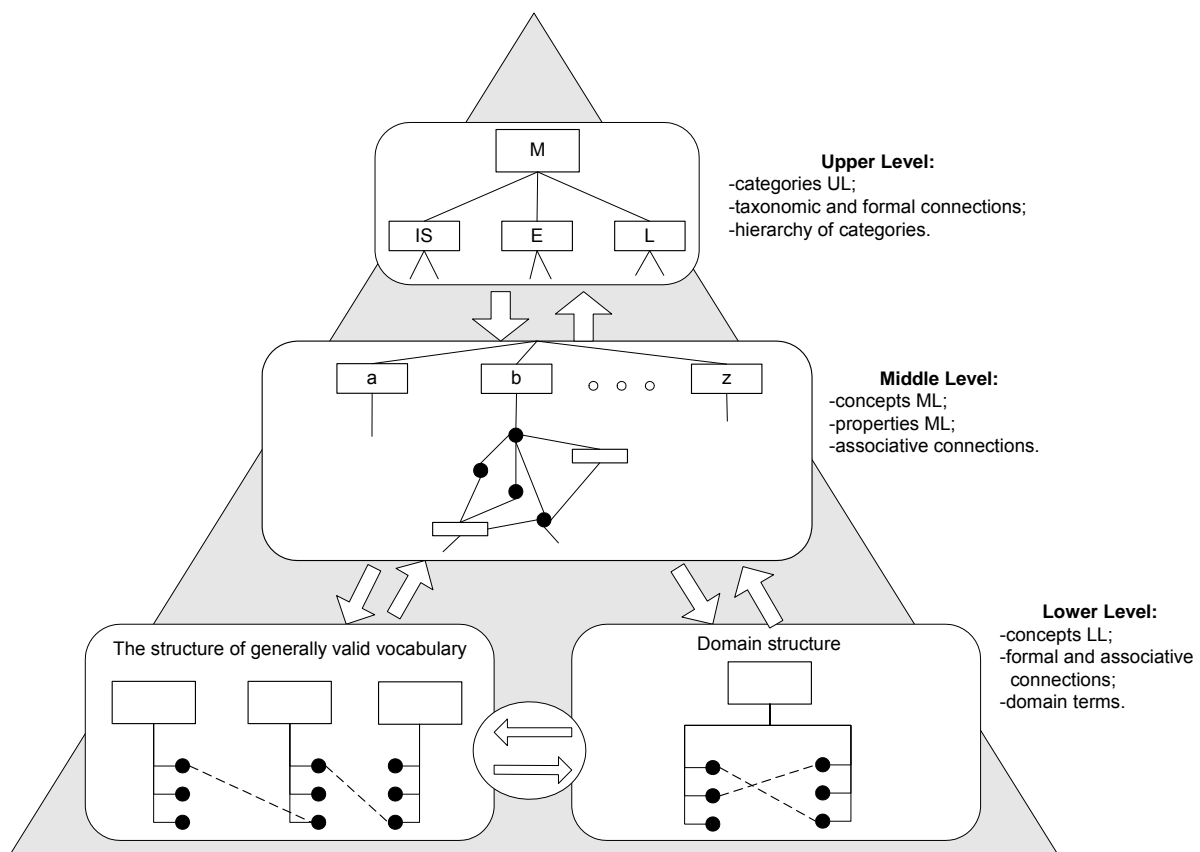


Fig1. The block diagram of three-level associations ontology.

Choice of categories for upper level of ontology

As it was specified above, within the limits of offered ontology the general picture of a Universe is subordinated to materialistic idea by V.I. Vernadsky. Apex of hierarchy is the philosophical category the *Matter*. It shows itself in

exhaustive manner as *Inert* and *Alive substance*. On the other hand, the Inert substance may become the forms either *Substance* or *Energy* (which are passing each another under the Preservation Law). Therefore, we realize trichotomy of the *Matter* on: *Substance* (inert), *Energy* and *Life* (the substance of *Alive*). In this case dividing is made on base "The Form existence of Matter". Each of three categories is presented by a number of subcategories, as is shown on Fig.2. It is necessary to add some explanations to the resulted ontological structure of upper level.

The first. At a pyramid of upper level, there are no such general categories of knowledge, as, for example: the Being, the Consciousness, the Measure, the State, the Property, the Quantity, the Quality and others. In **HiO** some of them are transferred on the middle level, owing to what they are released from philosophical sense and "work" as a terms of natural sciences.

The second. It is possible to show, that **HiO** upper level possesses property of completeness of conceptual volume. Really: a) *Substance, Energy* and *Life* are the forms of the *Matter* realization; b) the *Space* and *Time* serve as the forms of the *Matter* distribution and c) the *Reason* is the way of the *Matter* reflection. Summary these three metacategories are exhausting a metacategory "Being of the Matter". If to accept the given statement for an axiom, **HiO** covers all known (or real) properties of the Matter. As a result, the given categorical system on upper level is complete and closed.

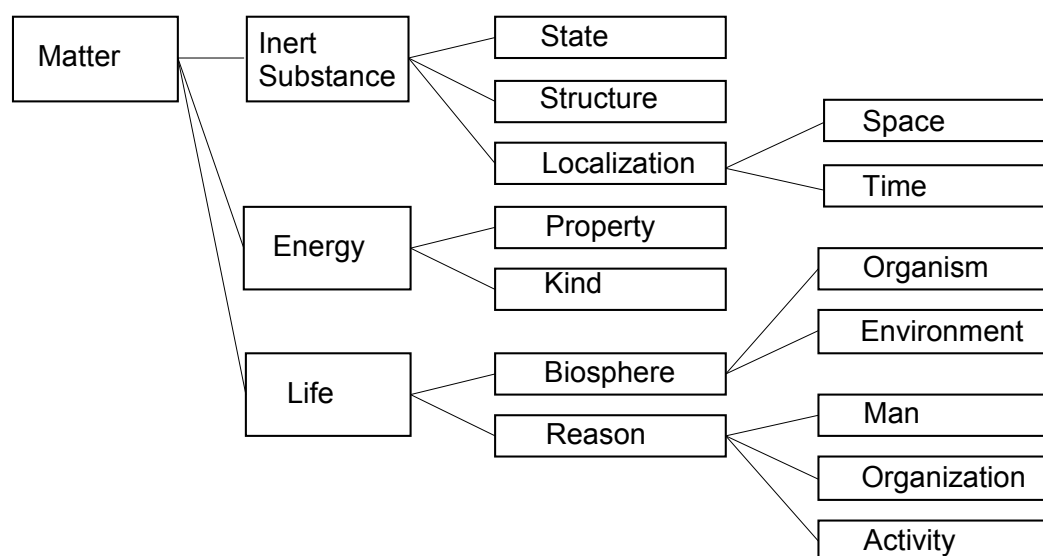


Fig. 2. Structure of top level ontology

In view of proposed above universal categories, concepts and comments the continuation of the ontological scheme may be formed.

The principles building the middle level of ontology

Purpose of the middle level of ontology (MLO) is, on the one hand, to disclose all categories of upper level - to give them semantic filling, and with another – to form a semantic environment for the concordance with the lower level's concepts.

MLO represents such level of knowledge, which is common to a various areas; that is interdisciplinary knowledge. Per se, the middle level of hierarchy fixes itself a layer of valid human knowledge, which is generalized by collective experience in science, culture, practice – out of professional sphere. It operates by generally accepted

words. A material to this level is formed by the knowledge engineer. The middle level is "conservative": it is a "constant" **HiO**s component. At the given level informal (associative) connections of type "object – property", which (in opinion of the expert) bear the helpful information for disclosing internal structure of ontology, are actively used. It is necessary to emphasize, that occurrence doubtful connections is not lack of associations ontology, quite the contrary – they open an opportunity to additional adjustment.

Brief description MLO. Middle level of ontology represents set of network structures: as a name (and initial node) of each structure serves a category of upper level; internal nodes are the concepts of the middle level; internal connections between concepts disclose the important characteristic properties of category to be done.

Here, with a view of place economy, concepts and the full structures making middle level of **HiO** are not presented. However, as an example, the structure of the *Organization* category from the *Reason* cluster (see Fig. 3) is shown.

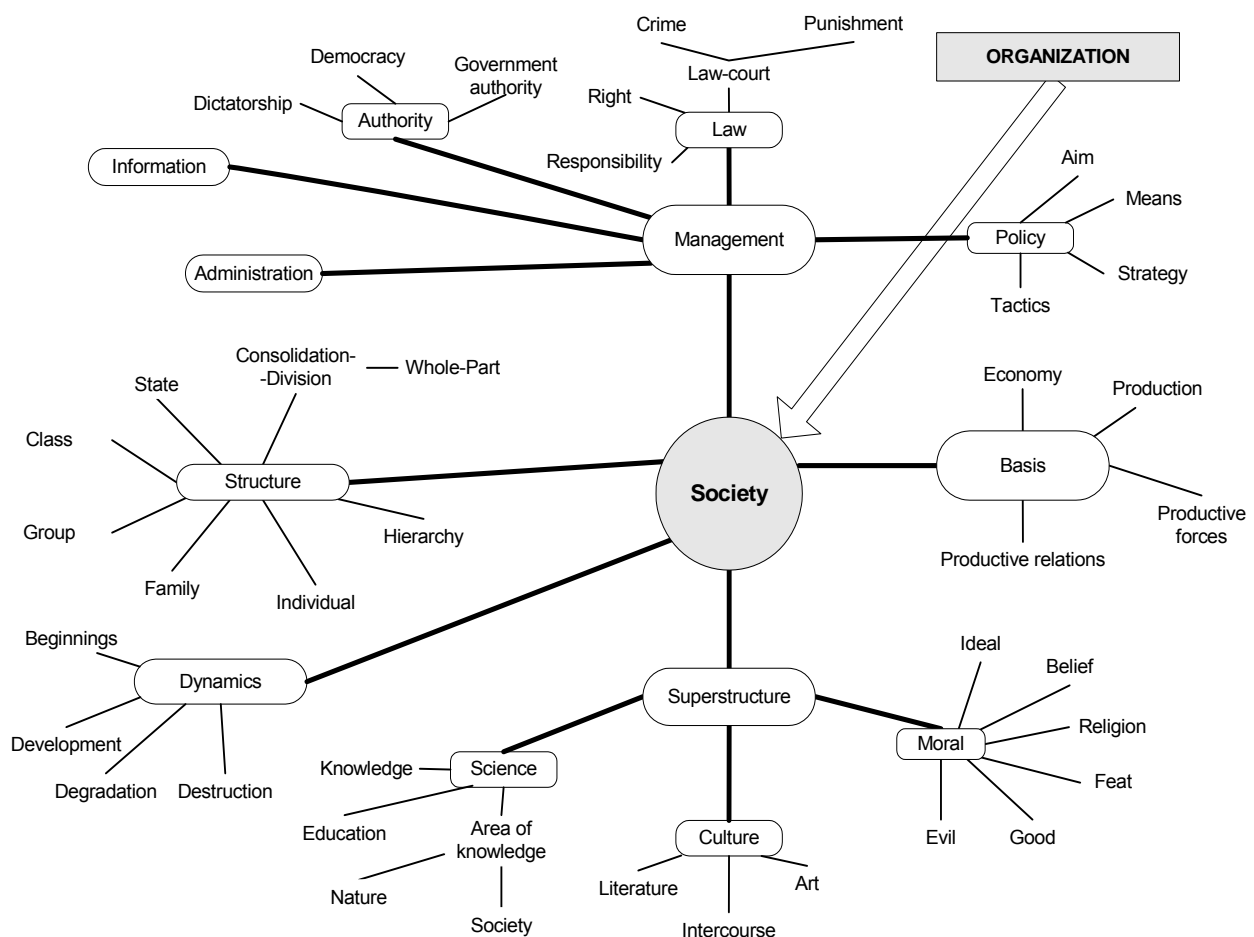


Fig. 3. A fragment of associations ontology structure of a middle level for a category "Organization".

The note. The interpreting opportunities of semantic analysis system, as a whole, essentially depend on successful construction of a middle part of the ontological structure. On the other hand, due to interaction of the middle level with upper one, text interpretation becomes more deeply and full: the explanatory resource **HiO** is used. If a new domain will be added, the middle level can be corrected.

The lower level of ontology

The **HiO** lower level is produced to bind together text keywords to concepts of the middle level of ontology: here the concepts of the middle level can be determined by means of text words. On this level, two blocks of the interconnected concepts are stipulated.

One of them – *the interdisciplinary block* – is intended for text processing of the general-thematic discourses. Other – *the domain block* – is synthesized under concrete domain. These two blocks are connected mutually and correlate to concepts of middle level too. When on an input of ontology appears some text interdisciplinary block always works; the domain block becomes more active on the professional text.

The choice of concepts field and connections between them for the lower level makes by engineer on knowledge and domain expert together. There are technologies to process linguistic resources, for example, the software complex "RuThes" and other [1, 2, 4] which allow to synthesize domain ontology. However, if to be oriented on more simple problem – the localization of the document theme – much more simple procedures can be offered.

One of them consists in indexing: the words and terms of common and professional glossaries (which are represented in database) can be connected with concepts of lower level of ontology. The other way is proposed by T.Taran: some situation or scene is determined by concept lattices [15].

After construction the lower level of ontology, **HiO** synthesis comes to an end. However, received ontology represents only the theoretic-descriptive scheme, instead of the analysis system. For a text processing, it is necessary to connect ontology with a special dictionary of natural language, which is contained in ones memory – database.

Text-processing procedure

The binding together of the text and the ontology is made through the dictionary of Russian. The dictionary reflects lexicon of a natural language. When the real NL text is analysed, first of all the set of keywords is discover. These keywords should make active some elements of the lower level of ontology. Which elements will be made active, will specify the list of the indexes, appropriated to dictionary elements. This list in an obvious kind sets associative connections between the given elements and concepts of the lower level. Hence, the concrete text word through the dictionary stimulates a subset of concepts and connections of the lower level, and through these concepts the signal transfers to middle and upper levels of **HiO**. As a result, in all ontological network automatically some semantic trajectory of an entrance word is localised.

This trajectory is possible to use in the text-processing system, namely: a) for deeper interpretation of the text, or b) – as an initial material for repeated, purposeful disclosing a theme in enriched context.

Conclusion

As a result of researches is developed hierarchical structure of three-level associations ontology, which differs by the following:

- unites in uniform structure the general categories of the description of the world (on upper level) with the conceptual environment of interdisciplinary knowledge (on a middle level) and with the topical concepts at the lower level. Ontology supposes inclusion of the new blocks – models of domains – without alteration of upper level and with expansion of a middle level of a network;
- the network model of associations ontology is the simple and constructive scheme, which allows to trace in the text the theme that was given. Synthesis HiO practically excludes greater expenditures of labour on viewing of the texts collection, because a priori is based on known natural-scientific knowledge;

– the HiO serves as a construct with well-founded basis of scientific general categories. At a level metaontology the bases of categories dichotomy are precisely well founded and their conceptual completeness is proved. For a semantic description of environment, informal (associative) connections are widely used. The semantic trajectories of the conceptual analysis, received as a result the text-processing, help to interpret any theme in a context of universal human knowledge.

Based on hierarchical three-level associations ontology the new version of "Konspekt" system is developing.

The literature

1. Dobrov B.V., Loukachevitch N.V. Linguistic ontology of natural sciences and technologies for applications in sphere of information search. Scientific notes of the Kazan state university. Series Physical and mathematical sciences. – 2007. Volume 149, book 2. pp. 49-73. In Russian.
2. Loukachevitch N.V., Nevzorova O.A. Aviaontology-2004: the analysis of a modern state of a resource // Computer linguistics and intellectual technologies: Works of the International seminar Dialogue'2004 (June, 2-7nd 2004r.) // Edited by I.M.Kobozeva, A.S.Narinjani, V.P.Selejev. – M.: Nauka – 2004. – Vol.2 - pp. 424-430. In Russian.
3. Palagin O.V., Petrenko M.G. Construction of abstract model of language- ontological information system. // Mathematical machines and systems. - 2007. - №1. - pp.42-50. In Ukrainian.
4. Artemieva I.L. Multilevel ontologies for domains with complicated structures. // Proceedings of the XIII-th International Conference "Knowledge-Dialogue-Solution" – Varna, 2007 Volume 2 Sofia, Institute of Information Theories and Applications FOI ITHEA, Bulgaria– 2007 pp. 403-410
5. Morkovkin V.V. Ideographic dictionaries. M.: Mosc. st. uni-ty, 1970. – 71 p. In Russian.
6. Baranov O.S. The Ideographic dictionary of Russian, M. – 2002, – 1200 p. In Russian.
7. Miller G., Beckwith R., Fellbaum C., Gross D. and Miller K. Five papers on WordNet. - CSL Report 43. Cognitive Science Laboratory, Princeton University, 1990.
8. Mahesh, Kavi. Ontology development for machine translation: ideology and methodology. Memoranda in Computer and Cognitive Science MCCS–96–292, New Mexico State University. 1996.
9. SUO, (2001), The IEEE Standard Upper Ontology web site, <http://suo.ieee.org>.
10. Sowa, John F. Knowledge Representation: Logical, Philosophical, and Computational Foundations, Brooks Cole Publishing Co., Pacific Grove, CA, © 2000. - 594pp.
11. Guarino N. Formal Ontology and Information Systems. In N. Guarino (ed.) Formal Ontology and Information Systems. //Proceedings of FOIS'98. - Trento, Italy. - 1998. - 6-8 June. – IOS Press, Amsterdam. – pp.3-15.
12. Grishina O.V. Information search in Internet // Intellectual technologies and systems. Collected articles of post-graduate and students. Issue 2 // Edited by Yu.N.Filippovich. – M.: Publishing house MGUP, – 1999. – pp. 18-24. In Russian.
13. Gladun V.P., Velychko V. Yu., Svyatogor L.A. Thematic analysis of natural language texts. // Computer linguistics and intellectual technologies: Works of International Conference "Dialogue 2006" (Bekasovo, 31 May – 4 June, 2006) // Edited by A.S.Narinjani, V.P.Selejev. – M.: RSUH Publishing Centre. – 2006. pp. 115-118. In Russian.
14. Kaziev V.M. Introduction in the analysis, synthesis and modelling of systems. – INTUIT.RU, Binom. Laboratory of knowledge. – 2006. – 248p. In Russian.
15. Taran T.A., Zubov O.A. Artificial intelligence: the theory and applications. Lugansk. – 2006. – 240 p. In Russian.

Authors' Information

Victor Gladun, Vitalii Velychko, Leonid Svyatogor – V.M.Glushkov Institute of cybernetics of NAS of Ukraine, Prospekt akad. Glushkova 40, 03680 Kiev, Ukraine; e-mail: aduis@rambler.ru

TRAINED NEURAL NETWORK CHARACTERIZING VARIABLES FOR PREDICTING ORGANIC RETENTION BY NANOFILTRATION MEMBRANES

Arcadio Sotto, Ana Martinez, Angel Castellanos

Abstract: *Many organic compounds cause an irreversible damage to human health and the ecosystem and are present in water resources. Among these hazard substances, phenolic compounds play an important role on the actual contamination. Utilization of membrane technology is increasing exponentially in drinking water production and waste water treatment. The removal of organic compounds by nanofiltration membranes is characterized not only by molecular sieving effects but also by membrane-solute interactions. Influence of the sieving parameters (molecular weight and molecular diameter) and the physicochemical interactions (dissociation constant and molecular hydrophobicity) on the membrane rejection of the organic solutes were studied. The molecular hydrophobicity is expressed as logarithm of octanol-water partition coefficient. This paper proposes a method used that can be used for symbolic knowledge extraction from a trained neural network, once they have been trained with the desired performance and is based on detect the more important variables in problems where exist multicolinearity among the input variables.*

Keywords: *Neural Networks, Radial Basis Functions, Nanofiltration; Membranes; Retention.*

ACM Classification Keywords: *K.3.2 Learning (Knowledge acquisition)*

Introduction

Phenolic compounds are commonly used as raw materials in the manufacture of polymers, plasticizers, hydraulic fluids and various industrial chemicals. Therefore, there are many wastewater effluents contain discharges amounts of these recalcitrant organic compounds.

Nanofiltration (NF) is a viable treatment for the removal of dissolved organic pollutants for production of drinking water and as combined method with advanced and traditional water treatment process [Van der Bruggen et al 2003] [Hellebrand et al 1997]. Many reported studies indicate that several physical phenomena can play a role in the solute transport through nanofiltration membranes: solution-diffusion, convection (sieving), electrostatic (charge) repulsion and dielectric exclusion. In addition, NF strongly depends on the feed water composition, membrane and solute properties, and operational conditions [Bellona et al 2004]. Therefore retention of organic compounds is influenced either by pore size and charge of membrane, or by the molecular size, hydrophobicity and ionization constant of solutes [Boussu K. et al 2008] [Arsuaga et al 2008].

Neural network is proposed as suitable tool to prediction the membrane performance on solute retention and detect the more important variables when the input variables exist high correlation. Artificial Neural networks perform adaptative learning. This advantage can be used to improve the knowledge acquisition in knowledge engineering. This paper proposes extracting knowledge from a neural network that has learned using sensitivity analysis used to determinate which are the most important variables for the prediction. These will guide the process of create one model for prediction with a few variables, at least the most important variables.

Characteristics about the Forecast Method

Neural networks [Anderson, James A. and Edward Rosenfield., 1988] are non-linear systems whose structure is based on principles observed in biological neuronal systems [Hanson, Stephen J. and David J. Burr. 1990]. A neural network could be seen as a system that can be able to answer a query or give an output as answer to a

specific input. The in/out combination, i.e. the transfer function of the network is not programmed, but obtained through a training process on empiric datasets. In practice the network learns the function that links input together with output by processing correct input/output couples. Actually, for each given input, within the learning process, the network gives a certain output that is not exactly the desired output, so the training algorithm modifies some parameters of the network in the desired direction. Hence, every time an example is input, the algorithm adjusts its network parameters to the optimal values for the given solution: in this way the algorithm tries to reach the best solution for all the examples. These parameters we are speaking about are essentially the weights or linking factors between each neuron that forms our network.

There is a great number of Neural Networks [Anderson, James A. 1995] which are substantially distinguished by: type of use, learning model (supervised/non-supervised), learning algorithm, architecture, etc. Multilayer perceptrons (MLPs) are layered feed forward networks typically trained with static backpropagation. These networks have found their way into countless applications requiring static pattern classification. Their main advantage is that they are easy to use, and that they can approximate any input-output map. In principle, backpropagation provides a way to train networks with any number of hidden units arranged in any number of layers.

The research community has developed several different neural network models, such as, radial basis function, growing cell structures and self-organizing feature maps. A common characteristic of the mentioned models is that they distinguish between learning and a performance phase. Neural networks with radial basis functions have proven to be an excellent tool in approximation with few patterns. Most relevant research in theory, design and applications of radial basis function neural networks is due to Moody and Darken [Moody and Darken, 1989]. Radial basis function (RBF) neural networks provide a powerful alternative to multilayer perceptron (MLP) neural networks to approximate or to classify a pattern set. RBFs differ from MLPs in that the overall input-output map is constructed from local contributions of Gaussian axons, require fewer training samples and train faster than MLP. The most widely used method to estimate centers and widths consist on using an unsupervised technique called the k-nearest neighbour rule. The centers of the clusters give the centers of the RBFs and the distance between the clusters provides the width of the Gaussians.

The object of the present study is to ascertain whether the membrane type NF90 has a quantitative effect on the values of the retention for different components analyzed, and affects the relationships between the different variables considered as input in the model propose for prediction retention. Retention behavior of the phenolic compounds by NF90 membrane was investigated in order to clarify the influence of the molecular weight (MW), size (diameter), acid dissociation constant (pK_a) and molecular hydrophobicity ($\log P$) of selected compounds on membrane performance. This paper proposes a method in order to detect the importance of the input variables. In multivariate analysis problems, when there exists correlation among different variables of forecasting, the importance and the sequence when adding variables in the model, can be detected from the knowledge stored in NN, and must be taken into account when the study of the correlations detect relationships among a set of variables

Neural networks can predict any continuous relationship between inputs and the target; artificial neural networks develop a gain term that allows prediction of target variables for a given set of input variables., we use neural networks models with analysis of sensibility, this model predict more accurately the relationship existing between variables, and is a suitable way to find the individual effects of forecasting variables over the variable to forecast , and the way to find a set of forecasting variables to include in the new model.

The addition of a given variable into a forecasting model does not implies that this variable will have an important effect over the response of the model, that is, if a researcher identifies a set of forecasting variables, he must check if they really affect the response. A frequent problem is that some of the forecasting variables are

correlated. If the correlation is small, then consequences will be less important. However, if there is a high correlation between two or more forecasting variables, then the model results will be ambiguous but not for obtain a bad prediction, the problem is the high correlation between variables (high linear association) decrease in a drastic way the individual effect over the response for each correlation variable and sometimes is difficult to detect and is not possible measure the real effect for each variable over the output.

The process of finding relevant data components is based on the concept of sensitivity analysis applied to trained neural networks. Two ANN models predict changes for certain combinations of input variables, detecting the most important influence in the output variable. We have studied different analysis for detecting relationships between molecular diameter, molecular weight, $\log P$ and pK_a in the two membranes during the process of nanofiltration. Retention organics compounds by correlated with characteristics of membrane and also with physico-chemical properties of organic solutes. In order to study the relationships between different variables it has been used neural networks models with a single hidden layer and Tanh as transfer function in both cases. One ANN model uses MLP (multilayer perceptron) and the other ANN model uses a normal radial basis function (RBF) for model development.

Two ANNs models have been implemented with four input neurons: molecular weight, molecular diameter, pK_a and $\log P$ to estimate the membrane solute retention. The MLP network uses a sigmoid activation function with a single hidden layer with four neurons. The general form of a feed-forward neural network expresses a transformation of the expected target as a linear combination of non-linear functions of linear combinations of the inputs. A normalized radial basis function (RBF) network is a feed-forward network with a single hidden layer using in this case, the same function sigmoid (Tanh), in the hidden layer with 15 clusters and one output layer. In contrast to MLP, each basis function is the ratio of a bell-shaped Gaussian surface. For all the learning process has been performed with the momentum algorithm. Unsupervised learning stage is based on 100 epochs and the supervised learning control uses as maximum epoch 10000, and threshold 0.001. We have performed an initial study using 17 patterns in training set.

Materials and methods

Seventeen phenolic compounds were selected to carry out membrane retention experiments. Table 1 summarizes the most important properties of selected compounds.

Thin-film composite polyamide membrane, NF90 supplied by Dow/Filmtec was evaluated in this study. It's classified as nanofiltration membrane. According to the manufacturers, NF90 membrane is polyamide thin-film composite with a microporous polysulfone supporting layer. A cross flow system (SEPA CF II, Osmonics) was used for membrane retention measurements. Organic solution concentrations were fixed at 100 mg L^{-1} and system temperature was maintained constant in all experiments at 25°C . It was controlled by circulating feed water through a stainless-steel coil immersed in the thermostatic bath. Quantitative analysis of the organic compounds was carried out by means of their respective absorptions in the ultraviolet region, using a Varian Cary 500 Scan UV-VIS-NIR spectrophotometer. Concentration of PEGs and saccharides were measured with a Total Organic Carbon (TOC) analyzer (model TOC-V CSN Shimadzu). Regression factor (R^2) obtained for calibrations within the range of experimental concentration used was greater than 0.99.

Retention R (%) of a solute was calculated using the expression:

$$R = 1 - \frac{C_p}{C_r} \times 100\% \quad (1)$$

where C_p and C_r are the concentrations for the permeate and retentate, respectively.

Compound	Formula	Molecular diameter (nm)	Molecular Weight (g \cdot mol $^{-1}$)	pK $_a$	logP
Phenol	C $_6$ H $_6$ O	0.1945	94.11	9.86	1.48
Resorcinol	C $_6$ H $_6$ O $_2$	0.1948	110.11	9.45	0.76
Hydroquinone	C $_6$ H $_6$ O $_2$	0.1908	110.11	10.33	0.66
Cathecol	C $_6$ H $_6$ O $_2$	0.2160	110.11	9.5	0.88
3-Nitrophenol	C $_6$ H $_5$ NO $_3$	0.2142	139.11	8.33	1.93
3-Chlorophenol	C $_6$ H $_5$ ClO	0.2134	128.56	9.00	2.40
2-Chlorophenol	C $_6$ H $_5$ ClO	0.2157	128.56	8.5	2.04
2-Nitrophenol	C $_6$ H $_5$ NO $_3$	0.2112	139.11	7.14	1.71
4-Chlorophenol	C $_6$ H $_5$ ClO	0.1915	128.56	9.47	2.43
4-Nitrophenol	C $_6$ H $_5$ NO $_3$	0.1849	139.11	7.23	1.57
Pirogallol	C $_6$ H $_6$ O $_3$	0.2154	126.11	9.12	0.29
Phloroglucinol	C $_6$ H $_6$ O $_3$	0.2331	126.11	7.97	0.06
Oxalic acid	C $_2$ H $_2$ O $_4$	0.1148	90.04	1.38	-0.24
Maleic acid	C $_4$ H $_4$ O $_4$	0.1291	116.07	3.15	0.04
Malonic acid	C $_3$ H $_4$ O $_4$	0.1378	104.06	2.92	-0.31
Acetic acid	C $_2$ H $_4$ O $_2$	0.1218	60.05	4.79	-0.17
Formic acid	CH $_2$ O $_2$	0.1335	46.03	3.74	-0.37
Ribose	C $_5$ H $_{10}$ O $_5$	0.20856	150.13	12.46	-2.39
Glucose	C $_6$ H $_{12}$ O $_6$	0.28356	180.16	12.45	-3.169
Sucrose	C $_{12}$ H $_{22}$ O $_{11}$	0.38956	342.3	12.81	-3.484
Raffinose	C $_{18}$ H $_{32}$ O $_{16}$	0.50256	504.42	12.81	-6.76

The solute permeation (B) was calculated from retention values and defined as follows:

$$B = \frac{1 - R}{R}$$

Results and Conclusions: determining the important inputs for the model

This example is based on detect the more important variables when exist multicollineality.

Multilayer feedforward networks are often used for modeling complex relationships between the data sets. Deleting unimportant data components in the training sets could lead to smaller networks and reduced-size data vectors. The process of finding relevant data components is based on the concept of sensitivity analysis applied to a trained neural network. ANN models predict changes for certain combinations of input variables, detecting the most important influence in the output variables.

After work with both neural network MLP and RBNF, in both case the variable Mw is the less signification above the model which propose for prediction of the retention B , and is consequence of high correlation between Diameter and MW .If we are looking for a model for prediction the retention of the membrane, the most important is the variable diameter as the first to include in the model forward the variable logP .

Analysis of the results obtained about the weight importance in percent is listed in the tables. . MLP results are in Table 2, and in table 3 have been shown the RBF results.

Table 2 Multilayer Perceptron results (MLP)

Sensitivity of criterion %			
Variables			
MW	Diameter	pK_a	$\log P$
10.182	42.634	18.170	29.014
16.357	42.664	–	38.979
–	50.497	17.565	31.938
Diameter		$\log P$	
51.689		48.311	

Table 3 Radial Basis Function results (RBF)

Sensitivity of criterion %			
Variables			
MW	Diameter	pK_a	$\log P$
11.709	34.042	13.720	40.529
18.939	56.182	–	24.879
–	36.545	18.174	45.281
Diameter		$\log P$	
65.599		34.401	

Tables 2 and 3 show how there is correspondence between the importance of the variables in percentage and the membrane retention for both variable and trained neural network.

It can be seen that, from tables 2 and 3, how the most important variable in percent % is the diameter followed by the $\log P$. The pK_a is not very important and finally the MW has no influence, but this last variable is correlated with the diameter and in some type of membrane it is possible obtain confuse measure about the importance over the output. It can be also seeing how the diameter is the most important variable through the different possible combinations of models, and error is softly decreasing.

The General performance probe displays the Mean Squared Error (MSE), the Normalized Mean Squared Error (NMSE), the Correlation Coefficient (r), and the Percent Error.

Table 3 General performance probe

MLP

All variables	without MW	without pK_a	without MW and pK_a
<pre>Active Performance MSE 0.001999154156 NMSE 0.005930583475 r 0.997030854625 % Error 19.890180033978 AIC -55.655528865992 MDL -70.240362065290</pre>	<pre>Active Performance MSE 0.002737718553 NMSE 0.008121568994 r 0.995933106324 % Error 19.296495772173 AIC -58.310715956590 MDL -70.561975844000</pre>	<pre>Active Performance MSE 0.001998634538 NMSE 0.005929042005 r 0.997037338529 % Error 21.054691900583 AIC -63.659948060254 MDL -75.911207947664</pre>	<pre>Active Performance MSE 0.002875232110 NMSE 0.008529509336 r 0.995727887962 % Error 19.109421056453 AIC -65.477571859592 MDL -75.395258435114</pre>

RBF

All variables	without MW	without pK_a	without MW and pK_a																																																
<table border="1"> <tr><td>MSE</td><td>0.001998103640</td></tr> <tr><td>NMSE</td><td>0.005927467069</td></tr> <tr><td>r</td><td>0.997037970864</td></tr> <tr><td>% Error</td><td>15.808930683890</td></tr> <tr><td>AIC</td><td>136.335535616454</td></tr> <tr><td>MDL</td><td>65.744942931855</td></tr> </table>	MSE	0.001998103640	NMSE	0.005927467069	r	0.997037970864	% Error	15.808930683890	AIC	136.335535616454	MDL	65.744942931855	<table border="1"> <tr><td>MSE</td><td>0.005727543189</td></tr> <tr><td>NMSE</td><td>0.016991022371</td></tr> <tr><td>r</td><td>0.991468822555</td></tr> <tr><td>% Error</td><td>23.571149989472</td></tr> <tr><td>AIC</td><td>64.238033748885</td></tr> <tr><td>MDL</td><td>19.900140823021</td></tr> </table>	MSE	0.005727543189	NMSE	0.016991022371	r	0.991468822555	% Error	23.571149989472	AIC	64.238033748885	MDL	19.900140823021	<table border="1"> <tr><td>MSE</td><td>0.002886384182</td></tr> <tr><td>NMSE</td><td>0.008562592473</td></tr> <tr><td>r</td><td>0.995717740601</td></tr> <tr><td>% Error</td><td>23.459232183464</td></tr> <tr><td>AIC</td><td>-17.411762043373</td></tr> <tr><td>MDL</td><td>-41.330888490221</td></tr> </table>	MSE	0.002886384182	NMSE	0.008562592473	r	0.995717740601	% Error	23.459232183464	AIC	-17.411762043373	MDL	-41.330888490221	<table border="1"> <tr><td>MSE</td><td>0.002772065590</td></tr> <tr><td>NMSE</td><td>0.008223461071</td></tr> <tr><td>r</td><td>0.995879904046</td></tr> <tr><td>% Error</td><td>23.263600644600</td></tr> <tr><td>AIC</td><td>-18.098763118933</td></tr> <tr><td>MDL</td><td>-42.017889565780</td></tr> </table>	MSE	0.002772065590	NMSE	0.008223461071	r	0.995879904046	% Error	23.263600644600	AIC	-18.098763118933	MDL	-42.017889565780
MSE	0.001998103640																																																		
NMSE	0.005927467069																																																		
r	0.997037970864																																																		
% Error	15.808930683890																																																		
AIC	136.335535616454																																																		
MDL	65.744942931855																																																		
MSE	0.005727543189																																																		
NMSE	0.016991022371																																																		
r	0.991468822555																																																		
% Error	23.571149989472																																																		
AIC	64.238033748885																																																		
MDL	19.900140823021																																																		
MSE	0.002886384182																																																		
NMSE	0.008562592473																																																		
r	0.995717740601																																																		
% Error	23.459232183464																																																		
AIC	-17.411762043373																																																		
MDL	-41.330888490221																																																		
MSE	0.002772065590																																																		
NMSE	0.008223461071																																																		
r	0.995879904046																																																		
% Error	23.263600644600																																																		
AIC	-18.098763118933																																																		
MDL	-42.017889565780																																																		

Once the most important variables for the model have been determined, we can train again the neural network with three or two variables, in this case with diameter and $\log P$ we obtained a very good results Squared Error SME less 0.001 for prediction solute retention.

This paper presents a method for prediction. In this method, firstly the global problem is obtain the most important variables, extracted and finally the solution is globalize with a model or prediction. Two stages have been judiciously combined, which allow selected to be a more efficient, effective and easy to control process. The obtained results show that this mixed system could be applied to different situations other than the one considered in this paper, due to the general nature of the proposed solution.

Bibliography

- [Anderson 1995] Anderson, James A., 1995. An Introduction to Neural Networks Cambridge, MA: MIT Press.
- [Arsuaga et al 2008] Arsuaga J.M., Lopez-Muñoz M. J., Aguado J., Sotto A., 2008. Temperature pH and concentration effects on retention and transport of organic pollutants across thin-film composite nanofiltration membranes, Desalination 221. 253-258.
- [Bellona C. et al 2004] Bellona, Drewes J., P. Xu and G. Amy, 2004. Factors affecting the rejection of organic solutes during NF/RO treatment. Water Res. 38. 2795-2809.
- [Boussu K. et al 2008] K. Boussu C. Vandecasteele and B. Van der Bruggen, 2008. Relation between membrane characteristics and performance in nanofiltration, J. Membr. Sci. 310. 51-65.
- [Hellebrand R. et al 1997] Hellebrand R., D. Mantzavinos, I. S. Metcalfe y A.G. Livingston, 1997. Integration of Wet Oxidation and Nanofiltration for Treatment of Recalcitrant Organics in wastewater, Ind. Eng. Chem. Res. 36. 5054-5062.
- [Moody, J. and Darken C. (1989)]. Moody, J. and Darken C., 1989. Fast learning in networks of locally-tuned processing units. Neural Computation, 1:281-294 .
- [Van der Bruggen et al 2003] Van der Bruggen B., Vandecasteele C. 2003. Removal of pollutants from surface water and groundwater by nanofiltration: overview of possible applications in the drinking water industry, Environ. Pollut. 122. 435.

Authors' Information

Sotto A. - Department of Chemical and Environmental Technology, University Rey Juan Carlos. C/ Tulipán s/n, 28933-Móstoles, Madrid, Spain. arcadio.sotto@urjc.es

Martinez A. - Natural computing group. Universidad Politécnica de Madrid, Spain. ana.martinez@upm.es

Castellanos A. - Departamento de Ciencias Básicas aplicadas a la Ingeniería Forestal. Escuela de Ingeniería Técnica Forestal. Universidad Politécnica de Madrid, Avda. de Ramiro de Maeztu s/n 28040 Madrid, Spain. angel.castellanos@upm.es

ГИПЕРПЛОСКОСТИ В «МНОЖЕСТВАХ И РАССТОЯНИЯХ СООТВЕТСТВИЯ»: КЛАСТЕРИЗАЦИЯ

Николай Кириченко, Владимир Донченко

Аннотация: Рассматриваются общие проблемы кластеризации. Предложена концепция «множеств» и «расстояний соответствия» в построении кластеров, рассмотрены модели кластеризации, в которых «множествами соответствия» являются гиперплоскости, а «расстояниями соответствия» – различные варианты расстояний в связи с соответствующими гиперплоскостями. Развита аппарат псевдообращения по Муру – Пенроузу: приведены рекуррентные формулы возмущения для ортогональных проекторов и R -операторов, связанных с псевдообращением. Рекуррентные формулы возмущения использованы для построения алгебраического варианта Jack Knife'a. Приведена сборка важных для приложений результатов, касающихся псевдообращения.

Ключевые слова: кластеризация, кластеризация по гиперплоскостям, псевдообращение по Муру - Пенроузу, сингулярное представление (SVD), ортогональные проекторы, псевдообращение для возмущённых матриц, преобразование Хока.

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

Вступление

Статья посвящена алгебраическим аспектам задачи кластеризации (см., например, [Kohonen, 2001]) как задачи группирования информации. В дальнейшем будет обсуждаться вопрос о разбиении имеющихся элементов на два класса с тем, что процедуру такого разбиения можно запускать рекуррентно.

Важным, как представляется авторам, во всех методах кластеризации является представление о «множествах соответствия» и «расстояниях соответствия». Типичным представителем первых являются прототипы-представители (prototypes) классов в методе k -средних. Что касается «расстояний соответствия», то, это меры соответствия «множествам соответствия», в соответствии с которыми элемент относят к тому или иному классу: как правило, – по минимальному значению «расстояния». Как правило, такими расстояниями соответствия являются, евклидовы расстояния в соответствующих пространствах признаков.

Заметим также, что процедуры кластеризации построены на применении стандартной рекуррентной процедуры: последовательного объединения (merging), разбиения (splitting) или уточняющих друг друга разбиений.

Разделяют также процедуры кластеризации с учителем (обучение с учителем – supervised learning) и – без учителя (unsupervised learning). В первом случае имеющиеся элементы уже разделены на классы, во втором – следует выделить классы на основе анализа внутренней структуры совокупности $x(1), \dots, x(n)$ векторов из пространства признаков R^m

В задачах кластеризации следует также выделять этап обучения: построения соответствующих классов-кластеров (этап обучения), и этап использования построенного разбиения: отнесения каждого нового вектора признаков к одному из построенных классов.

В предлагаемой вниманию читателю работе речь идёт об использовании гиперплоскостей в качестве «множеств соответствия», о «расстояниях соответствия», построенным в связи с гиперплоскостями, а также об обеспечении рекуррентности применения процедуры обучения без учителя; о согласованности обучения с учителем и без него; об аппарате псевдообращения по Муру – Пенроузу ([Moore,1920], [Penrose,1955]); о важном продвижении и расширении возможностей аппарата псевдообращения: о теории возмущения псевдообратных матриц ([Кириченко, 1997]), а также – о её совершенствовании и применении в задачах кластеризации (см. также [Кириченко, Донченко, 2007 а),b])). Заметим, что важные примеры применения теории псевдообращения к исследованию классических прикладных задач, отличных от задач кластеризации, можно найти в работах [Кириченко, Лепеха, 2002], [Кириченко, Донченко, 2005] .

Заметим также, что важными вехами в развитии аппарата псевдообращения, в частности, в обеспечении эффективности построения соответствующих рекуррентных процедур и вычисления расстояний соответствия в них, – являются: прямые [Алберт, 1977] и обратные [Кириченко, 1997] формулы Гревилля; формулы псевдообращения для замены строки или столбца матрицы [Кириченко, Лепеха, 2002,], [Кириченко, Донченко, 2005]; также формулы возмущения для Z - и R -операторов [Кириченко, Донченко., 2007 b)]. Отметим также, что задача кластеризации по гиперплоскостям, порождённым пространствами значений подходящих аффинных операторов, как вариант применения преобразования Хока, рассматривалась в работе [Donchenko, 2003].

В первой части предлагаемой работы приводится подборка результатов, важных в технике применения аппарата псевдообращения.

Во второй части рассматривается собственно задача кластеризации: рассматриваются подходящие гиперплоскости в качестве «множеств соответствия», строятся подходящие «расстояния соответствия» в связи с введёнными в рассмотрение гиперплоскостями, рассматриваются проблема обеспечения рекуррентности в вычислении «расстояний соответствия» как внутри рекуррентного шага, так и между разными шагами.

Отметим, что аппарат псевдообращения позволяет выписывать явные формулы, как для «расстояний соответствия», так и явно описывать «множества соответствия» в терминах смещения и явного описания ортогональных проекторов соответствующих линейных подпространств (ср. с вычислительными процедурами [Varnik,1998] для статистических вариантов кластеризации на основе ковариационных). Вычислительные алгоритмы для расстояния от гиперплоскостей использовались, к примеру, также в работе [Найкин,1999].

Постановка задачи

Собственно, использование гиперплоскостей как аппарата решения задач группирования информации в статистической постановке, восходит к методу главных компонент: [Pearson , 1901] (другие названия метод Хётеллинга (*Hotelling*), метод Карунена-Лоева (*Karhunen-Loeve*)) и имеет в основе идею такого ортогонального преобразования имеющегося набора случайных величин, которое бы приводило матрицу ковариаций к главным осям. Ещё раз обратим внимание читателя на специфически статистический вариант постановки и применения метода кластеризации в виде метода главных компонент, связанный с анализом естественного матричного объекта, каковым является матрица ковариаций, и применению классического результата Сильвестра [Sylvester, 1889]. Псевдообращение позволяет анализировать матрицы произвольной размерности, а не обязательно квадратные; позволяет эффективно строить ортогональные проекторы, отвечающие «естественным подпространствам» линейного оператора:

подпространству значений и ядру оператора; описывать гиперплоскости, отвечающие всем решениям системы линейных алгебраических уравнений (СЛАУ), а также описывать необходимые и достаточные условия существования таких решений: описывать «наилучшие» приближенные решения (псевдорешения) СЛАУ; явно описывать невязку соответствующего приближения.

В последующем будем рассматривать задачу кластеризации в обучении без учителя для дихотомического варианта постановки задачи: для разбиения имеющейся совокупности $x(1), \dots, x(n)$ векторов: из пространства признаков R^m на две части. В качестве множеств соответствия для каждого из классов-кластеров будут рассматриваться две гиперплоскости $\Gamma(k) \subseteq R^m, k = 1, 2$:

$\Gamma(k) = x_k + L_k \subseteq R^m, k = 1, 2$, x – будем называть смещение гиперплоскости, L – подпространством гиперплоскости. Таким образом, решение задачи кластеризации в такой постановке включает в себя

- построение «множеств соответствия» в виде гиперплоскостей: описание их смещений и соответствующих подпространств;
- описание «расстояний соответствия»;
- разбиение векторов $x(1), \dots, x(n)$ обучающей выборки на две части в соответствии с минимумом «расстояния соответствия» на две части:

$$x(i_1), \dots, x(i_{n_1}) \in \Gamma(1), \quad x(j_1), \dots, x(j_{n_2}) \in \Gamma(2) :$$

$$\{i_1, \dots, i_{n_1}\} \cup \{j_1, \dots, j_{n_2}\} = \{1, 2, \dots, n\}, \quad n_1 + n_2 = n ;$$

- построение решающего правила, в соответствии с которым следует относить объект, не представленный в выборке, к одному из двух классов.

Естественным для рекуррентной процедуры построения классов-кластеров является получение и использование результатов, обеспечивающих рекуррентность.

Заметим, что вариантом указанной задачи кластеризации является такой, в котором дополнительно фиксируется общая размерность $s: s < m$ гиперплоскостей $\Gamma(k) = x_k + L_k, k = 1, 2$.

Напомним, что «гиперплоскости соответствия» подлежат определению на основе внутренней структуры имеющегося набора векторов $x(1), \dots, x(n)$.

Вспомогательные определения и утверждения

Псевдообращение и сингулярное (SVD –) представление. Псевдообращение – обозначается A^+ – по Муру - Пенроузу ([Moore, 1920], Penrose, 1955], см. также [Алберт, 1977]) для $m \times n$ матрицы A может определяться одним из нескольких эквивалентных способов, среди которых отметим определение через сингулярное представление матриц (SVD-разложение), когда псевдообращение определяется соотношением:

$$A^+ = \sum_{i=1}^r x_i y_i^T \lambda_i^{-1}, \quad (1)$$

которое определяется элементами SVD-представления исходной матрицы:

$$A = \sum_{i=1}^r y_i x_i^T \lambda_i, \quad (2)$$

в котором: $\lambda_1^2 \geq \dots \geq \lambda_r^2 > 0$ – общий набор ненулевых собственных чисел матриц $AA^T, A^T A$, $y_i, i = \overline{1, r}$ и $x_i, i = \overline{1, r}$, соответственно, – ортонормированные наборы собственных векторов этих матриц, а $r = \text{rank } A = \text{rank } A^T$.

Чаще всего сингулярное разложение матрицы A представляется в виде, определяемом следующей леммой.

Лемма 1. Для любой $m \times n$ матрицы A ранга r существуют $Y - m \times r$ и $X - r \times n$ с ортонормированными столбцами и строками соответственно, а также диагональная матрица $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_r), \lambda_1 \geq \dots \geq \lambda_r > 0$ такие, что

$$A = Y\Lambda X. \quad (3)$$

Представление (2) является эквивалентным вариантом представления (3), если через $x_i, i = \overline{1, r}$ обозначить столбцы (ортонормированные) матрицы X , а через $y_i^T, i = \overline{1, r}$ – строки (ортонормированные) матрицы Y . В таких обозначениях справедлива следующая лемма.

Лемма 2. Произведение $Y\Lambda X$ матриц в (3) может быть представлено через «столбцовое» для Y и «строчное» для X представление может быть представлено в виде

$$X\Lambda Y = \sum_{i=1}^r y_i x_i^T \lambda_i.$$

Собственно, указанное несложное утверждение вытекает из того, что произведение BC двух матриц со «столбцовым» представлением для первой и «строчным» представлением для второй:

$$B = (b(1), b(2), \dots, b(p)), \quad C = (c(1), c(2), \dots, c(p))^T \quad (4)$$

допускает представление в виде

$$BC = (b(1), b(2), \dots, b(p))(c(1), c(2), \dots, c(p))^T = \sum_{i=1}^p b(i)c(i)^T \quad (5)$$

В дальнейшем представление (4) будет использоваться и для других матриц. При этом обозначение обозначения $b(i), c(i)^T, i = \overline{1, p}$, будут использоваться для обозначения соответственно строк и столбцов необходимых матриц.

Основные ортогональные проекторы: P-проекторы, Z-проекторы. Псевдообращение позволяет в явном виде выписать пару ортогональных проекторов (ОП), – обозначим их $P(A), P(A^T)$, и будем называть P - проекторами, – на подпространства $L(A^T), L(A)$ значений операторов A^T, A соответственно: $P(A) = A^+ A, P(A^T) = A^{T+} A^T = AA^+$. Ортогональные проекторы, которые будем обозначать $Z(A), Z(A^T)$ и называть Z - проекторами – определим соотношениями: $Z(A) = E_n - P(A), Z(A^T) = E_m - P(A^T)$ соответственно. Очевидным образом, Z - проекторы являются ортогональными проекторами на подпространства $L_{A^T}^\perp, L_A^\perp$, ортогональные к подпространствам $L(A^T), L(A)$ соответственно. Заметим, что $L_{A^T}^\perp = \text{Ker } A, L_A^\perp = \text{Ker } A^T$.

Соответственно, $Z(A), Z(A^T)$ являются ортогональными проекторами на подпространства нулей $\text{Ker}A, \text{Ker}A^T$ операторов A, A^T соответственно.

Замечание 1. Обратим внимание также на то, что каждое из подпространств $L(A), L(A^T)$ является линейной оболочкой соответственно векторов-столбцов и векторов-строк матрицы A .

R-операторы. Важными в связи с определением расстояний соответствия и рекуррентными формулами псевдообращения: формулами позволяющими записывать соответствующий оператор при добавлении или вычёркивании строки или столбца матрицы, – являются также операторы, которые будем называть R - операторы. Их будем определять соотношениями:

$$R(A) = A^+ A^{T+}, R(A^T) = A^{+T} A^+.$$

Важную роль в реализации аппарата псевдообращения в прикладных задачах играют прямые (см., например, [Алберт, 1977]) и обратные [Кириченко 1997] формулы Гревилля(Greville), а также формулы возмущения псевдообращения [Кириченко, 1997]. И в том и в другом случае речь идёт о формулах, связывающих псевдообращение преобразованной матрицы с псевдообращением исходной. В первом случае (прямых или обратных формулах Гревилля) речь идёт о преобразовании матрицы введением или вычёркиванием дополнительного строки или столбца. Во втором – о преобразовании исходной матрицы аддитивной добавкой ab^T . Таким образом, в формулах псевдообращения для возмущённых матриц речь идёт о выражении псевдообращения возмущённой матрицы $(A + ab^T)^+$ через A, A^+, a, b .

Прямые и обратные формулы Гревилля приведены ниже. Формулы возмущения псевдообращения можно найти в уже цитированной работе [Кириченко 1997]. Ниже приведены полученные на их основе формулы возмущения для Z - и R -операторов.

Заметим, что комбинация прямых и обратных формул Гревилля позволяет получить формулы псевдообращения при замене строки или столбца исходной матрицы. Соответствующие представления можно найти в работах [Кириченко, Лепеха, 2002], [Кириченко, Донченко, 2005]. Там же приведены формулы, определяющие вид, Z - и R - операторов при замене строки или столбца матрицы, для которой они рассматриваются.

Прямые формулы Гревилля (Greville).

Напомним, что прямые формулы Гревилля – это формулы, определяющие вид псевдообращения матрицы при её дополнении строкой или столбцом. Они определяются соотношениями, в которых используется блочное представление псевдообращения расширенной матрицы: через $P - m \times n$ - матрицу и, $q - n \times 1$ - вектор для расширения матрицы строкой

$$\begin{pmatrix} A \\ a^T \end{pmatrix}^+ = (P : q) \quad (6)$$

и через $Q - n \times m$ - матрицу и, $q - m \times 1$ - вектор

$$(A : a)^+ = \begin{pmatrix} Q \\ q^T \end{pmatrix} \quad (7)$$

при дополнении матрицы столбцом.

Замечание 2. Обратим внимание читателя, что вектор a в (6) и (7) имеет разные размерности: размерность $n \times 1$ в первом и $m \times 1$ во втором.

Теорема 1. (прямые формулы Greville– дополнение строкой). В представлении (6)

$$\begin{cases} P = (E - qa^T)A^+ \\ q = \begin{cases} \frac{Z(A)a}{a^T Z(A)a}, a^T Z(A)a > 0 \text{(нез.)} \\ \frac{R(A)a}{1 + a^T R(A)a}, a^T Z(A)a = 0 \text{(зав.)} \end{cases} \end{cases}, \quad (8)$$

Теорема 2. (прямые формулы Greville– дополнение столбцом) В представлении (7)

$$\begin{cases} Q = A^+(E - aq^T) \\ q = \begin{cases} \frac{Z(A^T)a}{a^T Z(A^T)a}, a^T Z(A^T)a > 0 \text{(нез.)} \\ \frac{R(A^T)a}{1 + a^T R(A^T)a}, a^T Z(A^T)a = 0 \text{(зав.)} \end{cases} \end{cases}, \quad (9)$$

Замечание 3. Вид вектора q в прямых формулах Гревилля определяется линейной зависимостью вводимого вектора a^T или a от, соответственно, строк или столбцов матрицы A . Линейная независимость обеспечивается нулевым значением квадратичной формы (с соответствующей матрицей Z -оператором) на векторе a .

Обратные формулы Гревилля. Как и в прямых формулах Гревилля, вид выражений, связывающих псевдообращения исходной и преобразованной матрицы, выписывается в рамках блочного представлением (6) или (7), и так же – определяется линейной зависимостью или независимостью вычёркиваемой строки или столбца: сохранением или падением ранга преобразованной матрицы. Изменяется только вид соответствующего условия. Теперь условием независимости является условие $a^T q = 1$.

Теорема 3. (обратные формулы Гревилля – вычёркивание строки) В обозначениях (6) имеет место соотношение

$$A^+ = \begin{cases} \left(I_n - \frac{qq^T}{\|q\|^2} \right) P, a^T q = 1, \text{(нез.)}, & \text{ранг падает} \\ \left(I_n - \frac{qa^T}{1 - a^T q} \right) P, a^T q < 1 \text{(зав.)}, & \text{ранг сохраняется} \end{cases}. \quad (10)$$

Теорема 4. (обратные формулы Гревилля – вычёркивание столбца) В обозначениях (7) имеет место соотношение

$$A^+ = \begin{cases} Q \left(I_m - \frac{qq^T}{\|q\|^2} \right), a^T q = 1, (\text{нез.}), & \text{ранг падает} \\ Q \left(I_m - \frac{aq^T}{1 - a^T q} \right), a^T q < 1 (\text{зав.}), & \text{ранг сохраняется} \end{cases}.$$

Теорема 5 (формулы возмущения для Z- и R- операторов). При возмущении матрицы A матрицей $a \times b^T$ Z- и R- операторы для возмущённой матрицы определяется следующими соотношениями, вид которых определяется линейной зависимостью или независимостью векторов - составляющих возмущения от соответствующих составляющих матрицы A , а также от того, сохраняется или падает ранг возмущённой матрицы:

1) Для векторов a и b^T линейно не зависимых от, соответственно, столбцов и строк матрицы A , т.е. при выполнении условий $a^T Z(A^T)a > 0, b^T Z(A)b > 0$, справедливы следующие соотношения

$$Z(A + ab^T) = Z(A) + \frac{Z(A)bb^T Z(A)}{b^T Z(A)b};$$

$$Z((A + ab^T)^T) = Z(A^T + ba^T) = Z(A^T) + \frac{Z(A^T)aa^T Z(A^T)}{a^T Z(A^T)a};$$

$$R(A + ab^T) = R(A) - R(A) \frac{bb^T Z(A)}{b^T Z(A)b} - \frac{Z(A)bb^T}{b^T Z(A)b} R(A) - cA^+ ab^T Z(A) - cZ(A)ba^T A^{+T} +$$

$$+ \frac{A^+ aa^T A^{+T}}{a^T Z(A^T)a} + \frac{b^T R(A)ba^T Z(A^T)a + (1 + b^T A^+ a)^2}{a^T Z(A^T)a [b^T Z(A)b]^2} Z(A)bb^T Z(A),$$

где $c = \frac{1 + b^T A^+ a}{a^T Z(A^T)a b^T Z(A)b}.$

2) Для вектора a линейно зависимого от столбцов матрицы A , а вектора b^T – линейно не зависимого от строк матрицы таким образом, что, – для упрощения представления результата, – $b \perp L_{A^T}$, т.е. при выполнении условий $a^T Z(A^T)a = 0, b^T Z(A)b = \|b\|^2$, справедливы соотношения:

$$Z(A + ab^T) = Z(A) + \frac{k_{A,a,b} k_{A,a,b}^T}{\|k_{A,a,b}\|^2} - \frac{bb^T}{\|b\|^2},$$

где: $k_{A,a,b} = A^+ a - \frac{b}{\|b\|^2},$

$$Z((A + ab^T)^T) = Z(A^T + ba^T),$$

$$R(A + ab^T) = \left(I_n - \frac{kk^T}{\|k\|^2} \right) R(A) \left(I_n - \frac{kk^T}{\|k\|^2} \right).$$

3) Для векторов a и b^T одновременно линейно зависимых от соответственно столбцов и строк матрицы A , при условии падения ранга возмущённой матрицы: $\text{rank}(A + ab^T) = \text{rank}A - 1$, т.е. при выполнении условий: $a^T Z(A^T)a = 0, b^T Z(A)b = 0, b^T A^+ a = -1$, справедливы следующие соотношения:

$$Z(A + ab^T) = Z(A) + \frac{A^+ aa^T (A^+)^T}{a^T R(A^T)a},$$

$$Z((A + ab^T)^T) = Z(A^T + ba^T) = Z(A) + \frac{(A^+)^T bb^T A^+}{b^T R(A)b},$$

$$R(A + ab^T) = A^+(a, b)A^{+T}(a, b),$$

где: $A^+(a, b) = A^+ \frac{A^+ aa^T R(A^T)}{a^T R(A^T)a} \frac{R(A)bb^T A^+}{b^T R(A)b} + cA^+ ab^T A^+, c = \frac{b^T R(A)A^+ a}{a^T R(A^T)a b^T R(A)b}.$

4) Для векторов a и b^T одновременно линейно зависимых от, соответственно, столбцов и строк матрицы A , но при условии неизменности ранга возмущённой матрицы по сравнению с рангом A , т.е. при выполнении условий

$$a^T Z(A^T)a = 0, b^T Z(A)b = 0, b^T A^+ a \neq -1,$$

справедливы следующие соотношения:

$$Z(A + ab^T) = Z(A), \quad Z((A + ab^T)^T) = Z(A^T + ba^T) = Z(A^T),$$

$$R(A + ab^T) = R(A) \frac{A^+ ab^T R(A)}{1 + b^T A^+ a} \frac{R(A)ba^T A^{+T}}{1 + b^T A^+ a} + \frac{b^T R(A)b}{1 + b^T A^+ a} A^+ aa^T A^{+T}.$$

Основные элементы кластеризации по гиперплоскостям – множества соответствия

Как уже упоминалось, построение «множеств соответствия» в виде гиперплоскостей предполагает конструктивное описание их смещений и соответствующих линейных подпространств.

Смещение гиперплоскостей.

Смещения предлагается определять как средние векторов, принадлежащих к каждой из частей разбиения. Можно также выбрать в качестве смещения один из элементов разбиения.

Подпространства гиперплоскостей. При наличии смещений подпространства гиперплоскостей определяются как подпространства, натянутые на центрированные смещением (преобразованные вычитанием определённого вектора) векторы каждой из частей разбиения. В дальнейшем будет предполагаться, что центрирование каждой части разбиения производится соответствующими средними $\bar{x}_k, k = 1, 2$.

Конструктивное описание подпространств, натянутых на каждую из центрированных совокупностей векторов, обеспечивается построением для каждой из гиперплоскости подходящей матрицы $A_k, k = 1, 2$ так, чтобы подпространство - множество значений $L_k, k = 1, 2$ каждой из них совпадало с подпространством соответствующей гиперплоскости, т.е. с линейной оболочкой каждой из

центрированных групп векторов. В соответствии с замечанием 1 в качестве таких матриц можно выбрать матрицы, столбцами которых являются центрированные векторы каждой из частей разбиения соответственно. В этом случае ортогональными проекторами $P_{L_k}, k = 1, 2$ для каждого из подпространств гиперплоскостей будут P -проекторы для транспонированных к соответствующим матрицам:

$$P_{L_k} = P(A_k^T), k = 1, 2.$$

Таким образом, гиперплоскости $\Gamma_k, k = 1, 2$ определяются парами $(\bar{x}_k, A_k), k = 1, 2$:

$$\Gamma_k = \Gamma(\bar{x}_k, A_k), k = 1, 2 \quad (11)$$

Основные элементы кластеризации по гиперплоскостям – расстояния соответствия

В качестве «расстояний соответствия» векторов до каждого «множеств соответствия» $\Gamma_k = \Gamma(\bar{x}_k, A_k), k = 1, 2$ предлагается рассматривать евклидово расстояние векторов до гиперплоскостей $\Gamma_k = \Gamma(\bar{x}_k, A_k), k = 1, 2$, каковыми эти «множества соответствия» являются. Средства псевдообращения позволяют конструктивно описать соответствующие расстояния. Такое конструктивное описание возможно и в том случае, когда задаётся размерность $s: s \leq \min(\text{rank} A_k, k = 1, 2)$ – подпространств гиперплоскостей. Формулы, определяющие соответствующие расстояния, являются предметом следующей леммы.

Лемма 1. Для $\Gamma_k = \Gamma(\bar{x}_k, A_k), k = 1, 2$ расстояния соответствия $\rho(x, \Gamma_k), k = 1, 2$ произвольного вектора $x \in R^m$ до каждой из двух гиперплоскостей $\Gamma_k, k = 1, 2$ определяются соотношением:

$$\rho(x, \Gamma_k) = (x - \bar{x}_k)^T Z(U_s^T(k))(x - \bar{x}_k), k = 1, 2, \quad (12)$$

$$\text{Где } U_s(k) = \begin{cases} A_k = \sum_{i=1}^r y_i(k) x_i^T(k) \lambda_i(k) & \text{разм. } s \text{ не задана} \\ \sum_{i=1}^s y_i(k) x_i^T(k) \lambda_i(k) & \text{разм. } s \text{ задана} \end{cases}, r = 1, 2.$$

Кластеризации по гиперплоскостям – основные шаги алгоритма

Алгоритм кластеризации по гиперплоскостям состоит в последовательном, рекуррентном уточнении «множеств соответствия», каковыми являются гиперплоскости. На каждом рекуррентном шаге происходит уточнение набора элементов, порождающих «множества соответствия», построение пар $(\bar{x}_k, A_k), k = 1, 2$ отвечающих уточнённому разбиению, после чего происходит новое «уточнение разбиения» отбором в каждую часть разбиения векторов исходного набора по минимуму расстояний до вновь построенных гиперплоскостей. В общем, алгоритм состоит в выполнении следующих шагов.

1. На первом шаге производится разбиение на две совокупности произвольным образом.
2. На втором шаге для каждой из частей разбиения вычисляются:
 - смещения $\bar{x}_k, k = 1, 2$, как средние по векторам каждой из частей разбиения;

- матрицы $A_k, k = 1, 2$, как матрицы, построенные из центрированных соответствующими средними векторов каждой из групп как из столбцов.
3. На третьем шаге происходит «уточнение» разбиения: вычисляются «расстояния соответствия» каждого из векторов $x(1), \dots, x(n)$ до каждого из двух построенных «множеств соответствия»: до каждой из двух гиперплоскостей, – и происходит отнесение каждого из векторов $x(1), \dots, x(n)$ к той части разбиения, к которой он оказался ближе по «расстоянию соответствия» (12). В результате происходит формирование нового, «уточнённого» разбиения векторов $x(1), \dots, x(n)$ на две части.
 4. На четвёртом шагу происходит возвращение ко второму шагу алгоритма.

Кластеризация по гиперплоскостям – модификация расстояний

Расстояния до гиперплоскостей в лемме 1 определяются значениями квадратичных форм с матрицами

$$\sum_{i=1}^s y_i(k) x_i^T(k), k = 1, 2 \quad ,$$

Их можно рассматривать как взвешенное среднее матриц $y_i(k) x_i^T(k), i = \overline{1, r}, k = 1, 2$ с весами

$$\omega_i = \begin{cases} 1, & i \leq s \\ 0, & i = \overline{s+1, r} \end{cases}, \text{ соответственно, в нормированном варианте } \omega_i = \begin{cases} 1/s, & i \leq s \\ 0, & i = \overline{s+1, r} \end{cases}.$$

Рассмотрение взвешенных варианта сумм из $y_i(k) x_i^T(k), i = \overline{1, r}, k = 1, 2$ с нормированными весами $\lambda_i^2(k), i = \overline{1, r}, k = 1, 2$ даёт следующий вариант расстояний ρ_R до «множеств соответствия». Они для этого случая определяются соотношением:

$$\rho_R(x, \Gamma_k) = \frac{1}{\text{tr}R(A_k^T(k)A_k)} (x - \bar{x}_k)^T R(A_k^T(k)) (x - \bar{x}_k), k = 1, 2. \quad (13)$$

Использование в качестве «расстояний соответствия» расстояний, определяемых соотношением (13) приводит к очевидному изменению алгоритма кластеризации: в нём «уточнение» разбиения третьего шага происходит на основе «расстояний соответствия», определяемых соотношениями (13) вместо – (12).

Рекуррентные формулы для алгебраического Jack Knife'a

При проверке элементов совокупностей на соответствие вычислением расстояний по формулам (12) или (13) тестируемые элементы принимают участие в формировании гиперплоскостей, представляющих кластеры. Резонной является также построение такой процедура проверки соответствия, при которой тестируемый элемент кластера, исключается из числа объектов, которые его определяют. В статистике такая процедура исключения носит название «Jack Knife» (складной нож) [Эфрон, 1988]. Поэтому процедуру тестирования на принадлежность кластеру с исключением тестируемых элементов из описания кластера будем называть алгебраическим Jack Knife'ом.

Заметим, что естественным является вариант кластеризации, когда исключение элемента приводит к падению ранга матрицы $A(k), k = 1, 2$ (п.3 теоремы 5)). Псевдообращение даёт конструктивную явную формулу проверки соответствующего условия.

Исключение тестируемых элементов из кластера изменяет как сдвиг (центр кластера), так и линейное подпространство кластера. Формулы (12),(13) при таком исключении, очевидным образом, переписываются в виде, для изменённых смещений (будем считать их средними) и изменённых матриц: $x_k^{(0)}, A^{(0)}(k), k = 1, 2$ соответственно.

Лемма 1 даёт возможность эффективной организации процедуры «отсеивания», в которой критерий замены строится на основе леммы 1 и имеет вид, определяемый следующей теоремой.

Теорема 6. В условиях падения ранга (п.3 теоремы 5) расстояния до гиперплоскостей, после исключения элемента из числа порождающих элементов, определяется следующим соотношением для одного из значений $k=1$ или $k=2$

$$\rho(x_j(k), \Gamma_j^{(0)}(k)) = \frac{n_k^2}{\left\| \begin{pmatrix} E_m & q_j(k)q_j^T(k) \\ & \|q_j(k)\|^2 \end{pmatrix} \sum_{l \neq j} q_l(k) \right\|^2}, j = \overline{1, n_k}, k = 1, 2,$$

Где $x_j(k), \Gamma_j^{(0)}(k) j = \overline{1, n_k}, k = 1, 2$ – исключаемые элементы каждой из совокупностей и гиперплоскости, отвечающие «усечённым» совокупностям, а $q_j(k), j = \overline{1, n_k}, k = 1, 2$ столбцы с номером $j, j = \overline{1, n_k}$ в каждой из матриц $A_k^+, k = 1, 2$.

Заключение

В работе рассмотрены задачи кластеризации на основе концепции «множеств» и «расстояний соответствия», предложены варианты алгоритмов кластеризации, когда «множествами соответствия» являются гиперплоскости, а «расстояния соответствия» построены на основе вариантов расстояний до них. Применение аппарата псевдообращения позволяет описать все элементы соответствующих построений явными формулами, включая варианты алгебраического Jack Knife'a

Литература

- [Алберт, 1977] Алберт А. Регрессия, псевдоинверсия, рекуррентное оценивание. | М.: Наука, 1977.–305 с.
- [Donchenko, 2003] Donchenko V.S. Hough Transform and Uncertainty//Proceedings International Conference "Knowledge Dialog – Solution". – V. – June 16-23, 2003.–Varna (Bulgaria). – P.391-395.
- [Найкин, 1999] Neural networks. A comprehensive Foundation. – New Jersey. – 1999.– 842 p.
- [Кириченко, 1997] Кириченко Н.Ф. Аналитическое представление псевдообратных матриц//Киб. и СА. - №2, 1997– С.98-122.
- [Кириченко, Донченко, 2007, а)] Кириченко Н.Ф., Донченко. В.С. Псевдообращение в задачах кластеризации.// Киб. и СА. - №4, 2007.– С.98-122.
- [Кириченко, Донченко, 2007, b)] Кириченко Н., Донченко В. Алгебраический Jack Knife: кластеризация по гиперплоскостям// Proceedings: XIII-th International Conference "Knowledge –Dialog – Solution".–June 18-24, 2007, Varna (Bulgaria). – 2007. – V.1.– P.89-95.
- [Кириченко, Лепеха, 2002] Кириченко Н.Ф., Лепеха Н.П. Псевдообратные и проекционные матрицы в применении к исследованию задач управления, наблюдения и идентификации// Киб. и СА.- №4, 2002. – С.107-123.
- [Kohonen, 2001] Kohonen T., Self-Organizing Maps.– Third Extended Edition.– New York, 2001.– 501 p.

- [Кириченко, Донченко, 2005] Кириченко М.Ф, Донченко В.С. Задача термінального спостереження динамічних системи: множинність розв'язків та оптимізація //Ж. Обч. та пр. мат. – Вип..3 , 2005.– С. 63-78.
- [Moore, 1920] Moore E.H. On the reciprocal of the general algebraic matrix//Bull. Amer. Math. Soc. – 26, 1920. – P. 394-395.
- [Pearson, 1901] Pearson K., On lines and planes of closest fit to systems of points in space//Philosophical Magazine.–1901, N 2.– P. 559—572.
- [Penrose, 1955] Penrose R. A generalized inverse for matrices// Proc. Camb. Philosophical Soc.- 51, 1955.– P. 406-413.
- [Sylvester, 1889] Sylvester J.J. On the reduction of a bilinear quantic of the nth order to the form of a sum of n products by a double orthogonal substitution, Messenger of Mathematics, – 1889.– N19.– P., 42—46;
- [Vapnik, 1998] Vapnik V.N. Statistical Learning Theory.–New York: Wiley. – 1998.
- [Эфрон, 1988] Эфрон Б. Нетрадиционные методы многомерного статистического анализа. – М.: Фин. и стат. – 1988.– 263 с.
-

Информация об авторах

Кириченко Николай Ф.– Профессор, Институт кибернетики им. В.М.Глушкова НАН Украины, ведущий научный сотрудник

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

ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА КИРЛИАН ИЗОБРАЖЕНИЙ В ДИАГНОСТИЧЕСКОЙ СИСТЕМЕ

Владимир Калмыков, Виталий Вишневский, Татьяна Романенко

Аннотация: Для поддержки принятия решений в диагностической системе разработана информационная технология предварительной обработки Кирлиан изображений, содержащих изображения свечения десяти пальцев рук, полученных одновременно. Предварительная обработка включает выделение объектов – свечений каждого пальца на общей фотографии и формирование файлов изображений для каждого из объектов. Приведены описания алгоритмов и примеры обработки фотографий.

Ключевые слова: диагностическая система, изображения Кирлиан, сегментация

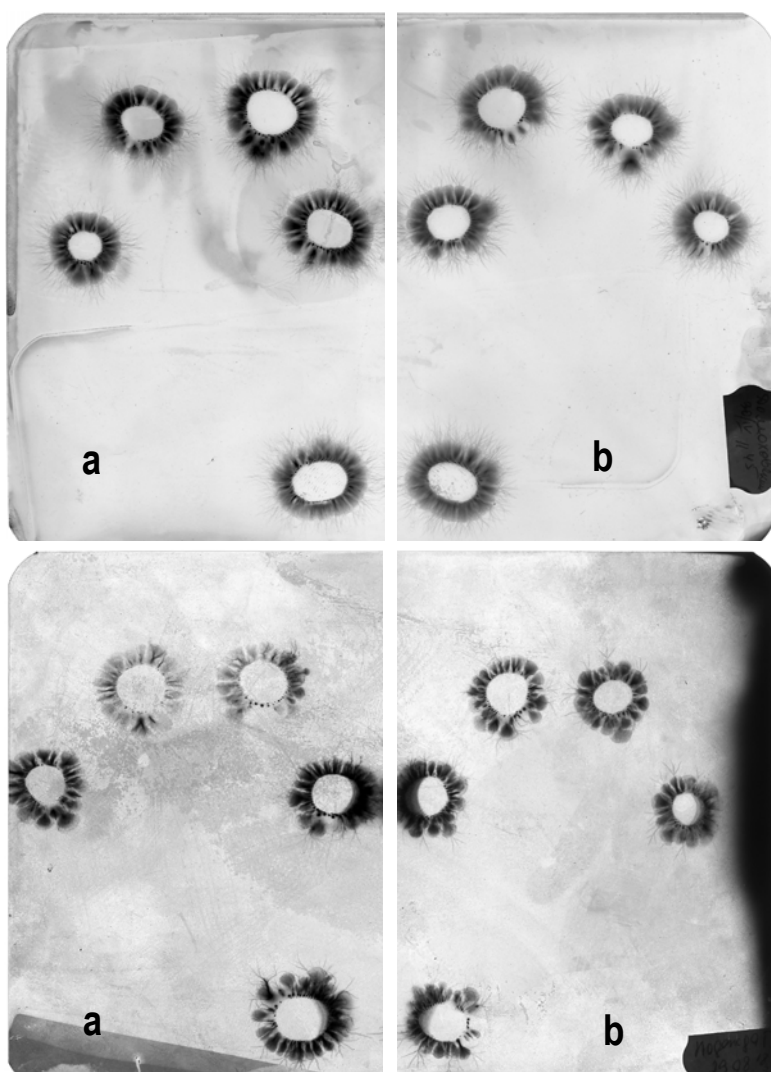


Рис. 1. Примеры изображений Кирлиан:
а – пальцы левых рук, б – пальцы правых рук

Введение

В данной работе представлены предварительные результаты разработки информационной технологии по обработке изображений Кирлиан для использования в процессе принятия решений в диагностической системе.

Под изображением Кирлиан понимают зарегистрированное на фотоматериале или иным способом свечение газового разряда, возникающего вблизи поверхности объекта при помещении его в электрическое поле высокой напряженности. Кирлианография получила большое распространение в мире как метод экспериментальных исследований. Наибольший интерес вызвали исследования кирлианограмм биологических объектов, в основном организма человека.

Первые же исследования изображений Кирлиан показали, что вид кирлианограмм меняется при изменении состояния

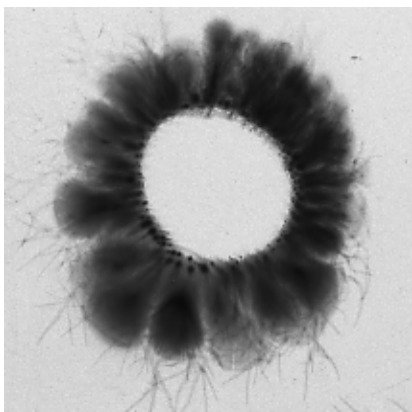


Рис.2. Изображение Кирлиан свечения одного пальца.

человека. Например, по виду кирлианограмм пальцев рук и ног оказалось возможным судить об общем уровне и характере физиологической активности организма, оценивать состояние отдельных его систем и следить за влиянием различных воздействий: препаратов, терапии и т.п.[Песоцкая,1]. Это позволило развить эффективные системы диагностики, основанные на использовании эффекта Кирлиан.

Эффект Кирлиан является в данный момент единственным инструментальным методом, позволяющим на физическом и энерго-информационном уровнях оценить состояние не отдельного органа или системы, а всего организма в целом во взаимоотношении отдельных частей друг с другом. В перспективе этот метод видится воплощенным в виде практического инструмента на столе любого врача. Создание

информационной технологии для автоматизированной обработки изображений Кирлиан является очередным шагом на пути к использованию метода в повседневной врачебной практике.

Задачи предварительной обработки Кирлиан изображений в диагностической системе

Изображения Кирлиан представляют собой снимки, полученные на специальной фотопленке, размером А5, на которых зафиксированы свечения от каждого из пяти пальцев обеих рук (рис.1). Следует отметить, что с позиций автоматической или автоматизированной обработки для изображений характерны нестабильность и неравномерность фона, значительное количество помех, которые по уровню яркости и величине сравнимы с объектами, неустойчивость формы и уровня яркости самих объектов. Хотя по диагностическому содержанию на настоящее время эти изображения могли бы считаться бинарными, однако, даже задача бинаризации таких изображений не может считаться тривиальной, не говоря уже о задачах дальнейшей обработки, в частности, задачах распознавания с целью диагностики. Изображение свечения каждого пальца имеет вид темного ореола, обрамляющего светлое пятно, по форме часто близкое к эллипсу. Ширина ореола может быть разной, даже для одного изображения. Часто ореол имеет не сплошную, а прерывистую форму, а также может состоять из отдельных фрагментов. Светлое пятно внутри ореола соответствует месту контакта пальца с пленкой и его яркость соответствует яркости фона изображения.

В настоящее время для диагностических целей используется стандартное прикладное программное обеспечение, поставляемое вместе с прибором, работающим по методу Короткова [Коротков,2] Этот прибор предусматривает последовательное получение изображений отдельно по каждому пальцу (рис.2). Программное обеспечение позволяет построить по Кирлиан изображениям десяти пальцев общую кирлианограмму организма пациента и поставить диагноз, то есть принять решение. Следует отметить, что в процессе последовательного получения изображений свечения по каждому пальцу состояние исследуемого пациента может существенно измениться, отчего окончательный диагноз может оказаться искаженным. Так что одномоментное получение кирлианограмм всех пальцев является более предпочтительным. Появляется возможность получения последовательностей кирлианограмм и изучения их диагностических возможностей в случае быстрого изменения состояния пациента. Чтобы использовать стандартное программное обеспечение для изображений (рис.1), необходимо предварительно сегментировать такие изображения, то есть из общего изображения выделить

изображение каждого пальца и повернуть его, чтобы оно соответствовало вертикальному направлению пальца.

Основные функции и алгоритмы предварительной обработки Кирлиан изображений

В процессе предварительной обработки Кирлиан изображений выполняется полуавтоматическая сегментация изображений Кирлиан пяти пальцев на изображения свечения от каждого пальца отдельно, автоматической корректировки ориентации каждого пальца. Для выполнения операции ориентации предусмотрена функция поворота изображения свечения для каждого пальца вокруг условного центра.

Программное обеспечение выполняет следующие функции:

- открытие файла полутонового изображения, отображение его на экране монитора для предварительной оценки экспертом и возможной корректировки в интерактивном режиме с целью удаления помех;
- интерактивное определение расположения свечения пальцев на изображении путем указания оператором центров условных центров свечения;
- присвоение номера каждому свечению пальцев по их взаимному расположению;
- определение центра ладони и условных углов поворота пальцев;
- определение параметров прямоугольников, включающих свечения отдельных пальцев с учетом угла поворота каждого прямоугольника;
- поворот изображений отдельных пальцев;
- формирование файлов свечения каждого пальца; формирование описания каждого пальца.
- интерактивный режим корректировки результатов работы функций.

Наиболее сложной является функция уточненного определения расположения свечения пальца на изображении, определения границ и условного центра пятна.

Исходя из общих представлений об изображениях Кирлиан, представилось целесообразным, с целью выделения свечения каждого пальца как объекта на изображении и определения его параметров, аппроксимировать внутренний контур свечения каждого пальца эллипсами, а внешние границы свечения каждого пальца – окружностями. Параметрами эллипса являются координаты его центра, размеры полуосей и угол наклона большой оси. Параметрами окружности, описанной вокруг темного ореола каждого пальца, являются центр и диаметр. Центры эллипса и окружности совпадают. Окружность должна охватить все свечение (темный ореол), по возможности исключая тонкие периферические линии – «дендриты». Предполагается, что центр ладони находится на середине отрезка, соединяющего центры первого и пятого пальцев. Центр ладони соединяют с центрами всех пальцев, для определения угла поворота каждого пальца относительно вертикали. Будем считать, что изображение Кирлиан каждого пальца ограничено квадратом, описанным вокруг окружности, охватывающей темный ореол, и повернутым на угол наклона линии, соединяющей центр ладони с центром свечения.

Исходная информация – предварительный список координат центров свечения $x_{1l}, y_{1l}, l=1,5$ – пяти пальцев.

Выходная информация – список параметров пяти окружностей, ограничивающих пятна, список параметров пяти эллипсов, ограничивающих внутренность пятна.

Для поиска параметров окружностей и эллипсов используют оптимизационные алгоритмы, реализующие метод градиентного спуска.

Для окружностей параметрами оптимизации являются координаты центра x_u, y_u и величина радиуса r .
Целевая функция

$$c = \max_{x_u, y_u, r} S_{\text{окр}}, \text{ при } \sum_{(x-x_u)^2 + (y-y_u)^2 < r^2} v(x, y) > \theta \cdot \Psi \quad (1)$$

где $\theta \approx 0.9$ – некоторый порог, определяемый экспериментально,

$v(x, y)$ – значение оптической плотности пиксела с целочисленными координатами x, y ;

$\Psi = \sum_{(i-x_u)^2 + (j-y_u)^2 < r_{\text{max}}^2} v(i, j)$ – суммарная оптическая плотность в пределах круга максимального радиуса

свечения r_{max} с координатами центра x_u, y_u

i, j – целочисленные координаты пиксела.

Для эллипсов параметрами оптимизации являются координаты центра x_u, y_u , величины большой a и малой b полуосей и угол поворота α . Целевая функция

$$c = \max_{x_u, y_u, a, b, \alpha} S_{\text{эл}}, \text{ при } \sum_{\frac{(x-x_u)^2}{a^2} + \frac{(y-y_u)^2}{b^2} < 1} v(x, y) < (1 - \theta) \cdot \Psi \quad (2)$$

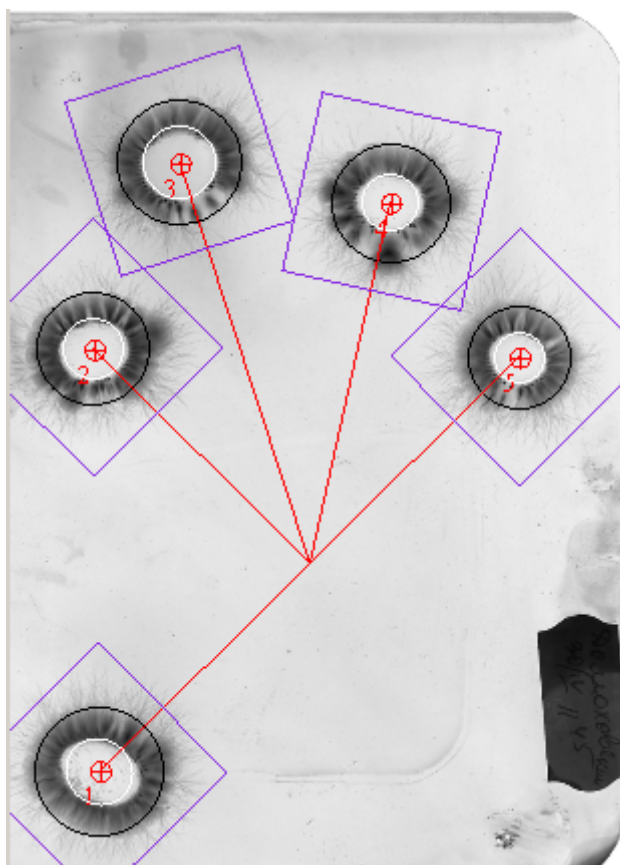


Рис.3 Изображение Кирлиан и выделенные на нем изображения свечений отдельных пальцев.

Алгоритм функционирования системы заключается в следующем (рис.3).

1. Оператор указывает курсором на экране монитора примерное место расположения центра изображения каждого пальца. Местоположение центра может быть указано приблизительно, но обязательно должно находиться в пределах светлого пятна. Система определяет координаты указанных центров.

2. По указанным координатам автоматически определяются параметры эллипса и окружности для изображения каждого пальца. При необходимости оператор может скорректировать центр изображения пальца, при этом пересчет параметров эллипса и окружности осуществится автоматически.

3. Автоматически определяются координаты центра ладони, значения углов поворота изображений пальцев и параметры описанных квадратов, ограничивающих изображения отдельных пальцев.

4. Изображения пальцев записываются в отдельные растровые файлы (рис.4.).

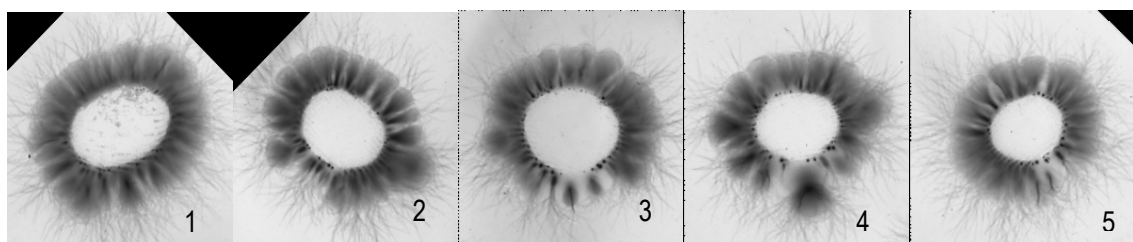


Рис. 4. Изображения отдельных пальцев, вырезанные из общего изображения Кирлиан и записанные в отдельные файлы.

Алгоритм автоматического определения параметров эллипса и окружности для каждого пальца состоит в следующем:

1. Определяют яркость фона в месте расположения свечения каждого пальца. В качестве яркости фона принимают среднюю яркость в квадрате, размером 10×10 пикселей, центр которого совпадает с центром изображения пальца.
2. Определяют параметры эллипса максимальной площади при выполнении ограничений (2), используя метод градиентного спуска.
 - 2.1. В нулевом приближении используем координаты центра эллипса, полученные в интерактивном режиме, размер полуосей равен 10 пикселям, угол поворота равен нулю градусов относительно вертикали.
 - 2.2. Изменяя значения параметров x_c , y_c , a , b , α , определяют их значения, максимизирующие площадь эллипса при выполнении ограничений по яркости (2).
3. Определяют диаметр окружности, охватывающей темный ореол. Координаты центра окружности совпадают с координатами центра эллипса.
 - 3.1. В качестве нулевого приближения выбирают радиус окружности, равный большей полуоси эллипса.
 - 3.2. Изменяя значения радиуса r , определяют его значение, максимизирующее площадь окружности при выполнении ограничений по яркости (1).

Заключение

1. Проведенные лабораторные испытания показали работоспособность и возможность использования созданной технологии обработки изображений Кирлиан для поддержки принятия решений в диагностических системах.
2. Одновременное получение Кирлиан изображений позволяет принимать решения в диагностических системах при быстрых изменениях состояния организма пациента, а также исследовать возможность использования последовательностей кирлианограмм в диагностических целях.

Литература

- [1] Песоцкая Л.А., Компаниец В.А. Современная Кирлиан диагностика в сб. статей Эффект Кирлиан. – Днепропетровск, Днепропетровский центр НТИ, 2008. с. 9-15.
- [2] Коротков К.Г. Основы ГРВ биоэлектрографии. – СПб, Изд-во СПбГИТМО, 2001. 360с.
-

Информация об авторах

Владимир Калмыков - старший научный сотрудник, кандидат технических наук, Институт проблем математических машин и систем, просп. акад. Глушкова 42, 03680, Киев 187, Украина;
e-mail: vl.kalmykov@gmail.com

Виталий Вишневецкий - заведующий отделением, старший научный сотрудник, кандидат технических наук, Институт проблем математических машин и систем, просп. акад. Глушкова 42, 03680, Киев 187, Украина; e-mail: vit@immssp.kiev.ua

Татьяна Романено - научный сотрудник, Институт проблем математических машин и систем, просп. акад. Глушкова 42, 03680, Киев 187, Украина; e-mail: romanenko@immssp.kiev.ua

METHOD OF FINDING HAMILTON ROUTES IN TRANSPORT NETWORK

Irina Garashchenko, Anatoliy Panishev

Abstract: This article discusses a solution method for Hamilton Problem, which either finds the task's solution, or indicates that the task is unsolvable. Offered method has significantly smaller requirements for computing resources than known algorithms.

Keywords: Hamilton traveling-salesman problem, minimal cost route, assignment problem, cyclic expansion.

ACM Classification Keywords: Algorithms

Introduction

The paper presents the method of Hamilton's Problem (HP) solution with significantly lesser requirements for computing resources, than for known methods.

Let's formulate the HP.

Let $H = (V, U)$ is a connected graph without closed loops and multiple edges, V – set of vertexes, $|V| = n$, U – set of edges. An edge $\{i, j\} \in U$ has the weight (cost) $d_{ij} \in Z_0^+$, $i, j = \overline{1, n}$, Z_0^+ is set of non-negative numbers. The symmetric matrix of weights $[d_{ij}]_n$ completely defines weighed graph $H = (V, U)$, and at this matrix $d_{ij} \in Z_0^+$ if $\{i, j\} \in U$ else $d_{ij} = \infty$, $i \neq j$, $d_{ii} = \infty$, $i = \overline{1, n}$.

Graph $H = (V, U)$ is Hamiltonian if it contains a Hamilton cycle, i.e. the simple cycle which is passing through all vertices of V exactly once. The cost $D(z)$ of the Hamilton cycle is equal to the sum of weights z edges – it is put in correspondence to this cycle.

The essence of HP is in finding a Hamilton cycle z^* of the minimum cost $D(z^*)$.

The task is NP - difficult [1]. Known methods of its solution are presented by schemes of the organization of exhaustive search of all cycles in graph H [2-5]. Practical implementation of these methods is problematic even with application of the most powerful computing systems.

The HP is also not always solvable. Therefore a unique method of finding a minimum cost Hamilton cycle z^* is the method which builds z^* if a HP it is solvable, or correctly indicates that graph $H = (V, U)$ is not Hamiltonian.

Let's setup the solution search in two stages. At the first stage sufficient conditions of non-Hamiltonian graph $H = (V, U)$ are checked. Complexity of checking each of them is estimated by a polynomial of a degree not above 3 from a size of task input data. A HP it is not solvable, if at least one of sufficient conditions of the non-Hamiltonian graph is fulfilled.

It is obvious, that if graph N contains final vertices the HP has no solutions. The condition of graph containing concatenation points being non-Hamiltonian is less obvious. Concatenation point is a vertex deleting which together with incidental edges, results in a disconnect graph [6]. Recognition of concatenation points in connected graph $H = (V, U)$ is fulfilled with complexity $O(|V| + |U|)$ [6]. It is not complicated to show, that graph is not Hamiltonian if it contains the bridge defined as such edge deleting which increases the number of connectivity components [6]. It is possible to fulfill recognition of bridges in time $O(|U| + |V|)$ in graph $H = (V, U)$.

At the second stage let's search the solution a HP z^* for the graph which is not containing final vertexes, points of a concatenation and bridges.

The search scheme of the minimum cost Hamilton path on a transport network.

The z^* searching algorithm is constructed according to the main scheme of the branch and bound algorithm. It calls procedure of solving an assignment problem (AP) for an evaluation of lower bounds of magnitude $D(z^*)$ [4]. At the same time, it has features inherent only in it which in the course of branching allow to define, what HP subtasks have no solution.

Let's consider a matrix of weights of HP $[d_{ij}]_n$. To calculate a lower bound of required magnitude $D(z^*)$, it is required to solve AP for this matrix. But AP with a matrix of weights $[d_{ij}]_n$ contains a part of elements $d_{ij} = \infty$ and may not have a solution. Therefore, for an evaluation of the lower bounds in the course of finding z^* the algorithm is required, which correctly discovers AP solution or strictly indicates, that AP has no solutions. Modification of Caen-Munkres algorithm works exactly in such way [4].

The algorithm of Caen-Munkres solves AP on a maximum with the assumption that all units $d_{ij} \neq \infty, i \neq j$. The input of the updated algorithm is the matrix $[d'_{ij}]_n$ where $d_{ij} = d - d_{ij}$ if $d_{ij} \neq \infty$ and $d'_{ij} = -\infty$ else, $d \neq \infty$ is a maximum element of matrix $[d'_{ij}]_n$. Then, if there is solution of AP on a maximum for a matrix $[d'_{ij}]_n$, it is the solution AP on a minimum for a matrix $[d_{ij}]_n$. Weights of solutions $C(\pi)$ and $C'(\pi)$ accordingly for matrixes $[d_{ij}]_n$ and $[d'_{ij}]_n$ are linked by equality

$$C(\pi) = nd - C'(\pi).$$

The updated algorithm of Caen-Munkres searches for an AP solution in a bipartite graph $K(X, Y, E)$, $|X| = |Y| = n$, $|E| = 2|U|$ corresponding to a matrix $[d'_{ij}]_n$ where the vertex $x_i \in X$ is connected to vertex $y_j \in Y$ by an edge (x_i, y_j) with weight $d(x_i, y_j) = d_{ij} \neq \infty$. The AP is solvable if the perfect matching π with the maximum sum of weights of edges is constructed in the graph $K(X, Y, E)$. The AP is unsolvable, if the graph $K(X, Y, E)$ does not contain perfect matchings.

The detailed description of Caen-Munkres algorithm modification is presented in [4]. Its main part is a known procedure of searching of a perfect matching in the bipartite non-weighted graph $H(X, Y, E)$, $|X| = |Y|$, $|E| = 2|U|$ with additional means of determining amount of units $d_{ij} = -\infty$ and their disposition in the matrix $[d'_{ij}]_n$ when AP is insoluble [4, 7]. Algorithm of Caen-Munkres and its modification both are characterized by the labor expenditures estimated by magnitude $O(n^4)$ [7].

The algorithm of finding z^* is fulfilled under the scheme of a branch and bound algorithm offered in [2] for symmetric TSP solution. A combination of this scheme with modification of algorithm of Caen-Munkres is applicable for a solution a HP as well.

Let's assume the perfect matching π is constructed. It supplies the target AP functional with a minimum $C(\pi)$, which is accepted as the lower estimate of weights of a required cycle z^* . Considering the matching π as permutation of columns of the weight matrix, we will present its cyclic expansion, i.e. as a set of non-intersecting cycles. Cyclic expansion of permutation π and the estimate $C(\pi)$ form the rout of a tree of branching. The permutation π presented by a unique cycle, is a z^* – HP solution. It's weight is $C(\pi)$.

Generally the AP solution contains some non-intersecting cycles. Let's select from them a cycle $\sigma = (v_1, v_2, \dots, v_k, v_1)$ with the minimum number of edges. Let's delete all solutions of AP, which contain a cycle σ , without losing any admissible solution z in an initial matrix. It is possible to present set of all solutions of an AP as a partition on k subsets because at least one of k edges $(v_1, v_2), (v_2, v_3), \dots, (v_k, v_1)$ should not be included in z . Let's designate AP with an initial matrix $[d_{ij}]_n$ as P_0 . Then P_0 is divided on k subtasks P_1, P_2, \dots, P_k . Each of those subtasks correspond in a one-to-one relationship to edges of cycle σ . Weights of edges of σ are set to ∞ in $[d_{ij}]_n$ matrix, all of the remaining weights are not modified. In the matrix $[d'_{ij}]_n$ of AP on maximum this edge's weight is assigned $-\infty$. Then if there exists a HP solution it belongs to a set of solutions of any of the subtasks P_1, P_2, \dots, P_k , which are presented by vertices of a branching tree, emerging from vertex P_0 .

In each subtask $P_i, i = \overline{1, k}$, it is possible to eliminate not only those solutions which containing the cycle σ , but also the solutions including cycles with vertices from set $S = \{v_1, v_2, \dots, v_k\}$. To achieve that, let's take the weight matrix of subtask P_i received from P_0 by assigning the element $d_{v_i, v_{i+1}}, i = \overline{1, k}, v_{k+1} = v_1$, of weight equal to ∞ . In this matrix let's set $d_{v_i, v_j} = \infty$ for all $v_i \in S \setminus \{v_j\}$. In the corresponding weight matrix of AP subtask on maximum each unit d_{v_i, v_j} gets weight $-\infty$.

For AP on maximum corresponding to subtask $P_i, i = \overline{1, k}$, let's apply modification of algorithm of Caen-Munkres to build a permutation π_i if P_i is solvable or find out that it has no solutions. If the subtask P_i is unsolvable the vertex corresponding to it in a tree of branching has no admissible prolongation. Let's suppose that from k subtasks P_i there are k_1 solvable subtasks $P_{i_s}, i \in \{1, 2, \dots, k\}, s = \overline{1, k_1}$, i.e. there were built optimal permutations π_{i_s} and values supplied by them were calculated $C(\pi_{i_s})$. Obviously, it is possible to limit the cost of the required Hamilton cycle z^* from below to magnitude

$$C = \min\{C(\pi_{i_s}) \mid i \in \{1, 2, \dots, k\}, s = \overline{1, k_1}\}.$$

Let's consider a subtask P_q for which $C(\pi_q) = C$. If the solution π_q is a Hamilton cycle it will be a solution of HP also. Otherwise permutation π_q produces several non-intersecting cycles. Then the node P_q of the solution tree is declared as top of branching [2]. The task P_q is divided into the subtasks which solutions do not contain a minimum length cycle σ from permutation π_q expansion. In solutions of subtasks all cycles generated on set of vertexes of σ are eliminated also. Having fulfilled modification of algorithm of Caen-Munkres for each received subtask P_{q_s} , we will define k_2 solvable subtasks, $q \in \{1, 2, \dots, k_1\}, s = \overline{1, k_2}$. Now a current lower bound of cost of an optimal Hamilton cycle z^* is magnitude

$$C = \min\left\{\min\{C(\pi_{q_s}) \mid q \in \{1, 2, \dots, k_1\}, s = \overline{1, k_2}\}, \min\{C(\pi_{i_s}) \mid i \in \{1, 2, \dots, k\}, s \in \{1, 2, \dots, k_1\}, i_s \neq q\}\right\},$$

corresponding to the task P_p . If all subtasks received from the task P_q are unsolvable the vertex of branching of a solution tree corresponds to the task P_p with a current estimation $C = \min\{C(\pi_{i_s}) \mid i \in \{1, 2, \dots, k\}, s \in \{1, 2, \dots, k_1\}, i_s \neq q\}$.

Further branching for vertex P_p is carried out in the same way as for vertex P_q . Finding the HP solution is completed in one of two cases. In the first case the algorithm finds a Hamilton cycle z^* for HP, which is a solution

of a AP with current value of C . In the second case algorithm determines that all finite vertexes of the solution tree are not subjects to further branching, and therefore the HP is unsolvable.

Graph H has no trailing vertexes and concatenation points. As a result of vertex-edges reorganization (VER) its only chain $(1, 8, 2)$ is replaced by an edge connecting vertexes 1 and 2. Therefore it is impossible to assert that graph H not Hamiltonian.

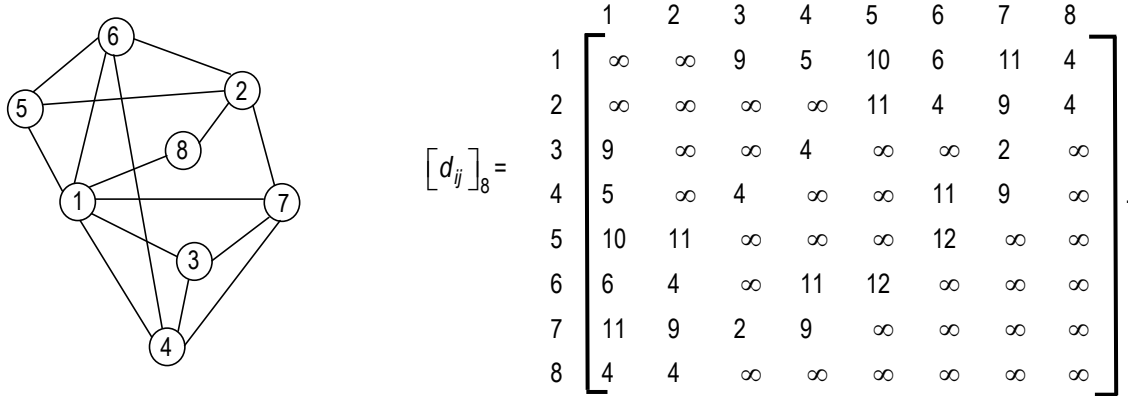


Fig. 1. Graph $H = (V, U)$ and a matrix of weights of its edges.

Let's apply modified algorithm of Caen-Munkres [4] to solve AP P_0 with an input weight matrix $[d'_{ij}]_8$.

The algorithm searches for the AP solution on maximum for a matrix

$$[d'_{ij}]_8 = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{matrix} & \begin{bmatrix} -\infty & -\infty & 3 & 7 & 2 & 6 & 1 & 8 \\ -\infty & -\infty & -\infty & -\infty & 1 & 8 & 3 & 8 \\ 3 & -\infty & -\infty & 8 & -\infty & -\infty & 10 & -\infty \\ 7 & -\infty & 8 & -\infty & -\infty & 1 & 3 & -\infty \\ 2 & 1 & -\infty & -\infty & -\infty & 0 & -\infty & -\infty \\ 6 & 8 & -\infty & 1 & 0 & -\infty & -\infty & -\infty \\ 1 & 3 & 10 & 3 & -\infty & -\infty & -\infty & -\infty \\ 8 & 8 & -\infty & -\infty & -\infty & -\infty & -\infty & -\infty \end{bmatrix} \end{matrix}$$

Optimal solution of the AP P_0 both on maximum and minimum is permutation $\pi = (4, 8, 7, 1, 6, 5, 3, 2)$, $C'(\pi) = 50$. The minimum of a target functional of AP P_0 is $C(\pi) = 46$. It limits from below weight of the required HP solution z^* . Cyclic expansion of permutation π looks like the following: $\sigma_1 = (1, 4, 1)$, $\sigma_2 = (2, 8, 2)$, $\sigma_3 = (3, 7, 3)$, $\sigma_4 = (5, 6, 5)$. Each cycle of expansion contains two edges. Thus for branching we will select any of four, for example $\sigma_3 = (3, 7, 3)$. On fig. 2 the branching tree of finding the Hamilton cycle z^* $C(\pi)$ is presented. All calculation results, which form the branching tree, are given in table 1.

The AP P_0 as a result of branching generates two tasks P_1 and P_2 on maximum with weight matrixes received from $[d'_{ij}]_8$. In matrix $[d'_{ij}]_8$ for task P_1 $-\infty$ is assigned to element $d'_{37} = 10$, and for task P_2 – to element $d_{73} = 10$. The modified algorithm of Caen-Munkres finds solutions π_1 and π_2 for these tasks and defines weights of received solutions $C(\pi_1) = C(\pi_2) = 49$. For branching it is possible to select any of subtasks with identical

estimates. Let's select task P_1 . Permutation π_1 is exhausted with two cycles, from which the cycle (3, 4, 7, 3) has the minimum length. Elimination of AP solutions containing this cycle and all cycles with vertices from set {3, 4, 7} produces three tasks P_{11}, P_{12}, P_{13} on maximum with corresponding weight matrices. In the first matrix $d'_{37} = d'_{34} = -\infty$, in the second $d'_{37} = d'_{43} = d'_{47} = -\infty$, in the third $d'_{37} = d'_{73} = d'_{74} = -\infty$. All three tasks are solvable. Weights of their solutions are $C(\pi_{11}) = 57, C(\pi_{12}) = 52, C(\pi_{13}) = 58$. By the current moment all final vertices of branching tree P_2, P_{11}, P_{12} are active P_{13} . As, $C = \min \{C(\pi_2) C(\pi_{11}) C(\pi_{12}) C(\pi_{13})\} = \min \{49, 57, 52, 58\} = 49$, the vertex P_2 appears as branching vertex.

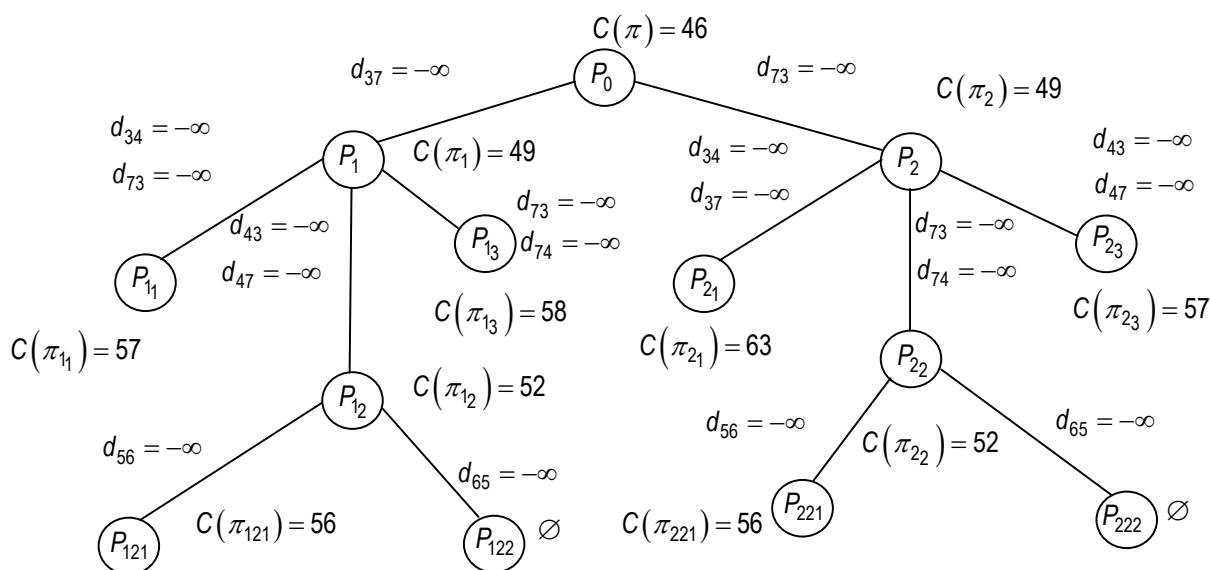


Fig. 2. The branching tree for HP with numerical data of an example 1.

Task P_2 is divided into three subtasks P_{21}, P_{22}, P_{23} with weights of solutions $C(\pi_{21}) = 63, C(\pi_{22}) = 52, C(\pi_{23}) = 57$. From current set of active vertexes $\{P_{11}, P_{12}, P_{13}, P_{21}, P_{22}, P_{23}\}$ of the branching tree we will select vertex with the minimum estimate $C = \min \{57, 52, 58, 63, 52, 57\} = 52$. Vertices P_{12}, P_{22} have equal rights to become the branching vertex. For further branching we will select subtask P_{12} . In cyclical expansion of permutation π_{12} the cycle (5, 6, 5) has the minimum length. Therefore the task P_{12} generates two subtasks P_{121} and P_{122} . The modified algorithm of Caen-Munkres builds AP solution of P_{121} in the form of a Hamilton cycle (1, 8, 2, 7, 3, 4, 6, 5, 1) with weight $C(\pi_{121}) = 56$ and determines that task P_{122} has no solutions.

Having built Hamilton cycle with weight 56 we no longer need to branch all active vertices with equal or greater estimates. There remains a single AP P_{22} whose weight of solution is equal to 52. Cyclical expansion of permutation π_{22} includes a cycle (5, 6, 5), calling two new subtasks P_{221} and P_{222} . AP P_{222} has no solution, and a solution of AP P_{221} is a Hamilton cycle (1, 5, 6, 4, 3, 7, 2, 8, 1) for which the total weight of edges $C(\pi_{221})$ is equal to 56.

The cost of the constructed Hamilton cycles is less, than a lower bound in any final vertex of the branching tree. Thus, the optimal solution of the HP are cycles $z_1^* = \pi_{121} = (1, 8, 2, 7, 3, 4, 6, 5, 1), z_2^* = \pi_{221} = (1, 5, 6, 4, 3, 7, 2, 8, 1)$.

□

Table 1.

AP	AP solution	Cycle expansion of AP solution
P_0	$\pi = (4, 8, 7, 1, 6, 5, 3, 2), C(\pi) = 46$	(1, 4, 1) (2, 8, 2) (3, 7, 3) (5, 6, 5)
P_1	$\pi_1 = (5, 8, 4, 7, 6, 2, 3, 1), C(\pi_1) = 49$	(1, 5, 6, 2, 8, 1) (3, 4, 7, 3)
P_2	$\pi_2 = (5, 8, 7, 3, 6, 2, 4, 1), C(\pi_2) = 49$	(1, 5, 6, 2, 8, 1) (3, 7, 4, 3)
P_{11}	$\pi_{11} = (4, 8, 1, 7, 6, 5, 3, 2), C(\pi_{11}) = 57$	(1, 4, 7, 3, 1) (2, 8, 2) (5, 6, 5)
P_{12}	$\pi_{12} = (8, 7, 4, 1, 6, 5, 3, 2), C(\pi_{12}) = 52$	(1, 8, 2, 7, 3, 4, 1) (5, 6, 5)
P_{13}	$\pi_{13} = (3, 7, 4, 3, 6, 5, 2, 1), C(\pi_{13}) = 58$	(1, 8, 1) (2, 7, 2) (3, 4, 3) (5, 6, 5)
P_{21}	$\pi_{21} = (8, 7, 1, 3, 6, 5, 4, 2), C(\pi_{21}) = 63$	(1, 8, 2, 7, 4, 3, 1) (5, 6, 5)
P_{22}	$\pi_{22} = (4, 8, 7, 3, 6, 5, 2, 1), C(\pi_{22}) = 52$	(1, 4, 3, 7, 2, 8, 1) (5, 6, 5)
P_{23}	$\pi_{23} = (3, 8, 7, 1, 6, 5, 4, 2), C(\pi_{23}) = 57$	(1, 3, 7, 4, 1) (2, 8, 2) (5, 6, 5)
P_{121}	$\pi_{121} = (8, 7, 4, 6, 1, 5, 3, 2), C(\pi_{121}) = 56$	(1, 8, 2, 7, 3, 4, 6, 5, 1)
P_{122}	AP is unsolvable	–
P_{221}	$\pi_{221} = (5, 8, 7, 3, 6, 4, 2, 1), C(\pi_{221}) = 56$	(1, 5, 6, 4, 3, 7, 2, 8, 1)
P_{222}	AP is unsolvable	–

Conclusion

The method has been implemented in C# programming language. For performance tests we have used Celeron 1.8GHz PC. Solution time for HP with cost matrix size around 60 in the worst case does not exceed 3 minutes.

Bibliography

- [1] Гэри М., Джонсон Д. Вычислительные машины и труднорешаемые задачи. – М.: Мир, 1982. – 416с.
- [2] Кристофидес Н. Теория графов. Алгоритмический подход. – М.: Мир, 1978. – 432 с.
- [3] Бондаренко М.Ф., Белоус Н.В., Руткас А.Г. Компьютерная дискретная математика. – Харьков: «Компания СМІТ», 2004. – 480с.
- [4] Панишев А.В., Плечистый Д.Д. Модели и методы оптимизации в проблеме коммивояжера. – Житомир: ЖГТУ, 2006. – 300с.
- [5] Вьялицин А. А. О перечислении гамильтоновых циклов // Дискретная математика. – 1991. – Т.3, вып. 3. – С. 46 – 49.
- [6] Рейнгольд Э., Нивергельт Ю., Део Н. Комбинаторные алгоритмы. Теория и практика. – М.: Мир, 1980. – 476 с.
- [7] Свами М., Тхуласираман К. Графы, сети, алгоритмы. – М.: Мир, 1984. – 454 с.

Authors' Information

Irina Vladimirovna Garashchenko – teacher-assistant of the chair of informatics and computer modelling, aspirant, Zhitomir State Technological University, Chernyahovskogo 103, Zhitomir, Ukraine, 10005; e-mail: irinagar@rambler.ru

Anatoliy Vasilyovich Panishev – head of the chair of informatics and computer modelling, professor, doctor of technical sciences, Zhitomir State Technological University, Chernyahovskogo 103, Zhitomir, Ukraine, 10005

ANALYSIS OF DERMATOGLYPHIC SIGNS FOR DEFINITION PSYCHIC FUNCTIONAL STATE OF HUMAN'S ORGANISM

**Anatoly Bykh, Elena Visotska, Vladimir Kobzev, Yury Mischeryakov,
Nataly Bilous, Irena Kulya, Alexander Kobzev, Nataly Makhaikina,
Irena Makartseva, Katarine Dyakonenko**

Abstract: Relation between dermatoglyphic signs and temperaments types is considered. An algorithm for papillary patterns classification and Izenk's two factors model are used for establishment relationship asymmetry signs with psychic functional state of human's organism.

Keywords: papillary patterns analyze, dermatoglyphic signs, temperament type

ACM Classification Keywords: J.3 Life and medical science. Medical information systems.

Conference topic: Information Systems

Introduction

Studies of papillary patterns have two goals. First goal is to discover distinction concrete human patterns from patterns all other people. To find common group's signs is the second goal.

Finger-print's patterns have several common features. They contain papillary lines which have the same anatomic nature, properties, like outward signs and details of structure.

It is known that skin integuments are descended from the same embryo sources that central nervous system's. Therefore dermatoglyphic signs may be considered as original markers of central nervous system's morphological organization.

In papillary patterns of hand's finger-print are distinguished five zones: distal (1), left lateral (2), central (3), right lateral (4) and basic (5).

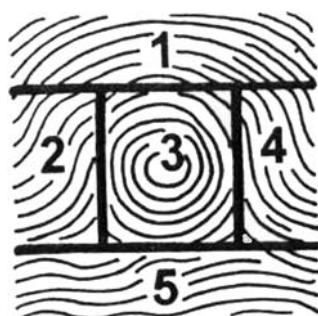


Fig 1 – Hand finger-print structure's zones

Across basic zone gone one stream of papillary lines, which usually cross nail phalange in horizontal direction. Left and right lateral zones are formed by streams arc like papillary lines. Distal zone is place above central zone, it separate from lateral streams by conditional line which touch extreme upper line of central drawing and is parallel to basic lines stream.

Maximum interest have present central zone. Its patterns contain one or two line's stream. In depend on stream's quantity patterns is divided for simple and double. By direct and stream line forms there are three global groups: arc-like, loop-like, and tangle wind-like.

Every group has several subgroups. Arc-like patterns may be simple, marquee, pyramidal or fir-treelike. Loop-like patterns contain ulnae, radial, half loops, closed loops and racket-like loops.

Circles, oval (ellipse-like), spirals, loop-spirals, tuber-like, snail loops and other come in wind-like (curl) patterns group. Sometimes papillary pattern's name is not corresponding to geometric figure one.

Usually in known algorithms for papillary pattern central zone analysis [1] have taken into account direct and density like («parallel») lines. In this way different version of pattern line's space frequency filtering are used. Discovered (filtered) distinctions are the basis for making decision in benefit one from standard papillary patterns. It's known, that across profile uniform papillary lines (summit and hollow altitude) well describes by next periodic function.

$$Z(\gamma, l, L) = \cos\left(\pi \frac{l}{L(\gamma)} + \varphi_0\right),$$

In this formula expression l is distinct from start point around line which across papillary lines with angle γ , $L(\gamma)$ is summit repetition period. Choice start point on line summit ($\varphi_0=0$) shorten parameter number in pattern analysis.

Summit repetition period L essentially depends from angle γ and in diapasons $0 < \gamma < \pi$ has been singular minimum which is reached in $\gamma = \pi/2$, that is with perpendicular across if papillary pattern lines. Patten's lines in central zone may have any direct. Therefore practice use will be enough to select only four most typical angle values γ from set $\{0, \pi/4, \pi/2, 3\pi/4\}$. On this statutes are based the next algorithm for central zone papillary pattern analysis.

1. Choice pattern's start point and division finger-point on 48 segments (6 circle stripes with equal width in 8 equal sectors).
2. Choice mean values M_i of grey color level $I(x, y)$ in the limit of segment number i .
3. Definition of differences $K=I(x, y)-M_i$ and standard deviations V_i in the limit of segment number i .
4. Definition of normalized finger image $N_i(x, y)$ in each segment number point.

$$N_i(x, y) = \begin{cases} M_0 + \sqrt{\frac{(v_0) \times (I(x, y) - M_i)^2}{v_i}}, & \text{if } I(x, y) > M_i, \\ M_0 - \sqrt{\frac{(v_0) \times (I(x, y) - M_i)^2}{v_i}}, & \text{otherwise.} \end{cases}$$

In this expression M_0 and V_0 are required mean and deviation values.

5. Division formalized finger image to four components by Gabor's space-frequency filtering [2] accordance to four directs which are defined by values $(0, \pi/4, \pi/2, 3\pi/4)$.

$$G(x, y, \gamma, f) = \exp\left\{-\frac{1}{2} \left(\frac{x'^2}{\delta_x^2} + \frac{y'^2}{\delta_y^2} \right)\right\} \times \cos(2\pi f x'),$$

$$x' = x \cdot \sin \gamma + y \cdot \cos \gamma,$$

$$y' = x \cdot \cos \gamma - y \cdot \sin \gamma.$$

In last expression f is sin-like wave frequency in direct γ , x' and y' are new coordinates for point (x, y) after turning coordinate system on angel γ , δ_x and δ_y describe Gauss rounding around axes X and Y .

6. Definition pattern's codes $F_{i\gamma}$ as standard deviation for four variants in 48 segments (in all 192 values)

$$F_{i\gamma} = \sqrt{\sum_{K_i} (G_i(x, y, \gamma) - M_{i\gamma})^2}.$$

In this expression $G_i(x,y,\gamma)$ is component image which correspondent to angle γ for segment number i , K_i – quantity of points in segment number i , M_{iy} – means of grey level in segment number i for concrete value of angle γ .

7. Analyzed pattern relate to one from five most typical kind (curl, left or right loop, arc, half sphere) accordance to maximum coincidence pattern's codes (192 measured signs vectors) with codes previously received etalons.

Next moments are used for most precision account disposition of the points which forms papillary pattern's lines.

If for reserved number of points are correct next expression

$$\left(r - \frac{\Delta}{2}\right)^2 \leq (x - x_0)^2 + (y - y_0)^2 \leq \left(r + \frac{\Delta}{2}\right)^2,$$

then this number of points forms circle with central coordinates x_0 and y_0 , radius r and circle line width Δ .

If for reserved number of points are correct two next expressions

$$\frac{(x - x_0)^2}{\left(a \pm \frac{\Delta}{2}\right)^2} + \frac{(y - y_0)^2}{\left(b \pm \frac{\Delta}{2}\right)^2} = 1 \quad \text{and} \quad b^2 = a^2 - c^2,$$

then this number of points forms ellipse with central coordinates x_0 and y_0 , half axes a and b , half focus distance c , ellipse line width Δ .

Increase distance from each new neighbour point of figure to some central point (x_0, y_0) which are accompanied simultaneous continuous increasing or decreasing of the turn angle straight line which going over new and central points testify to presence of spiral in papillary pattern.

This moments and classification papillary patterns results by described algorithm are the basis for research relationship hand finger nail phalange papillary pattern's type with psychic functional state of human's organism. Dermatoglyphic phenotypes of 78 persons with different temperaments were examined. Temperament type (choleric, sanguine, phlegmatic or melancholic) previously are defined by Izenk's two factors model. This model by means of basic index extraversion-introversion and neurotismus allows to estimate personality direction for internal or external world and to expose his emotional uneasiness level.

Papillary patterns were classified to five types: 1 – arcs (marquee arcs, fir-treelike), 2 – ulnae loops, 3 – another loops (radial, half loops, closed loops), 4 – curls (circle, ovals, spirals), 5 – another curls (tuber-like, snail loops). Described algorithm, partial analyses conditions and visual classification are used for this aims.

Results of these studies are presents into next table.

Sign	Presence dermatoglyphic signs of persons with different temperament type			
	Extravert		Introvert	
	Choleric	Sanguine	Melancholic	Phlegmatic
Asymmetry $\Delta r > \Delta l$ 1 finger	+	+	-	-
Asymmetry $\Delta r < \Delta l$ 2 finger	+	+	-	-
Asymmetry $\Delta r > \Delta l$ 2 finger	-	-	+	+

"+" means presence of sign, "-" means absence of sign

Accordingly to these results extraverts have next dermatoglyphic signs: asymmetry of three radiuses on thumb $\Delta r > \Delta l$ and asymmetry $\Delta r < \Delta l$ on forefinger.

Introvert's finger pattern allows to mark out dermatoglyphic type which is characterized by symmetric patterns on thumb and asymmetry $\Delta r < \Delta l$ on forefinger.

Next results are obtained in the sense relationship temperament's type with papillary patterns. The first, choleric have high probability of presence arcs patterns on the one from fingers. Second, melancholic have small probability of presence curl patterns, in particular circle.

Conclusion

Dermatoglyphic signs are marks central nervous system's types which are sensitive to external influence and this way they are connected with emotional uneasiness level.

As results studies of papillary patterns persons with various temperaments are ascertained that there are reasons to say about characteristic of them certain dermatoglyphic types.

Bibliography

1. Хрулёв А.А., Аполлонова И.А., Спиридонов И.Н. Автоматизированная обработка дерматоглифических изображений на аппаратно-программном комплексе «Малахит» / Биомедицинские технологии и радиозлектроника, 2004, №10. – с.50-54.
2. D. Gabor. Theory of communication. Journal of EEE, 93: 429-457, 1946.

Authors' Information

Anatoly Bykh – Doctor of Physics and Mathematics, professor, Head of Biomedical Electronic Devices and Systems Department of Kharkov National University of Radio Electronics, Lenina Av., 14, Kharkov, Ukraine, 61166.

Elena Visotska – PhD, lecturer of Biomedical Electronic Devices and Systems Department of Kharkov National University of Radio Electronics, Lenina Av., 14, Kharkov, Ukraine, 61166.

Vladimir Kobzev – PhD, lecturer of Information Control Systems Department of Kharkov National University of Radio Electronics, Lenina Av., 14, Kharkov, Ukraine, 61166; e-mail: vgkobzev@rambler.ru

Yuri Mischeryakov – PhD, lecturer of Systemotechnics Department of Kharkov National University of Radio Electronics, Lenina Av., 14, Kharkov, Ukraine, 61166; e-mail: eng.syst@gmail.com

Nataly Bilous – PhD, professor, lecturer of Automated Systems Software Department of Kharkov National University of Radio Electronics, Lenina Av., 14, Kharkov, Ukraine, 61166.

Irena Kulya – student of Electronic Devices Faculty of Kharkov National University of Radio Electronics, Lenina Av., 14, Kharkov, Ukraine, 61166.

Alexander Kobzev – student of Electronic Engineering Faculty of Kharkov National University of Radio Electronics, Lenina Av., 14, Kharkov, Ukraine, 61166.

Nataly Makhaikina – student of Computer Science Department of Kharkov National University of Radio Electronics, Lenina Av., 14, Kharkov, Ukraine, 61166.

Irena Makartseva – student of Computer Science Department of Kharkov National University of Radio Electronics, Lenina Av., 14, Kharkov, Ukraine, 61166.

Katarine Dyakonenko – student of Computer Science Department of Kharkov National University of Radio Electronics, Lenina Av., 14, Kharkov, Ukraine, 61166.

AUTOMATIC ECG ANALYSIS FOR PRELIMINARY AND DETAILED DIAGNOSTICS BASED ON SCALE-SPACE REPRESENTATION

Natalie Belous, Gleb Kobzar

Abstract: *A novel approach of automatic ECG analysis based on scale-scale signal representation is proposed. The approach uses curvature scale-space representation to locate main ECG waveform limits and peaks and may be used to correct results of other ECG analysis techniques or independently. Moreover dynamic matching of ECG CSS representations provides robust preliminary recognition of ECG abnormalities which has been proven by experimental results.*

Keywords: *electrocardiogram, ECG analysis, express-diagnostics, curvature scale-space, dynamic programming, dynamic time wrapping.*

ACM Classification Keywords: *I.5 Pattern Recognition.*

Introduction

Automatic diagnostics of cardiac diseases is the traditional task of medical cybernetics. At current large experience is accumulated in this area. In particular attempts of complete automation of heart activity diagnostics have failed to be enough robust comparatively to traditional human diagnostics. The most essential stage of traditional ECG analysis is recognition of its major elements, which consist of finding out the QRS-complex, selection of its characteristic points (tops of Q, R, S indents, scopes of indents and their borders).

The results of element analysis (annotation) and measuring of their parameters are used for ECG interpretation with purpose of correct diagnosis. Two basic categories of algorithms, applied in different systems of automatic diagnostics currently exist. Algorithms that design logic of doctor-diagnostician fall into the first category. The algorithms of the second category as a rule are based on the methods of multidimensional statistical analysis and theory of possibility.

Indisputable advantage of medical algorithms - in possibility rapid decision realization. It is conditioned by the fact that they concentrate experience of diagnostics accumulated in medicine and do not require preliminary teaching. Diagnostic possibilities of such algorithms are limited to the modern level of medicine development and quality of concrete algorithms – to the competence of persons assembling them as technicians and medics. Nevertheless location of necessary for traditional diagnostics points on ECG is not a trivial task.

Advantage of nonmedical diagnostic algorithms is that they can utilize any parameters of electrocardiogram representation. Due to it backlogs of information which in clinical practice remain unutilized appear accessible for such algorithms. The lack of these algorithms is complication of teaching and readiness. In case substantial difficulties are related to the selection of the well probed patients with diseases which an automat must learn to distinguish. Nevertheless, the algorithms of the second category are considered more perspective, because access to the new information should make diagnostics more effective.

Various automatic algorithms of the second group have been proposed, such as the threshold-crossing intervals X and the auto-correlation function [1] and algorithms based on neural-networks [2]. Time—frequency ($t - f$) analysis [3] and wavelet analysis [4] have also been used. All these algorithms suppose finding one special disease or detecting that ECG signal could be classified as normal.

This article focuses on building technique of ECG analysis that is able to support traditional diagnostics providing waveform limits and peaks and preliminary diagnostics to be performed at home or in non-medical environments by professional doctors or even users having no medical knowledge. The preliminary diagnostic results correspond to three possible binary decisions: normal ECG, disease, not known.

We propose a novel approach of normal ECG recognition which is based on scale-scale signal representation used for geometric object shape recognition previously - curvature scale-space (CSS) and our dynamic programming algorithm for matching ECG signal represented as CSS descriptors. Though the idea of scale-scale signal representation for ECG analysis is far not new [5] including the fact that wavelets are multi-scale by nature the approach proposed may have some valuable advantages over existing scale-scale approaches.

The article is organized in 4 main sections. The first section describes CSS briefly as it's not a well-known technology in medical signal processing. The second section introduces how to apply CSS to ECG signals. The third section describe waveform limits and peaks location technique. The fourth section focuses on dynamic programming matching algorithm proposed. And the last chapter describes some tests performed.

Curvature Scale Space representation of shapes

The Theory of Scale-Space Signal Representation was introduced by Vitkin and Coendric in 1983. The methodology consist in embedding a measured signal into a one-parametric family of derived signals, the scale-space, where the parameter, denotes scale parameter $\sigma \in \mathfrak{R}_+$, is intended to the current level of scale. For a signal $f : \mathfrak{R} \rightarrow \mathfrak{R}$, the scale-space representation $L : \mathfrak{R} \times \mathfrak{R}_+ \rightarrow \mathfrak{R}$ is defined as [6]:

$$L(x;0) = f(x). \quad (1)$$

And representation at coarser scales are given by convolution of the given signal with Gaussian kernels of successively increasing width:

$$L(x; \sigma) = g(x, \sigma) * f(x). \quad (2)$$

Curvature Scale-Space was introduced later by Mokharitan for geometric object shape representation as follows [7]. Having curvature of each closed contour $L(x; \sigma)$ points calculated, curvature zero-crossing points can be found easily. Then Curvature Scale-Space is built by locating zero-crossings in $(x; \sigma)$ space (u - normalized arc length, σ - Gaussian filter kernel size [7]). The resulting CSS can be represented as a binary image of CSS. Cross-sections of CSS by horizontal lines define position of zero-crossing points on the corresponding $L(x; \sigma)$.

It is obvious that CSS images of normalized curves are invariant to affine transformations and noise. Rotation of an object causes circular shift of it's CSS representation (fig.1a,b). The same effect is caused by the change of contour starting point. Because of normalization scaling also don't affect the view of CSS. In addition noise do not effects high scales of CSS (the threshold may be determined easily).

These properties of CSS image are used for effective representation and recognition of object shapes. It should be also noted that successful identification of shapes based on CSS representation don't require CSS Images themselves but arc maximums only []. The set of CSS maximums consists of pairs and form well-known CSS description which was selected as one of the main shape descriptors for MPEG-7 standard.

Application of CSS to ECG Representation

Curvature Scale-Space representation and matching process was introduced for closed curve contours only and are not suitable for signal analysis. We will show that it is possible to adapt CSS for ECG representation in this section and ECG matching based on such representation in the next section.

As curvature scale-space methodology has its roots in multi-scale signal representation area switching back from closed contours to signal segments representation is very easy. Signal curvature zero-crossing points may be acquired in the same way calculating the first and the second discrete differences (fig. 1a).

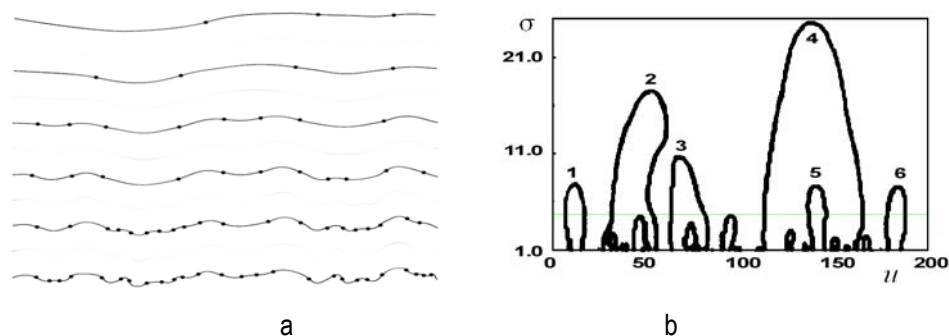


Fig. 1. Evolution of the signal and appropriate curvature zero-crossing points:
a – zero-crossings at different scales, b – CSS image and valuable arcs.

CSS images of such zero-crossings found on each scale of ECG signal evolving will look the same as for geometric object shape contour (fig.1b). The only difference here is in representation of CSS image maxima which affects matching process also. As closed contour may be started to detect from different points CSS image is also circular. In a difference an ECG signal starts and finishes at certain points. So CSS image for ECG and the set of maxima as well is not be shifted during acquisition and matching process.

ECG Waveform limits and peaks location

The most valuable result achieved using CSS for ECG representation is ability of physiologically correct and stable location of main waveform limits and peaks. The algorithm of waveform location based on wavelet QRS detection is described below.

Wavelet QRS detection algorithm is used first to locate largest peaks of ECG QRS complexes as it has been proven to provide the most precise and stable to noise QRS detection [9]. Unfortunately this algorithm is not able to provide correct waveform limit points automatically and require frequency parameter setup depending on the shape of waveforms (Fig.2).

Physiologically correct position of waveform limits are at points of maximum slope of the first order ECG derivative [10].

The CSS representation of ECG signal is built as described previously. First order ECG derivative is used instead of actual ECG signal to provide physiologically correct location of waveform limits at points of maximum slope which in the case of CSS are zero-crossings of the first order ECG derivative.

Having the largest QRS peak located using wavelet detector the cross-section scale level is then determined to locate the rest of waveform limits and peaks by the highest scale intersection point of QRS largest peak position and CSS arcs (green line at Fig.2).

The rest of points are located as follows. Correct QRS left and right limits lie at the previous and the next crossing points of the cross-section line accordingly.

Correct P wave maxima (positive or negative or maxima of bi-phases P-wave) and the left limit is located as two crossing points before the QRS left limit. Correct T wave maxima and two limits are located as three crossing after the point next to the QRS right limit.

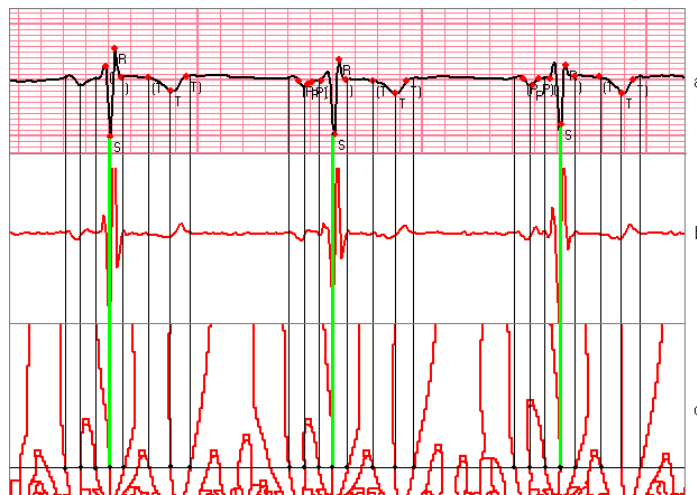


Fig. 2. ECG waveform limits and peaks location:
 a – ECG annotated using wavelet approach,
 b – ECG derivative,
 c – CSS of ECG derivative with a cross-section applied

Figure 2 shows that QRS limits located by the proposed algorithm are physiologically more correct than ones given by wavelet algorithm and allow correct and easy S and Q wave location. T and P wave limits are also shown to be more precisely located.

It is also seen that noise does not effect much correct waveform location by the algorithm which is caused by the nature of CSS representation. All noise affected CSS arcs are laying on the lowest scales of CSS (Fig.2c). Nevertheless large high-frequency noise may affect the results and addition high-pass filtration would be required then.

Matching CSS Representations of ECG for preliminary diagnostics

Matching algorithm for preliminary diagnostics is also proposed based on CSS representation of ECG. Every signal in the database is represented by the locations of the maxima of it's CSS images. Matching algorithm compares two sets of maxima and assigns a matching value to them which represent the similarity between the actual boundaries of objects (Fig.3a). Obviously as the nature of curves differs for contours and signals matching algorithm of CSS representations introduced by Mokhtarian with its main principle to shift maximums is not suitable for CSS representations of ECG signals [7]. Moreover the algorithm will fail to match curves with large dissimilarities of structure distances (heart rhythm change).

A different algorithm of matching CSS representations of ECG based on dynamic programming is proposed. It is organized in the same way as Dynamic Time Warping (DTW).

Having two sets of maxima $Q = q_1, q_2, \dots, q_b, \dots, q_n$ ($q_i = (s_i, x_i)$, where s is a scale parameter and x is position of the maxima) and $C = c_1, c_2, \dots, c_j, \dots, c_m$ ($c_i = (s_i, x_i)$) of length n and m respectively matching of these sets using proposed DP matching algorithm requires to construct an n -by- m matrix (DP matrix) where the (i^{th}, j^{th}) element of the matrix contains the distance $d(q_i, c_j)$ between the two points q_i and c_j (in a metric defined below). Each matrix element (i, j) corresponds to the alignment between the points q_i and c_j . This is illustrated in Figure 3b. A warping path W , is a continuous (in the sense stated below) set of matrix elements that defines a mapping between Q and C . The k^{th} element of W is defined as $w_k = (i, j)_k$ so we have:

$$W = w_1, w_2, \dots, w_k, \dots, w_K \quad \max(m, n), K < m+n-1$$

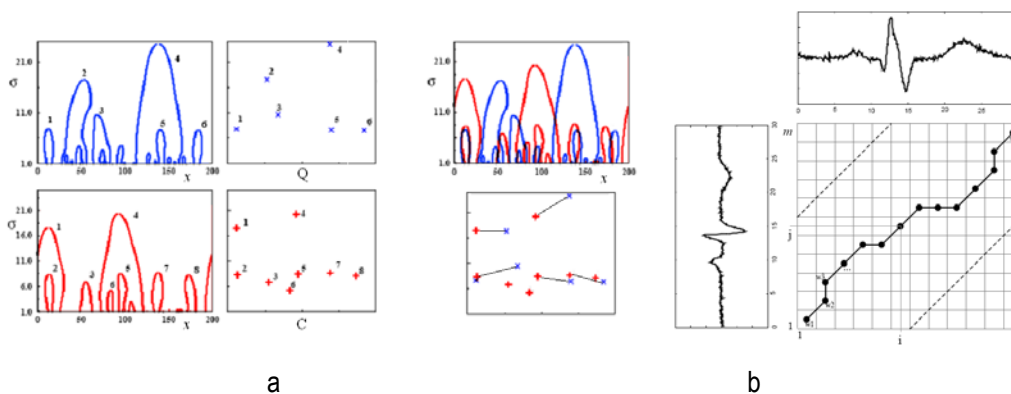


Fig. 3. Dynamic matching of two ECG signals: a – CSS maximas, b – dynamic matching

The warping path is typically subject to several constraints:

- Boundary conditions: $w_1 = (1, 1)$ and $w_K = (m, n)$, simply stated, this requires the warping path to start and finish in diagonally opposite corner cells of the matrix.
- Continuity: Given $w_k = (a, b)$ then $w_{k-1} = (a', b')$ where $a - a' \leq 1$ and $b - b' \leq 1$. This restricts the allowable steps in the warping path to adjacent cells (including diagonally adjacent cells).
- Monotonicity: Given $w_k = (a, b)$ then $w_{k-1} = (a', b')$ where $a - a' \geq 0$ and $b - b' \geq 0$. This forces the points in W to be monotonically spaced in time.

There are exponentially many warping paths that satisfy the above conditions, however we are interested only in the path which minimizes the warping cost:

$$\Omega(Q, C) = \min_{(Q', C', f)} \left\{ \sum_{i=1}^{|Q|} \sqrt{(x^{-1}(q_i) - x^{-1}(f(q_i)))^2 + (\sigma(q_i) - \sigma(f(q_i)))^2} + \sum_{q \in Q' / Q'} \sigma(q) + \sum_{q \in C' / C'} \sigma(c) \right\} \tag{3}$$

where $Q', C', |Q'| = |C'|$ - are subsets of Q, C accordingly,

$(x^{-1}(q_i) = x(q_i) - x(q_{i-1}), x(q_0) = 0$ - is a distance to previous CSS maxima (time interval).

The path can be found very efficiently using dynamic programming to evaluate the following recurrence which defines the cumulative distance $\gamma(i, j)$ as the distance $d(i, j)$ found in the current cell and the minimum of the cumulative distances of the adjacent elements:

$$\gamma(i, j) = d(q_i, c_j) + \min \{\gamma(i-1, j-1), \gamma(i-1, j), \gamma(i, j-1)\} \quad (4)$$

where $\gamma(i, j)$ - accumulated cost of w_k DP table cell,

$d(q_i, c_j)$ - wrapping cost in w_k which is calculated as follows:

$$\begin{aligned} d_{match}(q_i, c_j) &= \sqrt{x^{-1}(q_i) - x^{-1}(f(q_i))}^2 + (\sigma(q_i) - \sigma(f(q_i)))^2 \\ d_{skipi}(q_i, c_j) &= \sigma(q_i), d_{skipj}(q_i, c_j) = \sigma(q_i) \end{aligned} \quad (5)$$

where $d_{match}(q_i, c_j)$ - distance for matching two points or a diagonal movement,

$d_{skipi}(q_i, c_j)$ - distance for skipping the point q_i or a horizontal movement,

$d_{skipj}(q_i, c_j)$ - distance for skipping the point q_j or a vertical movement.

Such formulation of $d(q_i, c_j)$ has already been proven in practical application of geometric shape matching as a part of Mokhtarian's algorithm.

Finally the cost of matching could be found in $\gamma(n, m)$ using backtracking algorithm while building the optimal path W .

Experimental Results

We have also evaluated our algorithm in terms of important interval values P duration QRS duration and QT interval. The evaluation has been done in terms of the standard deviation of the differences between intervals estimated by our algorithm (10.2ms), intervals measured by wavelet technique (12.3ms) and intervals measured from the mean referee estimates of CSE database (9.6ms) which proves better performance of proposed algorithm. 120 beats was used provided for program testing and 30 expert annotated beats contained in CSE.

Also 51 ECG signals was selected for preliminary diagnostics testing from the international ECG database PhysioNet (Physikalisch-Technische Bundesanstalt – PTB, the National Metrology Institute of Germany).

Table 1. Distribution of classes

Diagnostic class	Number of ECG signals
Myocardial infarction	5
Cardiomyopathy/Heart failure	8
Bundle branch block	5
Dysrhythmia	4
Myocardial hypertrophy	5
Valvular heart disease	5
Myocarditis	5
Miscellaneous	4
Healthy controls	10
Total	51

Techniques of CSS representation and dynamic matching were realized as a computer program which was tested on the assembled database of ECG signals.

Following the well known technique of automatic ECG interpretation quality estimation [9] we assume that all database signals are verified (normal forms and pathological forms). Then counting up the general number of the followings events: correct classification of normal ECG (TN), improper classification of normal ECG as pathological (FP), improper classification of pathological ECG as normal (FN), correct classification of pathological ECG (TP).

In total, quality of ECG interpretation is calculated as "probability that classification is correct" according to the following formula:

$$TA = \frac{TP + TN}{TP + FP + TN + FN} 100\% \quad (6)$$

Thus according to experiment results TP=41, TN=9, FP=1, FN=0 and resulting TA=98%. After introduction of the third diagnostic result – "not known" FN=0 may be achieved. In the case if a user received "not known" result he can repeat the measurements. And if the same result achieved again it would be recommended to consult with professional doctor-cardiologist.

Conclusion

A novel approach of automatic ECG analysis based on scale-scale signal representation is proposed. The approach uses curvature scale-space representation to locate main ECG waveform limits and peaks and may be used to correct results of other ECG analysis techniques or independently. Moreover dynamic matching of ECG CSS representations provides robust preliminary recognition of ECG abnormalities.

The main advantage over existing waveform location techniques is physiologically correct and noise stable detection of waveform limits.

The main advantage of the approach over the existing preliminary diagnostic techniques is faster extraction process and matching process as it is done in CSS space. Experimental results show that the approach is quite robust for preliminary normal ECG recognition. It is planned to conduct larger tests on different ECG databases and comparative tests to compare the quality of diagnostics and the rapidness with other interpretation techniques.

Bibliography

- [1] R.H. Clayton, A. Murray, R.W.F. Campbell, Comparison of four techniques for recognition of ventricular fibrillation of the surface ECG, *Med. Biol. Eng. Comp.* 31 (1993), pp. 111—117.
- [2] K. Minami, H. Nakajima, T. Toyoshima, Real-time discrimination of ventricular tachyarrhythmia with Fourier transform neural network, *IEEE Trans. Biomed. Eng.* 46 (2) (1999), pp. 179—185.
- [3] V.X. Afonso, W.J. Tompkins, Detecting ventricular fibrillation, *IEEE Eng. Med. Biol.* 14 (1995) pp. 152—159.
- [4] A.S. Al-Fahoum, I. Howitt, Combined wavelet transformation and radial basis neural networks for classifying life threatening cardiac arrhythmias, *Med. Biol. Eng. Comput.* 37 (1999), pp. 566—573.
- [5] F. Jager; I. Koren; L. Gyergyek Multiresolution representation and analysis of ECG waveforms, *Computers in Cardiology Proceedings*. Volume 23-26 (1990), PP.:547 – 550
- [6] T. Linderberg *Scale Space Theory in Computer Vision*, Kluwer Academic Publishers. (1994), 440p.

- [7] F. Mokhtarian, S. Abbasi, J. Kittler Efficient and Robust Retrieval by Shape Content through Curvature Scale Space, Int. Workshop on Image DataBases and Multimedia Search (1996), pp. 35-42.
- [8] H. Strik, L. Boves Averaging physiological signals with the use of a DTW algorithm. In Proceedings of SPEECH'88, 7th FASE Symposium, Edinburgh, Book 3 (1988), pp. 883-890.
- [9] Y.C. Chesnokov, D. Nerukh D., R.C. Glen Individually adaptable automatic QT detector. In Proc. Of Computers in Cardiology (2006) pp. 337-340.
- [10] P. Laguna, R. Jané, P. Caminal Automatic detection of wave boundaries in multilead ECG signals: validation with the CSE database, In Proc. To Comput Biomed Res. Vol. 1 (1994), pp. 45-60.
-

Authors' Information

Natalie Belous – Ph.D., head of ITLCVS laboratory, professor of Kharkov National University of Radio Electronics, Kharkov, Ukraine; e-mail: belous@kture.kharkov.ua

Gleb Kobzar – Researcher; Kharkov National University of Radio Electronics, Kharkov, Ukraine; e-mail: belous@kture.kharkov.ua

ОТНОШЕНИЯ АССОЦИАЦИИ ДЛЯ АНАЛИЗА ПРЕДМЕТНЫХ ОБЛАСТЕЙ УЧЕБНЫХ КУРСОВ

Вера Любченко, Алексей Ярмула

Abstract: Different kinds of associative relations are described. There are defined measures for power of these associative relations.

Keywords: associative relation, associative map, measure

ACM Classification Keywords: I.2.4 Knowledge Representation Formalisms and Methods

Введение

Существует довольно обширный перечень вопросов, при решении которых представление, базирующееся на формализме ассоциативных сетей, оказывается весьма полезным. Одним из них является вопрос анализа предметных областей учебных курсов определенной учебной программы.

Предметная область любого учебного курса имеет свои базовые понятия – *концепты* – конструкции, описывающие все объекты в данной категории или классе сущностей, их взаимодействия, свойства и связи между ними. Каждый концепт характеризуется своим контекстом, задаваемым тремя компонентами:

- множеством учебных целей, для достижения которых необходимо изучение концепта,
- предусловием – множеством категорий сущностей, знание которых необходимо для изучения данного концепта,
- постусловием – категорией сущностей, которая изучается в данном концепте.

Анализ предметной области учебного курса должен определить силу смысловой связи между концептами и позволить выполнить его обоснованную грануляцию на тематические компоненты. Фактически, в ходе анализа проверяются структурированность и логическая связность учебного материала курса, являющиеся главными факторами при определении его качества.

В данной работе рассматривается множество отношений ассоциации, которые могут быть применены для анализа предметных областей учебных курсов.

Базовые отношения ассоциации

В [Любченко, 2007а] предложено рассматривать два типа ассоциативных связей:

- ассоциация по цели – отношение, которое показывает, что контексты концептов имеют общие элементы во множествах учебных целей;
- ассоциация по логике – отношение, которое показывает, что постусловие одного из них входит в предусловие второго.

Для каждого типа ассоциативной связи можно ввести количественный измеритель силы связи. Сила полной ассоциативной связи между двумя концептами может быть вычислена как сумма мер силы ассоциации по цели ass_G и ассоциации по логике ass_L :

$$ass(c_i, c_j) = ass_G(c_i, c_j) + ass_L(c_i, c_j),$$

где c_i и c_j – концепты, для которых рассчитывается сила полной ассоциативной связи.

Возможность идентифицировать парные ассоциации между двумя концептами позволяет ввести в рассмотрение более сложные конструкции, которые могут быть полезны для анализа содержания учебного материала. Дадим определение.

Ассоциативная карта – набор концептов учебного материала с определенными на них ассоциативными связями, математической моделью которой является неориентированный граф.

Используя ассоциативную карту, можно рассчитать значения двух показателей:

- *связность тематической компоненты* – сила ассоциативных связей между концептами данной компоненты, которую можно рассчитать как сумму весовых коэффициентов дуг между вершинами, соответствующими этим концептам;
- *сцепление тематической компоненты* – сила ассоциативных связей данной компоненты с остальными тематическими компонентами, которую можно рассчитать как сумму весовых коэффициентов дуг, соединяющих подграф данной тематической компоненты с остальными вершинами ассоциативной карты.

Дополнительно для тематической компоненты можно ввести в рассмотрение коэффициент разбиения – отношение значения сцепления к значению связности.

Относительно материала учебного курса связность между концептами дает возможность определить, является курс обзорным или специализированным. Можно предположить, что у специализированных курсов значение показателя связности выше, чем у обзорных. Сцепление показывает, насколько сильно учебный курс взаимодействует с другими курсами учебной программы. Можно предположить, что у обзорных курсов значение показателя сцепления выше, чем у специализированных.

Исследования поведения базовых отношений ассоциации

Исследования поведения базовых отношений ассоциации проводились для учебных курсов, информация о которых приведена в [Рекомендации, 2007]. Этот сборник содержит руководящие принципы создания учебных планов по программной инженерии. Рекомендации отражают обширный реальный опыт авторского коллектива и результаты обсуждения в рабочих группах. Таким образом, можно считать эффективность разбиения в данном случае эмпирически подтвержденной.

Информация о рекомендуемых курсах, их учебных целях и изучаемых концептах была собрана в базе фреймов [Любченко, 2007б], на основе которой строилась ассоциативная карта предметной области.

При построении ассоциативной карты основную сложность представляло построение списков связанных концептов, на основании которых рассчитывались силы ассоциативных связей между концептами. Для этого использовался рекурсивный алгоритм, использующий в качестве исходных данных список концептов предметной области. Для каждого концепта выполнялась такая последовательность действий:

1. Создается список предусловий текущего концепта, который образует список найденных концептов (СНК). Запоминается общее число найденных концептов
2. Определяется множество целей, с которыми связан концепт. Для каждой цели:
 - 2.1. Находятся все связанные с ней концепты и добавляются в СНК. Общее число найденных концептов увеличивается на количество концептов, добавленных в СНК.
 - 2.2. Находятся курсы, которым она принадлежит, и связанные с ней в этих курсах концепты. Список этих концептов добавляется в СНК. Общее число найденных концептов увеличивается на количество концептов, добавленных в СНК.

В результате выполнения алгоритма получается список всех концептов, связанных с текущим. Этот список использовался для расчета матрицы ассоциативных связей между концептами, на основании которой с использованием информации из базы фреймов рассчитывались значения показателей связности и сцепления.

Результаты расчетов показали, что для обзорных курсов значение показателя сцепления в 1,5-2 раза превышает значение показателя связности. Для специализированных курсов ситуация противоположна: значение показателя связности в 2-4 раза превышает значение показателя сцепления.

Дополняющие отношения ассоциации

Отметим тот факт, что базовые отношения ассоциации позволяют учесть не все связи, возникающие между концептами учебного курса.

Часто построение учебных курсов основано на рекомендуемом порядке изложения материала, однако, этот порядок не обусловлен логикой (соотношением множеств пред- и постусловий). Для учета этой информации следует ввести в рассмотрение ассоциацию по времени, позволяющую учесть для каждого концепта наличие зависимости от ранее изученных. Дадим определение.

Будем говорить, что между двумя концептами существует отношение временного порядка, если их рекомендуется изучать в определенной временной последовательности:

$$R_T(c_i, c_j), \text{ если } c_i \prec_t c_j,$$

где $c_i, c_j \in C$ — концепты. Это отношение антирефлексивно, антисимметрично и транзитивно.

Если вводится отношение временного порядка, то нужно ввести меру ассоциативной связи по времени. Эту меру предлагается определять следующим образом:

$$ass_T(c_i, c_j) = \frac{|Out_i \cap In_j| + 1}{|In_j \cup Pre_j|},$$

где In_j — множество предусловий концепта c_j , Out_i — множество постусловий концепта c_i , Pre_j — множество концептов, рекомендованных к изучению по времени раньше концепта c_j .

Базовые отношения ассоциации неявно предполагают, что концепты выполняют одинаковые функции в курсе. Однако, в зависимости от качественных характеристик взаимосвязи между учебными целями и концептами, можно выделить три типа концептов:

- основные концепты, изучение которых необходимо для достижения учебной цели курса;
- подготовительные концепты, которые необходимы для понимания и работы с учебными материалами основных концептов определенной учебной цели;
- родственные концепты, которые связаны с основными концептами специфичным образом, дополняя (расширяя) их.

Пусть Gr_i — множество подготовительных концептов, Rel_i — множество родственных концептов для концепта c_i . Дадим определения.

Будем говорить, что между двумя концептами существует ассоциация по подготовке, если концепт c_j является подготовительным для концепта c_i :

$$R_P(c_i, c_j), \text{ если } c_j \in Gr_i.$$

Будем говорить, что между двумя концептами существует ассоциация по расширению, если концепт c_j является родственным для концепта c_i :

$$R_R(c_i, c_j), \text{ если } c_j \in Rel_i.$$

Оба отношения антирефлексивны, антисимметричны и, в общем случае, нетранзитивны.

Соответственно, введем две меры для этих типов ассоциативной связи:

- $ass_P : F^2 \rightarrow [0,1]$ — вычисление меры ассоциативной связи по подготовке, которую предлагается определять как значение отношения

$$ass_P(c_i, c_j) = \frac{\max(1, |In_i \cap Gr_j|)}{|In_i \cup Gr_j|};$$

- $ass_R : F^2 \rightarrow [0,1]$ – вычисление меры ассоциативной связи по расширению, которую предлагается определять как значение отношения

$$ass_R(c_i, c_j) = \frac{|Cb_j|}{|Rel_i|},$$

где $Cb_j = \{c_k : c_j \in Rel_k, \forall k\}$ – множество концептов, для которых c_j является родственным.

После введения дополняющих отношений ассоциации сила полной ассоциативной связи может быть вычислена как сумма мер силы всех типов ассоциаций:

$$ass(c_i, c_j) = ass_G(c_i, c_j) + ass_L(c_i, c_j) + ass_T(c_i, c_j) + ass_P(c_i, c_j) + ass_R(c_i, c_j).$$

Заключение

В работе рассмотрено множество отношений ассоциации и их мер, которые могут быть использованы для анализа предметных областей учебных курсов. Собственно методика анализа учебного курса на основе ассоциативных карт, моделирующих предметную область курсов, в работе не рассматривается. Но можно утверждать, что анализ структуры целесообразно выполнять на основе гипотезы компактности [Загоруйко, 1999]. Соответственно, результаты вычислений силы ассоциативной связи можно использовать следующим образом:

- при изменении структуры учебных курсов на основе анализа изменений показателей сцепления и связности можно оценить величину эффекта данного изменения и принять решение о целесообразности этого изменения;
- при анализе учебных курсов определенной учебной программы (например, учебных курсов одной специальности) значительный разброс значений коэффициентов разбиения может служить индикатором неэффективного разделения учебного материала по курсам.

Список литературы

- [Загоруйко, 1999] Загоруйко. Н.Г. Прикладные методы анализа данных и знаний. – Новосибирск: Издательство Института математики, 1999.
- [Любченко, 2007а] Любченко В., Крисилов В. Использование ассоциативных сетей для грануляции материалов курса // Proceedings of the Second International Conference "Modern (e-) Learning" – Varna, 2007. – Sofia: FOI ITHEA, 2007. – P.61–64.
- [Любченко, 2007б] Любченко В.В., Нестеренко А.В. Фреймовая модель предметной области для уровневого представления учебного курса // Образование и виртуальность – 2007. Сборник научных трудов 11-й Международной конференции Украинской ассоциации дистанционного обучения. – Харьков-Ялта: УАДО, 2007. – С.122–128.
- [Рекомендации, 2007] Рекомендации по преподаванию программной инженерии и информатики в университетах. – М.: ИНТУИТ.РУ «Интернет-Университет Информационных Технологий», 2007. – 462 с.

Сведения об авторах

Вера Любченко – доцент, Одесский национальный политехнический университет, пр. Шевченко, 1, Одесса 65044, Украина; e-mail: vira.lyubchenko@gmail.com

Алексей Ярмула – магистр, Одесский национальный политехнический университет, пр. Шевченко, 1, Одесса 65044, Украина; e-mail: oleksiy.yarmula@gmail.com

МЕТОДЫ ИНТЕГРАЦИИ УЧЕБНЫХ МАТЕРИАЛОВ ДЛЯ ДИСТАНЦИОННОГО ОБУЧЕНИЯ

Игорь Шубин, Ярослав Святки, Ирина Белоус

Abstract: *The process of training is the most difficult for effective realization through information technologies. Is suggested the methods for the most complete implementation of original techniques of material description, ensuring versatility of development environment and functioning of interactive systems of training process. The given technology requires as the exact description of teaching model, as application of modern methods of development intelligent skills.*

Keywords: *Distance Learning, Computer Based Learning System, Fuzzy Logics, Fuzzy Nondetermined Automat.*

ACM Classification Keywords: *K.3.1 Computer Uses in Education*

Введение

Наиболее часто встречающейся проблемой, которая стоит перед авторами учебных курсов при компьютерном (дистанционном) обучении, является интеграция разнородных учебно-методических материалов в единый курс обучения. Актуальной является задача создания инструментальных систем поддержки интеграции учебно-методических материалов, к которым предъявляются жесткие требования, так как эффективность обучения существенно зависит от формы и качества предоставления учебных материалов. В основу подобных систем закладывается распределенная модель хранения информации.

Основу содержания современных учебных курсов должны составлять осваиваемые способы деятельности, а не конкретное предметное содержание. С точки зрения сетевых технологий, инструментальная система поддержки интеграции учебно-методических материалов должна выполняться на специально выделенном сервере обучения. Также необходимо отметить, что вся инструментальная система в целом и серверная часть в частности, должны отвечать минимальным требованиям безопасности, не допуская вмешательства обучаемых в процесс квазислучайной выдачи контрольных заданий и оценивания результатов его выполнения. Данное обстоятельство обуславливает необходимость многопоточной серверной реализации тех функций системы, которые связаны с проверкой ответов обучаемых и предоставлением доступа к системе и к базам данных учебных курсов.

Формализация структуры учебных курсов

Учебный курс представляется в виде направленного графа, в котором узлами служат документы специального вида – слайды, а ребрами – переходы между ними. Из каждого узла обучаемый может попасть в один из нескольких других узлов, непосредственно связанных с ним ребрами. Какой переход будет выбран, решает система на основе данных об обучаемом и текущего состояния учебного процесса. В первую очередь при решении используются протокол работы и модель обучаемого. Протокол представляет собой файл, в который автоматически заносится информация о всех действиях, которые обучаемый выполнял за терминалом компьютера. На основе анализа протокола, при необходимости, можно скорректировать модель обучаемого. Модель обучаемого должна содержать ответы на вопросы не только о том, что знает и умеет делать конкретный обучаемый, но и о том, к какому психологическому типу он относится и чего достиг за время обучения.

Некоторые узлы курса могут быть помечены как обязательные для посещения. Стартовый узел выбирается на основе уровня подготовки обучаемого, выявленного с помощью блока вступительного тестирования. Выбор конкретного перехода (направления дальнейшего развития учебного процесса) осуществляется на основе правил, связанных с каждым из ребер перехода.

Причем инструментальная система позволяет разработчикам вносить коррективы в эти правила с помощью средств визуального редактирования без необходимости ввода текста правил вручную.

Таким образом, универсальная КПУН, входящая в состав инструментальной системы поддержки интеграции учебно-методических материалов является интеллектуальной компьютерной обучающей системой (ИКОС) с накоплением информации, управляемой правилами. Обновление базы знаний КПУН производится с помощью инструментальной системы по мере надобности, в том числе и по результатам апостериорной статистической обработки матрицы результатов измерения знаний достаточно большого количества обучаемых. Использование такого подхода к проектированию инструментальных систем интеграции различных материалов в учебный курс позволяет достичь следующих целей:

- реализовать индивидуальный подход к объекту обучения. ИКОС являются особым классом обучающих систем, в структуру которых закладываются механизмы адаптации к конкретному объекту обучения. В состав ИКОС включаются вычислительные, классифицирующие, контрольно-диагностирующие средства, позволяющие оценивать параметры объекта обучения и вырабатывать оптимальную стратегию и тактику обучения на каждом этапе.
- проводить обучение в режиме реального времени. Процедуры адаптации должны выполняться с высокой скоростью, поскольку любые технологические задержки в работе системы можно рассматривать как возмущающие воздействия, влекущие за собой отклонения процесса обучения от оптимального.

Информация в узлах курса обучения должна быть разбита на небольшие смысловые части – слайды. Слайды могут быть двух типов: интерактивные и информационные. Информационные слайды состоят только из объектов отображения текстовой и графической информации, воспроизведения аудио- и видеофрагментов и кнопок управления ходом учебного процесса. Интерактивные слайды, кроме перечисленных объектов, могут также содержать объекты, реализующие различные виды тестовых заданий для организации обратной связи с обучаемым и оперативного контроля усвоения знаний, а также для приобретения обучаемым необходимых навыков. В текущей версии данной инструментальной системы применительно к интерактивным слайдам используется биномиальная модель педагогического теста. В рамках этой модели результат выполнения каждого задания оценивается по бинарной шкале: 1, если испытуемый успешно выполнил задание, содержащиеся в данном интерактивном слайде; 0 – в противном случае. Таким образом, в учебном курсе каждый интерактивный слайд имеет как минимум два исходящих ребра: одно соответствует успешному выполнению тестового задания, второе – неправильному решению задания. При таком подходе, количество интерактивных слайдов в блоке интерактивного обучения, а также количество вопросов в блоке заключительного тестирования определяется следующими параметрами: N – искомое количество интерактивных слайдов; p – вероятность успешного выполнения тестового задания (истинный балл обучаемого, характеризующий истинный уровень обученности); d – надежность определения параметра p ; Φ_0 – интеграл вероятностей (функция Лапласа); ϵ – точность определения истинного балла обучаемого p .

Задавая значения величин точности (ϵ), надежности (d) и, учитывая текущий уровень обученности (p), с помощью формул можно определить требуемое количество интерактивных слайдов в подготавливаемом курсе обучения. В случае возникновения необходимости проведения комплексной объективной аттестации с высокой степенью точности ($\epsilon = 0,05$) и надежности ($d = 0,9$), для испытуемых с истинным уровнем обученности $p = 0,7$ потребуется подготовить $N = 228$ интерактивных слайдов. Заметим, что

объективные зарубежные тесты для профессиональной аттестации (для адвокатов, врачей, бухгалтеров и др.) содержат, как правило, от 220 до 250 тестовых заданий, при p в диапазоне от 0,7 до 0,75. Однако для определения качества усвоения материалов учебного курса не требуется чрезвычайно высокая точность педагогического измерения. Для этой цели вполне подойдет средняя точность и надежность, поэтому величины ε и d можно положить равными 0,1 и 0,8 соответственно. В этом случае, для определения истинного балла $p = 0,6$ достаточно подготовить 40 интерактивных слайдов.

Нечеткая модель обучаемого

При личных контактах со студентом преподаватель имеет большую возможность определить личностные качества и индивидуальные особенности обучаемого. Это позволяет ему адекватно корректировать процесс обучения, изменяя форму и содержание материала, тактику и стратегию обучения, а иногда и методологию преподавания. Для этого преподаватель использует субъективно осознанную модель обучаемого. В случае компьютерного (дистанционного) обучения такой возможности у преподавателя практически нет.

Подобную форму обучения вполне логично рассматривать как человеко-машинную систему, в которой обучаемый и преподаватель выступают в роли человека-оператора. Если обучаемый управляет системой приобретения знаний, то преподаватель опосредованно управляет процессом обучения. Как и при проектировании человеко-машинной системы при создании ИСРКОС следует решать вопросы перераспределения функций между человеком и вычислительными средствами.

С этой целью предлагается использовать в процессе обучения индивидуальную модель обучаемого, которая строится на основе экспертных оценок [1]. При этом предполагается, что процесс обучения представляет собой последовательную совокупность освоения отдельных разделов учебного материала. Каждый этап обучения заканчивается тестированием. Обучаемый может управлять процессом своего обучения посредством выбора способов освоения учебного материала.

Под моделью обучаемого будем понимать нечеткий недетерминированный автомат вида:

$$A = \langle U, X, Y, s_0, \delta, \sigma \rangle, \quad (1)$$

где $U = \{U_1, U_2, \dots, U_m\}$ — конечное множество входов;

$X = \{X_1, X_2, \dots, X_n\}$ — конечное множество состояний;

$Y = \{Y_1, Y_2, \dots, Y_p\}$ — конечное множество выходов;

$\delta: X \times U \times X \rightarrow [0, 1]$ — функция переходов;

$\sigma: X \times Y \rightarrow L$ — функция выходов;

s_0 — начальное состояние.

Функция δ порождает множество нечетких матриц перехода:

$$T = \{\delta_{X_i, X_j}(U)\}, 1 \leq i \leq j \leq n; \quad (2)$$

функция порождает нечеткую матрицу выхода. Среди множества состояний автомата выделяется множество финальных (заключительных) состояний X_n . Нас будет интересовать такой тип автомата, для которого каждое состояние X_i , $i \in I = \{1, \dots, n\}$ зависит от предыдущего состояния X_{i-1} . Подобная зависимость может определяться последовательностью реализации подцелей, приоритетом выполнения и т.п.

В этом случае автомат можно задать как нечеткий граф:

$$G = \{\mu_G(X_{i-1}, X_i) \in M\}, \quad (3)$$

где M — множество принадлежностей элементов $X_{i-1} \times X$. При подобном рассмотрении цель обучения декомпозируется на i последовательных (по времени освоения материала) подзадач. Будем интерпретировать X_i как множество результатов i -го теста, Y_j , $j \in J = \{1, \dots, p\}$ — как множество интервалов

времени на обучение, L — как множество доходов, связанных с реализацией выбранного способа обучения (освоения материала) $u \in U$ на интервале времени Y_j . Очевидно, что в рассматриваемом типе автомата $m=p$ и может трактоваться как число последовательных этапов (шагов) достижения цели.

Управляющие решения и состояния подзадач, протекающие во времени, будем трактовать как нечеткие события на интервале Y_k , $1 \leq k \leq p$. При таком подходе функция переходов может задаваться экспертным путем и отражать уже имеющийся опыт обучения, исходя из практического опыта преподнесения учебного материала. При этом, естественно, не учитывается результат обучения в зависимости от времени его реализации, а также личностные особенности приобретения знаний конкретным индивидом. Для учета данного обстоятельства необходимо построить индивидуализированную функцию переходов. С этой целью используется исходная информация от обучаемого, которая включает прогноз применения того или иного способа освоения материала в зависимости от возможных результатов тестирования в виде функции $\mu: X_i, \times U_1 \rightarrow [0,1]$, а также прогноз перехода управляемого процесса обучения из исходного состояния S_0 на первом шаге решения в зависимости от ограничений на имеющиеся ресурсы. На основе этой информации программируется автоматная модель. Для этого на каждом шаге решается система композиционных уравнений вида [3]:

$$\left. \begin{aligned} \mu(X_i)/U_i &= \mu(X_{i-1})/U_{i-1} \circ \delta(X_{i-1}, X_i)/U_i, \\ \mu(U_i) &= \mu(X_{i-1}) \circ \mu(X_{i-1}, X_i)/U_i \end{aligned} \right\} \quad (4)$$

где "o"- знак операции «композиция»;

$\mu(X_i)/U_i, \mu(X_{i-1})/U_{i-1}$ - нечеткие оценки возможности управляемого процесса находится в состояниях X_i, X_{i-1} , при применении способов освоения учебного материала U_i и U_{i-1} , соответственно;

$\mu(U_i)$ - нечеткая оценка выбора обучаемым способа освоения материала U_i .

Полученные оценки группируются попарно, исходя из следующего условия:

$$\mu(U_i) \leq \mu(X_i)/U_i \quad (5)$$

Формирование пар по такому принципу согласуется с реальным выбором решения: результату теста с максимальной оценкой возможности должен соответствовать способ освоения учебного материала также с максимальной оценкой применения его обучаемым. Выделение пар позволяет выявить наиболее возможные связи по способам освоения материала между результатами тестирования каждого этапа обучения. При этом каждый способ освоения, маркирующий связь, характеризуется нечеткой оценкой использования его обучаемым и нечеткой оценкой цены обучения (нормированный доход), субъективно осознаваемой обучаемым, в зависимости, например, от времени или сложности освоения учебного материала.

Построение автомата осуществляется следующим образом [3]. Из исходного состояния в состояния первого шага принятия решений проводятся дуги, маркированные теми способами освоения учебного материала, использование которых, в соответствии с прогнозом, позволяет добиться результатов промежуточного тестирования на первом этапе обучения, характеризующихся наибольшей оценкой. В зависимости от означивания этих дуг и на основе сформированных пар «способ освоения материала» — «результаты тестирования» проводятся означенные дуги от первого до n -го этапа обучения.

В результате построений получаем нечеткий недетерминированный автомат, моделирующий поведение обучаемого при различных результатах промежуточного тестирования. Применяя подход, используемый в динамическом программировании, можно выделить классы стратегий обучаемого. В первую очередь нас будет интересовать тот класс, стратегии которого позволяют достичь цели обучения и характеризуются максимальными оценками связей между результатами тестов. Для выделения таких стратегий на множестве финальных результатов тестирования определяются результаты, соответствующие цели обучения. Далее выделяются результаты тестирования на $(n-1)$ -м шаге, переход из которых в целевые состояния n -го шага характеризуется способами освоения учебного материала с оценкой, равной

$$\alpha(U_{n-1}) = \max(\min(\mu U_{n-1}, \sigma_{X_{n-1}, U_{n-1}})) Y_{n-1} \quad (6)$$

Подобная процедура осуществляется для каждого шага решения, вплоть до состояния S_0 . Использование процедуры позволяет выделить возможные стратегии обучения, представляющие собой взвешенные пути на графе от вершины S_0 до вершин из множества X_n . Каждый q -й путь представляет собой взвешенную относительно способов освоения учебного материала последовательность.

Элементы (результаты итогового тестирования) множества X_n могут представлять для обучаемого различную ценность, что отражается заданием на X_n нечеткой цели с функцией принадлежности:

Очевидно, что стратегия, имеющая максимальную оценку наиболее соответствует индивидуальному стилю приобретения знаний обучаемым, но не всегда может соответствовать максимальной функции принадлежности результата тестирования относительно цели g . Полученного противоречия при построении модели обучения можно избежать, используя следующие способы:

- изменение предпочтений обучаемого;
- изменение исходных матриц переходов и выходов;
- использование такой организации сопровождения обучения, которая бы учитывала только конечную цель, т.е. состояние нечеткой цели с максимальной функцией принадлежности.

В последнем случае следует использовать смешанные стратегии, образованные частями стратегий из классов π . Необходимость в формировании смешанной стратегии определяет трансформацию стратегии из некоторого класса, выбранную, исходя из критерия максимальности в стратегию из другого класса.

Очевидно, что необходимость в трансформации возникает в тех случаях, когда выбранная исходная стратегия перестает быть эффективной по отношению к нечетко заданной цели. При этом важно определить те пространственно-временные точки графа (вершины), где подобная трансформация возможна. С этой целью введем понятие коэффициента свободы выбора обучаемого, под которым будем понимать отношение числа допустимых стратегий к общему числу стратегий из этого же класса. Здесь ε означает порог возможностей обучаемого в достижении поставленной им цели, при использовании индивидуального стиля освоения учебного материала. Чем меньше значение такого коэффициента, тем меньше у обучаемого вариантов в достижении поставленной цели без изменения индивидуальной стратегии. Поэтому обучающая система, построенная на основе модели поведения обучаемого, должна:

- рекомендовать после получения каждого результата тестирования те способы освоения материала, которые являются составными частями допустимых стратегий и не делают значение КСВ равным или близким нулю;
- в случае невозможности выполнения предыдущего пункта трансформировать индивидуальную стратегию обучаемого, переходя на смешанную стратегию обучения. Такая стратегия обучения будет иметь меньшие возможности в реализации индивидуального стиля, но останется допустимой по отношению к достижению поставленной цели;
- перепрограммировать модель обучения в ситуациях перехода от одного результата тестирования к другому в пределах одного и того же этапа обучения.

В последнем случае необходимость в перепрограммировании модели обучения определяется изменением функции доходов. Использование рассмотренной модели позволяет наметить пути автоматической коррекции индивидуальных стратегий обучаемого. Особенность предлагаемого подхода состоит в возможности перенесения модели обучаемого, сформированной преподавателем при личных контактах, в образовательную среду компьютерного (дистанционного) обучения.

Выводы

На данном этапе встает вопрос о соотношении в курсе обучения информационных и интерактивных слайдов. Исходя из современных положений о структуре учебного курса, рекомендуется количество интерактивных слайдов выбирать в диапазоне 0,3 ... 0,5 от количества информационных слайдов. Узлы, как и ребра, могут создаваться, модифицироваться и уничтожаться. При создании узла автор учебного курса имеет возможность непосредственно указывать файлы, находящиеся на локальном диске. Именно для обеспечения этой возможности инструментальная система должна функционировать непосредственно на сервере, поскольку все остальные функции программы могут выполняться и удаленно, через Интернет или локальную сеть. Более того, данный подход позволяет одновременно использовать инструментальную систему в процессе совместной работы над одним учебным курсом коллективу таких специалистов, как преподаватель, ведущий аналогичный курс, психолог, дизайнер, редактор и т.д.

Использование инструментальной системы поддержки интеграции учебно-методических материалов позволит не только существенно сократить время создания курсов компьютерного (дистанционного) обучения, но и в дальнейшем позволит оперативно обновлять материалы курсов. Внедрение таких систем позволит обеспечить более быстрый переход на единые форматы передачи информации сети INTERNET, что, в свою очередь, будет способствовать расширению сферы применения данной инструментальной системы.

В последнем случае необходимость в перепрограммировании модели обучения определяется изменением функции доходов. Использование рассмотренной модели позволяет наметить пути автоматической коррекции индивидуальных стратегий обучаемого. Особенность предлагаемого подхода состоит в возможности перенесения модели обучаемого, сформированной преподавателем при личных контактах, в образовательную среду компьютерного (дистанционного) обучения.

Библиография

1. Rashid El Baji, Olexandr Vyrodov, Igor Shubin. Adaptive Hypermedia Methods and Student Models for Computer Based Didactic Systems – The 5th International Conference «Internet. Education. Science (IES-2006) », Ukraine, Vinnitsia, 2006
2. Nataliya Bilous, Mikhail Bondarenko, Olexandr Vyrodov, Igor Shubin. The Web-Based Software Implementation of Learning Course Model – The Third International Conference on Advanced Engineering Design, Czech Republic, Prague, 2003
3. Nataliya Bilous, Mikhail Bondarenko, Igor Shubin. The Ukrainian e-Learning Region; In Proceedings of 10-th International LInE Conference New Partnerships and Lifelong Learning, Helsinki, Finland, 2008.

Информация об авторах

Igor Shubin – Professor of software Department, Kharkov National University of Radioelectronics, Lenin ave., 14, Kharkov, 61166, Ukraine; e-mail shubin@kture.kharkov.ua

Yaroslav Svyatkin – researcher of software Department, Kharkov National University of Radioelectronics, Lenin ave., 14, Kharkov, 61166, Ukraine

Iryna Bilous – researcher of Kharkov National Economical University, Lenin ave., 9-a, Kharkov, 61166, Ukraine

Decision Making

EXPERIENCE OF DEVELOPMENT AND APPLICATION OF EXPERT SYSTEMS FOR ECONOMIC MACROPARAMETERS FORECASTING AND OPERATION

Alexey Voloshyn, Victoria Satyr

Annotation: *In this article there are considered problems of forecasting economical macroparameters, and in the first place, index of inflation. Concept of development of synthetical forecasting methods which use directly specified expert information as well as calculation result on the basis of objective economical and mathematical models for forecasting separate "slowly changeable parameters" are offered. This article discusses problems of macroparameters operation on the basis of analysis of received prognostic magnitude.*

Key words: *decision tree method, index of inflation, expert information, operation.*

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

Introduction

In articles [Voloshyn, 2003, 2005, 2006, 2007, 2008] presented on "Knowledge-Dialog-Solution" (KDS) conferences there is developed concept of "subjective multideterminism" [Voloshyn 2006, 2007] for qualitative forecasting of economic macroparameters and in the first place of inflation index. This concept is based on the fact that effect is determined by the variety of interdependent reasons (objective and subjective). Reasons themselves (factors, parameters) and intensity of their interaction are determined subjectively (by expert judgment) and are presented by "illegible" decision tree (by its tops and arcs appropriately), "effect" (result of forecasting) is presented by leaves of the tree. There are developed instruments that allow to execute tree derivation (to insert and to eliminate tops, specify rating of arcs), to find more probable development of scenario of predictable process, to define magnitude of predicting parameter "illegibly". The construction of an applied system of support of decision-making is reduced to highlighting by experts of problems and sub problems (tops of the tree) and links between them (arcs of the tree). Experts determine the weights (probabilities) of transitions between tops. It is acceptable to get unclear expert estimations achieved by the Boolean variables described by the values of function of belonging (by vectors of the real numbers from 0 to 1). Every expert sets three estimations – optimistic, realistic and pessimistic, the scaling of which is effected taking into account the psychological type of the expert. The type is determined on the basis of psychological tests built into the system. Such psychological tests help to determine the coefficients of «veracity», «independence», «caution», etc.

The tree is built on the basis of collective estimations of experts using the method of pair comparisons. The construction of a resulting tree requires the use of the algebraic methods of treatment of expert information, and the Hemming metrics and measure of lacks of coincidence of grades of objects is used as distance between ranges. A resulting tree is determined as the Kemeny-Snella median or as a compromise [Voloshyn, 2005]. In the case of setting priorities in an unclear form the elements of matrix are set through the functions of belonging.

The algorithms of successive analysis of variants [Voloshyn, Panchenko, 2002] allowing to process trees with hundred tops are offered for determining optimum ways within the tree.

At KDS conferences there were voiced and published in [Voloshyn, 2007] predictable magnitude of inflation index I Ukraine diverging from actual for 3-5%.

Particularly in June of 2007 at KDS-07 inflation index on 01.01.2008 was set calculated in the first quarter of 2007: "National Bank of Ukraine predicts inflation in Ukraine in 2007 at the rate of 7%, government – 8%. Our prediction is 17.3%. We just need to wait for the beginning of KDS-08 conference and to compare our

predictions". As it's known actual rate Of inflation was 16.6%. If to take into account that inflation index was considered by us as "unclear" parameter and magnitude 17.3% - "the most probable" (maximizing membership function) received result is worth attention. And with previous "coincidences" mentioned above to the authors' opinion it is worth scrupulous attention.

At present (April, 2008) inflation in Ukraine is topic number one. For the first quarter of 2008 inflation in Ukraine has become the highest amongst all Former Soviet Union countries (CIS), specified in budget magnitude of 9.6% are already exceeded (9,7%). Cabinet of Ministers of Ukraine has worsened forecast of inflation growth according to the results of 2008 from 9, 6% to 15,3% which is mentioned in the Cabinet of Ministers of Ukraine Resolution No. 581 dated June 27, 2008 "Adoption of Amendment to the Cabinet of Ministers of Ukraine Resolution No. 976 dated July 27, 2007", posted on the site of the government as UNIAN broadcasts. The government also changed forecast of price index for industrial production manufacturers according t the result of that year from 19, 2% to 24, 6%. As it was notified in accordance with data of State Statistics Committee inflation in Ukraine from May to April was 1, 3% by May, 2007 – 31, 1%. In April, 2008 by April of the previous year it was 30, 2%, in January-April till analogous period of 2007- 24, 4%. From the beginning of current year till April growth of consumer prices was 13, 1%. Inflation in April contrary to that in March, 2008 was 3,1%. Growth of price index for manufacturers in Ukraine in January-May, 2008 contrary to January-May, 2007 was 31, 7%. According to data of State Statistics Committee growth of manufacturers' prices in May contrary to April, 2008 was 3, 7% and to December, 2007 – 24,2%. Cabinet of Ministers of Ukraine improved forecast of nominal GDP according to the results of 2008 from 921,2 to 956,8 billion hryvnas without change of its growth at the rate of 6,8%.

Operation of predictable parameters

While controlling socio-economic processes especially in the period of transition it is necessary to foresee inflation rate alongside with other factors. Inflation substantially and sometimes vitally influences on flow of real income of population, formation of investment climate, pace of industry and agriculture development, execution of budget's income, exchange rates and so on.

Forecasting models we have review earlier [Voloshyn, 2003, 2005, 2006, 2007] can be referred to the class of "Positive" that answer the question "What will be?". Though experts evaluating degree of parameters' interaction to some extent take into consideration its change at forecasting interval, these models can be considered "time-independent" (or stationary). "Normative" models answer the question "How shall it be?" and must recommend amending "stationary" scenario. In our models it is realized by the following reformations in "decision-making tree" - by determination of "narrow places" (arcs and tops that down to the limit influence on the result) with the purpose of recommendation of "correction" of causes degree interaction (in particular inserting and eliminating tops) for receiving desirable magnitudes of forecasting parameter (for example, value pf inflation index mustn't exceed 10%). In addition, suggested models and program instruments of their realization accordingly differ from suggested in the previous realizations at the following point. For "objectification" [Voloshyn, 2006, 2007] of decision-making process in predictable model some ratings of arcs are determined on the basis of relative economical and mathematical models (see further). From formally mathematical point of view "correction" of decision-making tree is analysis of its sensitivity [Voloshyn, 2006] but if correction of arcs' rating can be called traditional analysis insertion-elimination of tops needs development of ad hoc algorithms.

Forecasting and operation of inflation in "bottlenecks"

Hundreds of monographs and thousands of articles are devoted to the research of inflation. We will shortly describe main factors effecting inflation that were used by us in a varying degree for forecasting of inflation rate [Barro, 1998], [Greene, 2000], [Baumol, 2001], [Poroshenko, 2008].

First, worldwide tendency of the last years is impetuous increase in price for supplies and energy resources. Index of world prices for supplies that is calculated by investment bank Goldman Sachs in 2006 grew for 26%, and in 2007 for 41% in the connection with what bank's experts introduced the term of "agflation" (sudden increase in price for agricultural production).

Second. Increase in price for fuel is also long-term tendency. On 16th of April, 2008 oil price broke another record achieve USD 115 for barrel.

Third. Also more serious problem in Ukraine is impetuous expansion of foreign liabilities of bank. Only for 2007 amount of drawn by banks foreign credits grew almost twice – from USD 14 to USD 31 billions. As these means are used for consumer lending such debt is becoming powerful factor of inflation.

Forth. Inflation wouldn't be so noticeably sped-up if economy of Ukraine responded key criteria of market economy. In particular there are not used possibilities for progressive expansion of labour productivity.

In the purpose of "objectiveness" in modified variant of expert system for forecasting parameters a priori change of which isn't spasmodic (for example, in demography, see figure 1) there are used regression models. Preliminary tests of models demonstrated that forecasting for trend equations doesn't present any correct predictions because dynamics of inflation rate changes is difficult. That is why single-factor regression models were chosen. Multiple factors should be taken into consideration during forecasting including government, bank and financial structures policy, real changes in economy, world markets situation, position of developed countries governments concerning Ukraine sand many other factors. Regression equation looks like $Y=a+bx$, where Y is composite Consumer Price Index in percentage by previous December (predictable rate); a- fragmenton equal to initial value of dependent variable; b – inclination that shows how much inflation rate is changing with the change of factor x per unit.

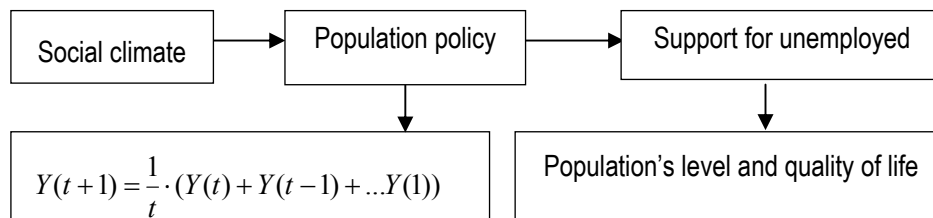


Figure 1.

The following regression models are used. The simplest model based on plain averaging is $Y(t+1)=(1/(t))*[Y(t)+Y(t-1)+...+Y(1)]$, and unlike the simplest "naïve" model which is conformed with "tomorrow it will be as it's today" principle this model is based on "tomorrow it will be as it is today for the latest time". Such model is for sure more resistant to fluctuations because mavericks about the mean are smoothed. However, this method is as simple ideologically as "naïve" models and it has almost the same defects. In the above presented formula it is supposed that row is averaged during quite long period of time. However, as a rule values of time series from the recent times characterize forecast than "older" values of this row. Then moving average $Y(t+1)=(1/(T+1))*[Y(t)+Y(t-1)+...+Y(t-T)]$ can be used for forecasting. Its meaning is that model sees only recent past (per T time counts deep) and being based only on these data it makes prediction. For forecasting exponential mean-value method was also used. Formula describing this model looks like $Y(t+1)=a*Y(t)+(1-a)*Y(t)$, where $Y(t+1)$ is a forecast for the following period of time; $Y(t)$ – operational significance at the period of time; $^aY(t)$ – previous forecast at the moment t; a – consonant of evening-out ($0 \leq a \leq 1$).

Inflation influences not only on debt relationship of such subjects like households, bank, entrepreneurial and financial sectors but also on debt relationship between state and households and entrepreneurial sector (national

domestic debt) and intergovernmental debt relationship. Consequently, the next connection we are going to consider will be inflation and national debt. Reasons of limited capacities of national debt loss of value in the modern conditions are:

- Increase of state short term debts portion;
- Increase of sensitivity rate (elasticity coefficient) of reaction of interest rate value on inflation rate;
- Provision with long-term and intermediate credits not in corpore at once but by tranches that allows considering inflation processes influence on loan amount.

There were considered influence of inflation on long-term loans. Let's discuss influence of inflation on contracts that determine salary level. Labour market supply and demand have common determinant (according to the neoclassical theory) which is actual salary level. Agreeing on its amount, employers and employees react in a definite way on conditions that have been formed at the labour market. If, for example, gross domestic product rate (GDP) and employment are high salary tends to increase. And vice versa – if GDP and employment decreases salary is increasing slowly. Phillips' wage curve represents connection between salary inflation (loss of value) and GDP decelerating and can be expressed by formula : $q = \lambda(Y - Y^*)$, where λ is sensitivity coefficient (elasticity) of salary response on changes of GDP amount, Y – GDP real volume, Y^* - potential.

Taking into account inflationary expectations equation of Phillips' curve changes into $q = P_e + \lambda(Y - Y^*)$, where P_e - expected inflation rate. There is presented fragment of mathematical model built in decision tree on the figure 2.

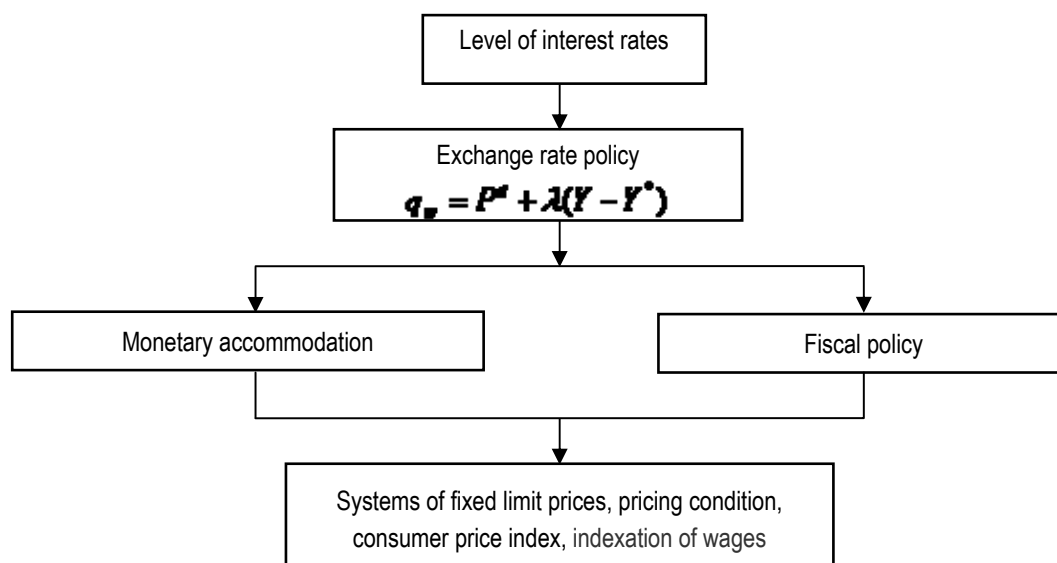


Figure 2

The last equation means that at any prescribed GDP level salary increases faster than anticipated inflation rate. It is supposed that nominal wage increases at 1% faster for every additional percent of anticipated inflation.

Said about demands answer to the question "Why do enterprises agree for such salary acceleration?" Such behavior of entrepreneurs is explained by the fact that increase of nominal wage will not be vulnerable for them if prices for their production are growing at the same rate. In such case both entrepreneurs and employees are in such conditions in which they would be if there is no inflation.

Consequently, wage-push expects anticipated inflation. But as we already know inflation can be unexpected. In such case employees realize wage-push and inflation of their salaries only after some time and will demand relevant compensation. In such situation compensation becomes recovery of losses only in case of unexpected

inflation that wasn't taken into consideration in previous labour contract. collective arrangements entered by trade unions that have requirement of wage indexing with adjustment for inflation also bring particular confusion on the matter if this compensation is for previous inflation or the expected one.

This condition is very important for the further forecasting of inflation rates. If labor rates for the next year expresses inflation rate of the previous year and prices are based on the actual salary so today inflation will express yesterday's one and inflation rates will change slowly.

- If for determining labor rates only anticipated inflation is taken into consideration the policy radical change is possible that changes expectation concerning fast change of inflation rate. Besides mentioned aspect it is important to remember another aspect of this problem. Salary can be indexed for 100% or only partially for example for 50-60%.
- There are two ways of salary partial indexation:
- Upper bound delimiting of prices increase exceeding of which provides compensation. For example, if inflation rate exceeds 5% so from 5,1% income begins to be indexed;
- Upper bound delimiting of compensation. This way limits values that are used for compensation of prices increase by beforehand fixed percentage. Making labour contracts for 2-5 years can't provide with complete certainty change of inflation rate especially in shaky economy. That is why for adjustment for inflation there are used two the most widespread ways:
- indexation of wages in accordance with Consumer Price Index and periodic (quarterly or once a half of year) salary revision regarding its increase according to the prices increase for this period;
- prediction of periodic salary increase announced beforehand considering increase of expected prices rates.

If inflation could be predicted with complete certainty both methods would have similar consequences. But as expectations are often mistaken it is considered that indexation based on actual rates of inflation guarantees stability of actual salary more reliably then when pay-outs are decided beforehand.

Inflation influences not only on share of debtors and creditors, employees and entrepreneurs but also it effects state interests. Yes, in conditions of inflation there occurs voluntary prolongation of payments for state budget by taxpayers that allows taxpayer to settle accounts with state by undervalued money. This phenomenon of inflationary taxing in scientific literature was called "Oliver-Tanzy effect".

This effect shortly can be explained as following: any inflation decreases tax weight. It develops with more strength with growth of inflation rate and prolongation of tax payment time.

At the same time it must be noticed that inflation weakening tax weight predetermined by explained above reasons creates another, opposite tendency.

Forecasting of inflation can be done on the basis of index of consumer price change with consideration of pent-up demand. pent-up demand is equal to population compulsory savings. They are calculated by progressive total for range of years taking into consideration of their year-on-year increase. At the same time index of required savings is equal to good circulation and services index. year-on-year increase and good circulation and services value ratio present hidden inflation or population cash income index and good circulation and services index ratio.

On the macrolevel inflation rate can be determined on the basis of Fisher's equation.

$$MV = PQ,$$

where M is amount of currency in circulation; V - velocity of money; P - price; Q - good sales and services value.

As inflation is caused by many factors its level can be forecasted by development of multifactor models. As factors there may be: change in the exchange rate, growth of money supply, change of refinancing rate and so

on. Popular way of forecasting inflation is also calculation of its level on the basis of GDP deflator. Thus inflation shall be explained as money's loss of value. One of the most obvious factors of inflation presence or absence, its depth is price index. As indirect indicator of inflation level are used statistic data of trade stock and amount of population money on deposit ratio (reduction of stock and increase of deposits give evidence about increase of inflationary tension rate). Forecasting of inflation is made on the basis of index of consumer prices change with taking into consideration of pent-up demand. On the macrolevel inflation rate can be determined on the basis of Fisher's equation.

Conclusions

In the development of qualitative forecasting concept on the basis of unclear decision tree it is offered to use "built-up" economical and mathematical models of forecasting separate parameters in the first place determining tree's "bottlenecks". On the basis of the previous analysis of received prognostic values of inflation index in Ukraine for 2008 there are suggested the following (in first priority decrease order) controlling influences with the purpose of correction of adverse forecast of inflation index values (our previous forecast is rated at 30%): salary indexation of budget, graduated income tax levy, backing of agricultural sector, state control of monopolies ("natural" and "artificial"), introduction of minimum hourly pay, reduction of interest for bank loans, refusal from pegging of national currency to USD.

Bibliography

- [Voloshin, 2003] Voloshin O.F., Panchenko M.V. The System of Quality Prediction on the Basis of a Fuzzy Data and Psychography of the Experts. International Journal "Information Theories & Applications", 2003, Vol.10, №3. -P.261-265.
- [Voloshin, 2005] Voloshin O.F. Problems of Decision Making in Socio-Economic Systems// «KDS-2005» conference works, Vol. 1, Sofia, 2005. - p.205-212.
- [Voloshin, 2006] Voloshin O.F. Decision Support Systems as Personal Smart Tools of Decision-Making Person // «KDS-2006» conference works, Sofia, 2006. - p.149-153.
- [Voloshyn, 2007] Voloshyn A, Satyr V. Problems of Forecasting of Economic Macroparameters // «KDS-2007» conference works, Sofia, 2007, vol. 1.-p.264-269.
- [Voloshyn, 2007] Voloshyn A, Satyr V. The inflation Index Prognosis Based on the Method of Decision-Making "Tree"// "Informations technologies & application", 2007, vol.14, N1.-P.63-67.
- [Voloshyn, 2008] Voloshyn A, Satyr V.. Experience of Development and Application of Expert Systems for Economic Macroparameters Forecasting and Operation // «KDS-2008» conference works, Sofia, 2008, Vol. 2.-C.1-6.
- [Barro, 1998] Barro R. Macroeconomics. – London: The MIT Press Cambridge, 1998. -867 p.
- [Greene, 2000] Greene W. Econometric Analysis. – New York: Prentice Hall, 2000.- 1004 p.
- [Baumol, 2001] Baumol. What Alfred Marshal Didn't Know: Contribution of XX Century to Economics. – Economic Issues. – № 2. – 2001. – p. 73-107.
- [Poroshenko, 2008] Poroshenko P. Antiinflationary Policy: Shock or Therapy? // «Zerkalo Nedeli», 2008, №15.

Authors' Information

Oleksiy Fedorovych Voloshyn, Kyiv Taras Shevchenko National University, Faculty of Cybernetics, Professor. Kyiv, Ukraine. E-mail: ovoloshin@unicyb.kiev.ua

Victoria Valeriyivna Satyr, Kyiv Taras Shevchenko National University, Faculty of Cybernetics, postgraduate student. Kyiv, Ukraine. E-mail: Brili@Ukr.net

НЕЛИНЕЙНАЯ СХЕМА КОМПРОМИССОВ В МНОГОКРИТЕРИАЛЬНЫХ ЗАДАЧАХ

Альберт Воронин, Юрий Зиатдинов

Аннотация. Определение многокритериального решения по своей природе компромиссно и принципиально основано на использовании субъективной информации. Возможность решения проблемы основана на гипотезе существования некоторой функции полезности. Традиционный подход линеаризации функции полезности обладает многими недостатками. Предлагается концепция нелинейной схемы компромиссов.

Ключевые слова: многокритериальные задачи, ситуация принятия решения, адаптация, модель функции полезности, нелинейная схема компромиссов.

ACM Classification Keywords: H.1 Models and Principles – H.1.1 – Systems and Information Theory; H.4.2 – Types of Systems.

Содержание проблемы

Всякое сужение области эффективных решений, а тем более выбор единственного из них принципиально требует привлечения дополнительной субъективной информации от лица, принимающего решение (ЛПР), или группы людей (экспертов), которые участвуют в решении многокритериальной задачи. Причина в том, что эффективные точки несравнимы между собой формально. Возможность решения проблемы основана на гипотезе существования некоторой функции полезности, возникающей в сознании ЛПР при решении конкретной многокритериальной задачи. Дополнительная информация заключается в ответе на вопрос: *сколькими единицами выигрыша по одному критерию можно, по мнению ЛПР, компенсировать неизбежный проигрыш единицы по другому (другим) в заданной ситуации?* На основании этой дополнительной информации формулируется конкретная схема компромиссов для данной многокритериальной задачи и в итоге находится искомое решение.

Таким образом, определение многокритериального решения по своей природе компромиссно и принципиально основано на использовании субъективной информации. Получив эту информацию и выбрав схему компромиссов, можно перейти от общего векторного выражения к скалярной свертке частных критериев, что является основой для построения конструктивного аппарата решения многокритериальных задач. Если используется способ скалярной свертки, то математически модель решения задачи векторной оптимизации для минимизируемых критериев представляется в виде

$$x^* = \arg \min_{x \in X} Y[y(x)],$$

или

$$x^* = \arg \min_{x \in X} Y[y_0(x)],$$

если вектор критериев $y(x)$ пронормирован вектором ограничений A :

$$y_0(x) = \{y_k(x) / A_k\}_{k=1}^s = \{y_{0k}(x)\}_{k=1}^s. \quad (1)$$

Здесь x – вектор аргументов оптимизации; X – допустимая область решений; $Y(y)$ – скалярная свертка вектора частных критериев, имеющая смысл целевой функции. Ее вид является отражением функции

полезности ЛПР и зависит от выбранной схемы компромиссов. По сути, практически все подходы к определению скалярной свертки сводятся к построению той или иной модели функции полезности ЛПР.

В большинстве случаев при решении многокритериальных задач ограничиваются линеаризованной моделью. Такой подход, обладая несомненным преимуществом простоты, характеризуется рядом недостатков, присущих методу линеаризации вообще. Так, линейная модель приводит к правильным результатам лишь в малых окрестностях рабочей точки, положение которой зависит от ситуации принятия многокритериального решения. Любое изменение ситуации приводит к необходимости перерасчета весовых коэффициентов модели. В серьезных многокритериальных задачах целесообразно строить *нелинейную* модель функции полезности ЛПР.

Содержательный анализ функции полезности ЛПР

Введем понятие *напряженности ситуации* как меры близости относительных частных критериев к своему предельному значению (единице):

$$\rho_k = 1 - y_{0k}, \rho_k \in [0; 1], k \in [1, s].$$

Если многокритериальное решение принимается в напряженной ситуации, то это значит, что в заданных условиях один или несколько частных критериев в результате решения могут оказаться в опасной близости к своим предельным значениям ($\rho_k \approx 0$). И если один из них достигнет предела (или выйдет за него), то это событие не компенсируется возможным малым уровнем остальных критериев (обычно не допускается нарушение любого из ограничений).

В этой ситуации необходимо всемерно препятствовать опасному возрастанию наиболее неблагоприятного (т.е. наиболее близкого к своему пределу) частного критерия, не очень считаясь с поведением в это время остальных. Поэтому в достаточно напряженных ситуациях (при малых значениях ρ_k) ЛПР, если и допускает ухудшение максимального (наиболее важного в данных условиях) частного критерия на единицу, то только компенсируя это большим количеством единиц улучшения остальных критериев. А в очень напряженной ситуации (первый полярный случай: $\rho_k=0$) ЛПР вообще оставляет в поле зрения только этот один, наиболее неблагоприятный частный критерий, не обращая внимания на остальные. Следовательно, адекватным выражением схемы компромиссов в случае напряженной ситуации является минимаксная модель (эгалитарный принцип)

$$x^* = \arg \min_{x \in X} \max_{k \in [1, s]} y_{0k}(x), \text{ т.е. } Y[y_0(x)] = \max_{k \in [1, s]} y_{0k}(x).$$

В менее напряженных ситуациях необходимо возвращаться к одновременному удовлетворению и других критериев, учитывая противоречивое единство всех интересов и целей системы. При этом ЛПР варьирует свою оценку выигрыша по одним критериям и проигрыша по другим в зависимости от ситуации. В промежуточных случаях выбираются схемы компромиссов, дающие различные степени частичного выравнивания частных критериев. С уменьшением напряженности ситуации предпочтения по отдельным критериям выравниваются.

И, наконец, во втором полярном случае ($\rho_k \approx 1$) ситуация настолько спокойная, что частные критерии малы и не возникает никакой угрозы нарушения ограничений. ЛПР здесь считает, что единица ухудшения любого из частных критериев вполне компенсируется равнозначной единицей улучшения любого из остальных. Этому случаю соответствует экономичная схема компромиссов, обеспечивающая минимальные для заданных условий суммарные потери по частным нормированным критериям. Такая схема выражается моделью интегральной оптимальности (утилитарный принцип)

$$x^* = \arg \min_{x \in X} \sum_{k=1}^s y_{0k}(x), \text{ m.e. } Y[y_0(x)] = \sum_{k=1}^s y_{0k}(x).$$

Нелинейная схема компромиссов

С точки зрения формализации целесообразно задачу выбора схемы компромиссов заменить эквивалентной задачей синтеза некоторой *единой* скалярной свертки частных критериев, которая в различных ситуациях выражала бы разные принципы оптимальности. Требования к синтезируемой функции $Y(y_0)$:

- она должна быть гладкой и монотонной;
- в напряженных ситуациях она должна выражать принцип минимакса;
- в спокойных условиях – принцип интегральной оптимальности;
- в промежуточных случаях должна приводить к парето-оптимальным решениям, дающим различные меры частичного удовлетворения критериев.

Иными словами, такая универсальная свертка должна быть выражением схемы компромиссов, *адаптирующейся* к ситуации. Для этого необходимо, чтобы в выражение для скалярной свертки в явном виде входили характеристики напряженности ситуации ρ .

Из возможных функций, отвечающих перечисленным требованиям, выберем простейшую:

$$Y(\alpha, y_0) = \sum_{k=1}^s \alpha_k [1 - y_{0k}(x)]^{-1}; \alpha_k \geq 0, \sum_{k=1}^s \alpha_k = 1, \quad (2)$$

где $\alpha_k = \text{const}$ – формальные параметры, определенные на симплексе и имеющие двоякий физический смысл. С одной стороны – это весовые коэффициенты, выражающие *предпочтения* ЛПР по отдельным критериям. С другой – это коэффициенты регрессии *содержательной регрессионной модели функции полезности ЛПР*, построенной на основе концепции нелинейной схемы компромиссов.

Весовые коэффициенты рассчитываются по формуле

$$\alpha_k = \frac{f_k}{\sum_{j=1}^s f_j}, k \in [1, s], \quad (3)$$

где f_k – оценка важности (приоритетности) k -го критерия, данная экспертом по *шкале баллов*.

Таким образом, нелинейной схеме компромиссов соответствует модель векторной оптимизации, в явном виде зависящая от характеристик напряженности ситуации ρ :

$$x^* = \arg \min_{x \in X} \sum_{k=1}^s \alpha_k [1 - y_{0k}(x)]^{-1}.$$

В отличие от линейной модели, определенной в малой окрестности рабочей точки, нелинейная модель функции полезности ЛПР определена на всей допустимой области решений X и не требует перерасчета коэффициентов α_k при изменениях ситуации.

Нормализация скалярной свертки

При многокритериальном оценивании альтернатив часто возникает необходимость получения не только аналитической, но и *качественной* оценки. Для этого следует выражение скалярной свертки $Y(\alpha, y_0)$

нормировать и полученное значение Y_0 соотнести с градациями обращенной нормированной фундаментальной шкалы:

Категория качества	Интервалы обращенной нормированной фундаментальной шкалы оценок y_0, Y_0
Неприемлемое	1,0 – 0,7
Низкое	0,7 – 0,5
Удовлетворительное	0,5 – 0,4
Хорошее	0,4 – 0,2
Высокое	0,2 – 0,0

Конструкция нелинейной схемы компромиссов позволяет нормировать скалярную свертку не к максимальному (обычно неизвестному), а к *минимальному* значению. Положив в выражении для нелинейной скалярной свертки (2) идеальные (нулевые) значения минимизируемых критериев $y_{0k}(x) = 0$ и учитывая нормировку весовых коэффициентов α , получим $Y_{0\min} = 1$ и формула нормированной минимизируемой скалярной свертки имеет вид

$$Y_0 = 1 - \frac{1}{Y} \quad (4)$$

Модельный пример

Покажем здесь возможности нелинейной схемы компромиссов в задаче *многокритериального анализа*, а именно, оценки качества по нескольким критериям глиссадного спуска при посадке самолета.

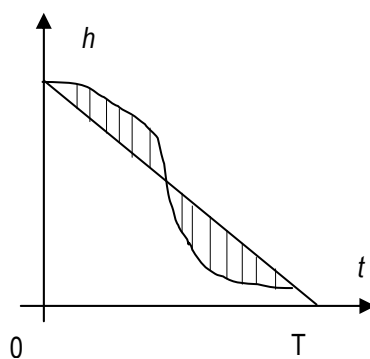


Рис.1

На рис.1 в координатах $(h-t)$ схематически представлено изменение высоты самолета в процессе глиссадного спуска. Предполагается, что в момент времени $t=T$ высота $h=0$.

Аналогичным образом может быть представлено изменение положение самолета b относительно центральной линии взлетно-посадочной полосы (ВПП) в боковой плоскости в процессе глиссадного спуска.

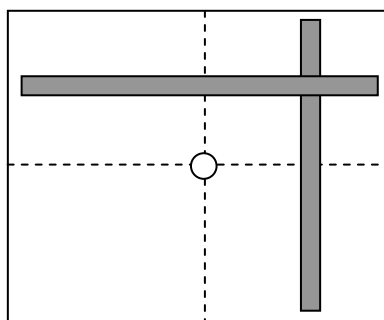


Рис.2

В процессе глиссадного спуска летчик управляет самолетом с помощью директорного прибора, схематически изображенного на рис.2. Положение планок, показанное на рисунке, означает, что самолет находится выше глиссады и правее осевой линии ВПП. Управление заключается в совмещении перекрестия планок с центральной точкой прибора.

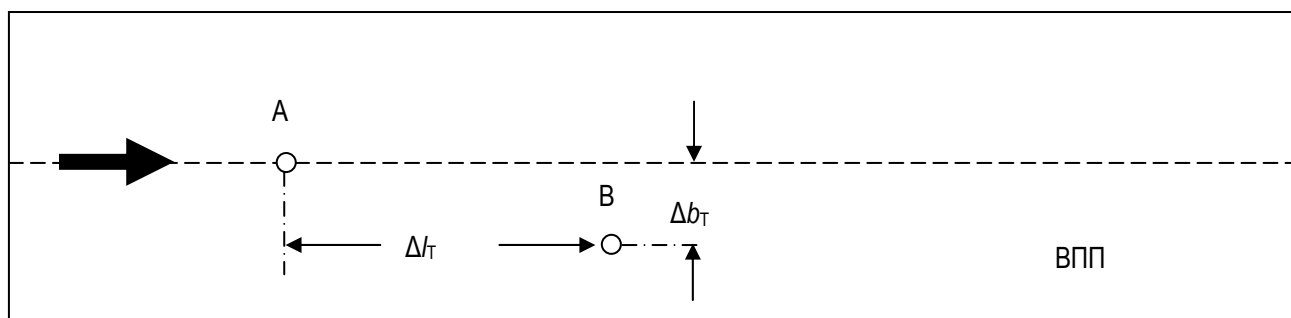


Рис.3

На рис.3 показано, что в момент времени $t=T$ самолет коснулся ВПП в точке В, находящейся на расстоянии Δl_T от расчетной точки А и на расстоянии Δb_T от осевой линии ВПП.

Для оценки качества посадки самолета используются три *терминальных* ($t=T$) критерия качества:

$$y_1 = |\Delta l_T| < A_1 \text{ – модуль отклонения от расчетной точки касания в продольной плоскости;}$$

$$y_2 = |\Delta b_T| < A_2 \text{ – модуль отклонения точки касания от продольной оси ВПП в боковой плоскости;}$$

$$y_3 = V_h^{(T)} < A_3 \text{ – вертикальная скорость в терминальной точке, а также два } \textit{интегральных} \text{ критерия:}$$

$$y_4 = \frac{1}{T} \int_0^T |\Delta h| dt < A_4 \text{ – среднее отклонение от глиссады в вертикальной плоскости;}$$

$$y_5 = \frac{1}{T} \int_0^T |\Delta b| dt < A_5 \text{ – среднее отклонение от глиссады в горизонтальной плоскости.}$$

Кроме этих критериев качество процесса посадки характеризуют еще y_6 – отклонение от посадочной скорости в терминальной точке; y_7 – курсовой угол в терминальной точке; y_8 – угол крена в терминальной точке; y_9 – угол тангажа в терминальной точке; y_{10} – средний расход рулей на глиссаде (интегральный критерий) и т.д. Будем считать, что последние критерии при всех посадках удовлетворяются.

Для расчета интегральных критериев воспользуемся приемом приближенного интегрирования.

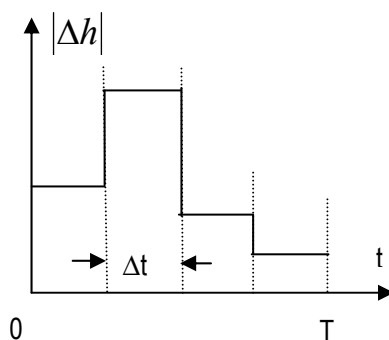


Рис.4

Прием иллюстрируется графиком рис.4. Интервал времени спуска по глissаде $[0, T]$ разбивается на N подинтервалов Δt , в течение каждого из которых величина $|\Delta h_i|, i \in [1, N]$ измеряется и полагается постоянной. Тогда в формуле для интегрального критерия мы можем перейти от интеграла к суммированию:

$$y_4 = \frac{1}{T} \int_0^T |\Delta h| dt \approx \frac{1}{T} \sum_{i=1}^N |\Delta h_i| \Delta t_i$$

Если все подинтервалы одинаковы, т.е. $\forall i \Delta t_i = \Delta t$, то $T = N\Delta t$ и

$$y_4 \approx \frac{\Delta t}{N\Delta t} \sum_{i=1}^N |\Delta h_i| = \frac{1}{N} \sum_{i=1}^N |\Delta h_i| \quad (5)$$

Аналогичным образом рассчитывается и критерий y_5 .

Теперь методом нелинейной схемы компромиссов оценим качество двух посадок самолета при различных числовых значениях частных критериев.

Посадка 1

Сначала рассчитаем значения интегральных критериев при $N=4$ и числовых данных, представленных таблицей

i	1	2	3	4
$ \Delta h , \text{ м}$	10	12	15	5
$ \Delta b , \text{ м}$	5	10	8	6

По формуле (5) получим: $y_4=10,5$ м и $y_5=7,25$ м.

Зададим для посадки 1 следующие числовые данные:

$$y_1=6 \text{ м}; A_1=15 \text{ м}; f_1=8.$$

$$y_2=3 \text{ м}; A_2=10 \text{ м}; f_2=7.$$

$$y_3=0,2 \text{ м/сек}; A_3=1 \text{ м/сек}; f_3=9.$$

$$y_4=10,5 \text{ м}; A_4=30 \text{ м}; f_4=5.$$

$$y_5=7,25 \text{ м}; A_5=20 \text{ м}; f_5=3.$$

По формулам (1) и (3) рассчитаем

$$y_{01}=0,4; \alpha_1=0,25;$$

$$y_{02}=0,3; \alpha_2=0,22;$$

$$y_{03}=0,2; \alpha_3=0,28;$$

$$y_{04}=0,35; \alpha_4=0,16;$$

$$y_{05}=0,36; \alpha_5=0,09.$$

Рассчитаем скалярную свертку критериев по нелинейной схеме компромиссов (формула (2)):

$$Y = 0,25 \frac{1}{1-0,4} + 0,22 \frac{1}{1-0,3} + 0,28 \frac{1}{1-0,2} + 0,16 \frac{1}{1-0,35} + 0,09 \frac{1}{1-0,36} = 1,47$$

Нормировка по формуле (4) дает $Y_0 = 1 - \frac{1}{1,47} = 0,32$.

Сопоставление этого значения с качественными градациями обращенной нормированной фундаментальной шкалы позволяет сделать вывод, что данную посадку можно оценить как *хорошую*.

Посадка 2

Рассчитаем интегральные критерии по числовым данным таблицы

i	1	2	3	4
$ \Delta h $	12	15	16	10
$ \Delta b $	7	12	14	9

По формуле (5): $y_4=13,25$ м и $y_5=10,5$ м.

Зададим для посадки 1 следующие числовые данные:

По формулам (1) и (5)

$$y_1=3 \text{ м}; A_1=15 \text{ м}; f_1=8.$$

$$y_{01}=0,2; \alpha_1=0,25;$$

$$y_2=4 \text{ м}; A_2=10 \text{ м}; f_2=7.$$

$$y_{02}=0,4; \alpha_2=0,22;$$

$$y_3=0,6 \text{ м/сек}; A_3=1 \text{ м/сек}; f_3=9.$$

$$y_{03}=0,6; \alpha_3=0,28;$$

$$y_4=13,25 \text{ м}; A_4=30 \text{ м}; f_4=5.$$

$$y_{04}=0,44; \alpha_4=0,16;$$

$$y_5=10,5 \text{ м}; A_5=20 \text{ м}; f_5=3.$$

$$y_{05}=0,52; \alpha_5=0,09.$$

По формуле (2):

$$Y = 0,25 \frac{1}{1-0,2} + 0,22 \frac{1}{1-0,4} + 0,28 \frac{1}{1-0,6} + 0,16 \frac{1}{1-0,44} + 0,09 \frac{1}{1-0,52} = 1,85$$

Нормировка по формуле (4): $Y_0 = 1 - \frac{1}{1,85} = 0,46$.

По обращенной нормированной фундаментальной шкале посадка 2 оценивается как *удовлетворительная*. К такому выводу наша процедура привела не в последнюю очередь из-за критерия $y_3=0,6$ м/сек (посадка жесткая).

Сведения об авторах

Воронин Альберт Николаевич – профессор, доктор технических наук, профессор кафедры компьютерных информационных технологий Национального авиационного университета, проспект Комарова, 1, Киев-58, 03058 Украина; e-mail: alnv@voliacable.com

Зиатдинов Юрий Кашафович – профессор, доктор технических наук, заведующий кафедрой компьютерных информационных технологий Национального авиационного университета, проспект Комарова, 1, Киев-58, 03058 Украина; e-mail: oberst@nau.edu.ua

АНАЛИЗ СВОЙСТВ МОДЕЛИ ЛЕОНТЬЕВА ПРИ НЕЧЁТКО ЗАДАНЫХ ПАРАМЕТРАХ КРИТЕРИАЛЬНОЙ ФУНКЦИИ МЕТОДОМ БАЗИСНЫХ МАТРИЦ

Владимир Кудин, Григорий Кудин, Алексей Волошин

Аннотация: Предложено применение метода базисных матриц для анализа модели Леонтьева (МЛ) с нечетко заданными некоторыми ее компонентами. МЛ можно интерпретировать, как задачу прогноза затрат-выпуска продукции на основе известной статистической информации при нечётко заданных значениях части элементов технологической матрицы, вектора ограничений и границах переменных. Такими элементами могут быть и цены на выпускаемую продукцию (вектор градиента целевой функции). Это существенно усложняет анализ МЛ.

Ключевые слова: модель Леонтьева, количественный и качественный анализ, нечёткое множество, базисная матрица, функция принадлежности.

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

Введение

Математический аппарат нечётких множеств является формой задания неопределённых параметров, значения которых неизвестны до момента принятия решения. Одним из механизмов устранения неопределённости в задании параметров при построении модели является включение в контур принятия решения экспертов (ЛПР). ЛПР призваны качественно определить структуру модели, указать механизм устранения неопределённости при ее формировании [Орловский, 1981]. Существенным осложнением модели Леонтьева (МЛ) [Леонтьев, 1972], [Гасс., 1961] есть включение ограничений на значение переменных [Орловский, 1981]. Одной из особенностей МЛ является то, что она включает в себе математические проблемы анализа ряда линейных систем таких как систем линейных алгебраических уравнений с квадратной матрицей ограничений (СЛАУ), линейных алгебраических неравенств с соответствующей матрицей ограничений (СЛАН), а также и задач линейного программирования (ЗЛП) [Волошин, 1993], [Войналович, 1987], [Войналович, 1988], [Кудин, 2002]. Проведения качественного анализа модели [Орловский, 1981] предопределяет включение и количественного анализа непротиворечивости ее структурных элементов [Волошин, 1993], [Войналович, 1987], [Войналович, 1988], [Кудин, 2002], а также влияние изменения ее компонент на свойства модели в целом. Можно выделить такие основные стадии анализа:

- проверки адекватности математического и машинного представления матрицы ограничений, нахождение величины ее ранга;
- направленной коррекции величины ранга матрицы ограничений изменением отдельных её элементов (при необходимости);
- выявление совместных свойств собственно МЛ и ограничений на переменные – разрешимости (неразрешимости);
- определение свойств ограничений МЛ (многогранного множества) и ограничений на переменные, проведения, при необходимости, направленных изменений;
- нахождение решений при разрешимости;

- установление свойств решений,
- анализ изменений в модели на структурные свойства областей (допустимости и решений).

Постановка задачи

Введем в рассмотрение варианты МЛ, к которым можно прийти в результате эквивалентных преобразований канонической модели:

1. СЛАР вида

$$Au = C, \quad (1)$$

2.. СЛАН вида

$$Au \leq C, \quad (2)$$

3. МЛП (модель линейного программирования) при наличии целевой функций вида

$$\max_{u \in R^m} Bu, \quad (3)$$

вида (2)-(3), в которой $A = \{a_{ij}\}_{i=1, \bar{m}}^{j=1, \bar{m}}$ невырожденная квадратная матрица размерности $(m \times m)$,

$a_j = (a_{j1}, a_{j2}, \dots, a_{jm})$, $j \in J = I = \{1, 2, \dots, m\}$ – строки матрицы A , $u = (u_1, u_2, \dots, u_m)^T$ – вектор переменных, $B = (b_1, b_2, \dots, b_m)$, $C = (c_1, c_2, \dots, c_n)^T$ – вектора градиента целевой функции и ограничений модели, $a_j u \leq c_j$, $j \in J$ полупространство, которое определено гиперплоскостью $a_j u = c_j$, $j \in J$.

Считаем, что компоненты вектора градиента (3) могут задаваться, как функции в виде $B(t)$, где $B(t) = (b_1(t), b_2(t), \dots, b_m(t))$, которые зависят от аргумента t . Каждая компонента вектора b_r становится равной $b_r(t)$ при каждом $t \in T = [t_H, t_B]$. (Все определенные функции являются зависимыми от аргумента, в общем случае, $t \in (-\infty, +\infty)$, из класса C^2). Нечеткий интервал $T = [t_H, t_B]$ определяется как множество аргументов, на которых функция принадлежности имеет уровень не ниже λ . Такие изменения в элементах модели можно интерпретировать, как зависимость цен на продукцию r на интервале $t \in [t_H, t_B]$. Механизм формирования уровня λ (устранение нечеткости в системе, т. е. выбора интервала $T = [t_H, t_B]$) может быть разнообразным. В частности, пусть в системе имеется $P = \{1, 2, \dots, p\}$ экспертов. Каждый эксперт формирует свою функцию принадлежности $\mu_p(t)$, $p \in P$ зависимой от параметра t . Эти функции являются кусочно-линейными, для которых эксперт устанавливает уровни значений $\lambda^{(p)}$, $p \in P$. Это означает, что при указании уровня $\lambda^{(p)}$, $p \in P$ ($1 \geq \mu_p(t) \geq \lambda^{(p)}$) каждым экспертом определяется интервал изменения значений T_p , $p \in P$, где $T_p = [t_{p(H)}^{(-)}, t_{p(B)}^{(+)}]$, $p \in P$ [Орловский, 1981]. Результирующий интервал изменения переменной $T = [t_H, t_B]$ согласованный по P экспертам может определяться, например, как $T = \bigcap_{p=1}^P T_p$, что соответствует некоторой функции принадлежности $1 \geq \mu(t) \geq \lambda$. При каждом $t \in T = [t_H, t_B]$

формируется свой вариант $B(t) = (b_1(t), b_2(t), \dots, b_m(t))$ для (2),(3). Модель исследуется в пространстве E^m . Наличие в контуре принятия решения экспертов (фаза качественного анализа моделей (1),(2)-(3)) предопределяет последующую фазу (количественного анализа) - исследования при указанных уровнях $\lambda^{(p)}$, $p = \{1, 2, \dots, p\}$ указанных экспертами влияние изменения элементов $B(t)u$ при $t \in T$ на ранее выбранное оптимальное решение.

В работе предложено развитие методологии последовательного анализа [Волошин, 1987] и метода базисных матриц (МБМ) [Кудин, 2002] для проведения количественного анализа влияния функциональных изменений в МЛ на свойства оптимальных решения исходной задачи (2)-(3) при изменении элементов модели (3) в виде $B(t)u$, $t \in T$.

Основные положения метода базисных матриц (МБМ)

В предлагаемом МБМ введены в рассмотрение строчные базисные матрицы [Войналович, 1987], [Войналович, 1988], [Кудин, 2002]. Базисные матрицы в ходе итераций решения задачи последовательно изменяются вводом-выводом из нее строк-нормалей ограничений. В общем случае в исследуемой модели количество ограничений превышает количество переменных вида (2), а в данном случае в МЛ $m = n$:

Определение 1. Матрицу A_a , составленную из m линейно независимых нормалей ограничений (2),

будем называть базисной, а решение соответствующей ей системы уравнений $A_a u_0^T = C^0$ базисным.

Две базисные матрицы отличающиеся одной строкой будем называть смежными.

Пусть: β_j , $i, j \in I = \{1, 2, \dots, m\}$ - элементы базисной подматрицы A_a , e_r - элементы матрицы

A_a^{-1} , обратной к A_a ; $e_k = (A_a^{-1})_k$ - столбец обратной матрицы. Решение $u_0 = (u_{01}, u_{02}, \dots, u_{0m})$

системы уравнений $A_a u^T = c^0$, где, в общем случае, c^0 - подвектор C , компоненты которого состоят из

правых частей ограничений (2), нормали которых образуют базисную матрицу A_a ;

$\alpha_r = (\alpha_{r1}, \alpha_{r2}, \dots, \alpha_{rm})$ - вектор разложения нормали ограничения $a_r u_1 \leq c_r$ за строками базисной

матрицы A_a , $\alpha_0 = (\alpha_{01}, \alpha_{02}, \dots, \alpha_{0m})$ - вектор разложения градиента целевой функции (3) по строкам

базисной матрицы A_a , $\Delta_r = a_r u_0^T - c_r$ - невязка r -го ограничения (2) в вершине u_0 ;

J_a, J_f , $J = J_a \cup J_f$ - множества индексов базисных и небазисных ограничений (2). В работе

[Войналович, 1987] приведены формулы связи базисного решения, коэффициентов разложения нормалей ограничений и целевой функции (3), коэффициентов обратной матрицы, невязок ограничений и значений

целевой функции при переходе к базисной матрице \bar{A}_a , которая образуется из матрицы A_a заменой ее

строки a_k на a_l , которая не входит в базисную матрицу A_a . В новой базисной матрице \bar{A}_a введенные

величины будем называть элементами метода базисных матриц и будем обозначать черточкой сверху,

т.е. $\bar{\beta}_j$, $\bar{\alpha}_r$, $\bar{\Delta}_k$, \bar{e}_{ri} , $\bar{\alpha}_0$. Пусть $a_{i1}, a_{i2}, \dots, a_{im}$ - нормали ограничений,

$a_j u^T \leq c_j$, $j \in J_a$, где $J_a = \{i_1, i_2, \dots, i_m\}$ - индексы ограничений, нормали которых образуют

строки базисной матрицы A_a , \dot{a}_l - нормаль ограничения $a_l u \leq c_l$, $\alpha_l = (\alpha_{l1}, \alpha_{l2}, \dots, \alpha_{lm})$ -

коэффициенты разложения вектора a_l по строкам базисной матрицы A_a .

Лемма 1. (Критерий линейной независимости системы векторов). Необходимым и достаточным условием линейной независимости строк матрицы модели $a_{i_1}, a_{i_2}, \dots, a_{i_{k-1}}, a_l, a_{i_{k+1}}, \dots, a_{i_m}$, образованных заменой строки a_{i_k} , которая занимает k -ю строку в базисной матрице A_a , строкой a_l , является выполнение условия $\alpha_{lk} \neq 0$.

Теорема 1. (О связи между смежными базисными матрицами). Между коэффициентами разложения нормалей ограничений (2) и целевой функции (3) за строками базисной матрицы, элементами обратных матриц, базисными решениями, невязками ограничений (2) и значениями целевой функции для двух смежных базисных матриц имеют место соотношения

$$\bar{\alpha}_{rk} = \frac{\alpha_{rk}}{\alpha_{lk}}, \quad \bar{\alpha}_{ri} = \alpha_{r^3} - \frac{\alpha_{rk}}{\alpha_{lk}} \alpha_{i^3}, \quad r = \overline{0, n}; \quad i = \overline{1, m}; \quad i^3 \neq k; \quad (4)$$

$$\bar{e}_{rk} = \frac{e_{rk}}{\alpha_{lk}}, \quad \bar{e}_{ri} = e_{r^3} - \frac{e_{rk}}{\alpha_{lk}} \alpha_{i^3}, \quad r = \overline{1, m}; \quad i = \overline{1, m}; \quad i^3 \neq k; \quad (5)$$

$$\bar{u}_{0j} = u_{0j} - \frac{e_{jk}}{\alpha_{lk}} \Delta_l, \quad j = \overline{1, m}, \quad (6)$$

$$\bar{\Delta}_k = -\frac{\Delta_l}{\alpha_{lk}}, \quad \bar{\Delta}_r = \Delta_r - \frac{\alpha_{rk}}{\alpha_{lk}} \Delta_l, \quad r = \overline{1, n}; \quad r \neq k; \quad (7)$$

$$\bar{B}u_0 = Bu_0 - \frac{\alpha_{0k}}{\alpha_{lk}} \Delta_l, \quad (8)$$

причем условием того, что матрица остаётся базисной при замещении вектором a_l k -й строки базисной матрицы A_a , есть выполнение условия $\alpha_{lk} \neq 0$, условием допустимости опорного базисного решения есть $\alpha_{lk} < 0$, роста значений целевой функции $\alpha_{0k} < 0$.

Доказательство леммы 1 и теоремы 1 основывается на теоретических положениях, изложенных в [Войналович, 1987], [Войналович, 1988], [Кудин, 2002].

Соотношения (4)-(8) будут основополагающими при построении алгоритма поиска не только оптимального решения, но проведения анализа свойств МЛ методом базисных матриц.

Определение 2. Допустимое базисное решение u_0 оптимальное, если $Bu_0 \geq Bu$ для всех u , которые удовлетворяют (2).

Теорема 2. Для оптимальности базисного решения u_0 необходимо и достаточно неотрицательности коэффициентов разложения вектора нормали целевой функции (3) по строкам базисной матрицы A_a , т.е. $\alpha_{ok} \geq 0$ для всех $k = \overline{1, m}$, причем задача (2),(3) с квадратной невырожденной матрицей ограничений имеет единственное решение тогда и только тогда, когда $\alpha_{0i} > 0$, $i = \overline{1, m}$, а необходимым и достаточным условием не единственности решения задачи есть $\exists i \in I$ таких, что $\alpha_{0i} = 0$, причём множество решений имеет ребра неограниченности.

Справедливость критерия оптимальности вытекает из формулы (8) теоремы 1.

Выводы

Применение симплексной идеологии на основе МБМ при анализе МЛ даёт возможность:

- исследовать свойства решений МЛП (2), (3) при изменениях в (3);
- проводить анализ свойств МЛ при изменении значений отдельных элементов и ее компонент;
- использовать решение исходной МЛ при анализе возмущенной модели;
- контролировать или направлено изменять величину ранга системы;
- находить решение квадратной системы уравнений за фиксированное количество шагов;
- строить начальные решения задач на основе тривиальных базисных матриц, которые исключают трудоёмкие начальные вычисления;
- применять схему анализа для задач, которые предусматривают многошаговость или многократность расчетов на моделях при изменениях в компонентах модели.

Литература

- [Леонтьев, 1972] Леонтьев В.В., Форд Д. Межотраслевой анализ воздействия структуры экономики на окружающую среду // Экономика и математические методы.- 1972.-Т.VII.-Вып.3.-С.370-400.
- [Гасс, 1961] Гасс С. Линейное программирование. Физматгиз,-1961.
- [Волошин, 1987] Волошин А.Ф. Метод локализации области оптимума в задачах математического программирования // Докл. АН СССР. - 1987. -293, N 3.- С. 549-553.
- [Орловский, 1981] Орловский С.А Принятие решения при нечёткой исходной информации.- М.: Наука,-1981.- 206с.
- [Волошин,1993] Волошин А.Ф. Войналович В.М., Кудин В.И. Предоптимизационные и оптимизационные схемы сокращения размерности задачи линейного программирования // Автоматика,N4, 1993.
- [Войналович, 1987] Волкович В.Л., Войналович В.М., Кудин В.И. Релаксационная схема строчного симплекс метода // Автоматика. - 1987. -N4.-С. 79-86.
- [Войналович, 1988] Волкович В.Л., Войналович В.М., Кудин В.И. Релаксационная схема двойственного строчного симплекс метода // Автоматика.-1988. -N 1.-С.39-46.
- [Кудин, 2002] Кудин В.И. Применение метода базисных матриц при исследовании свойств линейной системы // Вестник Киевского университета. Серия физ.-мат. науки. - 2002.-2.- С. 56-61.

Сведения об авторах

Владимир Кудин - д.т.н., с.н.с., Киевский национальный университет имени Тараса Шевченко, факультет кибернетики, Украина, E-mail: V_I_Kudin@mail.ru

Григорий Кудин - к.т.н., с.н.с., Киевский национальный университет имени Тараса Шевченко, факультет кибернетики, Украина, E-mail: Kuding@mail.univ.kiev.ua

Алексей Волошин - д.т.н., профессор, Киевский национальный университет имени Тараса Шевченко, факультет кибернетики, Украина, E-mail: ovoloshin@unicyb.kiev.ua

О ЗАДАЧЕ НАХОЖДЕНИЯ СТРОГОЙ РЕЗУЛЬТИРУЮЩЕЙ РАНЖИРОВКИ В ВИДЕ МЕДИАНЫ КЕМЕНИ-СНЕЛЛА

Павел Антосяк, Алексей Волошин

Аннотация: Рассматривается задача, возникающая во многих приложениях, построения на основе индивидуальных предпочтений экспертов результирующей (коллективной) ранжировки в виде медианы Кемени-Снелла. Приводятся теоретические результаты, полученные авторами, на основании которых предлагаются алгоритмы, позволяющие решать задачи большой размерности.

Ключевые слова: ранжирование, медиана Кемени-Снелла; принципы Кондорсе, Борда, Парето; последовательный анализ вариантов, декомпозиция.

ACM Classification Keywords: H4.2 Decision support

Введение

Задача построения на основе индивидуальных экспертных предпочтений результирующего (коллективного) порядка возникает во многих приложениях, в первую очередь, в социологии и политологии. Регулярно публикуются рейтинги высших учебных заведений (насчитывающие тысячи единиц), стран (как правило, от 150 до 200) по десяткам критериев. Один из основных рейтингов – индекс развития человеческого потенциала, учитывающий демографическую ситуацию в стране, развитие рынка труда, материальное благосостояние, уровень охраны здоровья, доступность и качество образования и т.д. Актуальной задачей для Украины, переживающей перманентное состояние выборов, является задача построения списка кандидатов в депутаты Верховного Совета из 450 индивидуумов от партий (блоков). Как понимают авторы, эта задача решается эвристически на основе неизвестных и непонятных принципов, хотя в идеале результирующий список должен отражать предпочтение каждого делегата съезда, насчитывающего сотни и тысячи участников.

В докладе приводятся в обзорной форме основные теоретические результаты, полученные авторами в последние годы, на основании которых предлагаются алгоритмические процедуры решения указанной задачи. В основе численных алгоритмов лежит методология последовательного анализа вариантов [Михалевич, 1965] [Волкович, Волошин, 1978] в конкретизации [Волошин, 1987]. Авторы придерживаются понятий и определений, приведенных в [Волошин, 2006].

1. Постановка задачи

Пусть на фиксированном множестве объектов $O = \{o_1, \dots, o_n\}$ экспертами, нормированные коэффициенты компетентности α_l , $l \in L = \{1, \dots, m\}$, которых известны, заданы матрицы строгих парных сравнений $P^{(l)}$, $l \in L$. Элементы $p_{ij}^{(l)} \in \{-1, 1\}$ матриц $P^{(l)}$ являют собой результат сравнения l -ым экспертом объектов o_i и o_j , $i, j \in I = \{1, \dots, n\}$, $i \neq j$:

$$p_{ij}^{(l)} = \begin{cases} 1, & \text{если объект } o_i \text{ лучше } o_j \text{ по мнению } l\text{-го эксперта,} \\ -1, & \text{если по мнению } l\text{-го эксперта } o_i \text{ хуже } o_j. \end{cases}$$

Одним из методов нахождения результирующей ранжировки является вычисление медианы Кемени-Снелла:

$$R^* \in \operatorname{Argmin}_{R \in \mathfrak{R}} \{d(R) = \sum_{l \in L} \alpha_l d(R, P^{(l)})\}, \quad (1.1)$$

где \mathfrak{R} – множество всех матриц, которые отвечают строгой ранжировке n объектов,

$$d(R, P^{(l)}) = \frac{1}{2} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n |r_{ij} - p_{ij}^{(l)}| \text{ – расстояние Хемминга между } R \text{ и } P^{(l)}.$$

В работе [Антосяк, 2007] задача (1.1) сводится к эквивалентной задаче вида:

$$(i_1^*, \dots, i_n^*) \in \operatorname{KS} = \operatorname{Argmax}_{(i_1, \dots, i_n) \in \Omega} \{F(i_1, \dots, i_n) = \sum_{k=1}^{n-1} \sum_{j=k+1}^n a_{ikj}\}, \quad (1.2)$$

где $a_{ikj} = \sum_{l \in L} \alpha_l p_{l_{ikj}}^1$; Ω – множество всех возможных перестановок множества индексов I ; (i_1, \dots, i_n) – элемент множества Ω (вариант задачи (1.2)).

2. Процедуры локализации интервалов изменения оптимальных рангов объектов

Определение 2.1. Для $\forall R \leq r$, $R, r \in I$, множество $\{R, R+1, \dots, r-1, r\}$ назовем интервалом $[R, r]$ изменения рангов.

Выберем произвольный объект с индексом $i \in I$.

Определение 2.2. Ранг $r_i \in I$ назовем оптимальным рангом i -го объекта, если существует оптимальный вариант $(i_1^*, i_2^*, \dots, i_{\eta-1}^*, i_{\eta}^* = i, i_{\eta+1}^*, \dots, i_{n-1}^*, i_n^*) \in \operatorname{KS}$.

Множество всех оптимальных рангов i -го объекта обозначим через \mathfrak{R}_i .

Пусть

$$r_i^* = \operatorname{argmax}_{j \in \mathfrak{R}_i} j, \quad R_i^* = \operatorname{argmin}_{j \in \mathfrak{R}_i} j,$$

соответственно, минимальный и максимальный оптимальные ранги i -о объекта.

Определение 2.3. Интервал изменения рангов $[R_i^*, r_i^*]$ назовем интервалом изменения оптимальных рангов i -го объекта.

Под локализацией интервала изменения оптимальных рангов i -го объекта будем понимать процедуру построения такого интервала $[R_i^{(k_i^{loc})}, r_i^{(k_i^{loc})}]$, что

$$I \supseteq \dots \supseteq [R_i^{(k_i^{loc}-1)}, r_i^{(k_i^{loc}-1)}] = [R_i^{(k_i^{loc})}, r_i^{(k_i^{loc})}] \supseteq [R_i^*, r_i^*].$$

На каждом k_i -ом ($k_i \geq 1$) шаге процедуры локализации интервала изменения оптимальных рангов i -го объекта будем рассматривать вопрос о возможности $R_i^{(k_i-1)} \in \mathfrak{R}_i$ и $r_i^{(k_i-1)} \in \mathfrak{R}_i$.

Пусть $V_{i(n-1)}^q$ ($1 \leq q \leq n-1$) – множество всех возможных размещений без повторений множества индексов $I \setminus \{i\}$ из $(n-1)$ -го элемента по q . Элемент множества $V_{i(n-1)}^q$ будем записывать как (i_1, \dots, i_q) .

Пусть *extr* – это *max* или *min*. Обозначим

$$S_{iq}^{(k_i)} = \text{extr}_{(i_1, \dots, i_q) \in DV_{iq}^{(k_i)}} \sum_{j=1}^q a_{ij}, \quad (2.1)$$

где $DV_{iq}^{(k_i)} = \{(i_1, \dots, i_q) | (R_i^{(k_i-1)} + j) \in D_{i_1 i_q}^{(k_j)}, j = \overline{1, q}\}$, $D_{i_1 i_q}^{(k_i)} = [R_j^{(k_j)}, r_j^{(k_j)}] \cap [R_i^{(k_i-1)} + 1, R_i^{(k_i-1)} + q]$, для $\forall j \in I \setminus \{i\}$, если

(2.1) – задача на максимум. Иначе, $DV_{iq}^{(k_i)} = \{(i_1, \dots, i_q) | (r_i^{(k_i-1)} - j) \in D_{i_1 i_q}^{(k_j)}, j = \overline{1, q}\}$,

$D_{i_1 i_q}^{(k_i)} = [R_j^{(k_j)}, r_j^{(k_j)}] \cap [r_i^{(k_i-1)} - q, r_i^{(k_i-1)} - 1]$, для $\forall j \in I \setminus \{i\}$.

Теорема 2.1 (Необходимые условия оптимальности задачи (1.2)). Для каждого оптимального варианта (i_1^*, \dots, i_n^*) задачи (1.2) для $\forall t, h \in I$, $t < h$, используются следующие неравенства:

$$\sum_{j=t+1}^h a_{i_j^*} \geq 0,$$

$$\sum_{j=t}^{h-1} a_{i_j^*} \leq 0.$$

Как следствие имеем

Утверждение 2.1 (Достаточные условия локализации).

- I. Для задачи (2.1) на максимум. Если $S_{iq}^{(k_i)} < 0$ хотя бы для одного $q \in \{1, r_i^{(k_i-1)} - R_i^{(k_i-1)}\}$, то $R_i^{(k_i-1)} \notin \mathfrak{R}_i$.
- II. Для задачи (2.1) на минимум. Если $S_{iq}^{(k_i)} > 0$ хотя бы для одного $q \in \{1, r_i^{(k_i-1)} - R_i^{(k_i-1)}\}$, то $r_i^{(k_i-1)} \notin \mathfrak{R}_i$.

3. Процедуры фиксации результирующего отношения между двумя объектами

Выберем произвольные индексы $i, j \in I$, $i \neq j$. Обозначим $o_i \succ_{KS} o_j$, если в коллективном порядке, который отвечает медиане Кемени-Снелла, объект o_i лучше, чем объект o_j .

Замечание 3.1. Следует отметить, что введенное обозначение используется нами тогда и только тогда, когда не существует ни одного оптимального варианта задачи (1.2) такого, что $(i_1^*, \dots, i_t^* = j, \dots, i_h^* = i, \dots, i_n^*)$, $1 \leq t < h \leq n$.

Известно [Литвак, 1982], что медиана Кемени-Снелла удовлетворяет условию Парето. То есть, если в индивидуальных предпочтениях экспертов $o_i \succ o_j$, то $o_i \succ_{KS} o_j$. Следовательно, приняв во внимание правило построения элементов a_{ij} , имеем

Утверждение 3.1. Если $a_{ij} = 1$, то $o_i \succ_{KS} o_j$.

Из определения интервала изменения оптимальных рангов следует

Утверждение 3.2. Если $r_i^{(k_i^{loc})} \leq R_j^{(k_j^{loc})}$, то $o_i \succ_{KS} o_j$.

Для любого $i \in I$ введем обозначение:

$$K_i^- = \{k \in I | a_{ik} < 0, k \neq i\}, \quad K_i^+ = \{k \in I | a_{ik} > 0, j \neq i\},$$

$$S_i^- = \sum_{k \in K_i^-} a_{ik}, \quad S_i^+ = \sum_{j \in K_i^+} a_{ij}.$$

Лемма 3.1. Если $a_{ij} > -S_i^-$, то $o_i \succ_{KS} o_j$.

Если положительную величину a_{ij} рассматривать как «выигрыш» i -го объекта в парных «поединках» с j -м объектом, то результат леммы 3.1 следующий: если «выигрыш» i -го объекта в парных «поединках» с j -м объектом больше, чем его суммарный «проигрыш», то $o_i \succ_{KS} o_j$.

Лемма 3.2. Если $a_{ij} < -S_i^+$, то $o_i \prec_{KS} o_j$. То есть, если «проигрыш» i -го объекта в парных «поединках» с j -м объектом меньше, чем его суммарный «выигрыш», то $o_i \prec_{KS} o_j$.

Если же обозначить

$$K_{ij}^- = \{k \in I \mid a_{ik} < 0, k \neq i, k \neq j, R_j^{(k_j^{loc})} < R_k^{(k_k^{loc})} < r_i^{(k_i^{loc})} \text{ або } R_j^{(k_j^{loc})} < r_k^{(k_k^{loc})} < r_i^{(k_i^{loc})}\},$$

$$K_{ij}^+ = \{k \in I \mid a_{ik} > 0, k \neq i, k \neq j, R_i^{(k_i^{loc})} < R_k^{(k_k^{loc})} < r_j^{(k_j^{loc})} \text{ або } R_i^{(k_i^{loc})} < r_k^{(k_k^{loc})} < r_j^{(k_j^{loc})}\},$$

$$S_{ij}^- = \sum_{k \in K_{ij}^-} a_{ik}, \quad S_{ij}^+ = \sum_{k \in K_{ij}^+} a_{ik},$$

то из доказательства лемм 3.1, 3.2 и утверждения 3.2 получаем справедливость следующих утверждений.

Утверждение 3.3. Если $a_{ij} > -S_{ij}^-$, то $o_i \succ_{KS} o_j$.

Утверждение 3.4. Если $a_{ij} < -S_{ij}^+$, то $o_i \prec_{KS} o_j$.

Тогда, приняв во внимание замечание 1, имеем

Утверждение 3.5. Если в результате утверждений 3.1–3.4 было установлено $o_i \succ_{KS} o_k$ и $o_k \succ_{KS} o_j$, то $o_i \succ_{KS} o_j$.

4. Процедуры декомпозиции

Пусть $\{l_1, \dots, l_q\}$, $2 \leq q \leq n$, некоторое разбиение множества индексов I такое, что $l_t \cap l_h = \emptyset$ при $t \neq h$ и $\bigcup_{k=1}^q l_k = I$. Понятие декомпозиции задачи (1.2) имеет следующее содержание.

В случае, когда для каждого оптимального варианта (i_1^*, \dots, i_n^*) задачи (1.2) выполняется:

$$\text{если } j \in l_h \text{ и } i_j^* = j, \text{ то } \sum_{k=0}^{h-1} \dim(l_k) < t \leq \sum_{k=0}^h \dim(l_k) \quad (l_0 = \emptyset), \quad (4.1)$$

будем говорить, что декомпозиция $\{l_1, \dots, l_q\}$ является необходимой. Если же существует оптимальный вариант (i_1^*, \dots, i_n^*) задачи (1.2), для которого выполняется (4.1), то будем говорить, что декомпозиция $\{l_1, \dots, l_q\}$ допустима.

Декомпозиция задачи (1.2) осуществляется на основании следующих утверждений.

Лемма 4.1. Пусть существует набор индексов $I^{(1)} \subset I$ такой, что $a_{ij} \geq 0$ для $\forall i \in I^{(1)}, j \in I \setminus I^{(1)}$. Тогда декомпозиция $\{I^{(1)}, I \setminus I^{(1)}\}$ является допустимой.

Лемма 4.2. Пусть существует набор индексов $I^{(2)} \subset I$ такой, что $a_{ij} > 0$ для $\forall i \in I^{(2)}, j \in I \setminus I^{(2)}$. Тогда декомпозиция $\{I^{(2)}, I \setminus I^{(2)}\}$ является необходимой.

В терминах интервалов изменений оптимальных рангов справедливыми являются:

Утверждение 4.1. Пусть существует набор индексов $I^{(3)} \subset I$ такой, что $r_i^{(k_{loc})} \leq R_j^{(k_{loc})}$ для $\forall i \in I^{(3)}, j \in I \setminus I^{(3)}$. Тогда декомпозиция $\{I^{(3)}, I \setminus I^{(3)}\}$ является необходимой.

Утверждение 4.2. Пусть существует набор индексов $I^{(4)} \subset I$ такой, что $r_i^{(k_{loc})} \leq R_j^{(k_{loc})}$ для $\forall i \in I^{(4)}, j \in I \setminus I^{(4)}$ что $a_{ij} < 0$. Тогда декомпозиция $\{I^{(4)}, I \setminus I^{(4)}\}$ возможна (то есть, она либо допустима, либо необходима).

Обозначим через $I^* \subset I$ наименьшее такое множество, для которого выполняются условия леммы 1 или леммы 2.

Кластеризация (или классификация) неформально определяется как процесс объединения объектов в группы с "похожими" признаками. В этом случае I^* - это класс «фактических лидеров» (они выигрывают в парных сравнениях у произвольного кандидата, который не входит в этот класс) для заданного профиля предпочтений.

Утверждение 4.3 (следствие из леммы 4.2). В условиях леммы 4.2 I^* содержит всех победителей Копленда для соответствующих предпочтений задачи (1.2).

5. Локальный поиск

Для нахождения локально-оптимальных решений задачи (1.3) на сокращенном множестве вариантов в работе [Антосяк, 2006] предлагается алгоритм нахождения строгого результирующего ранжирования на основе метода вектора спада.

В качестве окрестности фиксированной точки x рассматривается множество следующих инвертирований:

$$x(j_1, j_2, \dots, j_s) = (x_1, \dots, -x_{j_1}, \dots, -x_{j_2}, \dots, -x_{j_s}, \dots, x_N), \quad (5.1)$$

$$1 \leq j_1 < \dots < j_s \leq N, 1 \leq s \leq N.$$

Поскольку не всякое инвертирование вида (5.1) дает допустимый вариант, то вводится множество допустимых инвертирований:

$$DI = \{(j_1, \dots, j_s) \mid x(j_1, \dots, j_s) \in D, 1 \leq j_1 \leq \dots \leq j_s \leq N, 1 \leq s \leq N\}.$$

В качестве вектора спада задачи (1), в некоторой окрестности точки $x \in D$ будем рассматривать вектор:

$$R(x) = [\Delta(x, x(j_1, \dots, j_s))],$$

$$\Delta(x, x(j_1, \dots, j_s)) = F(x(j_1, \dots, j_s)) - F(x), (j_1, \dots, j_s) \in DI.$$

Основной проблемой реализации метода является построение на каждом k -м шаге i ($k = 1, 2, \dots$) точек окрестностей. Учитывая практический аспект, во-первых, предлагается ограничиваться рассмотрением

случая $s=2$; во-вторых, для нахождения точек окрестности предлагается процедура определения допустимых инвертирований, в-третьих, поиск в окрестности происходит до первого «улучшения».

«Двойственным» к рассмотренному подходу является поход, реализованный в работе [Антосяк, 2005], в которой рассматривается построение ациклического отношения, которое в некотором смысле является самым близким к «идеальной точке».

Пусть R - бинарное отношение, на котором целевая функция задачи (1.1) достигает своего минимального значения, но которое не является ациклическим. $R = (r_{ij})$, $i, j \in I$, – матрица данного бинарного отношения.

Поскольку $R \notin \mathfrak{R}$, то R не может выступать в качестве решения. Такое бинарное отношение назовем «идеальной точкой» задачи (1.1).

Известно [Макаров, 1982], что для случая, который рассматривается нами (в силу свойств матриц индивидуальных парных сравнений), отсутствие (наличие) циклов эквивалентно отсутствию (наличию) циклов длины 3.

Определение 5.1. Циклическостью элемента r_{ij} , $i, j \in I$, матрицы, которая отвечает бинарному отношению R , назовем количество циклов длины 3, в которых объекты o_i и o_j находятся в отношении $o_i R o_j$.

Определение 5.2. Элемент с ненулевой циклическостью назовем циклическим элементом бинарного отношения R .

Для фиксированного, но любого, объекта $o_k \in O$, $k \in I$, определим множество пар индексов:

$$\tilde{N}_k = \{(i, j) | r_{ij} = 1, i \in I_k^1, j \in I_k^{-1}\},$$

где

$$I_k^{-1} = \{i | r_{ki} = -1, i \in I \setminus \{k\}\}, I_k^1 = \{i | r_{ki} = 1, i \in I \setminus \{k\}\}.$$

Справедливы следующие утверждения.

Утверждение 5.1. Множество $C = \bigcup_{i \in I} C_i$ описывает множество всех циклических элементов бинарного отношения R .

Утверждение 5.2. Отношение R ациклическое тогда и только тогда, когда $C = \emptyset$.

Ациклическость бинарного отношения R предлагается достигать путем последовательного исключения его циклических элементов.

6. Медиана Кемени-Снелла и принципы Кондорсе и Борда

В данной работе в принцип Кондорсе вкладывается следующее. Положительная (отрицательная) величина a_{ij} (разница между «суммарной компетентностью» группы экспертов, которые отдали предпочтение объекту o_i над объектом o_j и «суммарной компетентностью» группы экспертов, которые отдали предпочтение объекту o_j над объектом o_i) рассматривается как «выигрыш» («проигрыш») – результат «победы» («поражения») объекта o_i над объектом o_j . Сильным победителем Кондорсе объявим тот объект, коотрый имеет «выигрыш» у каждого другого объекта. Если же некоторый объект не имеет «проигрышей» ни от одного из других объектов, то по определению он объявляется слабым победителем Кондорсе. В принцип Борда вкладывается следующее. За правило подсчета баллов берется величина

$$S(o_i) = \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij}.$$

Под коллективным порядком (сильным) Кондорсе понимают коллективное предпочтение, построенное по следующему принципу [Мулен, 1991]: сильный победитель Кондорсе исключается из рассмотрения, для кандидатов (объектов), которые остались, снова ищется победитель, который занимает второе место в коллективном предпочтении и так далее. Аналогичным способом можно определить коллективный порядок для слабых победителей Кондорсе (в случае отсутствия сильных) и коллективный порядок Борда. Обозначим через K и B – соответственно все строгие коллективные порядки Кондорсе (единственный сильный или же все слабые) и все коллективные порядки Борда для индивидуальных предпочтений задачи (1.1)

Утверждение 6.1 (О медиане Кемени-Снелла и коллективных порядках Кондорсе). Если $K \neq \emptyset$, то $K = KS$.

Как известно [Мулен, 1991], правила Кондорсе и обобщенное правило Борда в определенном смысле являются «несовместимыми» – существуют профили, при которых победитель Кондорсе не может быть победителем Борда ни при какой системе баллов.

Утверждение 6.2 (Достаточное условие совпадения медианы Кемени-Снелла с коллективным порядком Борда). Пусть для некоторого оптимального варианта $(i_1^*, i_2^*, \dots, i_n^*) \in KS$ выполняется условие:

$$i_k^* \in \underset{i \in \{i_k^*, \dots, i_n^*\}}{\text{Argmax}} \{S(o_i) = \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij}\}, \quad k = \overline{1, n-1}.$$

Тогда $(i_1^*, i_2^*, \dots, i_n^*) \in B$.

Утверждение 6.3 (О коллективном порядке Кондорсе, Борда и медиане Кемени-Снелла). Если $K \neq \emptyset$ и выполняются условия утверждения 6.2, то $(K = KS) \cap B \neq \emptyset$.

7. Медиана Кемени и коэффициент согласованности

Пусть m экспертов ($m > 2$) задали ранжировки альтернатив по индивидуальным предпочтениям. Для характеристики степени согласованности экспертов при парном сравнении альтернатив вводится коэффициент согласованности [Литвак, 1982]:

$$V = \frac{8 \sum_{i \neq j} C_{b_{ij}}^2}{m(m-1)n(n-1)} - 1,$$

где b_{ij} – число экспертов, отдавших предпочтение объекту o_i по сравнению с объектом o_j .

Значения V изменяются от 1, при полном совпадении ранжировок экспертов, до 0, когда согласованность оценок экспертов отсутствует.

В работах [Кузьмин 1974], [Тюрин, 1978] указывается на связь, которая существует между мерами близости и коэффициентами ранговой корреляции по Кенделлу [Кузьмин, 1974]. Коэффициент ранговой корреляции по Кенделлу для ранжировок $P^{(1)}$ и $P^{(2)}$ $\tau(P^{(1)}, P^{(2)})$, а также мера близости Кемени $d(P^{(1)}, P^{(2)})$, удовлетворяют при отсутствии связанных рангов следующему соотношению:

$$\tau(P^{(1)}, P^{(2)}) = 1 - \frac{2d(P^{(1)}, P^{(2)})}{n(n-1)}.$$

Рассмотрим задачу нахождения строгой результирующей ранжировки в виде медианы Кемени (это частный случай медианы Кемени-Снелла, когда не учитывается компетентность экспертов [Тоценко, 2002]). Тогда справедливым будет следующий результат.

Теорема 7.1. Для медианы Кемени и коэффициента согласованности V справедлива оценка:

$$\frac{n(n-1)}{2} \left[m - \sqrt{m(1+V(m-1))} \right] \leq d(R^*) \leq \frac{n(n-1)}{2} (m-1)(1-V).$$

Заключение

На основе предложенных процедур построен декомпозиционный алгоритм решения задачи (1.1). Осуществлена его программная реализация с использованием ресурса высокоэффективных кластерных систем СКИТ-1 и СКИТ-2 Института кибернетики им. В.М. Глушкова НАНУ. Получены первые результаты, позволяющие сделать вывод о его перспективности. Окончательные результаты будут опубликованы.

Библиография

- [Волошин, 2006] Волошин О.Ф., Машенко С.О. Теорія прийняття рішень .-Київ:ВПЦ КУ, 2006. –304с.
- [Антосяк, 2007] Антосяк П.П. До правил голосування Кондорсе і Борда// Вісник Київського університету. Сер.: Фізико-математичні науки. – 2007. – № 4. – С. 50–55.
- [Михалевич, 1965] Михалевич В.С. Последовательные алгоритмы оптимизации и их применение. I. III// Кибернетика. – 1965. – №1. – С. 45-55; – №2. – С. 85-88.
- [Волкович, Волошин, 1978] Волкович В.Л., Волошин А.Ф. Об одной схеме метода последовательного анализа и отсеивания вариантов// Кибернетика. – 1978. – №4. – С. 98-105.
- [Волошин, 1987] Волошин А.Ф. Метод локализации области оптимума в задачах математического программирования// Доклады АН СССР. – 1987. –т. 293, №3.– С. 549-553.
- [Литвак, 1982] Литвак Б.Г. Экспертная информация: Методы получения и анализа. – М.: Радио и связь, 1982. – 184 с.
- [Антосяк, 2006] Антосяк П.П. Алгоритм побудови колективного ранжування на основі методу вектора спаду// Вісник Київського університету. Вип.4 Серія: фіз.-мат. науки, Київ, 2006. – С.145-147.
- [Антосяк, 2005] Антосяк П.П., Волошин О.Ф. Алгоритм знаходження медіани Кемени-Снелла методом ідеальної точки// Вісник Київського університету. Вип.3 Серія: фіз.-мат. науки, Київ, 2005.
- [Макаров, 1982] Макаров И.М. и др. Теория выбора и принятия решений. – М.: Наука, 1982. – 328 с.
- [Мулен, 1991] Мулен Э. Кооперативное принятие решений: Аксиомы и модели. – М.: Мир, 1991. – 464 с.
- [Кузьмин, 1974] Кузьмин В.Б., Овчинников С.В. Об измерениях в порядковых шкалах//Автоматика, 1974, №11.
- [Тюрин, 1978] Тюрин Ю.Н. Непараметрические методы статистики. – М.: Знание, 1978.
- [Тоценко, 2002] Тоценко В.Г. Методы и системы поддержки принятия решений. Алгоритмический аспект. – Київ: Наукова думка, 2002. – 381 с.

Сведения об авторах

Антосяк Павел Павлович – Ассистент, Ужгородский национальный университет, математический факультет. Ужгород, Украина. E-mail: antosp@ukr.net

Волошин Алексей Федорович – Профессор, Киевский национальный университет имени Тараса Шевченко, факультет кибернетики. Киев, Украина. E-mail: ovoloshin@unicyb.kiev.ua

ПРИМЕНЕНИЕ НЕЧЕТКОЙ ЛОГИКИ ДЛЯ ЗАДАЧ КОЛЛЕКТИВНОГО ВЫБОРА

Николай Маляр

Abstract: Рассматриваются проблемы коллективного выбора. Обсуждаются преимущества и недостатки коллективного выбора, различные подходы для решения этих задач. Предлагается подход, позволяющие использовать аппарат нечеткой логики для моделирования задач коллективного выбора.

Keywords: коллективный выбор, нечеткая логика, принятие решений.

ACM Classification Keywords: H.1.1. Systems and Information Theory.

Введение

Очень часто при решении практических задач возникает ситуация, когда одно лицо, принимающее решение (ЛПР), не в силах принять верное решение и обязан обратиться за помощью к другим ЛПР, имеющих одинаковые намерения и полномочия. Такой процесс носит название коллективный, групповой или коллегиальный выбор.

Сформулируем задачу коллективного выбора следующим образом. Пусть имеется множество альтернатив, которое может быть задано дискретным или континуальным образом. Предположим, что это множество состоит из m объектов (проектов, планов, кандидатов и т.д.), обозначим его через $A = \{a_1, a_2, \dots, a_m\}$ и в дальнейшем будем называть множеством альтернатив.

Пусть также имеется группа из n ЛПР $I = \{i_1, i_2, \dots, i_n\}$, которыми на данном множестве альтернатив задано n различных индивидуальных предпочтений P_1, P_2, \dots, P_n .

Ставится задача о выработке некоторого нового отношения P , которое будет выражать в некотором смысле „общее мнение” и позволит выбрать (выделить) из множества альтернатив A одну или несколько равнозначных альтернатив.

Перед тем, как выработать коллективное решение, каждый участник из группы ЛПР должен ознакомиться со свойствами альтернатив и дать им соответствующую оценку, при этом несущественно, руководствуется ли он субъективными соображениями или учитывает объективные характеристики альтернатив, ведет себя как эгоист или как альтруист.

На основании этих оценок строятся предпочтения и с помощью определенного правила, которое называется функцией коллективного выбора, процедурой голосования, методом объединения, арбитражной схемой производится выбор единственной (или нескольких равнозначных) альтернативы.

Правила коллективного выбора

Рассмотрим часто используемые правила коллективного выбора [Вержбицкий, 1987; Мулен, 1991].

Широко распространенным способом получения отношения P являются процедуры голосования. В этих процедурах групповое предпочтение совпадает с правилом большинства. Различаются: правило простого большинства, когда принятой считается альтернатива, получившая предпочтение не менее, чем $[n/2]+1$ голосов членов группы. Чем в большей степени коллективное решение влияет на каждого ЛПР, тем в большей степени он заинтересован в сохранении своего права вето. Поэтому правило простого

большинства не приемлемо во многих ситуациях и вместо него используется правило квалифицированного большинства (например, две трети или три четверти голосов) или абсолютного большинства. Правило абсолютного большинства – это процедура, в которой с выбранной альтернативой согласны все члены группы. Это правило называется консенсусом или правилом единогласия и очень широко используется в современном международном праве. К этой группе правил относится правило взвешенного большинства, которое используется в международных организациях и различных акционерных обществах. Суть его заключается в том, что страна-участница или акционер обладают числом голосов, зависящих от размера его взноса или числа акций, которыми он обладает. Недостатком правила большинства является игнорирование мнения меньшинства.

Принцип диктаторства – это процедура, согласно которой принимается решение одним из членов группы. Она чаще всего применяется в военных действиях и чрезвычайных обстоятельствах.

Правило суммирования рангов учитывает мнения всех членов коллектива. Данный принцип широко используется на практике, например, для выставления оценок в спортивных состязаниях.

Принцип минимальных отклонений, согласно которому коллективным решением должно быть решение, минимизирующее расхождение между индивидуальными предпочтениями отдельных ее членов и предпочтениями группы в целом.

Как видим из приведенного выше, принятие коллективного решения является сложной процедурой и не существует универсального правила, которое может быть использовано при решении конкретной задачи. Таким образом, возникает необходимость в разработке новых подходов для решения данных проблем.

Построение нечетких множеств

Рассмотрим класс задач коллективного выбора, для которых свойства альтернатив могут оцениваться несколькими критериями эффективности. Пусть задано универсальное множество критериев эффективности, по которым может быть оценено множество альтернатив. Каждое ЛПР из группы может формировать свое подмножество критериев, по которым оно может оценить множество альтернатив, наперед не известное всем остальным членам группы.

Опишем задачу выбора для одного ЛПР. Задано множество альтернатив A и множество критериев U^i , по которым оцениваются альтернативы. Необходимо решить задачу ранжирования этих альтернатив.

Математическая модель данной задачи запишется в следующем виде:

Альт. \ Крит.	a_1	a_2	...	a_m
u_1^i	o_{11}	o_{12}	...	o_{1m}
u_2^i	o_{21}	o_{22}	...	o_{2m}
\vdots	\vdots	\vdots	\ddots	\vdots
$u_{l_k}^i$	$o_{l_k 1}$	$o_{l_k 2}$...	$o_{l_k m}$

где o_{lj} - числовая оценка j -ой альтернативы по l -му критерию. Число альтернатив равно m для всех членов группы, а количество критериев l_k - величина переменная, то есть для каждого участника - своя. Модель такого вида - это модель задачи выбора на языке матрицы решений, то есть задана матрица оценок альтернатив $O^i = (O_{lj}^i)$, $l = \overline{1, l_k^i}$, $j = \overline{1, m}$.

Не ограничивая общности, предположим, что все альтернативы принадлежат множеству Парето, а наилучшей считается альтернатива, для которой оценки достигают своего максимального или минимального значения. Пусть все оценки – положительные числа, иначе можно применить преобразование

$$O_{ij} = \left| \max_j O_{ij} - \min_j O_{ij} \right| - O_{ij}.$$

Таким образом, множество альтернатив A представляет собой некоторое подмножество точек из евклидова пространства E_{++} . Поскольку перед каждым ЛПР стоит задача ранжирования альтернатив, то для решения этой задачи используется метод, описанный в работах [Волошин, Маляр, 2005; Маляр, 2005, Маляр, 2006]. Опишем схему данного метода.

Каждый участник коллективного выбора выбирает из универсального множества критериев свое подмножество и делает по этому подмножеству оценку множества альтернатив. Для этой процедуры могут быть привлечены эксперты. Далее, каждый ЛПР обязан задать свою «точку удовлетворения» и описать нечеткое множество относительно этой точки. «Точкой удовлетворения» называется «абстрактная» альтернатива, оценки которой могут удовлетворить ЛПР. Нечеткое множество может быть описано, например, как множество альтернатив «близких», «лучших» и т.д. к «точке удовлетворения».

Результатом работы этого метода будет профиль, заданный функцией принадлежности данному нечеткому множеству. Таким образом, каждый индивидуум задает свой профиль ранжирования.

На основании индивидуальных профилей составим реляционное отношение $I \times X$ в виде матрицы:

	X_1	X_2	...	X_m
I_1	μ_{11}	μ_{12}	...	μ_{1m}
I_2	μ_{21}	μ_{22}	...	μ_{2m}
\vdots	\vdots	\vdots	\ddots	\vdots

где μ_{ij} - функция принадлежности множества, определенному i -ым ЛПР для j -ой альтернативы. Этот

профиль можно рассматривать как нечеткое отношение \tilde{R} между ЛПР и альтернативами.

Каждый ЛПР из группы может по-разному влиять на принятие коллективного решения. Это может быть связано с его компетентностью, авторитетом, количеством акций и т.д. Опишем, например, шкалу «степеней» владения профессиональными навыками для нечеткого отношения «компетентность»

<i>Степень знания</i>	<i>Числовой эквивалент</i>
<i>Очень высокая</i>	1
<i>Высокая</i>	0,8
<i>Достаточная</i>	0,6
<i>Средняя</i>	0,5
<i>Удовлетворительная</i>	0,4
<i>Недостаточная</i>	0,2
<i>Неудовлетворительная</i>	0

По этой шкалой, можем построить на множестве ЛПР I нечеткое множество \tilde{B} «компетентность». В дальнейшем воспользуемся нечетким логическим выводом [Кофман, 1982].

Определение. *Нечетким логическим выводом* называется получение заключения в виде нечеткого множества, соответствующего текущим значениям входов, с использованием нечеткой базы знаний и нечетких операций.

Основой нечеткого логического вывода является композиционное правило Заде, которое формулируется следующим образом: если известно нечеткое отношение \tilde{R} между входной (x) и выходной (y) переменными, то при нечетком значении входной переменной $x = \tilde{A}$, нечеткое значения выходной переменной определяется так: $y = \tilde{A} \circ \tilde{R}$, где \circ - максимная композиция. Таким образом, если за базу значений \tilde{R} взять реляционное отношение $I \times X$, а за нечеткое значение входной информации отношение \tilde{B} , то нечеткое значение выходной переменной определится как нечеткое множество

$$\tilde{C} = \tilde{B} \circ \tilde{R},$$

где \circ - композиция, например, максимное, минмаксное или максимультимпликативное произведения.

Воспользовавшись этим правилом, мы получили нечеткое отношение \tilde{C} индуцированное некоторым отображением из нечеткого отношения \tilde{B} .

Выводы

Таким образом, нечеткое множество \tilde{C} будем считать коллективным выбором, его функцию принадлежности используем как функцию коллективного выбора. По этой функции будет проводиться ранжирование альтернатив. Суть этого подхода состоит в том, что каждый ЛПР имеет возможность выбрать свое подмножество критериев и сравнивать альтернативы не между собой, а только со своей «точкой удовлетворения». Количество сравнений уменьшается в $(m - 1)!$ раз для каждого ЛПР.

Библиография

- [Вержбицкий, 1987] Вержбицкий А.Л. Переговоры и посредничество в условиях конфликта: Многосторонняя рациональность и интерактивные процессы // Сист. исслед., 1987. –М., - С. 76-98.
- [Мулен, 1991] Мулен Э. Кооперативное принятие решений: Аксиомы и модели. -М.: Мир. 1991.-464 с.
- [Волошин, Маляр, 2005] Волошин А. Ф., Маляр Н. Н. Нечеткие модели многокритериального коллективного выбора // Proceedings XI – th International Conference “Knowledge – Dialogue – Solution” – Sofia, 2005. – Vol. 1. – P. 247-250.
- [Маляр, 2005] Маляр М.М. Описання задач вибору на мові розмитих множин // Вісник Київського університету. Вип.4: Серія: фіз.-мат. науки, Київ, 2005.- С.197-201.
- [Маляр, 2006] Маляр М.М. Задача вибору та підхід до її розв'язання // Вісник СевДТУ. Вип.50: Інформатика, електроніка, зв'язок: Зб. наук. пр. – Севастополь: Вид-во СевДТУ, 2006.- С. 98-104.
- [Кофман, 1982] Кофман А. Введение в теорию нечетких множеств. -М: Радио и связь, 1982, 432с.

Сведения об авторе

Маляр Николай Николаевич – Декан математического факультета, заведующий кафедрой кибернетики и прикладной математики Ужгородского национального университета, кандидат технических наук, доцент, Украина, Ужгород, ул. Подгорная, 46; e-mail: cyber@mail.uzhgorod.ua

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ ОЦЕНКИ РИСКА БАНКРОТСТВА ПРЕДПРИЯТИЙ УКРАИНЫ

Юрий Зайченко, Светлана Рогоза, Владимир Столбунов

Аннотация: Рассмотрена проблема анализа риска банкротства предприятий. Изложен классический метод дискриминантного анализа риска банкротств, предложенный Е. Альтманом, проанализированы его достоинства и недостатки. Дается оценка возможности его применения в Украине. Далее излагается нечетко-множественный подход к оценке риска банкротства. Наконец, описан предложенный метод оценки риска банкротства предприятий на основе использования нечетких нейросетей с различными алгоритмами нечеткого вывода. Приводятся результаты сравнительного анализа различных методов в задаче оценки риска банкротства.

Ключевые слова: риск банкротства, метод Альтмана, нечетко-множественный метод, нечеткие нейронные сети, Мамдани, Цукамото, ANFIS

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

Conference topic: Knowledge- Dialogue-Solution"KDS-2008, Uzhgorod, Ukraine, 2008

Введение

Одной из актуальных проблем, связанных со стратегическим менеджментом и планированием является анализ финансового состояния и оценка риска банкротства предприятия (корпорации).

Своевременное выявление признаков возможного банкротства позволяет руководству принимать срочные меры по исправлению финансового состояния и снижению риска банкротства.

В настоящее время существует несколько общепризнанных методов и методик оценки риска банкротства. Наиболее известной и широко применяемой является методика профессора Альтмана [Altman, 1968].

Z- модель Альтмана представляет собой статистическую модель, которая на основе оценки показателей финансового состояния и платежеспособности компании позволяет оценить риска банкротства и разделить хозяйственные субъекты на потенциальных банкротов и небанкротов. Вместе с тем модель Альтмана имеет ряд недостатков, и ее применение для экономики Украины сопряжено с определенными трудностями. Поэтому в последние годы разрабатываются альтернативные подходы и методы, учитывающие специфику анализа и принятия решений в условиях неопределенности. К их числу относятся аппарат нечетких множеств и нечеткие нейронные сети.

Целью настоящей работы является анализ эффективности применения различных методов и подходов к оценке риска банкротства предприятий применительно к экономике Украины.

Модель Альтмана

Модель Альтмана построена с использованием аппарата мультипликативного дискриминантного анализа (МДА), который позволяет подобрать такие показатели, дисперсия которых между группами была бы максимальной, а внутри группы минимальной. В данном случае классификация проводилась по двум группам компаний одни из которых позднее обанкротились, а другие, наоборот, смогли выстоять и упрочить свое финансовое положение.

В результате МДА построена модель Альтмана (Z-счет), имеющая следующий вид [Altman, 1968]:

$$Z=1.2 K_1 +1.4 K_2 +3.3 K_3+0.6 K_4+2.0 K_5$$

где K_1 = собственный оборотный капитал/ сумма активов;

K_2 = перераспределенная прибыль/ сумма активов;

K_3 = прибыль до уплаты процентов/ сумма активов;

K_4 = рыночная стоимость собственного капитала/ стоимость заемного капитала;

K_5 = объем продаж/ сумма активов;

В результате подсчета Z – показателя для конкретного предприятия делается заключение :

если $Z < 1,81$ – очень высокая вероятность банкротства;

если $1,81 \leq Z \leq 2,7$ – высокая вероятность банкротства;

если $2,7 \leq Z \leq 2,99$ –возможно банкротство;

если $Z \geq 30$ – вероятность банкротства крайне мала.

Модель Альтмана дает достаточно точный прогноз вероятности банкротства с временным интервалом 1-2 года.

Примером применения модели Альтмана являются результаты исследований самим автором модели 86 компаний- банкротов в период 1969 – 1975 р.р., 110 компаний банкротов в период 1976 – 1995 р.р. и 120 обанкротились в период 1997 – 1999 годы. Используя пороговое значение 2,675, точность применения метода была в диапазоне от 82% до 96%. При повторном тесте модели, который базировался на одном финансовом периоде у, точность была в пределах 80-90%. [Altman, 1983].

Слабая сторона модели Альтмана состоит в том, что модель является чисто эмпирической, подогнанной по выборке, и не имеет под собой самостоятельной теоретической базы. Кроме того, приведенные коэффициенты должны определяться для различных отраслей промышленности и будут естественно, различаться.

В экономике Украины модель Альтмана пока не получила широкого применения по следующим причинам:

- 1) требуется вычисление соответствующих коэффициентов при показателях K_i , $i=1,5$, которые, естественно, отличаются от их значений для зарубежных стран;
- 2) информация о финансовом состоянии анализируемых предприятий, как правило, недостоверна, руководство ряда предприятий «сознательно» подправляет свои показатели в финансовых отчетах, что делает невозможным найти достоверные оценки коэффициентов в Z- модели.

Поэтому задача оценки вероятности риска банкротства должна решаться в условиях неопределенности, неполноты исходной информации, и для ее решения предлагается использовать адекватный аппарат принятия решений – нечеткие множества и нечеткие нейронные сети (ННС).

Нечетко-множественный метод оценки риска банкротства

Рассмотрим матричный метод прогнозирования банкротства корпораций, предложенный доктором эконом. наук О.А. Недосекиным [Недосекин].

1. Эксперт строит лингвистическую переменную со своим терм-множеством значений. Например, «Уровень менеджмента» может иметь следующее терм-множество значений «Очень низкий, Низкий, Средний, Высокий, Очень высокий».

2. Для того, чтобы конструктивно описать лингвистическую переменную, эксперт выбирает соответствующий количественный признак – например, сконструированный специальным образом показатель уровня менеджмента, который принимает значения от нуля до единицы.
3. Далее эксперт каждому значению лингвистической переменной которая по построению является нечетким подмножеством значений интервала $[0,1]$, ставит в соответствие функцию принадлежности того или иного нечеткого множества. Как правило, это трапецеидальная функция принадлежности. Верхнее основание трапеции соответствует полной уверенности эксперта в правильности классификации, а нижнее – уверенности в том, что никакие иные значения интервала $[0,1]$ не попадают в выбранное нечеткое множество. (см. рис.1.)

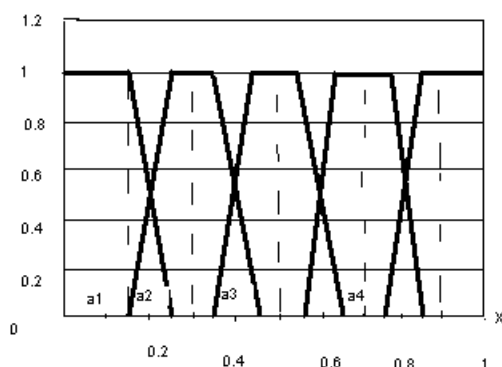


Рис.1. Трапецеидальная функция принадлежности

На этом описание лингвистических переменных заканчивается.

Нечетко-множественный метод, известный также, как матричный метод, состоит из следующих этапов:

Этап 1 (Лингвистические переменные и нечеткие множества)

а. Лингвистическая переменная **Е «Состояние предприятия»** имеет пять значений

E_1 – нечеткое подмножество состояний «предельного неблагополучия»;

E_2 – нечеткое подмножество состояний "неблагополучия";

E_3 – нечеткое подмножество состояний "среднего уровня";

E_4 – нечеткое подмножество состояний "относительного благополучия";

E_5 – нечеткое подмножество состояний "предельное благополучие".

б. Соответствующая переменной **Е** лингвистическая переменная **Г «Риск банкротства»** также имеет 5 значений:

G_1 – нечеткое подмножество состояний "предельный риск банкротства",

G_2 – нечеткое подмножество состояний "степень риска банкротства высокая",

G_3 – нечеткое подмножество состояний " степень риска банкротства средняя",

G_4 – нечеткое подмножество состояний « низкая степень риска банкротства»,

G_5 – нечеткое подмножество состояний "риск банкротства незначительный".

Носитель множества **Г** – показатель степени риска банкротства g – принимает значения от нуля до единицы по определению.

в. Для отдельного финансового показателя или показателя управления X_i задаем лингвистическую переменную V_i «уровень показателя X_i » на следующем терм-множестве значений:

V_{i1} - подмножество "очень низкий уровень показателя X_i ",

V_{i2} - подмножество "низкий уровень показателя X_i ",

V_{i3} - подмножество "средний уровень показателя X_i ",

V_{i4} - подмножество "высокий уровень показателя X_i ",

V_{i5} - подмножество "очень высокий уровень показателя X_i ".

Этап 2 (Показатели). Построим набор отдельных показателей $X=\{X_i\}$ общим числом N , которые по мнению эксперта, с одной стороны влияют на оценку риска банкротства предприятия, а с другой стороны, оценивают разные по природе стороны деловой и финансовой жизни предприятия. Например, в матричном методе используются такие показатели [Недосекин, Максимов]

- X_1 – коэффициент автономии (отношение собственного капитала к валюте баланса);
- X_2 – коэффициент обеспечения оборотных активов собственными средствами (отношение чистого оборотного капитала к оборотным активам);
- X_3 – коэффициент промежуточной ликвидности (отношение суммы денежных средств и дебиторской задолженности к краткосрочным пассивам);
- X_4 – коэффициент абсолютной ликвидности (отношение суммы денежных средств к краткосрочным пассивам);
- X_5 – оборотность всех активов за год (отношение выручки от реализации к средней выручке за период стоимости активов);
- X_6 – рентабельность всего капитала (отношение чистой прибыли к средней за период стоимости активов).

Этап 3 (Значимость показателей). Поставим в соответствие каждому показателю X_i уровень его значимости r_i . Для того, чтобы оценить этот уровень, необходимо поставить все показатели по порядку уменьшения их значимости так, чтобы выполнялось соотношение:

$$r_1 \geq r_2 \geq \dots r_N.$$

Если система показателей проранжирована в порядке уменьшения их значимости, то вес i -го показателя r_i необходимо определять по правилу Фишберна:

$$r_i = \frac{2(N-i+1)}{(N+1)N}. \quad (1)$$

Если все показатели имеют одинаковый вес, то $r_i = 1/N$.

Этап 4 (Классификация степени риска). Построим классификацию текущего значения g показателя степени риска как критерий разбиения этого множества на нечеткие подмножества (таблица 1):

Таблица 1. Классификация степени риска

Интервал значений g	Классификация уровня параметра	Степень оценочной уверенности (функция принадлежности)
$0 \leq g \leq 0.15$	G_5	1
$0.15 < g < 0.25$	G_5	$\mu_5 = 10 \times (0.25 - g)$
	G_4	$1 - \mu_5 = \mu_4$
$0.25 \leq g \leq 0.35$	G_4	1
$0.35 < g < 0.45$	G_4	$\mu_4 = 10 \times (0.45 - g)$
	G_3	$1 - \mu_4 = \mu_3$
$0.45 \leq g \leq 0.55$	G_3	1
$0.55 < g < 0.65$	G_3	$\mu_3 = 10 \times (0.65 - g)$
	G_2	$1 - \mu_3 = \mu_2$
$0.65 \leq g \leq 0.75$	G_2	1
$0.75 < g < 0.85$	G_2	$\mu_2 = 10 \times (0.85 - g)$
	G_1	$1 - \mu_2 = \mu_1$
$0.85 \leq g \leq 1.0$	G_1	1

Этап 5 (Классификация значений показателей). Построим классификацию текущих значений показателей X как критерий разбиения полного множества их значений на нечеткие подмножества вида B . Один из примеров такой классификации приведен ниже. В клетках таблицы стоят трапецеидальные нечеткие числа, которые характеризуют соответствующие функции принадлежности

Этап 6 (Оценка уровня показателей). Проведем оценку текущего уровня показателей и сведем полученные результаты в таблицу 2.

Таблица 2. Текущий уровень показателей

Показатель	Текущее значение
X_1	x_1
...	...
X_j	x_j
...	...
X_N	x_N

Этап 7 (Классификация уровня показателей). Проведем классификацию текущих значений x по критерию таблицы, построенной на этапе 5. Результатом проведенной классификации есть таблица значений λ_{ij} – уровней принадлежности носителя x_i нечетким подмножествам B_j .

Этап 8 (Оценка степени риска). Выполним вычислительные операции для оценки степени риска банкротства g :

$$g = \sum_{j=1}^5 g_j \sum_{i=1}^N r_i \lambda_{ij}, \quad (2)$$

где

$$g_j = 0.9 - 0.2 * (j-1), \quad (3)$$

Смысл применения формул (2) и (3) состоит в следующем. Сначала мы оцениваем вес того или иного подмножества из **B** в оценке состояния корпорации **E** и в оценке степени риска **G**. Эти веса далее принимают участие во внешней сумме для определения среднего значения показателя g , где g_j есть не что иное как средняя оценка g из соответствующего диапазона таблицы 1 этапа 4.

Этап 9 (Лингвистическое распознавание). Классифицируем полученное значение степени риска на базе данных таблицы 1. Результатом классификации являются лингвистическое описание степени риска банкротства и степени уверенности эксперта в правильности его классификации.

Основные достоинства нечетко-множественного матричного метода состоят в следующем:

- 1) возможность использования, кроме количественных, и качественных факторов;
- 2) учет неточной, приблизительной информации о значениях факторов.

Нечеткие нейронные сети для оценки риска банкротства

Преимущества систем с нечеткой логикой и ННС состоят в следующем [Недосекин], [Зайченко, 2004]:

- 1) описание условий и методы решения задач в терминах приближенных к естественному языку;
- 2) возможность оперирования с нечеткой информацией, в частности, с информацией заданной в виде интервальных оценок, использование знаний специалистов – экспертов, которые представляются в виде нечетких правил вывода «если...то».

Для анализа риска банкротства были применены нечеткие нейронные сети Цукамото, Мамдани и ANFIS [Зайченко, 2004]. Они использовались в качестве классификаторов. В качестве входных данных выступали разные наборы финансовых показателей предприятий, описанные ниже. Таким образом, число входных нейронов равняется количеству показателей в соответствующем наборе, а число выходных нейронов равняется количеству классов, т.е. имеем 2 исходных нейрона.

База правил состояла из нечетких правил вида «если-то» с гауссовскими функциями принадлежности. Выбор гауссовских функций принадлежности был обусловлен их дифференцируемостью в каждой точке, которая позволяет гарантировать сходимость процесса обучения (градиентного алгоритма). Были проведены эксперименты с разным количеством правил (20, 30, 40), параметры которых настраивались в процессе обучения. Первичная выборка разбивалась на обучающую и проверочную. Кроме того, для большей надежности процесс обучения осуществлялся на нескольких окнах данных.

Экспериментальные исследования

В данной работе для анализа эффективности прогнозирования методами дискриминантного анализа, анализа на основе нечетких множеств, предложенного Недосекиным, а также нечетких нейронных сетей были взяты финансовые показатели пятидесяти шести предприятий Украины, из которых 13 в 2006 году были признанные арбитражным судом банкротами. Другие предприятия выбирались случайно. Финансовые показатели взяты из годовых балансов и годовых отчетов о финансовых результатах за 2003, 2004 и 2005 года. Были проведены эксперименты по классификации предприятий относительно риска банкротства с использованием нечетких нейросетей с логическими выводами Мамдани, Цукамото и Сугено.

Сначала был проведен анализ риска банкротства на базе следующих показателей (Набор 1):

1. Отношение рабочего капитала к общей стоимости активов;
2. Отношение чистой прибыли к общей стоимости активов;
3. Отношение чистой выручки от реализации до общей стоимости активов;
4. Отношение уставного капитала к сумме задолженности;
5. Отношение дохода от реализации к общей стоимости активов.

Обучение осуществлялось в несколько этапов длиной в 100 эпох.

Итоговые результаты представлены на рис.2.

Далее набор показателей был изменен следующим образом (Набор 2):

1. Отношение чистого оборотного капитала к общей стоимости активов;
2. Отношение прибыли к собственному капиталу;
3. Коэффициент оборотных активов;
4. Норма прибыли.

Соответствующие результаты приведены на рис. 3.

Были проведены сравнительные эксперименты по оценки риска банкротства с использованием статистических методов- метода Альтмана, метода Спрингейта и метода Давыдовой-Беликова [Давыдова, Беликов,1999], нечетко-множественного метода Недосекина, а также нечетких нейросетей с выводом Мамдани, Цукамото и Сугено (ННС ANFIS), результаты которых приведены на рис. 4.

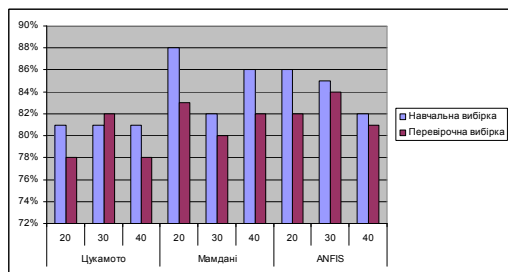


Рис. 2

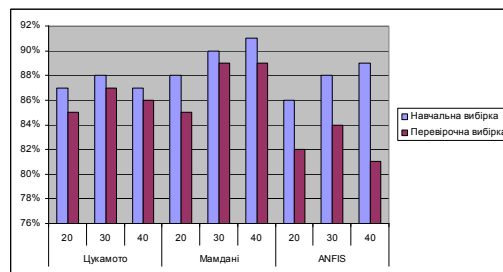
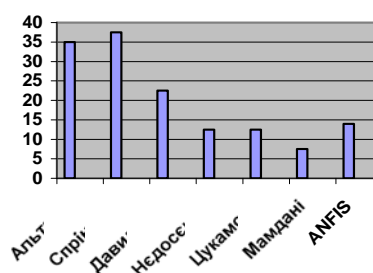
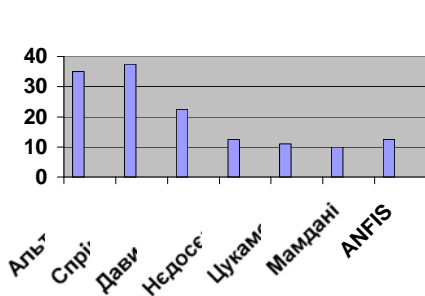


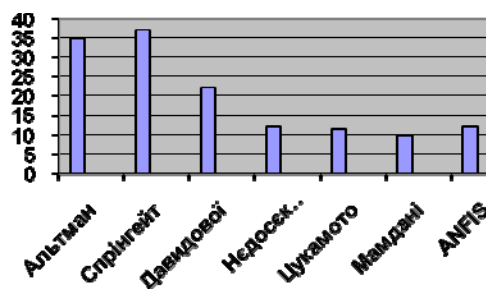
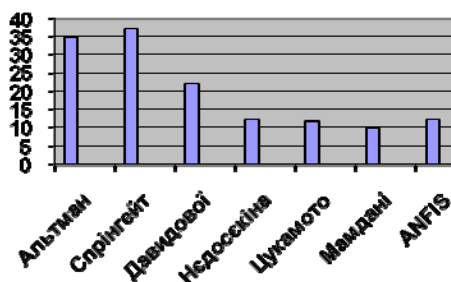
Рис. 3

набор 1

набор 2



а) 20 правил



б) 30 правил

Рис. 4

Анализ этих результатов позволяет сделать следующие **выводы**:

Наилучших результатов удалось достичь с помощью нечеткой нейронной сети Мамдани, точность классификации на которой при 40 нечетких правилах составила 97% на обучающей и 95% на тестовой выборках. На втором месте находятся результаты прогноза банкротства, полученные с использованием ННС Цукамото, и наконец, на третьем месте результаты, полученные матричным методом Недосекина.

Что же касается, статистических методов дискриминантного анализа, то они оказываются существенно хуже в сравнении с нечеткими методами, среди них наилучшим оказался метод Давыдовой-Беликова.

Выводы

1. Рассмотрены различные методы оценки риска банкротства предприятий: классический метод дискриминантного анализа Альтмана, нечетко-множественный метод Недосекина, нечеткие нейронные сети с различными алгоритмами вывода.
2. Проведены сравнительные экспериментальные исследования вышеуказанных методов применительно к анализу риска банкротства предприятий Украины.
3. Проведенные эксперименты показали, что наиболее предпочтительными для оценки риска банкротства для украинских предприятий оказались ННС с выводом Мамдани, за ними следует матричный метод проф. Недосекина, и, наконец, на последнем месте оказались статистические методы дискриминантного анализа.

Литература

- [Давыдова, Беликов, 1999] Давыдова Г.В., Беликов А.Ю. Методика количественной оценки риска банкротства предприятий//Управление риском, 1999, № 3, с. 13-20.
- [Зайченко, 2004] Зайченко Ю.П. Основы проектирования интеллектуальных систем. Навчальний посібник – К.: Видавничий Дім «Слово», 2004. – 352 с.
- [Недосекин, Максимов] Недосекин А.О. Максимов О.Б., Павлов Г.С. Анализ риска банкротства предприятия. Метод. указания по курсу «Антикризисное управление - На сайте http://sedok.narod.ru/sc_group.htm.
- [Недосекин] Недосекин А.О. Комплексная оценка риска банкротства корпорации на основе нечетких описаний.- На сайте: http://sedok.narod.ru/sc_group.htm.
- [Altman, 1983] Altman E.I. Corporate Financial Distress.- New York, John Wiley, 1983.
- [Altman 1968] Altman E.I. Financial Ratios, Discriminant Analysis and the prediction of Corporate Bankruptcy. // Journal of Finance, September, 1968, pp. 589-609.

Authors' Information

Зайченко Юрий Петрович, профессор, д.т.н., декан факультета «Институт прикладного системного анализа». НТУУ «КПИ», Киев, ул. Политехническая 14. тел: 38044-241-86-93, e-mail: zaych@i.com.ua

Рогоза Светлана, магистр «Институт прикладного системного анализа». НТУУ «КПИ», Киев, ул. Политехническая 14, e-mail: sweetlana_r@ukr.net

Столбунов Владимир, магистр «Институт прикладного системного анализа». НТУУ «КПИ», Киев, ул. Политехническая 14, atrabarba@gmail.com

ОПЕРАТОР МУТАЦИИ В ЭВОЛЮЦИОННОЙ ТЕХНОЛОГИИ РЕШЕНИЯ ЗАДАЧИ СОСТАВЛЕНИЯ РАСПИСАНИЙ

Елена Сипко

Аннотация: Статья посвящена исследованию используемого в генетических алгоритмах и задачах оптимизации оператора мутации. Рассмотрены проблемы, возникающие при составлении расписаний учебных занятий. В качестве примера использования случайной мутации приведен код фрагмента программы, применяемый при составлении расписания учебных занятий студентов Черкасского государственного технологического университета.

Ключевые слова: расписание, оператор мутации, популяция, кроссовер, инверсия.

ACM Classification Keywords I.2 Artificial Intelligence, H.4.1 Office Automation

Вступление

Генетические алгоритмы (ГА) ориентированы на применение операторов, которые изменяют генотип, или двоичное представление индивида. В зависимости от результатов работы операторов соответственно изменяется и фенотип. Одним из таких операторов является мутация. Мутация в ГА имеет полную аналогию в природе, когда происходит замена одного гена другим в ДНК под воздействием, например, радиоактивности [Воронин, 2004]. Считается, что мутация – это причина эволюции, и благодаря ей появляются новые виды. В ГА мутация способствует защите от преждевременной сходимости. Реализуется она путем выбора случайного гена или группы генов и их изменения по определенным правилам.

Оператор мутации моделирует естественный процесс мутации. Его применение в ГА обусловлено следующими рассуждениями. Начальная популяция, какой бы большой она не была, охватывает ограниченную область пространства поиска. Оператор скрещивания, безусловно, расширяет эту область, но все же до определенной степени, поскольку использует ограниченный набор значений, заданных начальной популяцией [Корнеев, 2003]. Внесение случайных изменений в генотип или фенотип индивида дает возможность отменить это ограничение и иногда значительно сократить время поиска и улучшить качество результата.

В бинарном случае оператор мутации заключается в инвертировании символов в случайно выбранных позициях. В случае работы с конечным алфавитом случайно выбранный символ заменяется на любой, отличный от него. Это может быть перестановка двух элементов в строке, замена элемента строки значением элемента из другой строки, в случае битовой строки может применяться инверсия одного из битов и др. Оператор мутации применяется не к каждому гену представителей в популяции. Обычно сначала задается вероятность мутации P_m и некоторый алгоритм осуществления мутации, например, такой [Кисляков, 2001]. Сначала нумеруются случайным образом все представители начальной популяции. Далее, начиная с первого гена первой хромосомы, пересматриваются элементы всей популяции, при этом выбираются случайные числа из интервала $[0, 1)$. Если на некотором шаге выбранное число оказывается меньше вероятности P_m , то текущий ген подвергается мутации.

Конкретное значение P_m зависит от решаемой задачи, но чаще всего вероятность мутации имеет достаточно малое значение. Так, в литературе встречаются оценки $P_m \approx 0,001$ в работе [De Jong, 1975], $P_m \in [0,005, 0,01]$ в работе [Schaffler, 1989], $P_m \approx 0,01$ в работе [Grefenstette, 1986].

Выбор метода осуществления мутации

Обычно при реализации ГА к скрещиваемым индивидам сначала применяют оператор скрещивания, а потом оператор мутации, хотя возможны и другие варианты. Существует мнение, что оператор мутации является основным поисковым оператором и известны алгоритмы, не использующие других операторов (кроссинговер, инверсия и т.д.), кроме мутации.

Исследования показали, что в простых задачах, используя ГА с мутацией (и без кроссовера), находят решение быстрее. Также для такого метода требуется меньший размер популяции. Если имеют место сложные многоэкстремальные функции, то лучше использовать ГА с кроссовером, поскольку этот метод более надежен, хотя и требует большего размера популяции.

Используя теорему шаблонов, можно утверждать [Исаев, 2003], что мутация только вредит росту количества представителей приемлемых шаблонов, поскольку лишней раз их разрушает. Однако мутация просто необходима для ГА с малым размером популяции. Сущность этого утверждения в том, что для малочисленных популяций свойственна преждевременная сходимость (premature convergence). Это та ситуация, когда в некоторых позициях все индивиды имеют один и тот же бит, но такой набор битов не соответствует глобальному экстремуму. При этом кроссовер практически не изменяет популяции, т.к. все индивиды почти одинаковы. В этом случае мутация способна инвертировать «застывший» бит у одного из индивидов и вновь расширить пространство поиска.

В качестве оператора мутации наибольшее распространение получили случайная и неравномерная мутация Михалевича [Michalewicz, 1995].

При случайной мутации ген, подлежащий изменению, принимает случайное значение из интервала своей области определения. В неравномерной мутации значение гена после оператора мутации рассчитывается по формуле:

$$c_i^* = \begin{cases} c_i + \delta(t, b_i - c_i) & \text{при } \chi = 0, \\ c_i - \delta(t, c_i - a_i) & \text{при } \chi = 1, \end{cases}$$

$$\delta(t, y) = y \left(1 - r \left(1 - \frac{t}{\varepsilon_{\max}} \right)^b \right),$$

где χ — целое случайное число, принимающее значение 0 или 1; $r \in [0, 1]$ — случайное вещественное число; ε_{\max} — максимальное количество эпох алгоритма; b — параметр, задаваемый исследователем.

Кроме того, если на протяжении достаточно большого числа поколений не происходит увеличения приспособленности, то применяются «малая» и «большая» мутации поколения. При «малой» мутации поколения ко всем особям, кроме 10% лучших, применяется оператор мутации. При «большой» мутации каждая особь либо мутирует, либо заменяется на случайно сгенерированную.

Число поколений до реализации «малой» и «большой» мутации во время работы алгоритма является постоянным. Также оператор мутации можно применять только в том случае, если к данной паре родительских особей не был применен оператор скрещивания. Можно выбирать некоторое количество точек в

хромосоме для инверсии, причем их число также может быть случайным. Допускается инвертирование сразу некоторой группы подряд идущих точек.

При увеличении вероятности мутации и при уменьшении влияния результатов отбора (например, за счет использования других стратегий отбора) размножение представителей приспособленных особей замедляется, но зато происходит интенсивный поиск других особей. Обратное, уменьшение вероятности мутации и увеличение влияния отбора ведет к интенсивному использованию найденных хороших особей, но тогда меньше внимания уделяется поиску новых [Holland, 1994].

Случайная мутация в задаче составления расписания учебных занятий

В задаче составления расписаний решением является оптимальный в некотором смысле вариант расписания. Оператор мутации предназначен для модификации одного из таких вариантов. Поскольку начальные расписания генерируются случайным образом, то такая генерация не обязательно будет обеспечивать поиск по всему пространству решений. Оператор мутации предназначен именно для того, чтобы внести в популяцию случайные решения, к которым трудно или вообще невозможно прийти, применяя только оператор скрещивания.

Поскольку в результате мутации генерируются новые решения и отсутствует внесение хаоса в популяцию, в исследовании используется вероятность возникновения мутации менее 0,01. Реализация мутации осуществляется с использованием нормального распределения: когда вероятность мутации оказывается менее 0,01 – в избранной позиции число изменяется на любое другое, но из определенного интервала [Бойко, 2006].

Реализация оператора мутации осуществляется посредством следующей процедуры:

procedure mutation;

begin

for n:=0 to a-1 do begin

for j:=0 to (5*5*t_g-1) do begin

for i:=0 to 1 do begin

if Random<0,01 then rez1[i,j,n]:=IntToStr(random(5)+1);

end;

if Random<0,01 then begin

temp1[2]:=IntToStr(random(t_g)+1);

if StrToInt(temp1[2])<10 then temp1[2]:='0'+temp1[2];

temp1[3]:=IntToStr(random(t_v)+1);

if StrToInt(temp1[3])<10 then temp1[3]:='0'+temp1[3];

temp1[4]:=IntToStr(random(t_d)+1);

if StrToInt(temp1[4])<10 then temp1[4]:='0'+temp1[4];

rez1[2,j,n]:=temp1[2];

rez1[3,j,n]:=temp1[3];

rez1[4,j,n]:=temp1[4];

end;

```

if Random<0,01 then begin //
  rez1[5,j,n]:=IntToStr(random(t_a)+1);
  if StrToInt(rez1[5,j,n])<10 then rez1[5,j,n]:='0'+rez1[5,j,n];
  if (t_a>=100) then
    if StrToInt(rez1[5,j,n])<100 then rez1[5,j,n]:='0'+rez1[5,j,n];
end; end;
for j:=0 to (5*5*t_g-1) do begin
  for i:=0 to 1 do begin
    if Random<0,01 then rez2[i,j,n]:=IntToStr(random(5)+1); end;
  if Random<0,01 then begin
    temp2[2]:=IntToStr(random(t_g)+1);
    if StrToInt(temp2[2])<10 then temp2[2]:='0'+temp2[2];
    temp2[3]:=IntToStr(random(t_v)+1);
    if StrToInt(temp2[3])<10 then temp2[3]:='0'+temp2[3];
    temp2[4]:=IntToStr(random(t_d)+1);
    if StrToInt(temp2[4])<10 then temp2[4]:='0'+temp2[4];
    rez2[2,j,n]:=temp2[2];
    rez2[3,j,n]:=temp2[3];
    rez2[4,j,n]:=temp2[4];
  end;
  if Random<0.01 then begin
    rez2[5,j,n]:=IntToStr(random(t_a)+1);
    if StrToInt(rez2[5,j,n])<10 then rez2[5,j,n]:='0'+rez2[5,j,n];
    if (t_a>=100) then
      if StrToInt(rez2[5,j,n])<100 then rez2[5,j,n]:='0'+rez2[5,j,n];
  end; end;
  for i:=0 to (5*5*t_g-1) do
    for j:=0 to 5 do begin
      new_hrom1:=new_hrom1+rez1[j,i,n];
      new_hrom2:=new_hrom2+rez2[j,i,n];
      y[j,i,2*n]:=rez1[j,i,n];
      y[j,i,2*n+1]:=rez2[j,i,n];
    end;
  new_pop:=new_pop+new_hrom1+new_hrom2;
end; end;

```

Предложенный алгоритм иллюстрирует случайную мутацию с вероятностью появления меньше 0,01 для создания расписания занятий в группе, причем предполагается, что оно составляется на пять дней и каждый день группа имеет пять учебных занятий.

Применение эволюционных алгоритмов при решении задачи составления расписаний сопряжено с необходимостью учета «проклятия размерности», поскольку среднестатистическое количество групп в среднестатистическом высшем учебном заведении составляет от 50 до 200. Заметим, что в таких алгоритмах выполняется значительное количество операций, не ведущих к нахождению оптимального решения, и их практическое тестирование и реализация наталкивается на проблемы, связанные со временем выполнения.

Сократить количество операций можно, используя некоторые дополнительные процедуры, оптимизирующие процесс поиска решения. Определенным образом они связаны с реализацией оператора мутации. Традиционно использовалась мутация, базирующаяся на разыгрывании случайной величины, имеющей равномерное распределение. Результаты проведенных экспериментов указывают на то, что «равномерный» выбор фрагмента хромосомы и «равномерная» его мутация не направлены на сокращение времени поиска оптимального решения, поскольку одинаковые шансы для модификации имеют как «перспективные», так и «неперспективные» решения. Причем мутация в первых может приводить к появлению «неперспективных» решений, а вторых – к появлению «перспективных», что указывает на случайный ненаправленный характер поиска оптимального решения.

Уменьшим количество «ложных» шагов алгоритма, используя место равномерного распределения нормального, что и предлагается в настоящей статье. Реализация такого подхода имеет следующие этапы. Вначале определяются участок потенциальных решений, который является одинаковым в каждом из них и которому соответствует максимальное или близкое к нему значение функции принадлежности. Напомним, что фенотип этого участка представляет собой действительное число, принадлежащее интервалу, который указывается в начале процесса кодирования элементов начальной выборки или начальной популяции решений [Бойко, 2006]. На следующем этапе разыгрывается случайное число, имеющее нормальное распределение со средним, соответствующем фенотипу участка, и среднеквадратическим отклонением, таким, чтобы интервал $(m - 3\sigma, m + 3\sigma)$, где m - фенотип соответствующего участка, σ - среднеквадратическое отклонение, совпадал с априорным интервалом изменения фенотипа. Этот интервал играет определенную роль при расчете величины мутации. Отметим также, что среднеквадратическое отклонение в процессе поиска решения изменяется, а именно, уменьшается, что указывает на то, что по мере приближения к оптимальному или квазиоптимальному решению вероятность значительных мутаций уменьшается.

Результаты экспериментов указывают на 30-50% сокращение времени поиска решения по сравнению с использованием равномерного распределения.

Заключение

Предложенная технология реализации оператора мутации при решении задачи составления расписаний направлена на сокращение времени поиска ее решения. Отметим, что выбор параметров реализации как алгоритма, так и операции мутации имеет значительное влияние на сходимость и скорость сходимости алгоритма, поэтому требует проведения дальнейших исследований и проведения экспериментов.

Библиография

- [Воронин, 2004] Воронин О., Дьюдни А. Дарвинизм в программировании // Мой компьютер. – 2004. – № 35 (310). – С. 3-12.
- [Корнеев, 2003] Корнеев В.В., Гареев А.Ф., Васютин С.В., Райх В.В. Базы данных. Интеллектуальная обработка информации. – М: изд-во „Нолидж”. – 2003 г. – 400 с.
- [Кисляков, 2001] Кисляков А.В. Генетические алгоритмы: операторы скрещивания и мутации // Информационные технологии. – 2001. – №1. – С. 29-34.
- [De Jong, 1975] De Jong, K.A. An analysis of the behavior of a class of genetic adaptive systems // Doctoral thesis. Dept Computer and Communication Sciences. – University of Michigan, Ann Arbor. – 1975.
- [Schaffler, 1989] Schaffler J.D., Caruana R.A., Escherman L.J., Das R. A study of control parameters affecting online performance of genetic algorithm for function optimisation // Proceedings of the Third International Conference on Genetic Algorithms and their Applications. – San Mateo, CA: Morgan Kaufmann Publishers. – 1989. – Pp. 51-60.
- [Grefenstette, 1986] Grefenstette J.J. Optimization of control parameters for genetic algorithms // IEEE Transaction on systems, man and cybernetics. SMC-16(1). – 1986. – Pp. 122-128.
- [Исаев, 2003] Исаев С.А. Популярно о генетических алгоритмах (<http://algolist.manual.ru/ai/ga/ga1.php>).
- [Michalewicz, 1995] Michalewicz Z. Genetic Algorithms, Numerical Optimization and Constraints, Proceedings of the 6th International Conference on Genetic Algorithms, Pittsburgh, July 15-19, 1995. - P. 151-158.
- [Holland, 1994] Holland J. H. Adaptation in natural and artificial systems. An introductory analysis with application to biology, control, and artificial intelligence. – London: Bradford book edition, 1994 – 211 p.
- [Бойко, 2006] Бойко О.М. Еволюційна технологія розв'язування задачі складання розкладів навчальних занять // Искусственный интеллект. – 2006. – №3. – С. 341-348.

Информация об авторе

Елена Сипко – аспирант, Черкасский государственный технологический университет, бул. Шевченко, 460, Черкассы, 18006, Украина; e-mail: boyko-e-n@mail.ru

ТЕХНОЛОГИЯ НЕЧЕТКОГО ПРОГНОЗИРОВАНИЯ ХАРАКТЕРИСТИК СЛОЖНЫХ ОБЪЕКТОВ И СИСТЕМ

Виталий Снитюк, Сергей Говорухин

Аннотация: В статье выполнена общая постановка задачи прогнозирования выходной характеристики сложных объектов по измеренным значениям входных факторов. Осуществлен анализ технологии решения поставленной задачи с использованием множественной линейной регрессии и методов кластеризации. Предложена альтернативная технология, использующая аппарат нечеткой логики, и выполнен сравнительный анализ результатов.

Ключевые слова: прогнозирование, регрессия, кластеризация, нечеткая логика.

ACM Classification Keywords I.2 Artificial Intelligence

Введение

В современном машиностроении важную роль играют задачи измерения различных характеристик, играющие важную роль при диагностике состояния колесных пар, контроле износа стенок трубопроводов, прогнозировании жесткости металлических пластин и т. д.

Разработка технологии повышения эффективности и качества измерений остается актуальной научной проблемой. На практике повышение качества измерений достигается использованием большого количества измерений. Сложность задачи точного измерения характеристик пластин обусловлена следующими причинами: большая площадь измерений, сравнительно небольшая площадь контакта сенсора и объекта на измеряемом участке, ошибки измерений, вносимые субъектами, которые их выполняют, ошибки, обусловленные случайными факторами.

Анализ релевантных источников

Для решения задач такого типа традиционно используются методы кластеризации и регрессии. Поскольку данные о сложных объектах и процессах имеют многоплановый и разнотипный характер, то это значительно усложняет, а часто делает невозможным выявление и построение их адекватных моделей. С целью преодоления таких трудностей на первом этапе необходимо выполнить идентификацию топологической структуры таких объектов. Для этого применяют методы кластеризации, основными недостатками которых являются [Gordon, 1999]: директивное априорное предположение о количестве кластеров, сложность выбора оптимальной метрики для определения близости объектов и кластеров, необходимость большого количества парных сравнений объектов, недостаточное статистическое обоснование результатов кластеризации и соответственно значительное количество возможных вариантов разбиений на кластеры.

На втором этапе для решения задач идентификации зависимостей традиционно используют методы линейной множественной регрессии (ЛМР). Использование данного метода базируется на ряде допущений и ограничений, главным из которых является предположение о наличии линейной зависимости в многомерной структуре данных между выходной характеристикой и входными факторами. Однако такое предположение не всегда является правоммерным.

Уменьшить влияние указанных недостатков традиционных методов при разработке технологии прогнозирования характеристик сложных объектов предлагается использовать методы нечеткой логики.

Постановка задачи

Рассмотрим постановку задачи в общем виде.

Пусть Y – результирующая характеристика, X_1, \dots, X_n – входные факторы. С помощью оригинального сенсора проведены измерения n параметров m объектов. Для каждого объекта осуществлено k измерений. Таким образом, получим обучающую выборку данных, представленную матрицей M , которая имеет $n + 1$ столбцов и $m \cdot k$ строк (табл. 1).

Таблица 1

№ объекта	№ измерения	Значения входных факторов			Выходная характеристика
		X_1	...	X_n	Y
1	1	x_{11}^1	...	x_{1n}^1	y_1^1

	k	x_{k1}^1	...	x_{kn}^1	y_k^1
...
m	1	x_{11}^m	...	x_{1n}^m	y_1^m

	k	x_{k1}^m	...	x_{kn}^m	y_k^m

Предположим, что информация, представленная в табл. 1, утеряна. Причиной этого могла бы стать также поломка или утеря сенсора. Тогда возникает необходимость использования другого сенсора, что приводит к осуществлению значительного количества измерений, причем значения результирующей характеристики оказываются смещенными.

Сделаем упрощающие предположения. Пусть при использовании других сенсоров получены малые выборки данных при небольшом количестве измерений для объектов. Априорно считаем, что значения Y утрачены, за исключением:

1. l объектов, $l \ll m$, причем для каждого объекта выполнено по одному измерению.
2. r объектов, $r \ll \ll m$, для которых выполнено по несколько измерений p_r .

Тогда $l \approx p_r \cdot r$. Причем и в первом, и во втором случае значения Y известны.

Необходимо по результатам измерений восстановить значения Y для всех новых объектов.

Технология решения задачи с использованием МЛР

Рассмотрим следующий алгоритм построения прогнозирующей модели МЛР при известных допущениях.

- Шаг 1. Рассчитать матрицу парных корреляций входных факторов: $M = cor(X_i, X_j)$.
- Шаг 2. Выполнить процедуру удаления зависимых факторов по одному из алгоритмов:

1. Исключить те факторы, для которых $|cor(X_i)| > K > 0$, где K - пороговое значение.
 2. Исключить те факторы, для которых $\sum_i |cor(X_i, X_j)| = \max_j \sum_i |cor(X_i, X_j)|$, $i \neq j$, а также те, для которых $\sum_i |cor(X_i, X_j)| > L$, где L – некоторая константа.
- Шаг 3. Вычислить вектор корреляций выходной характеристики с каждым из оставшихся входных факторов: $cor(Y, X_i)$.
- Шаг 4. Выбрать p факторов, для которых $|cor(Y, X_i)| > l_y$, где l_y – положительная константа, причем $p = 7 \pm 2$.
- Шаг 5. Выполнить проверку оставшихся факторов на мультиколлинеарность.
- Шаг 6. По данным обучающей выборки для выбранных p факторов построить уравнение множественной линейной регрессии: $Y_1 = F(X_1, X_2, \dots, X_p)$.
- Далее используем данные, полученные при использовании другого сенсора.
- Шаг 7. Разыграть l равномерно распределенных случайных чисел.
- Шаг 8. Выбрать соответствующие строки из таблицы.
- Шаг 9. По МНК построить второе уравнение: $Y_2 = F_2(X_1, X_2, \dots, X_p)$.
- Шаг 10. Построить графики функций Y_1 и Y_2 , проходящие через l выбранных точек (пример приведен на рис. 1).
- Шаг 11. Далее определить значения ошибок регрессии Y_2 одним из двух способов:
1. Рассчитать значение ошибки как среднее ошибок во всех точках: $d = \frac{1}{l} \sum_{i=1}^l d_i$, в этом случае значение ошибки будет одним и тем же во всех точках.
 2. Запомнить вектор значений ошибок во всех точках: $D = (d_1, d_2, \dots, d_l)$, в этом случае каждой точке соответствует единственное значение ошибки.
- В обоих случаях $d_i = |Y_1(O_i) - Y_2(O_i)|$, $i = \overline{1, l}$.
- Шаг 12. Для нового объекта выполнить измерения входных факторов (получаем новое измерение O') и подставить их значения в Y_2 . Тогда в зависимости от выбранного способа на шаге 11 соответственно рассчитываем значения Y_1 :
1. $Y_1 = Y_2 + d$;
 2. Во второй выборке найти то измерение, которое наиболее «близко» к данному в пространстве выбранных признаков. Для этого использовать один из известных методов кластерного анализа [Мандель, 1988], [Айвазян, 1989], [Jain, 1988]:
 - а) выбрать метод расчета расстояния между измерениями – меру близости объектов: $q(O_i, O_j)$;

- b) в малой обучающей выборке определить измерение O_i , расстояние до которого от данного измерения O' минимально: $q(O', O_i) = \min_j q(O', O_j)$, $i = \overline{1, l}$, $j = \overline{1, l}$.
- c) тогда $Y_1 = Y_2 + d_i$.

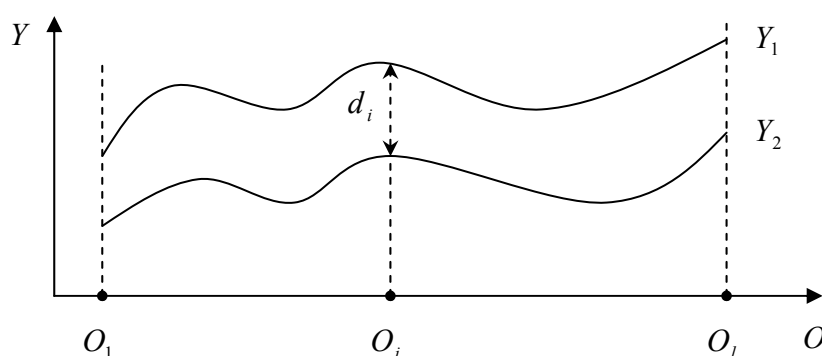


Рис. 1. Графики ЛМР, построенные по данным первой выборки (Y_1) и по фрагменту другой выборки (Y_2)

Нечеткая идентификация

Используем для решения поставленной задачи методы теории нечетких множеств.

Пусть A_{ij} , $i = \overline{1, m}$, $j = \overline{1, n}$, - нечеткие термы (лингвистические переменные (ЛП)), которые строятся на множестве значений фактора X_j объекта i , то есть на множестве значений $x_{1j}^i, \dots, x_{lj}^i$, $l = \overline{1, k}$, которое является универсальным множеством данной ЛП. Пусть $T_{ij} = (T_1, T_2, \dots, T_s)$ - терм-множество ЛП A_{ij} , каждому элементу которого соответствует ФП $\mu_{T_i}(X_j)$, $i = \overline{1, s}$ (рис. 2).

В качестве функций принадлежности для элементов терм-множества предлагается использовать такие: $\mu_{T_1}(X)$ и $\mu_{T_s}(X)$ - линейные, $\mu_{T_i}(X)$, $i = \overline{2, s-1}$ - треугольного вида с параметрами (a, c) .

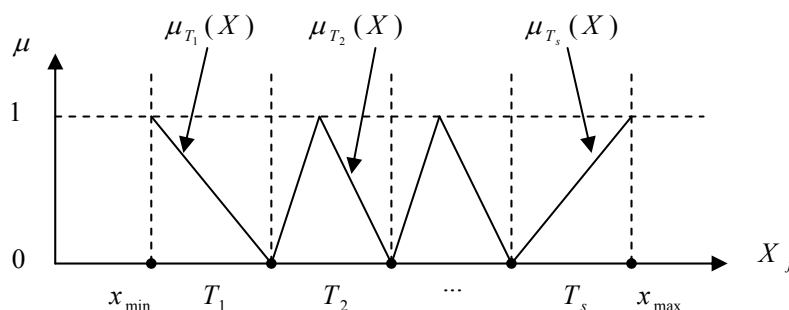


Рис. 2. ФП терм множества ЛП

Формируем нечеткую базу знаний (НБЗ). Пусть модель прогнозирования выходной характеристики задается набором правил R_i , $i = \overline{1, m}$:

Если $X_1 \in A_{i1}, X_2 \in A_{i2}, X_n \in A_{in}$ то $Y_i = C_i$,

где $A_{ij}, i = \overline{1, m}, j = \overline{1, n}$ – ЛП, $C_i = const$.

Так как в данной НБЗ antecedentes заданы нечеткими множествами, а консеквенты заданы константами, то нечеткий логический вывод (НЛВ) производится по синглтонной базе знаний, которая является частным случаем НЛВ Сугено [Штовба, 2007].

В соответствии с правилами НЛВ Сугено получаем следующий алгоритм прогнозирования выходной характеристики Y при заданном входном объекте $X' = (X'_1, X'_2, \dots, X'_n)$:

Шаг 1. Рассчитать степени выполнения посылки i -го правила для входного вектора X' :

$$\mu_i(X') = \min\{\mu_1(X'_1), \mu_2(X'_2), \dots, \mu_n(X'_n)\}, i = \overline{1, m}.$$

В данном случае использована операция минимума (t -норма).

Шаг 2. В результате выполнения предыдущего шага для всей НБЗ получаем нечеткое множество, которое соответствует входному вектору X' :

$$\tilde{y} = \left(\frac{\mu_1(X')}{C_1}, \frac{\mu_2(X')}{C_2}, \dots, \frac{\mu_m(X')}{C_m} \right).$$

Шаг 3. Выполнить дефаззификацию $\tilde{y} \rightarrow y$ одним из способов:

$$1. \text{ Найти взвешенное среднее: } y = \frac{\sum_{i=1}^m \mu_i(X') \cdot C_i}{\sum_{i=1}^m \mu_i(X')} ;$$

$$2. \text{ Найти взвешенную сумму: } y = \sum_{i=1}^m \mu_i(X') \cdot C_i .$$

Далее необходимо решить задачу построения модели для прогнозирования характеристик объектов при малой обучающей выборке, полученной с помощью нового сенсора. Один из подходов заключается в использовании НБЗ, в которой заключения заданы в виде линейных целевых функций. Основной недостаток данного подхода заключается в допущении о линейной зависимости выходных характеристик.

Другой подход базируется на идее использования методологии кластерного анализа. В [Аверкин, 1986] предлагается метод варьирования прототипов, который относится к косвенным методам построения ФП для одного эксперта. Пусть P - прототип, который характеризуется параметрами x_1, \dots, x_n . Введем меру расстояния между объектом и прототипом: $Q(P, X) = \|P - X\|$. Для оптимизации вычисления расстояния от объекта до разных прототипов вводится штрафная функция $d(P)$. Пусть имеем m прототипов $P_1 \dots P_m$. Тогда для каждого объекта вычисляется функция расстояния до ближайшего прототипа: $sim(O_i) = \min_j (Q(O_i, P_j) + d(P_j)), j = \overline{1, m}$.

Значения ФП объекта O прототипу P вычисляются по формуле:

$$\mu_p(O_i) = 1 - \frac{sim(O_i)}{\max_j sim(O_j)}, j = \overline{1, m}.$$

Таким образом, для каждого прототипа получим ФП объектов прототипу. Для вычисления выходной характеристики для каждого нового объекта $O' = (x_1, x_2, \dots, x_n)$ находим взвешенное среднее:

$$Y' = \frac{\sum_{i=1}^m \mu_{P_i}(O') \cdot Y_i}{\sum_{i=1}^m \mu_{P_i}(O')}.$$

Заключение

Предложенная технология решения поставленной задачи, которая базируется на использовании статистических методов, множественной линейной регрессии и кластерного анализа, является классической для решения задач такого типа. Идея использования аппарата нечеткой логики является перспективной [Асаи, 1993] и позволит повысить точность идентификации и прогнозирования жесткости пластин. В этом направлении планируется проводить дальнейшие теоретические и практические исследования.

Один из основных недостатков практического применения полученной нечеткой модели заключается в необходимости построения большой НБЗ. Например, в задаче прогнозирования жесткости металлических пластин $n = 73$, то есть одно правило состоит из 73 предусловий, а всего обучающая НБЗ имеет 98 правил. Решение задачи при таких условиях требует больших вычислительных затрат. Поэтому актуальной становится задача снижения размерности входных факторов при построении нечеткой модели, одним из возможных решений которой является использование методов нахождения главных компонент.

Перспективными для решения поставленной задачи являются также использование методов искусственного интеллекта, таких как нейронные сети и генетические алгоритмы.

Библиография

- [Gordon, 1999] A.D. Gordon. Classification. – Boca Raton, 2nd ed., CRC Press LLC, 1999.
- [Наконечный, 1998] С.И. Наконечный, Т.О. Терещенко, Т.П. Романюк. Эконометрия – К.: КНЕУ, 1998.
- [Мандель, 1988] И.Д. Мандель. Кластерный анализ. – М.: Финансы и статистика, 1988.
- [Айвазян, 1989] Прикладная статистика: Классификация и снижение размерности / С.А. Айвазян, В.М. Бухштабер, И.С. Енюков, Л.Д. Мешалкин. – М.: Финансы и статистика, 1989.
- [Штовба, 2007] С.Д. Штовба. Проектирование нечетких систем средствами MATLAB. – М.: Телеком, 2007.
- [Аверкин, 1986] А.Н. Аверкин и др. Нечеткие множества в моделях управления и искусственного интеллекта / Под ред. Д.А. Поспелова. - М.: Наука, 1986.
- [Асаи, 1993] Прикладные нечеткие системы / Под ред. Т. Тэрано, К. Асаи, М. Сугэно. – М.: Мир, 1993.

Информация об авторах

Снитюк Виталий – Черкасский государственный технологический университет, зав. кафедрой информационных технологий проектирования; бул. Шевченко, 460, Черкассы, Украина;
e-mail: snytyuk@gmail.com

Говорухин Сергей – Черкасский государственный технологический университет, аспирант факультета информационных технологий и систем; бул. Шевченко, 460, Черкассы, Украина;
e-mail: govorukhin@gmail.com

ИНТЕГРИРОВАННЫЙ МОДЕЛИРУЮЩИЙ КОМПЛЕКС ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ В КОМПЛЕКСНОЙ КОРПОРАТИВНОЙ СИСТЕМЕ УПРАВЛЕНИЯ МАГИСТРАЛЬНОЙ ГАЗОТРАНСПОРТНОЙ СЕТЬЮ

Виктор Борисенко, Юрий Пономарев, Татьяна Борисенко

Abstract: *The approach to construction, structuring class of tasks, technology of the use and stages of modernization systems support of decision-making in composition the operative-dispatch control computer-based system by the gas-transport system of Ukraine is considered.*

Keywords: *корпоративная автоматизированная система управления, магистральная газотранспортная сеть, распределенная многоуровневая система оперативно-диспетчерского управления, система поддержки принятия решений, интегрированный комплекс моделирования и оптимизации сетей магистральных газопроводов.*

ACM Classification Keywords: *H.4.2 Types of Systems*

Введение

В настоящее время разработана и одобрена Правительством Украины "Энергетическая стратегия Украины до 2030 года", которая определяет цели и основные направления развития газовой промышленности в среднесрочной и долгосрочной перспективе.

Основными целями энергетической стратегии Украины в газовой отрасли являются:

- Обеспечение надежного, экономически эффективного и экологически безопасного транспорта газа потребителям стран Центральной и Западной Европы.
- Обеспечение стабильного, бесперебойного и экономически эффективного удовлетворения спроса на природный газ на внутреннем рынке Украины.
- Увеличение объемов собственной добычи природного газа и уменьшение зависимости внутреннего рынка Украины от внешних поставок этого энергоресурса.
- Интеграция газотранспортной системы (ГТС) Украины в Европейскую газотранспортную систему.

Газотранспортная отрасль является одной из важнейших составляющих топливно-энергетического комплекса Украины. Ее влияние на другие отрасли экономики и на энергетическую безопасность Украины является широкомасштабным, глобальным и во многом определяющим фактором эффективности экономики страны в целом.

Магистральная газотранспортная сеть Украины представляет собой единый технологический комплекс и выполняет две основные функции:

- Транспорт, хранение и распределение природного газа потребителям Украины.
- Транзит природного газа по территории Украины в страны Центральной и Западной Европы.

Главными технологическими элементами ГТС Украины являются шесть многониточных магистральных газопроводов (МГ), суммарная длина которых составляет 37,2 тыс. км, 72 компрессорные станции (КС), 112 компрессорных цеха (КЦ), 786 газоперекачивающих агрегата (ГПА), более 1420 газораспределительных станций (ГРС) и 225 газовых месторождений.

Значительная пространственная распределенность, высокая размерность и техническая сложность, существенная экологическая опасность, присущая объектам магистральной газотранспортной сети приводят к тому, что главным стратегическим направлением повышения качества, надежности и эффективности управление процессами транспорта, хранения и распределения природного газа в настоящее время является создание комплексной корпоративной автоматизированной системы управления ГТС.

В качестве единой концептуальной и организационно-методологической базы поэтапного построения и внедрения КАСУ ГТС специалистами отраслевого Научно-исследовательского и проектного института транспорта газа (ИТГ) были разработаны и утверждены в 2006-2007 годах на научно-технических советах и правлении национальной акционерной компании (НАК) "Нефтегаз Украины" комплекс базовых нормативно-методических документов (НМД) [1-5].

Комплексное решение проблемы создания КАСУ приводит к необходимости существенной модернизации ГТС Украины в трех основных направлениях.

- 1) Целенаправленная и последовательная реконструкция ГТС с заменой устаревшего технологического оборудования и внедрением современных информационных технологий, систем автоматического управления (САУ) и АСУ технологическими процессами (АСУ ТП).
- 2) Внедрение комплексной автоматизированной системы управления бизнес процессами класса ERP (Enterprise Resource Planning).
- 3) Модернизация автоматизированной системы диспетчерского управления (АСДУ) ГТС Украины .

Одно из центральных мест в КАСУ ГТС занимает автоматизированная система оперативно-диспетчерского управления (АСОДУ), имеющая разветвленную 3-х уровневую архитектуру. Она должна получать данные реального времени из АСУ ТП технологическими объектами ГТС и передавать информацию систему управления бизнес-процессами класса ERP.

Таким образом, в настоящее время весьма важной и актуальной проблемой является поэтапная модернизация разработанной и внедренной в 80- 90-х годах автоматизированной системы оперативно-диспетчерского управления (АСОДУ) ГТС Украины.

В данной работе представлен подход к решению одной из базовых задач модернизации АСОДУ на основе разработки и внедрения комплексной системы поддержки принятия решений, ориентированной на класс задач оперативно-диспетчерского управления режимами работы ГТС и базирующейся на едином интегрированном комплексе моделирования и оптимизации (ИКМО) магистральных газотранспортных сетей.

Модернизации АСДУ на основе создания системы поддержки принятия решений

Основными особенностями ГТС Украины с точки зрения построения автоматизированной системы оперативно-диспетчерского управления являются :

- Большое количество межсистемных перетоков газа, когда газ перетекает из одного магистрального газопровода в другой.
- Наличие магистральных газопроводов с реверсными газовыми потоками.
- Цикличность изменения режимов транспортировки газа, вызываемых:
 - циклическим характером изменения температуры окружающей среды, т.к. изменение температуры окружающего воздуха на один градус вызывает скачек в потреблении газа на 30 млн. м³;

- цикличностью изменений планов поставки газа потребителям (сезонные колебания до 30 %)
- Непрерывность процессов функционирования ГТС.
- Инертность процессов транспортировки газа.
- Сильно связанное технологическое взаимодействие всех основных элементов системы (транспорт, хранение, распределение, экспорт и импорт) на всем протяжении ГТС, которые практически нельзя разделять и рассматривать локально.
- Большие текущие запасы газа в магистральных газопроводах определяют аккумулирующую способность газопроводов, что необходимо учитывать при прогнозировании режимов работы ГТС.

Одним из перспективных путей повышения качества и эффективности функционирования АСДУ является создание системы поддержки принятия решений (СППР).

В рамках задачи СППР основными функциональными комплексами задач оперативного управления ГТС являются:

- Поддержка принятия решений в аварийных и нестандартных ситуациях.
- Поддержка принятия решений при обосновании расчета режимов транспортировки газа для различных технологических объектов ГТС..

Необходимость создания и внедрения СППР определяется высокой организационно-технологической сложностью объектов управления, а также практической потребностью в обеспечении постоянной информационной поддержки и обучения недостаточно опытного диспетчерского персонала. Создание СППР ориентировано в первую очередь на снижение затрат времени на подготовку принятия решений и существенное снижение рисков принятия ошибочных решений оперативно - диспетчерским персоналом ГТС.

Практическое использование СППР в управлении ГТС направлено на оказание компьютеризированной поддержки работы специалистов диспетчерских служб:

- При многофакторном анализе текущих режимов транспортировки газа.
- Для идентификации разрывов магистральных трубопроводов и выдаче рекомендаций по проведению мероприятий по локализации этих разрывов.
- Для выдачи рекомендаций по поддержке живучести, т.е. максимально возможному сохранению работоспособности как отдельных участков магистральных газопроводов, так и ГТС в целом.

Принципиальным отличием СППР в составе модернизированной АСДУ является активное использование комплексов моделирования и оптимизации режимов работы ГТС, создание интеллектуальных процедур поддержки принятия решений по оперативно-диспетчерскому управлению, а также разработка, реализация и внедрение функций современных систем исполнительного управления производством класса MES (Manufacturing Execution System).

Информационное обеспечение СППР состоит из:

- Базы фактографических данных.
- Базы знаний.

База фактографических данных состоит из интегрированной распределенной базы данных (ИРБД) комплексной системы управления бизнес-процессами класса ERP и реляционной базы оперативных данных системы класса MES, которые обмениваются между собой информацией при помощи инструментальных средств промышленной интеграционной платформы.

База знаний о технологических ситуациях ГТС представляет собой гетерогенную семантическую сеть, объединяющую основные информационные сущности (классы и соответствующие им объекты), их свойства и типовые отношения между ними, а также сеть продукционных правил.

С формальной точки зрения, ГТС относится к классу целенаправленных, многомерных, многосвязных нелинейных стохастических систем с распределенными параметрами, для которых характерны сетевая многоуровневая структура, значительная пространственная распределенность, наличие ЛПР в контуре управления, наличие непрерывных и дискретных управляющих воздействий, высокий уровень неопределенности целей, структуры, параметров и состояний, а также воздействий со стороны окружающей среды.

До настоящего времени не нашли достаточно полного, комплексного, системного решения проблемы единого формализованного описания, поддержки динамического компьютерного моделирования и оптимизации режимов работы ГТС. из-за сложности адекватного описания реакции сложной системы (модели ГТС) на множество детерминированных и стохастических внутренних и внешних возмущений, а также из-за сложности описания самих этих возмущающих воздействий.

Структуризация класса задач оперативно-диспетчерского управления

Задача оперативного планирования решается в условно реальном масштабе времени средствами АСОДУ на базе филиала "Оперативно-диспетчерское управление" (ОДУ) компании "Укртрансгаз" с детализацией по всем нижестоящим подчиненным газотранспортным предприятиям.

Решение задачи стабилизации режима осуществляется в реальном времени средствами локальной автоматики (САУ и АСУ ТП технологических объектов и установок ГТС).

Процесс оперативного управления ГТС является циклическим и включает ряд основных этапов.

1. Этап оперативного контроля режима работы ГТС и состояния оборудования:

- сбор информации о текущем режиме работы ГТС;
- ведение архива режимных данных и архива событий;
- -автоматизированное формирование отчетной документации
- визуализация режима на технологической схеме ГТС с учетом ее иерархической структуры;
- восстановление потокораспределения в ГТС;
- расчет технико-экономических показателей хода ТП (запас газа, выполнение поставок газа, затраты энергоресурсов, ТТР);
- диагностика технического состояния ТО КС и трубопроводов.

2. Оперативный анализ фактического режима ГТС:

- анализ архивных данных о режимах и событиях;
- расчет фактического изменения запаса газа;
- сведение фактического баланса транспорта и распределения газа;
- расчет запаса ресурсов управления ГТС;
- выбор системы критериев эффективности и надежности режима работы ГТС;
- многокритериальный анализ фактического режима работы ГТС.

3. Оперативное планирование режима работы ГТС:

- прогнозирование внешних возмущений (изменения объемов подачи и потребления природного газа всеми категориями внешних и внутренних потребителей в зависимости от влияния трех основных групп факторов: хронологических, метеорологических и организационных);
- прогнозирование внутренних возмущений (ремонт, изменение технического состояния);
- прогнозирование возникновения и развития аварийных ситуаций;
- формирование согласованных по всем уровням целей управления и ограничений на режимы;
- многокритериальная оптимизация режима ГТС;

4. Оперативное управление режимом работы ГТС:

- оценивание возможности, риска и целесообразности перевода ГТС с фактического режима на оптимальный;
- моделирование процесса смены режима работы ГТС по заданному перечню команд управления;
- многокритериальный поиск оптимального перечня команд управления для смены режима работы ГТС;
- передача управляющих команд подсистемам нижнего уровня управления;
- контроль выполнения перехода на плановый режим работы ГТС;
- оперативная коррекция планового режима при изменении технологических условий.

Результаты анализа текущего состояния ГТС, ежедневных заявок стран-импортеров природного газа и внутренних потребителей, а также зависимости объемов потребления природного газа от температуры окружающей среды по каждому из регионов Украины и стран-импортеров, позволяют сформулировать целевую функцию и граничные условия для постановки и решения задачи оперативного планирования режима работы ГТС Украины на следующие сутки.

После нахождения оптимального планового режима необходимо принять решение о целесообразности и возможности перевода ГТС на этот режим, а также сформировать соответствующий диспетчерский график и перечень команд управления по переходу на новый режим.

В настоящее время в рамках АСДУ ГТС Украины используется множество программных комплексов, которые в большей или меньшей степени решают основные задачи, сбора, анализа, документирования информации, а также моделирования режимов работы ГТС.

Однако независимая разработка и эксплуатация множества различных программно-технических средств привели к неоправданному усложнению внутренней архитектуры АСДУ.

Нерациональное дублирование функциональных возможностей и отсутствие прямых связей между комплексами привело к снижению эффективности и надежности комплексной корпоративной системы управления ГТС.

Решение этой проблемы состояло в разработке единых технических требований к каждой компоненте создаваемой СППР и стандартизованных способов их информационного и технического взаимодействия.

Создание единого информационного пространства СППР ГТС Украины привело к использованию интегрированной системы распределенных баз данных и баз знаний, базирующаяся на единой промышленной интеграционной платформе. В настоящее время разработана и внедрена в компании "Укртрансгаз" подсистема сбора, хранения и передачи оперативной и нормативно-справочной информации, которая охватывает все уровни управления и настраивается на существующие технические каналы связи.

Подсистема ведения архивов оперативных данных и формирования отчетной документации реализована в виде программного комплекса «Журнал диспетчера», который имеет единый интерфейс для всех уровней управления ГТС и позволяет представлять данные в табличном и графическом виде.

Полнофункциональный графический редактор технологических схем ГТС позволяет не только вводить и отображать нормативно-справочную информацию о структуре ГТС и параметрах её технологического оборудования, но и в наглядном виде представлять оперативные и расчетные значения параметров газовых потоков и режимов работы объектов ГТС.

Интегрированный комплекс моделирования и оптимизации режимов работы ГТС

Принципиальным недостатком всех используемых на Украине и в странах СНГ комплексов моделирования и оптимизации (КМО) ГТС является применение детерминированных моделей и методов, которые не учитывают неполноту и недостоверность априорной информации о структуре и параметрах ГТС, а также неопределенность реальных условий функционирования системы. Именно это определяет недопустимо низкую степень технологической устойчивости получаемых решений, т.е. любые внешние и внутренние, даже не очень значительные, возмущения приводят к тому, что оптимальные решения, полученные с использованием этих комплексов, становятся недопустимыми.

Центральным ядром СППР ГТС Украины является интегрированный комплекс моделирования и оптимизации (ИКМО) квазистационарных и существенно нестационарных неизотермических режимов транспорта и распределения природного газа в ГТС []. Моделирование и оптимизация режимов работы ГТС возможны как для всей ГТС, отдельных предприятий (управлений магистральных газопроводов – УМГ, линейных производственных управлений магистральных газопроводов – ЛПУ МГ), так и для любого выделенного фрагмента системы. Расчет осуществляется для любого набора граничных условий и при разных способах задания компонентного состава природного газа (12 компонент, 3 компоненты или только плотность газа).

Особенностью СППР на базе ИКМО является то, что результаты решения всех функциональных задач, основанных на использовании математических моделей объектов и процессов, сопровождаются расчетами их статистических свойств в зависимости от статистических свойств характеристик исходных данных [6-9].

Анализ фактических режимов работы ГТС и оптимизация плановых режимов осуществляется по двум группам критериев.

1. Критерии качества функционирования ГТС:

- степень выполнения контрактных условий по объемам и физико-химическим параметрам поставляемого газа;
- средний и суммарный дефицит газа в системе.

2. Критерии эффективности функционирования ГТС:

- суммарные затраты на транспорт и распределение природного газа в ГТС в энергетическом или стоимостном выражении;
- эквивалентная или обобщенная товарно-транспортная работа ГТС;
- непродуктивный износ оборудования, обусловленный «лишними» переключениями и избыточными нагрузками;
- величина разбаланса газа;

- степень технологической устойчивости режима работы ГТС при заданном уровне неопределенности процессов подачи и потребления газа;
- объем запаса газа в трубопроводах ГТС.

При стохастическом подходе в качестве оценок этих критериев используются математические ожидания или вероятности соответствующих величин.

Для повышения эффективности метода динамического баланса и оценивания непроизводительных потерь природного газа используются три методики определения фактического запаса газа в ГТС:

- стандартизированная методика ОДУ ДК «Укртрансгаз»;
- методика с учетом фактического состояния запорной арматуры, но без восстановления фактического потокораспределения в ГТС;
- методика с учетом фактического состояния запорной арматуры и восстановлением фактического потокораспределения в ГТС.

Принципиальной особенностью постановки и решения задачи оперативного планирования режимов работы ГТС является то, что граничные условия о физических параметрах газовых потоках на входах и выходах ГТС задаются в виде математических ожиданий и дисперсий соответствующих случайных процессов. Результатом решения задачи является диспетчерский график в виде уставок для АСУ ТП и систем автоматизации нижнего уровня.

Опыт использования СППР в АСДУ ГТС

Использование АСОДУ в практике диспетчерских службы сводится к последовательному выполнению ряда этапов, что существенно сокращает время выработки и принятия управленческих решений.

1. Формирование классификатора типовых задач оперативно-диспетчерского управления ГТС.
2. Формирование сценариев решения задач управления ГТС.
3. Оценивание текущего состояния ГТС и выбор сценария решения задач управления ГТС.
4. Отработка сценария решения конкретной задачи в режиме off-line.
5. Анализ и тестирование полученного решения.
6. Реализация полученного решения задачи.
7. Контроль реализации и оценка эффективности принятого решения.

Заключение

Создание и внедрение рассматриваемой системы поддержки принятия решений как центрального звена автоматизированной системы оперативно-диспетчерского управления базируется на использовании интегрированного комплекса моделирования и оптимизации режимов транспорта и распределения природного газа в ГТС Украины.

Принципиальной особенностью постановки и решения задачи оперативного планирования режимов работы ГТС с использованием СППР на базе ИКМО является то, что граничные условия для физических параметров газовых потоках на входах и выходах ГТС задаются в виде математических ожиданий и дисперсий случайных процессов. Результатом решения задачи является диспетчерский график, обоснованно сформированный на основе результатов моделирования, и реализуемый в виде уставок для АСУ ТП и САУ нижнего уровня.

Бібліографія

- [1] Борисенко В.П., Борисенко Т.І. та інші. Концепція створення комплексної автоматизованої системи керування основними бізнес-процесами НАК "Нафтогаз України". Ч.1. Загальні положення.(ред. 2). – Харків - Київ: НДПІАСУтрансгаз – НАК "Нафтогаз України", 2007 р.
- [2] Борисенко В.П., Борисенко Т.І. та інші. Концепція створення комплексної автоматизованої системи керування основними бізнес-процесами НАК "Нафтогаз України". Ч.2. Корпоративна методологія створення та впровадження системи - Харків: НДПІАСУтрансгаз . - 2006 р
- [3] Борисенко В.П., Борисенко Т.І. та інші. Концепція створення комплексної автоматизованої системи керування основними бізнес-процесами НАК "Нафтогаз України" Ч.3. Програми створення системи (ред. 2). – Харків - Київ: НДПІАСУтрансгаз – НАК "Нафтогаз України" - 2007 р.
- [4] Борисенко В.П., Борисенко Т.І. та інші. Інвестиційне обґрунтування створення комплексної автоматизованої системи керування основними бізнес-процесами НАК "Нафтогаз України" (ред. 2).. - Харків - Київ: НДПІАСУтрансгаз – НАК "Нафтогаз України"-2007 р
- [5] Старовойтов В.Г, Коток В.Б. та інші. Концепція створення розподіленої автоматизованої системи керування технологічними процесами у нафтогазовій галузі України. – Харків - Київ: НДПІАСУтрансгаз – НАК "Нафтогаз України" - 2007 р.
- [6] Тевяшев А.Д., Тевяшева О.А., Фролов В.А. Стохастичний підхід до оцінювання ступеня технологічної стійкості режимів роботи газотранспортних систем.// Науково-виробничий журнал України "Нафтова і газова промисловість" - 2006 р., №4. - С.49-52., 4
- [7] Химко М.П., Фролов В.А., Павленко В.А., П'янило Я.Д., Притула Н.М. Розрахунок параметрів газотранспортних систем. // Науково-виробничий журнал України "Нафтова і газова промисловість" - 2006 р. №3 - С.33-37.,
- [8] Тевяшев А.Д., Коток В.Б., Выходцев Е.И., Пшеняник И.А. Об одном классе задач нестационарного неизотермического режима транспорта газа по ЛУ МГ. // Проблемы нафтогазовой промышленности: Сб. научных трудов. Вып.3.- Киев, 2006 – С.302-314.
- [9] Артюх А.Ю., Тевяшев А.Д., Фролов В.А. Стохастические модели и методы оптимизации режима работы компрессорных станций. // Всеукраинский межведомственный научно-технический сборник. «Автоматизированные системы управления и приборы автоматики», Харьков, ХНУРЭ, 2005, Выпуск 133, С.26-36.

Інформація об авторах

Виктор Борисенко – начальник ІТ-отдела, учений секретарь Совета института, к.т.н.,
Научно-исследовательский и проектный институт транспорта газа, ул. Маршала Конева, 16,
г. Харьков, 61004, Украина; e-mail: vborisenko@itransgaz.com

Юрий Пономарев – заместитель директора по научно-исследовательским работам, к.т.н., Научно-исследовательский и проектный институт транспорта газа, ул. Маршала Конева, 16, г. Харьков, 61004, Украина; e-mail: ponomar@itransgaz.com

Татьяна Борисенко – ведущий инженер-программист, к.т.н. Научно-исследовательский и проектный институт транспорта газа, ул. Маршала Конева, 16, г. Харьков, 61004, Украина;
e-mail: tiboris@itransgaz.com

ПРИМЕНЕНИЕ МЕТОДА ПАРАМЕТРИЧЕСКОЙ ОПТИМИЗАЦИИ ДЛЯ ЗАДАЧ ФОКУСИРОВКИ ПУЧКОВ ИОНОВ В ЛИНЕЙНЫХ УСКОРИТЕЛЯХ

Федор Гаращенко, Игорь Харченко

Аннотация: Предложен конструктивный метод параметрической оптимизации для проектирования систем ускорения и фокусировки ионов в линейных резонансных ускорителях. На примере математической модели процесса ускорения и фокусировки пучка ионов решена актуальная задача увеличения интенсивности пучка заряженных частиц для медицинского ускорителя заряженных частиц на 3 Мэв.

Ключевые слова: системы ускорения и фокусировки, структурно параметрическая оптимизация, динамические системы, линейный резонансный ускоритель.

ACM Classification Keywords: G.1.6 – Optimization – Nonlinear programming

Введение

Системы ускорения и фокусировки широко используются в научных исследованиях и различных отраслях народного хозяйства. В последнее время актуальным является проектирование линейных ускорителей для медицинских целей. Такие ускорители имеют ряд существенных преимуществ особенно для локального облучения пораженных участков. Методы проектирования таких систем базируются на основе физических принципов ускорения и фокусировки [Башняков, 2000]. На данный момент актуальными являются задачи проектирования ускоряюще-фокусирующих систем с оптимальными характеристиками пучка ионов на выходе ускорителя, которые позволяли бы при одном и том же уровне расходов получить пучки с большей энергией и с большей плотностью заряда [Бублик, 1985]. Параметрическое представление полей в таких задачах дает возможность определить оптимальные режимы для структур, которые подлежат практической реализации [Ладиков-Роев, 1996].

Постановка задач

Анализ различных характеристик ускоряюще-фокусирующих систем показывает, что их потенциальные возможности для получения оптимальных выходных характеристик пучка ионов используются не полностью. Для получения таких характеристик, например, в ускорителях с дрейфовыми трубками важно определить положение трубок (их длину, конфигурацию, напряжение поля в ускоряющих зазорах и т.п.), при которых достигается максимальная интенсивность пучка с заданными разбросами параметров его конечного состояния [Garashchenko, 2008].

Задача оптимизации пучка ионов в линейном резонансном ускорителе является очень сложной математической проблемой. Поэтому данную задачу при ее исследовании можно разбить на несколько подзадач:

- разработка методов расчета оптимальных ускоряющих структур для продольного движения;
- расчет динамики пучка при радиальной фокусировке ионов;
- исследование допусков на параметры системы ускорения и фокусировки;

– учет собственного электростатического поля пучка;
 – разработка быстродействующих численных алгоритмов расчета внешних электромагнитных полей и т.д.
 Для расчета и проектирования линейного ускорителя при учете параметров продольного движения, при заданном начальном разбросе по энергии и фазе, одна из основных задач заключается в определении структуры ускорителя таким образом, чтобы на выходе энергетический и фазовый разбросы пучка были минимальны.

Уравнение движения частицы будем рассматривать без учета сил кулоновского взаимодействия [Бублик, 1985]:

$$\frac{d\gamma}{d\xi} = \alpha(\xi) \cos \varphi,$$

$$\frac{d\varphi}{d\xi} = \frac{2\pi\gamma}{\sqrt{\gamma^2 - 1}},$$

$$\frac{d^2x}{d\xi^2} = \frac{\gamma}{\gamma^2 - 1} \left[\left(-\frac{1}{2} \frac{d\alpha}{d\xi} + g(\xi) \right) x - \alpha(\xi) \frac{dx}{d\xi} \right] \cos \varphi,$$

$$\frac{d^2y}{d\xi^2} = \frac{\gamma}{\gamma^2 - 1} \left[\left(-\frac{1}{2} \frac{d\alpha}{d\xi} - g(\xi) \right) y - \alpha(\xi) \frac{dy}{d\xi} \right] \cos \varphi.$$

Методика решения задачи

Будем считать, что известны точки переключения ξ_i и амплитуда напряженности поля, которая задается кусочно-постоянной функцией $\alpha(\xi)$.

Рассмотрим задачу выбора такого фокусирующего поля, чтобы для начальных условий $\gamma_0, \varphi_0, x_0, y_0, \dot{x}_0, \dot{y}_0$ выполнялся следующий критерий качества:

$$\min_{g(\xi)} (x^2(T) + y^2(T)).$$

Также будем полагать, что функция $g(\xi)$ принадлежит классу кусочно-постоянных функций. Тогда нормализованные уравнения движения частицы в ускоряющем зазоре линейного резонансного ускорителя имеют вид:

$$\frac{d\gamma}{d\xi} = \alpha(\xi) \cos \varphi,$$

$$\frac{d\varphi}{d\xi} = \frac{2\pi\gamma}{\sqrt{\gamma^2 - 1}},$$

$$\frac{dx_1}{d\xi} = x_2,$$

$$\frac{dx_2}{d\xi} = \frac{\gamma}{\gamma^2 - 1} \left[g(\xi)x_1 - \alpha(\xi) \frac{dx_1}{d\xi} \right] \cos \varphi,$$

$$\frac{dy_1}{d\xi} = y_2,$$

$$\frac{dy_2}{d\xi} = \frac{\gamma}{\gamma^2 - 1} \left[-g(\xi)y_1 - \alpha(\xi) \frac{dy_1}{d\xi} \right] \cos \varphi,$$

а уравнение движения частицы в дрейфовой трубке представлены следующими соотношениями :

$$\frac{d\gamma}{d\xi} = \alpha(\xi) \cos \varphi,$$

$$\frac{d\varphi}{d\xi} = \frac{2\pi\gamma}{\sqrt{\gamma^2 - 1}},$$

$$\frac{dx_1}{d\xi} = x_2,$$

$$\frac{dx_2}{d\xi} = 0,$$

$$\frac{dy_1}{d\xi} = y_2,$$

$$\frac{dy_2}{d\xi} = 0.$$

При этом скачки соответствующих фазовых координат в точках переключения системы ускорения имеют вид :

$$x_2(\xi_i + 0) = x_2(\xi_i) - \frac{1}{2} \frac{\gamma(\xi_i)}{\gamma^2(\xi_i) - 1} \cos \varphi(\xi_i) x_1(\xi_i) [\alpha(\xi_i + 0) - \alpha(\xi_i)],$$

$$y_2(\xi_i + 0) = y_2(\xi_i) - \frac{1}{2} \frac{\gamma(\xi_i)}{\gamma^2(\xi_i) - 1} \cos \varphi(\xi_i) y_1(\xi_i) [\alpha(\xi_i + 0) - \alpha(\xi_i)].$$

С целью оптимизации критерия качества поставленной задачи запишем функцию Гамильтона [Понтрягин, 1976]

$$\begin{aligned} H(\psi, x, y, \alpha, g, \gamma, \varphi, \xi) = & \psi_1(\xi) \alpha(\xi) \cos(\varphi) + \psi_2(\xi) \frac{2\pi\gamma}{\sqrt{\gamma^2 - 1}} + \psi_3(\xi) x_2 + \\ & + \psi_4(\xi) \frac{\gamma}{\gamma^2 - 1} \left[\left(-\frac{1}{2} \frac{d\alpha}{d\xi} + g(\xi) \right) x_1 - \alpha(\xi) x_2 \right] \cos \varphi + \psi_5(\xi) y_2 + \\ & + \psi_6(\xi) \frac{\gamma}{\gamma^2 - 1} \left[\left(-\frac{1}{2} \frac{d\alpha}{d\xi} - g(\xi) \right) y_1 - \alpha(\xi) y_2 \right] \cos \varphi. \end{aligned}$$

Тогда соответствующая система для расчета сопряженных переменных [Бублик, 1975] $\psi(\xi) = (\psi_1(\xi), \psi_2(\xi), \psi_3(\xi), \psi_4(\xi), \psi_5(\xi), \psi_6(\xi))^T$, которые будут далее использованы при решении оптимизационной задачи в ускоряющем зазоре, имеет следующий вид :

$$\frac{d\psi}{d\xi} = -\text{grad}_z H(\psi, x, y, \alpha, g, \gamma, \varphi, \xi), \quad z = (\gamma, \varphi, x_1, x_2, y_1, y_2)^T,$$

$$\frac{d\psi_1}{d\xi} = \frac{2\pi}{(\gamma^2 - 1)^{3/2}} \psi_2(\xi) + \psi_2(\xi) \frac{\gamma^2 + 1}{(\gamma^2 - 1)^2} [g(\xi)x_1 - \alpha(\xi)x_2] \cos \varphi + \psi_6(\xi) \frac{\gamma^2 + 1}{(\gamma^2 - 1)^2} [-g(\xi)y_1 - \alpha(\xi)y_2] \cos \varphi,$$

$$\frac{d\psi_2}{d\xi} = \psi_1(\xi)\alpha(\xi)\sin \varphi + \psi_4(\xi) \frac{\gamma}{\gamma^2 - 1} [g(\xi)x_1 - \alpha(\xi)x_2] \sin \varphi + \psi_6(\xi) \frac{\gamma}{\gamma^2 - 1} [-g(\xi)y_1 - \alpha(\xi)y_2] \sin \varphi,$$

$$\frac{d\psi_3}{d\xi} = -\psi_4(\xi) \frac{\gamma}{\gamma^2 - 1} g(\xi) \cos \varphi,$$

$$\frac{d\psi_4}{d\xi} = -\psi_3(\xi) + \psi_4(\xi) \frac{\gamma}{\gamma^2 - 1} \alpha(\xi) \cos \varphi,$$

$$\frac{d\psi_5}{d\xi} = \psi_6(\xi) \frac{\gamma}{\gamma^2 - 1} g(\xi) \cos \varphi,$$

$$\frac{d\psi_6}{d\xi} = -\psi_5(\xi) + \psi_6(\xi) \frac{\gamma}{\gamma^2 - 1} \alpha(\xi) \cos \varphi$$

с начальными условиями

$$\psi(T) = -(0, 0, 2x_1(T), 0, 2y_1(T), 0)^T.$$

Система для расчета сопряженных переменных в трубке дрейфа имеет вид :

$$\frac{d\psi_1}{d\xi} = \frac{2\pi}{(\gamma^2 - 1)^{3/2}} \psi_2(\xi),$$

$$\frac{d\psi_2}{d\xi} = 0,$$

$$\frac{d\psi_3}{d\xi} = 0,$$

$$\frac{d\psi_4}{d\xi} = \psi_3(\xi),$$

$$\frac{d\psi_5}{d\xi} = 0,$$

$$\frac{d\psi_6}{d\xi} = \psi_5(\xi)$$

с начальными условиями

$$\psi(T) = -(0, 0, 2x_1(T), 0, 2y_1(T), 0)^T.$$

Скачки в точках переключения для сопряженной системы переменных имеют вид :

$$\psi_1(\xi_i + 0) = \psi_1(\xi_i) - \frac{1}{2} \psi_4(\xi_i) \frac{\gamma^2(\xi_i) + 1}{(\gamma^2(\xi_i) - 1)^2} \cos \varphi(\xi_i) x_1(\xi_i) [\alpha(\xi_i + 0) - \alpha(\xi_i)] -$$

$$- \frac{1}{2} \psi_6(\xi_i) \frac{\gamma^2(\xi_i) + 1}{(\gamma^2(\xi_i) - 1)^{3/2}} \cos \varphi(\xi_i) y_1(\xi_i) [\alpha(\xi_i + 0) - \alpha(\xi_i)],$$

$$\begin{aligned} \psi_2(\xi_i + 0) = \psi_2(\xi_i) - \frac{1}{2}\psi_4(\xi_i) \frac{\gamma(\xi_i)}{\gamma^2(\xi_i) - 1} \sin \varphi(\xi_i) x_1(\xi_i) [\alpha(\xi_i + 0) - \alpha(\xi_i)] - \\ - \frac{1}{2}\psi_6(\xi_i) \frac{\gamma(\xi_i)}{\gamma^2(\xi_i) - 1} \sin \varphi(\xi_i) y_1(\xi_i) [\alpha(\xi_i + 0) - \alpha(\xi_i)], \end{aligned}$$

$$\psi_3(\xi_i + 0) = \psi_3(\xi_i) + \frac{1}{2}\psi_4(\xi_i) \frac{\gamma(\xi_i)}{\gamma^2(\xi_i) - 1} \cos \varphi(\xi_i) [\alpha(\xi_i + 0) - \alpha(\xi_i)],$$

$$\psi_5(\xi_i + 0) = \psi_5(\xi_i) + \frac{1}{2}\psi_6(\xi_i) \frac{\gamma(\xi_i)}{\gamma^2(\xi_i) - 1} \cos \varphi(\xi_i) [\alpha(\xi_i + 0) - \alpha(\xi_i)].$$

Градиент от критерия качества по параметрам g_i для поставленной оптимизационной задачи записывается в следующем виде :

$$\text{grad}_{g_i} I = - \int_{\xi_i}^{\xi_{i+1}} \text{grad}_{g_i} H(\psi, x, y, \alpha, g, \gamma, \varphi, \xi) d\xi,$$

где

$$\text{grad}_{g_i} H(\psi, x, y, \alpha, g, \gamma, \varphi, \xi) = \psi_4(\xi) \frac{\gamma}{\gamma^2 - 1} x_1 \cos \varphi - \psi_6(\xi) \frac{\gamma}{\gamma^2 - 1} y_1 \cos \varphi.$$

Выводы

Алгоритмы управления с использованием градиентных методов в классе кусочно-непрерывных функций для многих прикладных задач использовать трудно в силу плохой сходимости итерационных процедур. Поэтому применяют методику параметризации управления и, таким образом, переходят к конечномерным оптимизационным задачам [Горбунов, 1979]. Решение таких задач в силу неявной зависимости решений от параметров также представляет достаточно сложную оптимизационную задачу, особенно в случае большого количества оптимизационных параметров. Поэтому для конкретных задач, изучая их природу, на первом этапе пытаются ввести минимальное количество оптимизационных параметров, возможно и формальных, которые бы приближенно определяли большее количество реальных параметров и с достаточной адекватностью описывали функционирование объекта. Далее находят оптимальный режим в пространстве меньшего количества параметров, который дает возможность определить приближение для большего количества оптимальных характеристик пучка в линейном резонансном ускорителе. Большое количество оптимальных параметров может стать основой для определения начального приближения функции управления, например, в классе кусочно-непрерывных функций и т.д. Такой подход является конструктивным и связанным с усложнением математической модели, как по управлению, так и по ее структуре. С другой стороны, использование метода параметрической оптимизации дает возможность определить режимы с оптимальной физической реализацией [Яковлев, 1996].

На основе указанной методики в данной работе разработан и протестирован алгоритм оптимального выбора фокусирующего поля для проектирования линейных резонансных ускорителей ионов для медицинских целей на 3 МэВ.

Литература

- [Башняков, 2000] Башняков А.Н., Гаращенко Ф.Г., Пичкур В.В. Практическая устойчивость и структурная оптимизация динамических систем. – К.: Издательско-полиграфический центр “Киевский университет”, 2000.
- [Бублик, 1975] Бублик Б.Н., Кириченко Н.Ф. Основы теории управления. – К.: Высшая школа, 1975.
- [Бублик, 1985] Бублик Б.Н., Гаращенко Ф.Г., Кириченко Н.Ф. Структурно–параметрическая оптимизация и устойчивость динамики пучков. – К.: Наукова думка, 1985.
- [Горбунов, 1979] Горбунов В.К. Метод параметризации задач оптимального управления // Журнал вычислительной математики и математической физики. – 1979. – т. 19, №2. – С. 292-303.
- [Ладиков-Роев, 1996] Ладиков-Роев Ю.П., Самойленко Ю.И. Структурная оптимизация регулирующих сред // Проблемы управления и информатики, – 1996, – №1-2, – С. 101-108.
- [Понтрягин, 1976] Понтрягин Л.С., Болтянский В.Г, Гамкрелидзе Р.В., Мищенко Е.Ф. Математическая теория оптимальных процессов. – М.: Наука, 1976.
- [Яковлев, 1996] Яковлев О.С. Метод структурного синтеза нелинейных регуляторов // Проблемы управления и информатики, – 1996, – №1-2, – С. 211-223.
- [Garashchenko, 2008] Garashchenko F, Kharchenko I. Applied aspects of mathematical modeling and optimization of dynamics of charged beams //Decision Making and Business Intelligence Strategies and Techniques/ International Books Series “INFORMATION SCIENCE & COMPUTING”, Number 3, P. 23-29.

Информация об авторах

Федор Гаращенко – Профессор, Киевский национальный университет имени Тараса Шевченко, факультет кибернетики, заведующий кафедрой моделирования сложных систем, Киев, Украина, e-mail: garash@unicyb.kiev.ua

Игорь Харченко – Доцент, Киевский национальный университет имени Тараса Шевченко, факультет кибернетики, кафедра моделирования сложных систем, Киев, Украина, e-mail: ihar@unicyb.kiev.ua

ИСПОЛЬЗОВАНИЕ АЛГОРИТМА УСКОРЕННОГО ВЕРОЯТНОСТНОГО МОДЕЛИРОВАНИЯ В СХЕМЕ ТАБУ-ПОИСКА

Леонид Гуляницкий, Александр Турчин

Аннотация: Рассматривается метаэвристический метод комбинаторной оптимизации, основанный на использовании алгоритмов табу-поиска и ускоренного вероятностного моделирования. Излагается общая вычислительная схема предложенного метода, названного алгоритмом GS-tabu. Приведены результаты серии вычислительных экспериментов по решению известных задач коммивояжера и квадратичных задач о назначении.

Keywords: combinatorial optimization, stochastic local search, tabu search.

ACM Classification Keywords: G.1.6 Numerical Analysis; G.2.1 Discrete Mathematics; I.2.8 Artificial Intelligence.

Conference: Decision Making

Формальная постановка задач комбинаторной оптимизации

Задачи комбинаторной оптимизации (ЗКО) возникают при исследовании многих сложных процессов – например таких, как исследование организации сложных систем. Такие задачи – наряду с методами их решения – имеют ключевое значение при решении проблем проектирования и размещения объектов, планирования экспериментов, управления процессами обработки данных, принятия решений в экономике и бизнесе и др. [1-2].

В настоящее время общепринятым является определение ЗКО, предложенное Пападимитриу и Стайглицем [2]: необходимо найти $x^* \in X$ такое, что

$$x_* = \arg \min_{x \in D \subseteq X} f(x), \quad (1)$$

где X – конечное (или, возможно, счетное бесконечное) пространство решений задачи, D – его подпространство, определяемое ограничениями задачи, $f: X \rightarrow \mathbb{R}^1$ – заданная целевая функция задачи.

Такое определение однозначно относит к ЗКО все проблемы оптимизации на конечных множествах, однако в случае бесконечных пространств не позволяет четко классифицировать задачи оптимизации по структуре элементов пространства решений. К.Берж [3] предложил формализовать понятие комбинаторной конфигурации следующим образом: пусть имеем m, n – натуральные, а также два множества $U = \{1, \dots, m\}$, $V = \{v_1, \dots, v_n\}$, причем на V задан некий строгий порядок $v_1 < \dots < v_n$ (т.е. V – цепь).

Определение 1 [3]. Комбинаторной конфигурацией называется отображение $\varphi: U \rightarrow V$, которое удовлетворяет некоторому комплексу ограничений Λ .

Из определения следует, что при фиксированных m и n число комбинаторных конфигураций конечно. Выбор ограничений в Λ позволяет описывать различные комбинаторные конфигурации.

В работе [4] предлагается следующее обобщение схемы Бержа. Пусть заданы $Y = \{1, \dots, m\}$, Z – дискретное, в частности, конечное пространство (назовем его образующим), φ – гомоморфизм,

$\varphi: Y \rightarrow Z$, удовлетворяющий некоторой системе ограничений Ω . Напомним, что под дискретным пространством понимается множество, состоящее из изолированных точек.

Определение 2. Под комбинаторным объектом κ будем понимать триаду $\kappa = (\varphi, \tilde{X}, \Omega)$, где \tilde{X} – базовое пространство.

Определение 3. Назовем комбинаторными объектами 1-го порядка такие комбинаторные объекты, у которых базовое пространство совпадает с образующим:

$$\kappa = (\varphi, X_{(1)}, \Omega), \quad X_{(1)} \equiv Z.$$

Нетрудно убедиться, что если Z – это конечная цепь, то такие комбинаторные объекты совпадают с комбинаторными конфигурациями в смысле Берга [5].

Определение 4. Комбинаторными объектами k -го порядка ($k > 1$) назовем комбинаторные объекты $\kappa = (\varphi, X_{(k)}, \Omega)$, где $X_{(k)} \subseteq X_{(k-1)} \cup X^k$.

Возвращаясь к оптимизационной задаче (1), дадим следующее

Определение 5. Задача (1) называется задачей КО, если пространство ее решений X – это пространство, элементами которого являются комбинаторные объекты.

Специфика большинства ЗКО типа (1) состоит в наличии большого количества экстремумов целевой функции. Кроме того, как правило эти задачи относятся к числу NP -трудных, так что точное их решение весьма проблематично даже с использованием современных и перспективных компьютеров.

Среди наиболее эффективных подходов к решению ЗКО следует выделить использование *метаэвристических алгоритмов*, или *метаэвристик*. Существует несколько подходов к классификации метаэвристических алгоритмов, среди которых наиболее удачным представляется деление на *метаэвристики 1-го рода* и *метаэвристики 2-го рода*; в свою очередь, метаэвристики 1-го рода представлены классами *траекторных* и *популяционных* метаэвристик. Траекторными метаэвристиками являются методы имитационного отжига, табу-поиск [5], алгоритмы ускоренного вероятностного моделирования [6]. Класс популяционных метаэвристик представляют генетические алгоритмы, алгоритмы оптимизации муравьиными колониями, иммунные алгоритмы [7,8].

Метаэвристический алгоритм (GS-tabu)

В работе [9] предложен подход к построению метаэвристического метода на основе алгоритмов табу-поиска и ускоренного вероятностного моделирования. В качестве базового был выбран метод ускоренного вероятностного моделирования (*GS-method* [10]), который реализует стратегию «золотого сечения». Для построенного алгоритма предложено название GS-tabu.

Алгоритмы ускоренного вероятностного моделирования были разработаны для преодоления таких недостатков методов локального поиска, как преждевременная сходимость алгоритма и большая зависимость конечного результата от выбора начального приближения решения. А среди алгоритмов, которые принадлежат к классу ускоренного вероятностного моделирования, следует выделить G-алгоритмы, которые успешно применялись для решения многих сложных практических задач [6].

Главная идея алгоритма табу-поиска состоит в постоянной локальной модификации текущего решения задачи при запоминании произведенных изменений – чтобы предотвратить алгоритм от повторного построения неэффективных решений и зацикливания процедуры поиска решения [5]. Осуществленные модификации запоминаются в специальных банках памяти, которые называются «табу-списками» (tabu list). Общая схема предлагаемого метаэвристического метода GS-tabu приведена на рис.1.

При практической апробации GS-tabu было установлено, что наилучшие результаты были получены при размере банка памяти около $n/3$, где n -размерность задачи. Кроме того, в нашем случае величина μ_t определяется как меньшая (левая) из двух точек, которые реализовывают «золотое сечение» отрезка $[\mu_t, 1]$. Таким образом, правило «золотого сечения» задает динамику приближения левой границы этого отрезка к 1 [9].

```

procedure GS-tabu(x)
begin
   $x^0$  := некоторый начальный допустимый вариант решения из X;
   $\mu_0 := 0$ ;  $h := 0$ ;  $t := 0$ ;  $x_{rec} := x^0$ ;  $f_{rec} := f(x^0)$ ; ТабуСписок= $\emptyset$ ;
  while окрестность текущего решения  $L(x^h)$  не просмотрена полностью do begin
    while не выполнено условие равновесия do begin
       $y :=$  ГенерированиеСледующейТочкиОкрестности  $L(x^h)$ ;
      ПрохождениеПроверкиТабуСписка( $y$ );
      Вычисление  $\Phi(x^h, y)$ ;
       $p := (1 - \mu_t) \Phi(x^h, y)$ ;
       $\xi := random[0, 1]$ ;
      if  $p \geq \xi$  then
         $h := h + 1$ ;  $x^h := y$ ;
        if  $f_{rec} > f(x^h)$  then  $x_{rec} := x^h$ ;  $f_{rec} := f(x^h)$  end if
      end if;
      ОбновлениеТабуСписка;
    end;
    ФормированиеОчередногоЗначения  $\mu_{t+1}$ ;  $t := t + 1$ ;
  end;
  return  $x = x_{rec}$ ;
end

```

Рис. 1. Вычислительная схема алгоритма GS-tabu.

Правилом останова может служить: окончание перебора всех точек в окрестности без реализованного перехода в новую точку; ограничение по продолжительности работы алгоритма; достижение требуемой точности – при известной нижней границе целевой функции. Еще одно из распространенных правил – сравнение разности максимального и минимального значений целевой функции с максимальным значением изменения этой функции при данном значении μ : если это соотношение стремится к единице, то вычисления завершаются.

Исследование практической эффективности GS-tabu

Для исследования практической эффективности GS-tabu по сравнению с алгоритмом GS-method были выбраны два наиболее известных класса ЗКО: задачи коммивояжера из библиотеки TSPLIB95 [11] и квадратичные задачи о назначениях из библиотеки QAPLIB [12]. Результаты решения серий реальных и тестовых задач алгоритмами GS-method и GS-tabu приведены в табл.1 (задача коммивояжера) и табл. 2 (квадратичная задача о назначениях). Для каждой задачи приведено ее библиотечное имя [11,12], размерность (n) и наилучшее из известных решений.

Задача	n	Наилучшее известное решение	Методы			
			GS-method	Время,с	GS-tabu	Время,с
burmal4	14	3 323	3 340	1,2	3 323	1,2
ulysses16	16	6 850	6 876	2,1	6 891	2,1
gr17	17	2 085	2 114	1,9	2 092	1,9
ulysses 22	22	7 013	7 116	14,0	7 054	14,0
gr24	24	1 272	1 282	14,5	1 286	14,5
fri26	26	937	959	19,5	943	20,5
bays29	29	2 020	2 052	28,8	2 020	24,8
sui ss42	42	1 273	1 278	42,3	1 286	42,3
dantzig42	42	699	710	50,0	699	50,0
att48	48	10 628	10 930	72,6	10 670	76,6
hk48	48	11 461	11 464	78,4	11 603	78,4
eil51	51	426	431	106,0	430	106,0
brazil58	58	25 395	25 927	123,1	25 460	111,0
st 70	70	675	688	112,3	676	114,3
pr76	76	108 150	110 571	144,1	109 940	147,1
rat 99	99	1 211	1 244	173,0	1 251	153,3
kroA100	100	21 282	22 243	152,1	21 826	160,1
kroB100	100	22 141	22 609	154,0	22 540	159,0
kroC100	100	20 749	22 095	183,0	21 148	175,5
pr124	124	50 030	52 067	241,3	52 124	252,3
pr152	152	73 682	76 533	288,8	76 107	293,8
rati95	195	2 323	2 412	320,8	2 380	323,8

Таблица 1. Результаты решения задачи коммивояжера

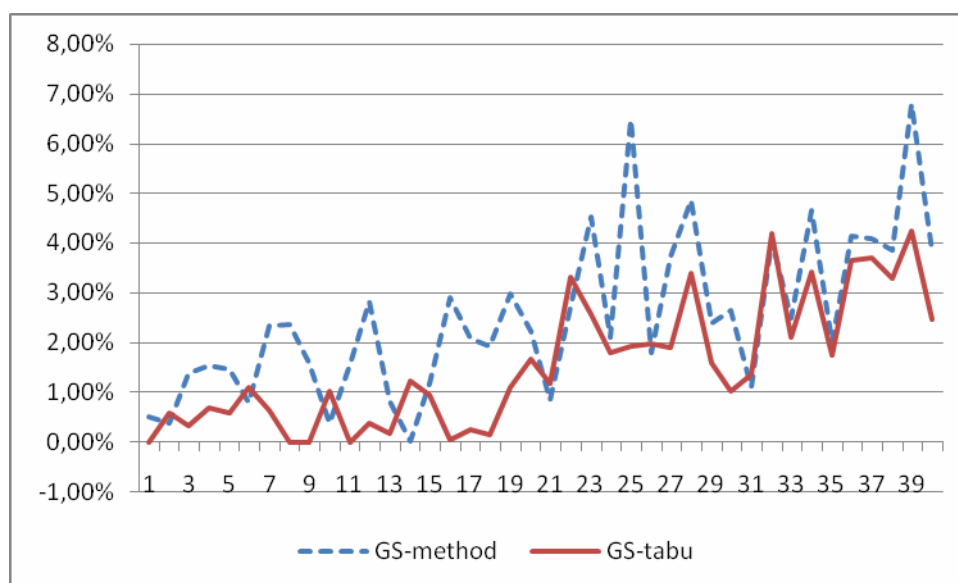


Рис. 2. Относительная точность решений

Задача	n	Наилучшее известное решение	Методы			
			GS-method	Время,с	GS-tabu	Время,с
chr15a	15	9 896	10 112	1,8	9 956	1,8
chr15b	15	7 990	8 064	3,2	8 011	2,9
chr15c	15	9 504	9 611	2,5	9 504	2,8
bur26d	26	3 821 225	3 929 672	21,2	3 911 457	23,6
bur26e	26	5 386 879	5 478 936	22,8	5 502 452	25,9
bur26f	26	3 782 044	3 802 312	28,3	3 788 565	34,2
esc64a	64	116	121	89,1	118	92,0
lipa70a	70	169 755	175 341	107,4	170 753	112,5
lipa70b	70	4 603 200	4 781 819	111,3	4 711 844	132,8
lipa80a	80	253 195	255 221	129,5	254 198	144,4
lipa80b	80	7 763 962	7 945 390	141,9	7 878 543	150,1
lipa90a	90	360 630	361 717	166,5	363 991	182,1
lipa90b	90	12 490 441	12 832 091	170,4	12 774 910	188,1
sko100a	100	152 002	156 748	179,9	154 340	202,2
sko100b	100	153 890	154 120	212,0	157 463	195,5
sko100c	100	147 862	155 543	200,4	149 341	225,7
sko100d	100	149 576	153 288	192,1	151 922	200,1
sko100e	100	149 150	154 012	188,4	149 781	211,4

Таблица 1. Результаты решения квадратичной задачи о назначениях

На рис. 2 приведен график отклонений по точности найденных метаэвристическим методом и GS-алгоритмом вариантов решения от наилучших из известных решений, упорядоченных по возрастанию размерностей задач.

Вычислительный эксперимент проводился на ПК Athlon 64 X2 4Ghz, 2 Gb оперативной памяти. Программный комплекс, реализующий оба метода, написан с помощью среды разработки Delphi.

Заключение

Результаты вычислительного эксперимента свидетельствуют о том, что применение предложенного метаэвристического метода, построенного на базе алгоритмов GS-method и табу-поиска, позволило улучшить точность получаемых решений по сравнению с использованием отдельных алгоритмов, входящих в его схему.

Направление дальнейших исследований будет сосредоточено на исследовании как практической, так и теоретической эффективности нового метода. Особое внимание предполагается уделить возможности построения с помощью GS-tabu новых эффективных метаэвристик 2-го рода, предназначенных для решения широкого круга прикладных ЗКО.

Литература

- [1] Сергиенко И.В. Математические модели и методы решения задач дискретной оптимизации. – К.: Наукова думка, 1985. – 384 с.
- [2] Пападимитриу Х., Стайглиц К. Комбинаторная оптимизация. Алгоритмы и сложность. – М.: Мир, 1985. – 512 с.
- [3] Berge С. Principes de combinatoire. – Paris: Dunod, 1968. – 146 p.
- [4] Гуляницкий Л.Ф., Сергиенко И.В. Метаэвристический метод деформируемого многогранника в комбинаторной оптимизации // Кибернетика и системный анализ. – 2007. – № 6. – С. 70–79.
- [5] Glover F. Heuristics for Integer Programming using surrogate constraints // Decision Sciences. – 1977. – N 8. – P.156-166.
- [6] Гуляницкий Л.Ф. Решение задач комбинаторной оптимизации алгоритмами ускоренного вероятностного моделирования // Компьютерная математика. – Киев: Ин-т кибернетики им. В.М. Глушкова НАН Украины, 2004. – №1. – С. 64–72.
- [7] Blum С., Roli А. Metaheuristics in Combinatorial Optimization: Overview and Conceptual Comparison // ACM Computing Surveys. – 2003. – 35, No. 3. – P. 268–308.
- [8] Hoos Н.Н., Stützle Т. Stochastic Local Search: Foundations and Applications. – San Francisco: Morgan Kaufmann Publ., 2005. – 658 p
- [9] Турчин О. Використання алгоритму імовірнісного моделювання у схемі методу табу-пошуку // Пр. Між. симп. «Питання оптимізації обчислень (ПОО-XXXIII)». – К.: ІК ім.В.М.Глушкова НАН України, 2007.– P. 73–77.
- [10] Hulyanitsky L., Turchin O. "Golden section" rule in probabilistic modeling algorithms // Вісник Нац. ун-ту "Львівська політехніка". – 2001. – №415. – С. 50-53.
- [11] <http://softlib.rice.edu/pub/tsplib>
- [12] <http://www.opt.math.tu-graz.ac.at/~karisch/qaplib>

Сведения об авторах

Леонид Гуляницкий (*Hulianytskyi*) – д.т.н., ведущий научный сотрудник Института кибернетики им. В.М.Глушкова НАН Украины, пр-т Глушкова, 40, Киев, 03680, Украина. e-mail: lh_dar@hotmail.com

Александр Турчин – Институт кибернетики им. В.М.Глушкова НАН Украины, пр-т Глушкова, 40, Киев, 03680, Украина. e-mail: turchin@ua.fm

ALGORITHM OF DECISION SUPPORT IN EMERGENCY

Oleksandr Kuzomin, Valentina Davidenko, Illya Klymov

Abstract: Constant increase of human population result in more and more people living in emergency dangerous regions. In order to protect them from possible emergencies we need effective solution for decision taking in case of emergencies, because lack of time for taking decision and possible lack of data. One among possible methods of taking such decisions is shown in this article.

Keywords: emergency, decision taking, time management

ACM Classification Keywords: H.3.4 Systems and Software

Introduction

Rapid growth of population on our planet inevitably leads to more and more people located on the territories which are affected by that or another dangers. According to United Nations statistics, more than 6% of world population is living on territories, annually suffering from emergency situations. Taking that into account, effective prediction, control and prevention measures for emergencies become extremely actual.

The task of complex development for preventing emergency situations is quite complex task of multi criteria optimization. For person, taking decisions, a problems of human risks and material losses minimization and also money losses are very actual. Taking decision in case of very limited time also leads to risk of errors and human factor impact.

Attempts to apply these or another automated complexes of providing solutions in emergency situations often lead to a problem of internal system complex representation (complex of links "operator-available resources-current situation-complex of measures to prevent emergency"), where change of one parameter can greatly affect selection of technical and human means.

So, the task of **context-dependent measures complex development in emergency situations** becomes important. In fact, the primary **aim** of this article is research of possible appliances of context-depended solutions in such cases.

Problem statement

As a parameter for our system there is a current situation **Sit**, which could be represented as set of substances' features and interconnections between them.

$$\text{Sit} = \text{cont}_k = \{A, L\},$$

where $A = \{a_1, \dots, a_m\}$ – set of environment substances and, m – their amount; $L = \{l_1, \dots, l_s\} = \{(a_i, a_j, c)\}$ – is a set of connections between pairs of environment substances. $i = 1, m; j = 1, m; c \in \{0, 1\}$.

In role of substances in current context depending on system task may appear different nature objects (wood array, sea, mountain hill), natural phenomenons (air stream, underwater streams), subjects (rescue rangers, scientific workers, researchers), material wealth (buildings, strategical military objects).

Each substance of context have a set of properties:

$$\forall a_i \in A = \{P_{a_i 1}, \dots, P_{a_i n_i}\},$$

where P_{a_i} – is a subset of properties for substance a_i . k_i – is a count of properties of substance a_i . Examples of substance “air flow” properties can be “velocity”, “direction” and so on.

As another substance of the system we will present a set of solutions $D = \{d_1, \dots, d_n\}$, which describes all available means for preventing and protecting from emergency.

As a data storage for taken earlier decisions in that or another context we will use a knowledge base (KB). Format of data storage will be triple system “context-solution-effectiveness”. Coefficient of effectiveness q is calculated depending on material and non-material losses in emergency context based on taken decision. In fact minimization of that parameter is the main criteria of system effectiveness.

Algorithm of optimum decision finding

For finding optimal decision in unknown context (we specify context as unknown if it isn't found in our KB) let's create a conversion function: $d'_k = F(cont_k)$, where $d'_k \in D'$ - set of new obtained from KB solutions; $F(x)$ – function of solution taking for context x . Principal scheme of that method implementation is presented on figure 1.

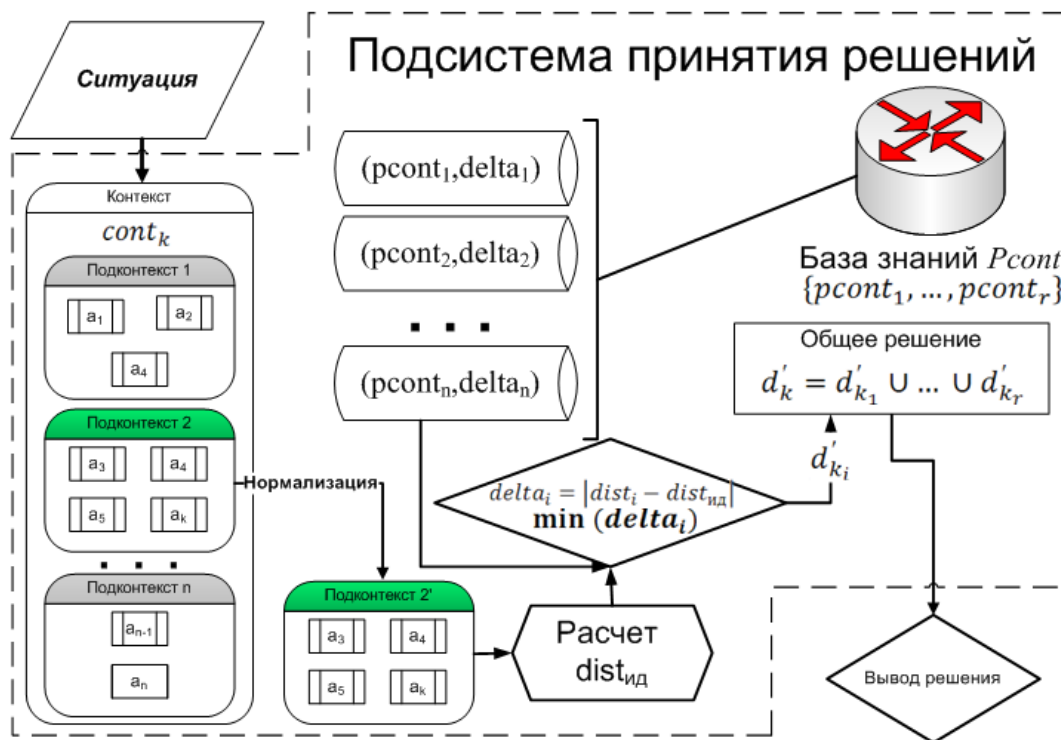


Fig. 1. Principal scheme of algorithm.

$F(x)$ execution algorithm:

1. Separate the environment context to sub-contexts. Separation will be performed according to common meaning of that situation. Sub-contexts are obtained both in current contexts and contexts, extracted from KB.
2. Create a set of sub-contexts $Pcont = \{pcont_1, \dots, pcont_r\}$, where $pcont_i = \{a_1, \dots, a_{n_2}, l_1, \dots, l_{s_2}\}$, where $\forall i, k: pcont_i \cap pcont_k \geq \emptyset$; r – amount of sub-contexts;

General solution for all context is composed from solutions for separate sub-contexts $d_k^i = d_{k_2}^i \cup \dots \cup d_{k_r}^i$, where $d_{k_i}^i$ - solution for sub-context $pcont_i$.

3. For each sub-context perform a properties normalization using formula: $x_n = \frac{x}{x_{max}}$, where x – current value; x_{max} – maximal possible value of that property; x_n – normalized value. A parameter q in KB is also normalized.
4. Introduce a virtual situation, where in presented sub-context losses will be minimal ($q=0$). For each sub-context of new description we find in KB nearest contexts, which are close to ideal situations.

Let's introduce to our system parameter $[A]$ – vector of coefficients attached to independent variables, which define logical distance from ideal situation to nearest context in KB with optimal solution. Independent variables are the values of substance properties for current sub-context and parameter “ q ” for current solution (taken from KB).

Define $[P^q]$ – as a vector of properties for sub-context substance, where last element is parameter q , which describes losses of current sub-context in KB; we receive it using conversion of properties set and adding element q to them.

Include set $Dist = \{dist_1, \dots, dist_N\}$, where $dist_i = [A] * [P_i^q]$, N – total amount of records in KB.

Criteria of how current sub-context is close to current (k) sub-context is $delta_k$:

$delta_k = |dist_i - dist_{id}|$, where $dist_{id}$ – characteristics of ideal state. We add a set $Delta = \{delta_1, \dots, delta_z\}$, where z – amount of parameters $delta$ for current sub-context.

After creating such sets we normalize their values.

After finding $delta$ for all sub-context we find minimal parameter $delta_{min}$ and add it to the set $Delta_{min} = \{delta_{min_1}, \dots, delta_{min_r}\}$, where r – amount of sub-contexts. In that case we store solution, where we reach $delta_{min}$, and add it to general solution.

5. We assume that there is a predefined information about structure information about solution interactions: $G = \{(g_i, g_j, c)\}$, where g_i, g_j – different solutions, $c \in \{-1, 0, 1, 2\}$, where
 - if $c = -1$ – events g_i and g_j are not compatible;
 - if $c = 0$ – events g_i and g_j can be performed parallel;
 - if $c = 1$ – event g_i must follow before event g_j ;
 - if $c = 2$ – event g_j must follow before event g_i .

According to this rules we perform a pair check of all proposed events (in term “event” we describe a possible mean to prevent emergency) and according to results of such comparison we perform a topological sort in order to receive a schedule of performing actions.

Conflicts resolution and system learning

Provided by $F(x)$ solution may content conflicting data. For example system may recommend to perform a forced avalanche launch, and do not allow it in another sub-solution. We will use a set Δ_{min} to eliminate such conflicts.

Selecting from composed set $d_{i_2}^i$ for context $cont_x$, we choose positive oriented solutions of same type and sum parameters of same type Δ_{min} (let call it k_{Σ}^+), so we perform the same for negative oriented recommendations (receive k_{Σ}^-). So, if $k_{\Sigma}^+ > k_{\Sigma}^-$, we save a positive solution, if $k_{\Sigma}^+ < k_{\Sigma}^-$, we save only negative oriented solution. If $k_{\Sigma}^+ = k_{\Sigma}^-$ we need an expert opinion.

System learning is performed by finding optimal system parameters in matrix $[A]$. This process performs by using classical scheme by learning with teacher, where we use a teacher to correct provided data.

So we learn system as follows:

1. We provide a context to system input.
2. System provides a current estimation.
3. Expert analyzes provided solution and performs corrections. If there are corrections we perform correction of matrix $[A]$ elements, in such way that $[A] * [P_{optimal}^q]$ was "arranged left" to $[A] * [P_{expert}^q]$.

Authors' Information

Oleksandr Kuzomin – Head of Innovation marketing department, professor in informatics dept.; Kharkiv National University of Radio electronics, e-mail: kuzy@kture.kharkov.ua

Valentina Davidenko – Postgraduate, National Aero Space University in the name of N.E. Jukovskiy; e-mail: davidenko.valentina@gmail.com

Iliya Klimov – Postgraduate, Informatics dept.; Kharkiv National University of Radio Electronics; e-mail: ilia.klimov@gmail.com

MODELING OF EFFECTIVE PROCESS OF NETWORK MAINTAINING BASED ON STATISTICAL DATA

Oleksandr Kuzomin, Nataliya Goloviy, Yaser Dayoub

Abstract: *This paper is dedicated to modelling of network maintaining based on live example – maintaining ATM banking network, where any problems are mean money loss. A full analysis is made in order to estimate valuable and not-valuable parameters based on complex analysis of available data. Correlation analysis helps to estimate provided data and to produce a complex solution of increasing network maintaining effectiveness.*

Keywords: *effectiveness estimation, statistical processes, correlation analysis*

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

Introduction

Choosing equipment of this or that supplier, as a rule, pay attention to following parameters: the price, productivity, reliability and convenience. For a network of ATM where breakage or device stop means the lost money, service is the extremely important.

Equipment downtimes because of malfunction and repair, breaking business process, are capable to worsen economic indicators of banks sharply. Therefore, corresponding requirements also are shown to service services - to provide competitive advantage at modern level it is necessary to contain in an appropriate technical condition the equipment at the minimum losses of manufacture.

For the organization service and warranty service the system representing a complex of organizational-technical actions which realization would allow to make the basic part of works in advance known terms providing the prevention of progressing wear process of details technicians in operation and reduction of losses because of its malfunction and repair should be created.

At the heart of such system creation attempt of process modelling which usually not to do without application empiric and statistical the data about refusals and changes of parameters of a condition of elements and specifications of time for performance of preventive operations taking into account specificity of manufacture and increasing complexity of the applied equipment lies.

Problem Statement

It is necessary to generate general idea about the complex character of the studied phenomenon expressed, in particular, in interrelations and interconditionality of separate features. To conduct research of the internal reasons forming specificity of the studied phenomenon, and to reveal the generalised factors which stand up for corresponding concrete indicators. Proceeding from the formulated problem definition, for an investigated subject domain it is necessary to study dependence between the features describing a condition of an ATM network.

An **objective** of this research is modelling of service process of an ATM network on the basis of statistical data for maintenance of its reliability and stability.

Data matrix forming

During conducted research as set of investigated objects the ATM network, being on service of the company which have some feature set (indicators, characteristics) is considered. The features characterising investigated

objects, carry both quantitative, and qualitative character. On the basis of the researches conducted and also expert appraisal, the features characterising each unit of a banking network were conducted.

Parameter	Definition	Type	Parameter	Definition	Type
$m1$	Service Life	quantitative	$m13$	Load	quantitative
$m2$	Operating time	quantitative	$m14$	Quantity of ATM's in region	quantitative
$m3$	Service refusals	quantitative	$m15$	Quantity of engineers in region	quantitative
$m4$	Overdue failures	quantitative	$m16$	Quantities of temperature parameters infringement	quantitative
$m5$	Time in service	quantitative	$m17$	Quantities of power losses	quantitative
$m6$	Average time of malfunctioning	quantitative	$m18$	Quantities of grounding parameters losses	quantitative
$m7$	Quantity of spare parts replacements	quantitative	$m19$	Quantities of installation problems	quantitative
$m8$	Qualifications of the engineer	quantitative	$m20$	Quantities of usage problems	quantitative
$m9$	Quantity of installations	quantitative	$m21$	Client	qualitative
$m10$	Novelty	quantitative	$m22$	Type (Rent/Owned)	qualitative
$m11$	Reaction Time	quantitative	$m23$	Warranty	qualitative
$m12$	Repair time	quantitative	$m24$	Region	qualitative

Table 1. Characteristics of ATM

Here we have n objects (ATM), each of them is characterized by set of m features. Let's define as x_{ij} value of j -th feature for i -th object. In that case data could be represented as a table called as "data matrix".

<u>Data matrix</u>						
Object number	Feature number					
	1	2	3	m
1	x_{11}	x_{12}	x_{13}	x_{1m}
2	x_{21}	x_{22}	x_{23}	x_{2m}
3	x_{31}	x_{32}	x_{33}	x_{3m}
...
n	x_{n1}	x_{n2}	x_{n3}	x_{nm}

Table 2. Data matrix sample

It consist of n lines (according to number of objects) and m columns (according to number of features), analysed sets of data consists of 908 objects.

Correlation analysis

Studying of dependences between the features describing the difficult phenomenon or process, invariably brings an attention to the question: what reasons which have caused given structure of communications? Difficult dependences of such system of features are rather difficult for "untangling", as interrelations of features (correlations measured, for example, in the factors) can be interpreted not only as dependences of one features on others, but also as dependences on the certain latent parameters defining change of the whole groups of correlated features.

We will generate some statistical hypotheses about dependence of the chosen features. We believe that maintenance of reliability of a network of cash dispenses from outside the service organization consists in time reduction of ATM malfunction and reduction of average time of malfunction. We receive that reliability of a network is defined by indicators $m5$ - $m6$. In turn these parameters depend on quantity of demands, delay of works and replaced spare parts (see Table 1). We investigate their dependence on other indicators for what we will carry out the correlation analysis.

We will calculate for existing quantitative features factors of correlation and we will construct a correlation matrix.

The structure of the experimental data which features in a context of a solved problem are described by mathematical model, is reflected by means of two basic categories of mutual relations between elements of a matrix of the data – similarity and distinction categories.

The matrix sets the relation "feature-featyre" and represents a two-dimensional symmetric square matrix of the size $m \times m$

$$R = \begin{pmatrix} r_{11}, \dots, r_{1m} \\ r_{21}, \dots, r_{2m} \\ \dots \\ r_{m1}, \dots, r_{mm} \end{pmatrix},$$

where R_{ij} — measure of connection between features m_i и m_j .

The considerable quantity of measures of communication between features is known. They differ both volume of calculations, and those aspects of communication which they reflect. Considering specificity of an investigated problem, we will consider two representation liaison groups between features. In the first group the covariation principle, and in the second — a principle of an associativity of features [2] is used.

Proceeding from the first principle, the conclusion about communication presence between variables becomes in that case when the increase in value of one variable is accompanied by steady increase or reduction of values another. In mathematical expression the problem is reduced to covariation calculation, that is accompanying change of numerical values of signs. Here to concern factor of correlation of Pirson (r) which represents product of the moments and is a measure of linear communication of two variables. It is calculated under the formula

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1)S_x S_y},$$

where x_i, y_i - values of two variables, \bar{x}, \bar{y} - their average values, S_x, S_y - their average square deviation, n – amount of pairs of values.

Many measures of communication differ from the resulted factor of correlation of Pirsona the external form, but are, as a matter of fact, the algebraic transformation of this factor considering specificity (type) of compared signs. For example, the factor of rank correlations of Spirmena (r_s), often applied to the analysis of numeric variables, represents algebraic simplification of factor of Pirsona r .

The same it is possible to tell about dotbiserial correlation factor (r_{pb}) which serves as a communication measure between dichotomizing and quantitative variables. Some other factors, in particular tetrachoric (r_{tet}) and biserial correlation factor (r_{bis}), it is possible to interpret correlation factor as approximation for certain types of signs [5]. The more correlation factor on the module, the more strongly linear dependence. Value of factor of correlation lies between -1 and +1. If the tendency of increase of one size is observed at growth another speak about positive correlated sizes if the tendency of increase in one size is observed at reduction another speak about negative correlated sizes.

We will exclude from consideration parameters which unequivocally do not depend from each other. The greatest interest represents calculation and the correlation analysis between dependent (m3-m7) and independent (m1 - m2, m8-m20) parameters. It is possible to consider insignificant factors as zero and to take into consideration only the significant. We will reject insignificant parameters :quantity of installations and infringement of parameters of operation. If the correlation factor is calculated on the basis of the selective data it is not excluded that its nonzero value is not reflexion of the valid communication between signs, and is simply received as a result of specificity of the given sample (whereas the correlation factor is equal in a general totality to zero, i.e. linear communication between signs is not present).

Thus, it is necessary to understand, how r value is far from zero. If value of factor of correlation is calculated under the selective data, for an estimation of its value in a general totality as usually, the confidential interval is under construction.

For construction of a confidential interval the error of factor of correlation r under the formula is calculated:

$$\sigma_r = \frac{\sqrt{1-r^2}}{\sqrt{n-2}}$$

Then correlation coefficient error is multiplied by parameter t , which depends on probability P (for probability 95% $t = 2$), in order to found maximal error.

At last, the confidential interval $r - \sigma_r$ is under construction and check is spent, whether zero value will get to this interval. If the zero does not get to a confidential interval, means, with high probability in a general totality there can not be a zero value of factor of correlation, i.e. communication between signs exists and in a general totality. In that case the correlation factor is statistically significant.



Figure 2. Confidential interval excluding zero

If the zero gets to a confidential interval, means, with high probability in a general totality there can be a zero correlation, i.e. absence of communication. In that case the correlation factor is statistically insignificant.

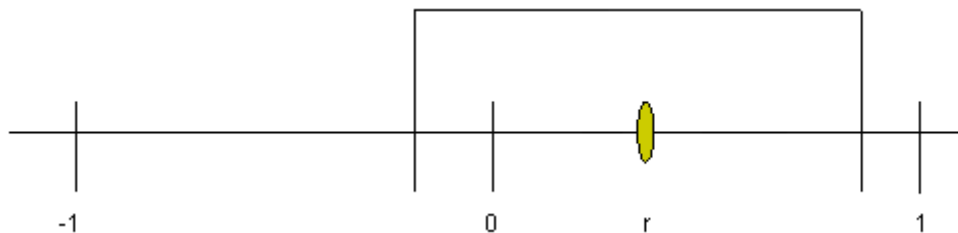


Figure 3. Confidential interval including zero

Conclusions

On the basis of the received results it is possible to draw following conclusions:

- The quantity of refusals depends on parameters:
 - Service life (-)
 - Operating time (+)
 - Qualifications of the engineer (-)
 - Novelty (-)
 - Reaction Time (+)
 - Performance Time (+)
 - Quantities of temperature parameters infringement (+)
 - Quantities of pressure parameters infringement (+)
 - Quantities of the unit parameters infringement (+)
- Quantity of the delayed demands depends on :
 - Service life (-)
 - Operating time (+)
 - Novelty (-)
 - Load (-)
 - Quantities of engineers in region (+)
 - Quantities of cash dispenses in region (+)
 - Quantities temperature parameters infringement (+)
 - Quantities of pressure parameters infringement (+)
- Time of ATM repair most depends on following parameters:
 - Quantities of cash dispenses in region, where ATM is located (+)
 - Quantities of engineers in region (+)
 - Load(-)
- Average time of malfunction depends on:
 - Service life (+)
 - Operating time (-)
 - Qualifications of the engineer (-)
 - Novelty (+)
 - Loading (-)
 - Quantities of engineers in region (+)

- Quantities of cash dispenses in region (+)
- Quantity of spare parts replacements depends on:
 - Service life (-)
 - Operating time (+)
 - Qualifications of the engineer (-)
 - Novelty (-)
 - Loading (-)
 - Quantities of engineers in region (-)
 - Quantities of cash dispenses in region (-)
 - Quantities temperature parameters infringement (+)
 - Quantities of pressure parameters infringement (+)
 - Quantities of grounding parameters infringement(+)

The feature near to each parameter testifies about sizes correlation – positive and negative.

Provided information allows greatly reducing losses due to network processing and estimating the common plan of ATM service works

Authors' Information

Oleksandr Kuzomin – Head of Innovation marketing department, professor in informatics dept.; Kharkiv National University of Radio electronics, e-mail: kuzy@kture.kharkov.ua

Nataliya Goloviy – Postgraduate, Informatics dept.; Kharkiv National University of Radio electronics, e-mail: rica1982@mail.ru

Yaser Dayoub – Postgraduate, Informatics dept.; Kharkiv National University of Radio electronics, e-mail: yaser_daeob@hotmail.com

EMERGENCY PREVENTION AND SITUATION MONITORING IN CARPATHIAN REGION

Oleksandr Kuzomin, Dmytro Malytsky, Asanbek Toroev

Abstract: Carpathian region is well known as tectonically active zone. So, in addition to common problems of such region, as common water floods, possible mudflows and landslides a local seismic activity must be taken in account. In this paper a main points of situation monitoring in Carpathian region and ways how they help in emergency prevention are described. A short overview of existing solutions and future approach is being made.

Keywords: emergency prevention, situation monitoring, early warning systems

ACM Classification Keywords: H.3.4 Systems and Software

Introduction and current situation overview

Carpathian region extending over the Folded Carpathians, foremountain and intermountain troughs and the adjacent parts of the platform, i.e. the parts of Transcarpathian, Chernivtsi, Ivano-Frankivsk and Lviv regions of Ukraine, is known as tectonically active zone. Its modern activity is attributed to processes forming the Carpathians, Transcarpathian and Forecarpathian Troughs and manifests itself through local seismicity.

Maximum intensity of earthquakes in the Carpathian region reaches 7⁰-8⁰ in the MSK scale (Tyachiv, 1781, 1870, MSK I₀ = 6⁰-7⁰; Rokoshyno, 1797, MSK I₀ = 6⁰-7⁰; Dovhe, 1872, MSK I₀ = 7⁰; Svalyava, 1908, MSK I₀=7⁰; Teresva, 1926, MSK I₀ = 6⁰-7⁰; Drahovo, 1937, MSK I₀ = 7⁰; Berehove, 1965, MSK I₀ = 7⁰; Zalishchyky, 1832, MSK I₀ = 7⁰; 1835, MSK I₀ = 7⁰-8⁰; Velyki Mosty, 1875, MSK I₀ = 6⁰-7⁰). That constitutes a serious danger to population, civil and industrial installations located here.

According to the present map of seismic hazard there are two zones in the Transcarpathian region – Mukachiv and Tyachiv – where the risk is considerably high.

Earthquakes with intensity of up to 8⁰ in the MSK scale and energy class K=14 (K=lgE) recur almost every 160 years in the Mukachiv zone. The most recent MSK intensity 7⁰ earthquake occurred to the South of Svalyava on January the 5th, 1908, and intensity 6⁰ earthquake – on October 24th, 1965 (near Berehove). There have been no significant tremors in the Mukachiv zone since a few decades what may indicate the increasing probability of strong earthquake in the zone in the nearest future.

The Tyachiv zone is remarkable for more frequent but of considerably lower intensity earthquakes.

In 1965 Lviv geophysicists organized the Carpathian Integrated Geodynamic Test Site in the Transcarpathian region to study connections between geophysical fields, structure and dynamics of lithosphere and local seismicity. In the site the network of 6 permanent geophysical stations (PGS) ("Mukachiv", "Berehove", "Trosnyk", "Nyzhnye Selyshche", "Bushtyno", and "Brid") was set up. Geophysical station "Ivano-Frankove" was located in the Lviv region. Continuous registration of the local earthquakes, microseisms, crustal deformations, borehole temperature and temporal changes of a broad set of geophysical fields is held at PGS. Integrated analysis of data series observed within the period of 1991-1995 shows a possibility of predicting the total seismic energy release in the region within the nearest 2-3 months.

There are a number of existing stations, specified on map as 1 including:

MUK – Mukachiv. Observations: seismological, meteorological, electromagnetic emission (em pulses from rock massifs);

BRD – Brid. Observations: seismological, geomagnetic (geomagnetic field modulus);

NSL – Nyzhnye Selyshche. Observations: seismological, geomagnetic, multicomponent magnetovariational, meteorological;

BST – Bushtyno. Observations: seismological;

TRS – Trosnyk. Observations: seismological, geomagnetic, meteorological, borehole acoustoemissive and geotemperature;

KOR – Koroleve, TRS division. Observations: seismological, extensometric, meteorological;

BRG – Berehove. Observations: seismological, geomagnetic, extensometric, tiltmetric, meteorological.

Seismic Stations:

UZH – Uzhgorod;

MEZ – Mizhgyrya.

Also three new stations, mostly in mountains are planned:

DOB – Dobron;

UKL – Uklin;

VBR – Velykyi Bereznyi.

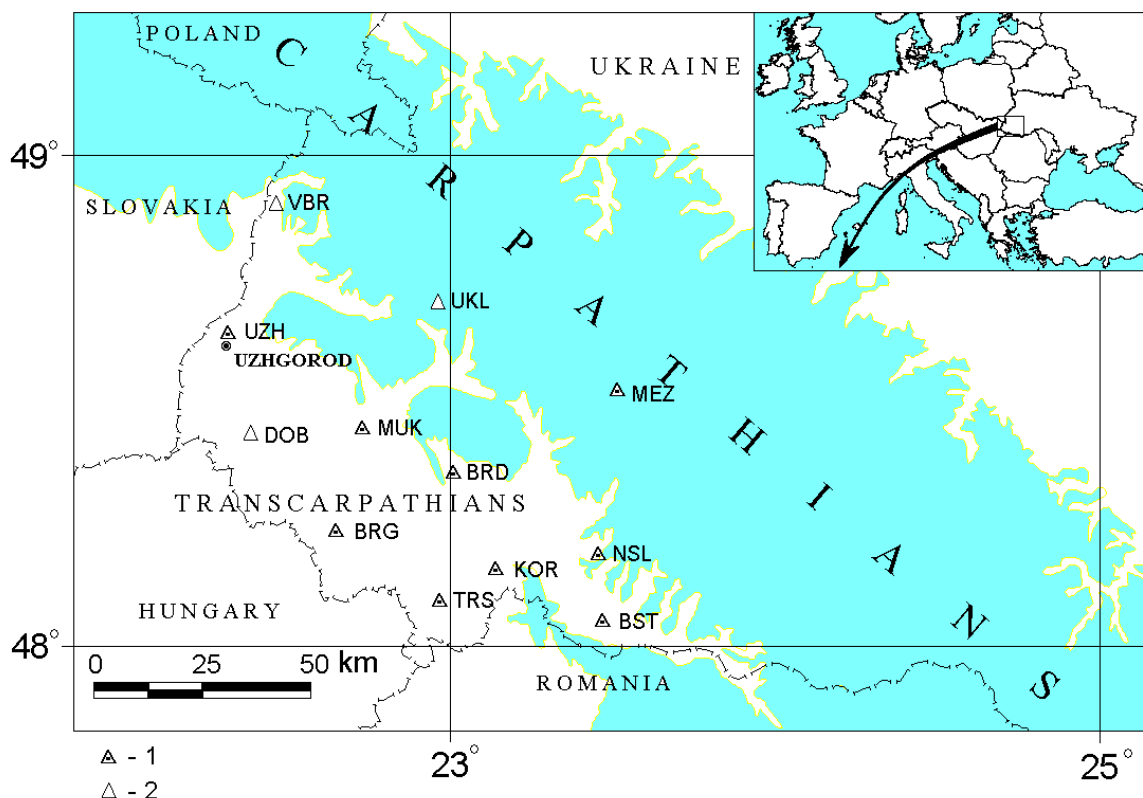


Fig. 1. Carpathian Integrated Geodynamic Test Site

Main task of the system

Gas blowing, coal mining, mudflows and landslides, tectonic activity and other emergencies greatly influence on rock arrays and result in rearranging tensions and deformations in them. Depending on scale of this process it can result in physical characteristics change of rock arrays and in generation of acoustic and seismic impulses.

Having a possibility to control such changes in time and space (especially parameters of seismic waves and acoustic emission) allows implementing a totally new ability of emergency prognosis and prevention (so called "early warning system"). By monitoring parameters of such waves we can precisely estimate physical parameters of rock arrays and their current state using non-intrusive, and so, way more cheap, methods

The primary task of the project – is development of theoretical and practical approaches for seismic and geophysical monitoring of earth's crust high layers, which are heavily affected by technogenic factors and estimation of possible emergency on that base.

The primary steps for this task are:

1. Analyze and develop mathematical model of seismic waves sources, which are created in zones with high tension concentrations (including earthquakes and acoustic emission);
2. Perform a model selection for such source, which maximally close describes the process of seismic waves creation (an analysis of other geophysical and geological fields may be required);
3. Develop the algorithm of seismic signals calculation based on selected model;
4. Estimate the influence of tension change in rock arrays on seismic signals from that array;
5. Analyze possible wave effects which may appear during seismic wave passing rock arrays with micro fissures.
6. Develop a set of recommendations for seismic and geophysical emergency monitoring development
7. Implement the following recommendation on practice
8. Estimate effectiveness of developed system and implement learning processes in it.

Proposed method

The main idea of proposed method is to establish a **constant network of monitoring devices**, based on existing and newly created stations, which will provide a current state of environment to **knowledge base**, which will contain previous estimations, the emergencies happened and what actions were taken order to prevent or overcome them. In other words each time a sensor request is being made our situation can be described as the following set:

$$S = s_1 \cup s_2 \cup \dots \cup s_n$$

$$s_k = \langle (x, y, z), P \rangle, k = \overline{1..n}$$

where S is a general situation composed from situations $s_1 \dots s_n$ measured on stations. Each that "small" situation can be described as a subset of coordinates where measurement was made, and current parameters P of sensors on that station. Proposed technology of the prognostication of emergency danger is the information complex, which consists of three basic blocks:

1. **knowledge base**- is intended for collection, storage and initial processing of the data of **seismic** and **weather** services, provided sensors information from stations, which contain information about tensions modifying in surface and rock arrays;
2. Mathematical and algorithmic **guarantee** - is the collection of mathematical methods and approaches, on base of which is produced the simulation and the prognostication of emergencies danger. The

prognostication of emergency situation is characterized by four basic parameters: by place, by type, by time and with its degree of power. Each of the characteristics has available their mathematical, algorithmic and program apparatuses.

3. block of results assignment for different levels of users - the obtained forecasts are analyzed by experts, after which they are transferred for modification to the system of decision making support for the purpose of use with the correction of anti-emergency measures and to the elimination of the consequences of gathering emergency.

The structure of proposed solution is showed on figure 2.

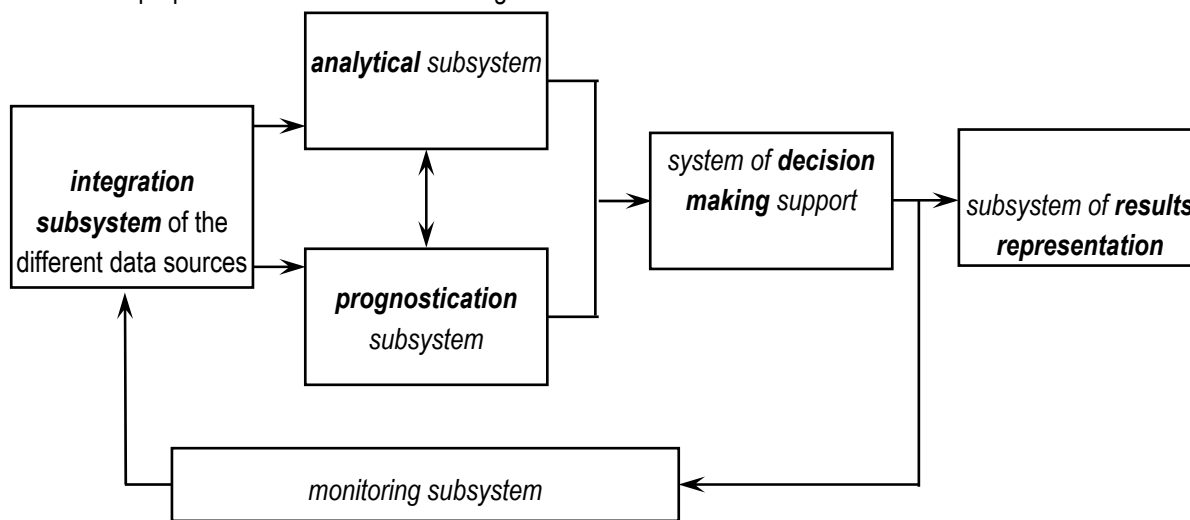


Fig. 2. Proposed system structure.

The system is easily extensible and allows a quickly reaction on all kinds of emergencies including technological situations. The main point of making system working is implementation of knowledge base, which will store all sets of different situations, occurred during last 10 years. This task is required to be solved on government level in order to create integral structure for easy storage, access, and managing such data.

Conclusions

Proposed system demonstrates a possibility of effective emergency situation management in Carpathian region. This involves all steps of early warning system development including technical means (set of monitoring stations), program solutions (knowledge base implementation) and mathematical and algorithmic approaches for such system modeling.

Authors' Information

Oleksandr Kuzomin – Head of Innovation marketing department, professor in informatics dept.; Kharkiv National University of Radio electronics, e-mail: kuzy@kture.kharkov.ua

Dmytro Malytskyy – Head of seismic and tectonical researches dept., Institute of Geo Physics of Ukraine National Academy of Science, e-mail: dmytro@cb-igph.lviv.ua

Asanbek Toroev – director of joint-stock company “Computational techniques and automation systems”

ИНДИВИДУАЛЬНО-ОПТИМАЛЬНЫЕ РАВНОВЕСИЯ В ИГРАХ ДВУХ ЛИЦ

Сергей Мащенко

Аннотация: Рассматривается принцип индивидуальной оптимальности, который представляет собой обобщение принципов оптимальности по Нешу, Берже и Парето. На его основе проводится исследование стабильности различных типов равновесий в играх двух лиц.

Keywords: равновесие по Нешу, равновесие по Берже, стабильность равновесий, индивидуально-оптимальные равновесия.

ACM Classification Keywords: H4.2 Decision support

Введение

Индивидуально-оптимальные равновесия представляют определенный интерес как новый принцип оптимальности в некооперативных играх, позволяющий построить общую модель, в которой каждый тип равновесия характеризуется уровнем его стабильности [Мащенко, 2007, 1]. В статьях [Мащенко, 2006, 2007, 1] рассматривается концепция индивидуальной оптимальности для произвольных игр n лиц, в статьях [Мащенко, 2007, 1; 2008] она конкретизируется для игр с вогнутыми и дифференцируемыми функциями выигрыша игроков. В данной работе рассматриваются условия индивидуальной оптимальности и оценки стабильности равновесий в играх двух лиц.

Индивидуально-оптимальные равновесия

Рассмотрим игру двух лиц DG в нормальной форме $DG = \langle X_1, X_2, u_1, u_2 \rangle$, где X_1, X_2 - множества стратегий игроков; $u_1(x), u_2(x)$ - функция их выигрыша, которые определены на множестве ситуаций игры $X = X_1 \times X_2$ и принимают действительные значения. Каждый из игроков стремится получить по возможности большее значения своей функции выигрыша.

Для определения равновесных ситуаций и исследования их стабильности будем использовать принцип индивидуальной оптимальности [Мащенко, 2007, 1], в соответствии с которым каждый игрок выбирает свои стратегии индивидуально, но учитывает функции выигрыша остальных игроков. Принцип индивидуальной оптимальности базируется на специальном отношении доминирования по Нешу.

Будем говорить, что ситуация $y \in X$ игры DG с вектором выигрыша $U(y) = (u_1(y), u_2(y))$ сильно доминирует по Нешу [Мащенко, 2007, 1] ситуацию x с вектором выигрыша $U(x)$, и обозначим это через $y \succ^{\text{NE}} x$, если ситуация y получена из ситуации x изменением каким-то одним игроком $i = 1, 2$ своей стратегии, т. е. $y = (y_i, x_j), j = 1, 2; j \neq i$, и $U(y) \succ U(x)$ ($U(y) \succ U(x) \Leftrightarrow u_i(y) > u_i(x), i = 1, 2$).

Ситуация x^* называется слабым индивидуально-оптимальным равновесием (множество этих равновесий обозначим через $WIOE(DG)$), если не существует другой ситуации, которая сильно доминировала бы по

Нешу x^* , т. е. $x^* \in WIOE(DG) \Leftrightarrow \neg \exists x \in X : x \succ^{\text{NE}} x^*$. Таким образом, можно сказать, что в индивидуально-оптимальном равновесии каждый игрок находит такой компромисс с остальными игроками, который не выгодно никому нарушать.

Среди основных свойств индивидуально оптимальных равновесий следует отметить следующие:

- если множество X ситуаций игры является непустым компактом, а функции выигрыша игроков - непрерывны (или - в случае конечной игры), всегда существуют индивидуально-рациональные (значения выигрышей игроков не менее гарантированных) и коллективно-рациональные (оптимальные по Парето) индивидуально-оптимальные равновесия [Мащенко,2007,2];
- множество индивидуально-оптимальных равновесий включает в себя [Мащенко,2007,1;2007,2]:

$$NE(DG) = \{x^* \in X \mid u_i(x^*) \geq u_i(x_i, x_j^*), \forall x_i \in X_i; i, j = 1, 2; j \neq i\} - \text{множество равновесий Неша};$$

$$BE(DG) = \{x^* \in X \mid u_i(x^*) \geq u_i(x_i^*, x_j), \forall x_j \in X_j; i, j = 1, 2; j \neq i\} - \text{множество равновесий Берже};$$

$$SO(DG) = \{x^* \in X \mid \neg \exists U(x) \succ \succ U(x^*), \forall x \in X\} - \text{множество слабо оптимальных по Парето (оптимальных по Слейтеру) ситуаций игры } DG.$$

Необходимые и достаточные условия индивидуальной оптимальности для игр n лиц в общем случае были рассмотрены в [Мащенко,2006;2007,1], для игр с вогнутыми и дифференцируемыми функциями выигрыша – в [Мащенко,2007,2;2008].

Условия индивидуальной оптимальности

Обозначим через $S(x^*) = \max_{i,j=1,2;j \neq i} \sup_{x_j \in X_j} (u_i(x_i, x_j^*) + u_2(x_j, x_i^*))$ верхнюю границу суммы выигрышей игроков в

ситуациях, которые получаются из x^* изменением игроками $i = 1, 2$, в отдельности, своих стратегий.

Теорема 1. Если ситуация $x^* \in WIOE(DG)$ и, без ограничения общности, $u_i(x^*) > 0, i = 1, 2$, то всегда существуют такие $\mu_i \in [0, S(x^*)], i = 1, 2$, в частности:

$$\mu_i^* = u_i(x^*) + (S(x^*) - u_i(x^*) - u_2(x^*)) / 2, i = 1, 2 \quad (1)$$

что для $\forall i, j = 1, 2; j \neq i$, и для $\forall y_i \in X_i$ справедливы неравенства:

$$\min(u_i(x) - \mu_i, u_j(x) - S(x^*) + \mu_i) \geq \min(u_i(y_i, x_j) - \mu_i, u_j(y_i, x_j) - S(x^*) + \mu_i). \quad (2)$$

Любое решение x^* системы неравенств (2) при заданных $\mu_i \in [0, S(x^*)], i = 1, 2$, является слабым индивидуально-оптимальным равновесием.

Доказательство. Докажем достаточность. Пусть x^* удовлетворяет неравенствам (2). Отсюда следует, что для $\forall i, j = 1, 2; j \neq i$, и для $\forall y_i \in X_i$ выполняются неравенства:

$$\begin{cases} u_i(x^*) - \mu_i \geq \min(u_i(y_i, x_j^*) - \mu_i, u_j(y_i, x_j^*) - S(x^*) + \mu_i), \\ u_j(x^*) - S(x^*) + \mu_i \geq \min(u_i(y_i, x_j^*) - \mu_i, u_j(y_i, x_j^*) - S(x^*) + \mu_i). \end{cases}$$

Тогда, если $u_i(y_i, x_j^*) - \mu_i \leq u_j(y_i, x_j^*) - S(x^*) + \mu_i$, то получим $u_i(y_i, x_j^*) \leq u_i(x^*)$. В противном случае,

выполнится $u_j(y_i, x_j^*) \leq u_j(x^*)$. Таким образом, $\neg \exists x \in X : x \succ \succ^{NE} x^*$, откуда следует $x^* \in WIOE(G)$.

Докажем необходимость. Пусть $x^* \in WIOE(G)$, тогда $\neg \exists x \in X : x \succ \succ^{NE} x^*$, поэтому для $\forall i, j = 1, 2; j \neq i$, и для $\forall y_i \in X_i$ выполняется по крайней мере одно из неравенств: $u_i(x^*) \geq u_i(y_i, x_j^*)$ или $u_j(x^*) \geq u_j(y_i, x_j^*)$. Определим μ_1, μ_2 по формулам (1). Несложно убедиться, что $\mu_i \in [0, S(x^*)], i = 1, 2$.

Из первого неравенства следует $u_i(y_i, x_j^*) - \mu_i^* \leq u_i(x^*) - \mu_i^* = -(S(x^*) - u_1(x^*) - u_2(x^*)) / 2$. А из второго неравенства очевидно следует, что $u_j(y_i, x_j^*) - S(x^*) + \mu_i \leq u_j(x^*) - S(x^*) + \mu_i = u_j(x^*) - S(x^*) + u_i(x^*) + (S(x^*) - u_1(x^*) - u_2(x^*)) / 2 = -(S(x^*) - u_1(x^*) - u_2(x^*)) / 2$. Поскольку правые части этих неравенств всегда равны между собой, то $\min(u_i(y_i, x_j) - \mu_i, u_j(y_i, x_j) - S(x^*) + \mu_i) \leq -(S(x^*) - u_1(x^*) - u_2(x^*)) / 2 = \min(u_i(x) - \mu_i, u_j(x) - S(x^*) + \mu_i), j = 1, 2; j \neq i$. Отсюда следует, что ситуация x^* удовлетворяет неравенствам (2). ♦

Таким образом, теорема 1 дает возможность конструктивно описать множество слабых индивидуально-оптимальных равновесий $WIOE(DG)$.

Следует отметить, что параметры μ_1, μ_2 , которые фигурируют в неравенствах (2), имеют определенный игровой смысл. Пусть x^* является слабым индивидуально-оптимальным равновесием игры DG . Тогда, по определению, ситуация x^* представляет собой для каждого игрока $i = 1, 2$ такой компромисс между желанием максимизировать свою собственную функцию выигрыша и функцию выигрыша другого игрока, от которого ему не выгодно отклоняться. Поскольку, на основании теоремы 1, ситуация x^* является решением системы неравенств (2), по крайней мере при $\mu_i^* = u_i(x^*) + (S(x^*) - u_1(x^*) - u_2(x^*)) / 2, i = 1, 2$, то легко заметить, что большему выигрышу игрока i в компромиссе соответствует большее значение параметра μ_i^* . С другой стороны, из системы неравенств (2) видно: чем больше $\mu_i \in [0, S(x^*)]$, тем большее предпочтение i -й игрок отдает собственной функции выигрыша в достигнутом компромиссе x^* , и, соответственно, на больший выигрыш он может рассчитывать.

Для индивидуально-оптимального равновесия x^* обозначим через $M_i(x^*)$ - множество значений параметра $\mu_i \in [0, S(x^*)]$, при котором x^* удовлетворяет системе неравенств (2). Отметим, что из теоремы 1 следует $M_i(x^*) \neq \emptyset, i = 1, 2$.

Лемма. Система неравенств (2) эквивалентна:

$$\begin{cases} \min(u_i(y_i, x_j^*) - u_i(x^*), u_j(y_i, x_j^*) - u_i(x^*) - S(x^*) + 2\mu_i^*) \leq 0, \\ \min(u_j(y_i, x_j^*) - u_j(x^*), u_i(y_i, x_j^*) - u_j(x^*) + S(x^*) - 2\mu_i^*) \leq 0. \end{cases} \quad (3)$$

$$\begin{cases} \min(u_i(y_i, x_j^*) - u_i(x^*), u_j(y_i, x_j^*) - u_i(x^*) - S(x^*) + 2\mu_i^*) \leq 0, \\ \min(u_j(y_i, x_j^*) - u_j(x^*), u_i(y_i, x_j^*) - u_j(x^*) + S(x^*) - 2\mu_i^*) \leq 0. \end{cases} \quad (4)$$

Доказательство. Несложно убедиться, что неравенства (2) будут иметь место тогда и только тогда, когда

$$\begin{cases} u_i(x^*) - \mu_i^* \geq \min(u_i(y_i, x_j^*) - \mu_i^*, u_j(y_i, x_j^*) - S(x^*) + \mu_i^*), \\ u_j(x^*) - S + \mu_i^* \geq \min(u_i(y_i, x_j^*) - \mu_i^*, u_j(y_i, x_j^*) - S(x^*) + \mu_i^*), \end{cases}$$

что эквивалентно (3), (4). ♦

Приведенные выше рассуждения о смысле параметров μ_1, μ_2 развивает далее следующая теорема.

Теорема 2. Справедливы следующие утверждения:

$$x^* \in NE(DG) \Leftrightarrow \exists \mu_i^* \in M_i(x^*) : \mu_i^* = S(x^*), i = 1, 2, \quad (5)$$

$$x^* \in SO(DG) \Rightarrow \exists \mu_i^* \in M_i(x^*) : \mu_1^* + \mu_2^* = S(x^*), \quad (6)$$

$$x^* \in BE(DG) \Leftrightarrow \exists \mu_i^* \in M_i(x^*) : \mu_i^* = 0, i = 1, 2. \quad (7)$$

Доказательство. Сначала докажем отношение (5). Пусть $x^* \in NE(DG)$. Тогда $u_i(x^*) \geq u_i(y_i, x_j^*), \forall y_i \in X_i; i, j = 1, 2; j \neq i$. При таких условиях неравенства (3) будут справедливы при $\forall \mu_i \in [0, S(x^*)], i = 1, 2$, в том числе, и при $\mu_i^* = S(x^*), i = 1, 2$. Неравенства (4) выполняются при значениях $\mu_i^* = S(x^*), i = 1, 2$, т.к. $u_i(y_i, x_j^*) - u_j(x^*) + S(x^*) - 2\mu_i^* = u_i(y_i, x_j^*) - u_j(x^*) - S(x^*) \leq 0$ для $\forall y_i \in X_i; i, j = 1, 2; j \neq i$. Отсюда, на основании леммы, получим $S(x^*) \in M_i(x^*), i = 1, 2$.

Пусть $S(x^*) = \mu_i^* \in M_i(x^*), i = 1, 2$. Тогда, согласно лемме, при значениях $\mu_i^* = S(x^*), i = 1, 2$ должны выполняться неравенства (3), (4). Поскольку в (3) выражение $u_j(y_i, x_j^*) - u_i(x^*) - S(x^*) + 2\mu_i^* = u_j(y_i, x_j^*) - u_i(x^*) + S(x^*) > 0$ для $\forall y_i \in X_i; i, j = 1, 2; j \neq i$, то неравенства (3) будут эквивалентны $u_i(x^*) \geq u_i(y_i, x_j^*), \forall y_i \in X_i; i, j = 1, 2; j \neq i$. Следовательно $x^* \in NE(DG)$.

Докажем отношение (6). Пусть $x^* \in SO(DG)$. Тогда $\neg \exists x \in X : U(x) \succ U(x^*)$, т.е. $\neg \exists x \in X : u_i(x) > u_i(x^*), i = 1, 2$. В частности, не существует отклонения любого, но только одного, игрока $i = 1, 2$ от ситуации x^* , при котором в полученной ситуации $(y_i, x_j^*), j = 1, 2; j \neq i$, выполняются неравенства $u_1(y_i, x_j^*) > u_1(x^*)$ и $u_2(y_i, x_j^*) > u_2(x^*)$. Таким образом, $x^* \in WIOE(DG)$. Отсюда, согласно теореме 1, при $\mu_i^* = u_i(x^*) + (S(x^*) - u_1(x^*) - u_2(x^*)) / 2, i = 1, 2$, ситуация x^* удовлетворяет неравенствам (2). Тогда очевидно $\mu_1^* + \mu_2^* = S(x^*)$.

Докажем отношение (7). Пусть $x^* \in BE(DG)$. Тогда $u_i(x^*) \geq u_i(x_i^*, y_j), \forall y_j \in X_j; i, j = 1, 2; j \neq i$. При таких условиях неравенства (4) будут справедливы при $\forall \mu_i \in [0, S(x^*)], i = 1, 2$, в том числе, и при $\mu_i^* = 0, i = 1, 2$. Неравенства (3) выполняются при значениях $\mu_i^* = 0, i = 1, 2$, поскольку $u_j(y_i, x_j^*) - u_i(x^*) - S(x^*) + 2\mu_i^* = u_j(y_i, x_j^*) - u_i(x^*) - S(x^*) \leq 0$ для $\forall y_j \in X_j; i, j = 1, 2; j \neq i$. Отсюда, на основании леммы, получим $0 \in M_i(x^*), i = 1, 2$.

Пусть $0 = \mu_i^* \in M_i(x^*), i = 1, 2$. Тогда, на основании леммы, при значениях $\mu_i^* = 0, i = 1, 2$, должны выполняться неравенства (3), (4). Поскольку в (4) выражение $u_i(y_i, x_j^*) - u_j(x^*) + S(x^*) - 2\mu_i^* = u_j(y_i, x_j^*) - u_i(x^*) + S(x^*) > 0$ для $\forall y_i \in X_i; i, j = 1, 2; j \neq i$, то неравенства (4) будут эквивалентны $u_i(x^*) \geq u_i(x_i^*, y_j), \forall y_j \in X_j; i, j = 1, 2; j \neq i$. Следовательно $x^* \in BE(DG)$. ♦

На основании теоремы 2 можно сделать следующие выводы. Если ситуация x^* - равновесие Неша, которое не выгодно нарушать каждому из игроков в отдельности, то она стабильна для обоих игроков и является индивидуально-оптимальным равновесием с оценкой $(\mu_1 + \mu_2) / S(x^*) = 2$. Оптимальная по Парето ситуация, которая стабильна лишь для пары игроков, которые действуют сообща, может быть получена как индивидуально-оптимальное равновесие при $(\mu_1 + \mu_2) / S(x^*) = 1$, а равновесие Берже, которое нестабильно для каждого игрока, будет иметь наименьшую оценку $(\mu_1 + \mu_2) / S(x^*) = 0$. Остальные индивидуально-оптимальные равновесия имеют промежуточные значения величины $(\mu_1 + \mu_2) / S(x^*)$ на интервале $[0, 2]$ в зависимости от уровня толерантности игроков.

На основании приведенных выше выводов можно выдвинуть гипотезу, что желание каждого игрока увеличить свой индивидуальный выигрыш в достигнутом компромиссе, может быть оценено величиной

$(\mu_1 + \mu_2)/S(x^*)$, которая характеризует стабильность индивидуально-оптимального равновесия. Поскольку индивидуально-оптимальное равновесие может быть одновременно и равновесием Неша, и Берже, и Парето-оптимальной ситуацией, то целесообразно оценить как максимальный, так и минимальный уровни его стабильности.

Максимальный уровень стабильности индивидуально-оптимального равновесия представляет собой агрегированную характеристику того, насколько бескомпромиссными могут быть игроки для удержания равновесия. С другой стороны, минимальный уровень стабильности - это агрегированная характеристика того, насколько толерантным должны быть игроки для удержания равновесия. Эти предельные характеристики могут дать ценную информацию для более глубокого исследования ситуации равновесия. Так, например, если некоторое индивидуально-оптимальное равновесие является равновесием Неша и равновесием Берже (один из очень интересных и особенно стабильных типов равновесия), то согласно теореме 2, величина $\max(\mu_1 + \mu_2)/S(x^*) = 2$, а $\min(\mu_1 + \mu_2)/S(x^*) = 0$.

Оценка стабильности равновесий

Для нахождения максимального (минимального) уровня стабильности индивидуально-оптимального равновесия x , который мы обозначим через $\hat{\mu}^{\max}(x)$ (обозначим через $\hat{\mu}^{\min}(x)$), сформулируем оптимизационную задачу относительно $\mu_i, i = 1, 2$, в которой ситуация x фигурирует как параметр:

$$\hat{\mu}^{\max}(x) = \max(\mu_1 + \mu_2)/S(x) \quad (\hat{\mu}^{\min}(x) = \min(\mu_1 + \mu_2)/S(x)), \quad (8)$$

$$\min(u_i(x) - \mu_i, u_j(x) - S(x) + \mu_i) \geq \min(u_i(y_i, x_j) - \mu_i, u_j(y_i, x_j) - S(x) + \mu_i), j = 1, 2; j \neq i \quad (9)$$

$$\mu_i \in [0, S(x^*)], i = 1, 2. \quad (10)$$

Функцию $\hat{\mu}^{\max} : WIOE(G) \rightarrow [0, n]$ ($\hat{\mu}^{\min} : WIOE(G) \rightarrow [0, n]$) назовем критерием максимальной (минимальной) стабильности индивидуально-оптимального равновесия x игры DG . Ограничение (9), (10) описывают, согласно теореме 1, условия слабой индивидуальной оптимальности ситуации $x \in X$. Поскольку заранее, как правило, не известно, какая ситуация игры является индивидуально-оптимальным равновесием игры DG , то ситуации, которые не индивидуально-оптимальны, будут приводить к несовместности ограничений (9), (10). Если считать, что при несовместной системе ограничений для $k = 1, 2$ игроков в ситуации $x \in X$, функции $\mu^{\max}(x) = -k$ и $\mu^{\min}(x) = -k$, а при совместной системе ограничений (9), (10) $\mu^{\max}(x) = \hat{\mu}^{\max}(x)$ и $\mu^{\min}(x) = \hat{\mu}^{\min}(x)$, то область определения функций $\mu^{\max}(x)$ и $\mu^{\min}(x)$ будет расширена до всего множества ситуаций игры. Следует отметить, что конструктивно реализовать такое расширение области определения функций можно посредством аппарата штрафных функций. Таким образом, $\mu^{\max} : X \rightarrow [-2, 2]$ и $\mu^{\min} : X \rightarrow [-2, 2]$ будут соответственно критериями максимальной и минимальной стабильности ситуаций $x \in X$.

Введенные нами критерии максимальной и минимальной стабильности ситуаций игры DG являются агрегированными характеристиками их стабильности для обоих игроков в целом, что предусматривает коллективную процедуру их применения. Поскольку при некооперативном поведении игроков возможности любых коллективных процедур могут быть очень ограниченными, то стоит использовать индивидуальные оценки стабильности ситуаций игры.

Обозначим через $D_i(x)$, $i \in N$, множества допустимых решений задач (8) (10). Обратим внимание на то, что при фиксированном $x \in WIOE(G)$ задачи (8) - (10) естественно декомпозируются на пары независимых задач математического программирования:

$$\hat{\mu}_i^{\max}(x) = \max\{\mu_i/S_i(x) \mid \mu_i \in D_i(x)\}, i=1,2, \quad \hat{\mu}_i^{\min}(x) = \min\{\mu_i/S_i(x) \mid \mu_i \in D_i(x)\}, i=1,2.$$

Расширяя область определения функций $\hat{\mu}_i^{\max}(x)$, $\hat{\mu}_i^{\min}(x)$ получим для $i=1,2$:

$$\mu_i^{\max}(x) = \begin{cases} \hat{\mu}_i^{\max}(x), & D_i(x) \neq \emptyset, \\ -1, & D_i(x) = \emptyset, \end{cases} \quad \mu_i^{\min}(x) = \begin{cases} \hat{\mu}_i^{\min}(x), & D_i(x) \neq \emptyset, \\ -1, & D_i(x) = \emptyset. \end{cases} \quad (11)$$

Если функции $\mu^{\max}(x)$ и $\mu^{\min}(x)$ представляют собой суммарные оценки стабильности ситуаций, то функции $\mu_i^{\max}(x): X \rightarrow [-1,1]$ и $\mu_i^{\min}(x): X \rightarrow [-1,1]$ могут интерпретироваться как соответственно максимальная и минимальная индивидуальные оценки стабильности отдельно для игрока $i=1,2$.

Для общей игры n лиц задачи поиска индивидуальных оценок стабильности типа (11) достаточно сложны [Мащенко, 2007,1]. В игре двух лиц появляется возможность их решения в аналитическом виде.

Теорема 3. Пусть, без ограничения общности, $u_i(x) > 0, i=1,2; \forall x \in X$. Тогда максимальная и минимальная оценки стабильности ситуации $x \in X$ игры DG определяются соответственно:

$$\mu_i^{\max}(x) = \begin{cases} 1, & u_i(x) \geq u_i(y_i, x_j), \forall y_i \in X_i, \\ \frac{1}{2} \left(1 + \min_{y_i \in X_i} \{u_i(x) - u_j(y_i, x_j) \mid u_i(x) < u_i(y_i, x_j)\} / S(x) \right), & u_i(x) \geq u_j(y_i, x_j), \\ -1, & \exists y_i \in X_i : u_i(x) < u_i(y_i, x_j), u_j(x) < u_j(y_i, x_j), \end{cases} \quad (12)$$

$$\mu_i^{\min}(x) = \begin{cases} 0, & u_j(x) \geq u_j(y_i, x_j), \forall y_i \in X_i, \\ \frac{1}{2} \left(1 - \min_{y_i \in X_i} \{u_j(x) - u_i(y_i, x_j) \mid u_j(x) < u_j(y_i, x_j)\} / S(x) \right), & u_i(x) \geq u_i(y_i, x_j), \\ -1, & \exists y_i \in X_i : u_i(x) < u_i(y_i, x_j), u_j(x) < u_j(y_i, x_j), \end{cases} \quad (13)$$

где $i, j = 1,2; j \neq i$.

Доказательство. Рассмотрим ситуацию $x \in X$. Согласно теореме 1 $x \in WIOE(DG)$ тогда и только тогда, когда существуют такие значения параметров $\mu_i \in [0, S(x)], i=1,2$, что для $\forall i=1,2$ и для $\forall y_i \in X_i$, имеют место неравенства (9). Они, в свою очередь, согласно лемме эквивалентны неравенствам (3), (4).

Рассмотрим для $\forall i=1,2$ следующие возможные случаи.

Пусть $\exists y_i \in X_i : u_i(x) < u_i(y_i, x_j), u_j(x) < u_j(y_i, x_j)$, где $j=1,2; j \neq i$, тогда $(y_i, x_j) \gg^{\text{NE}} x$, поэтому ситуация $x \notin WIOE(DG)$ и множество допустимых решений оценочных задач (11) $D_i(x) = \emptyset$. Расширяя область определения функций $\hat{\mu}_i^{\max}(x)$, $\hat{\mu}_i^{\min}(x)$ получим $\mu_i^{\max}(x) = -1, \mu_i^{\min}(x) = -1$.

Пусть для некоторого фиксированного $y_i \in X_i$ имеет место неравенство $u_i(x) \geq u_i(y_i, x_j)$, где $j=1,2; j \neq i$. Тогда неравенство (4) будет справедливым для любых $\mu_i \in [0, S(x)]$. Поэтому система неравенств (3), (4) будет эквивалентна (3). В свою очередь, неравенство (3) будет справедливым для любых $\mu_i \in [0, S(x)]$ при условии $u_i(x) \geq u_i(y_i, x_j)$ и будет эквивалентным неравенству $\mu_i \geq (S(x) + u_j(x) - u_i(y_i, x_j))/2$. Таким образом, если $u_i(x) \geq u_i(y_i, x_j)$ для $\forall y_i \in X_i$, то система

неравенств (3), (4), при условии $\mu_i \in [0, S(x)]$, будет эквивалентна неравенству $\mu_i \leq S(x)$. Если $u_i(x) < u_i(y_i, x_j)$ для некоторых $y_i \in X_i$, то система неравенств (3), (4) будет эквивалентна $\mu_i \leq \frac{1}{2} \left(S(x) + \min_{y_i \in X_i} \{ (u_i(x) - u_j(y_i, x_j)) | u_i(x) < u_i(y_i, x_j) \} \right)$. Отсюда следует, что максимальная индивидуальная оценка (11) стабильности отдельно для игрока $i = 1, 2$ будет иметь вид (12).

Если для некоторого фиксированного $y_i \in X_i$ имеет место неравенство $u_j(x) \geq u_j(y_i, x_j)$, где $j = 1, 2; j \neq i$, тогда неравенство (3) будет справедливым для любых $\mu_i \in [0, S(x)]$. Поэтому система неравенств (3), (4) будет эквивалентна (4). В свою очередь, неравенство (4) будет справедливо для любых $\mu_i \in [0, S(x)]$ при условии $u_j(x) \geq u_j(y_i, x_j)$ и будет эквивалентно неравенству $\mu_i \geq (S(x) - u_j(x) + u_i(y_i, x_j)) / 2$. Таким образом, если $u_j(x) \geq u_j(y_i, x_j)$ для $\forall y_i \in X_i$, то система неравенств (3), (4), при условии $\mu_i \in [0, S(x)]$, будет эквивалентна неравенству $\mu_i \geq 0$. Если $u_j(x) < u_j(y_i, x_j)$ для некоторых $y_i \in X_i$, то система неравенств (3), (4) будет эквивалентна $\mu_i \geq \frac{1}{2} \left(S(x) - \min_{y_i \in X_i} \{ (u_j(x) - u_j(y_i, x_j)) | u_j(x) < u_j(y_i, x_j) \} \right)$. Отсюда следует, что минимальная индивидуальная оценка (11) стабильности ситуации $x \in X$ отдельно для игрока $i \in N$ будет иметь вид (13). Теорема доказана. ♦

Заключение

Принцип индивидуальной оптимальности обобщает классические принципы оптимальности в некооперативных играх и расширяет класс конфликтно разрешимых игр. Применение этого принципа обосновано, если конфликт между игроками не может быть разрешен согласно классическим принципам оптимальности и игроки соглашаются идти на компромисс ради его решения.

Литература

- [Мащенко, 2007, 1] Мащенко С.О. Исследование стабильности равновесий на основе принципа индивидуальной оптимальности // Кибернетика и системный анализ. - 2007. - 4. - С. 162-169.
- [Мащенко, 2006] Мащенко С. О. Слабкі індивідуально- оптимальні рівноваги // Вісник Київського університету. Серія: фіз.-мат. науки. - 2006. - 2. - С. 169-177.
- [Мащенко, 2007, 2] Мащенко С. О. Індивідуально-оптимальні рівноваги в некооперативних опуклих іграх // Вісник Київського університету. Серія: фіз.-мат. науки. - 2007. - 4. - С. 133-139.
- [Мащенко, 2008] Мащенко С. О. Локальні умови слабкої індивідуальної оптимальності рівноваг // Вісник Київського університету. Серія: фіз.-мат. науки. - 2008. - 1. - С. 127-136.

Информация об авторе

Мащенко Сергей Олегович – Доцент, кандидат физ.-мат. Наук, Киевский национальный университет имени Тараса Шевченко; Пр. Академика Глушкова, 6, Киев – 207, Украина; e-mail: msomail@yandex.ru

О КЛАССИФИКАЦИИ ПРИБЛИЖЕННЫХ МЕТОДОВ КОМБИНАТОРНОЙ ОПТИМИЗАЦИИ

Сергей Сиренко

Аннотация: В работе предлагается классификация приближенных методов комбинаторной оптимизации, которая обобщает и дополняет существующие подходы.

Ключевые слова: комбинаторная оптимизация, классификация, приближенные методы.

ACM Classification Keywords: G.1.6 [Numerical Analysis] Optimization – Stochastic programming, G.2.1 [Discrete Mathematics] Combinatorics – Combinatorial algorithms, I.2.8 [Artificial Intelligence]: Problem Solving, Control Methods, and Search – Heuristic methods.

Введение

Развитие методов комбинаторной оптимизации (КО) уже превысило уровень первичного накопления знаний и идет путем обобщения полученных результатов. Это проявляется, в частности, в увеличении внимания к исследованию существующих подходов, анализе их общих и отличных характеристик, а также концептуальных оснований позволяющих достичь повышения эффективности алгоритмов. Описание общих характеристик происходит путем формулирования определенных обобщенных поисковых процедур, конкретизацией отдельных блоков которых обозначаются классы приближенных алгоритмов. Выделение принципиальных отличий может происходить, например, путем классифицирования алгоритмов.

В ряде работ рассматриваются вопросы классификации приближенных алгоритмов КО в целом и отдельных их классов, в частности [Birattari et al., 2001; Blum and Roli, 2003; Laguna, 2002]. В данной работе предлагается обобщение и дополнение вышеупомянутых подходов к классификации. Предлагаемая классификация будет изложена, после короткого описания понятия задачи КО.

Определение задачи комбинаторной оптимизации

Приведем формальное определение задачи КО [Гуляницкий, 2008]. Пусть заданы $Y = \{1, \dots, m\}$, Z – дискретное пространство, φ – гомоморфизм, $\varphi: Y \rightarrow Z$, удовлетворяющий некоторой системе ограничений Ω .

Определение 1. Под комбинаторным объектом κ будем понимать триаду $\kappa = (\varphi, Z, \Omega)$.

Определение 2. Задачей КО, называется задача нахождения такого $x_* \in X$, что

$$x_* = \arg \min_{x \in D \subseteq X} f(x),$$

где X – пространство решений задачи, элементами которого являются комбинаторные объекты, $D \subseteq X$ – подпространство допустимых решений, определяемое некоторым предикатом Π , $f: X \rightarrow \mathbb{R}$ – заданная целевая функция задачи.

Классификация приближенных методов комбинаторной оптимизации

При классифицировании алгоритмов КО прежде всего следует разделять точные и приближенные методы. Первые владеют свойством гарантировано находить оптимальное решение. Однако, почти все практически важные задачи КО являются NP-полными [Пападимитриу и Стайглиц, 1985]: для них не известно ни одного точного алгоритма с полиномиальной сложностью. Кроме того, постановки задач часто бывают с данными, которые имеют определенные погрешности, что окончательно делает значительные вычислительные затраты для нахождения точного решения неоправданными. В противоположность точным методам, современные приближенные алгоритмы КО разрешают получать решения "высокого качества" за приемлемое (с практической точки зрения) время. По этим, а также по другим причинам, все больше внимания в научной литературе уделяется приближенным методам КО. За последние десятилетия предложены десятки различных подходов, обладающие теми или иными свойствами и характеристиками. Для выделения характерных отличных черт существующих приближенных методов КО и предложена данная классификация (см. Таблицу 1). В примерах, приведенных в Таблице 1, ссылки указаны для тех методов которые далее в тексте не упоминаются.

Таблица 1. Классификация приближенных методов комбинаторной оптимизации

Классы / Характеристики	Примеры
По принципу принятия решений	
Детерминированные	<i>Локальный поиск</i>
Недетерминированные (стохастические)	<i>Имитационный отжиг</i>
По структуре	
Простой алгоритм	<i>Локальный поиск</i>
Гибридный алгоритм	–
Метаэвристика	<i>Метод муравьиных колоний, меметический алгоритм</i>
Гибридная метаэвристика	–
По типу используемых пространств	
Последовательные (конструктивные)	<i>Эвристика "иди к ближайшему"</i>
Итерационные	Траекторные <i>Локальный поиск, G-алгоритм [Гуляницкий, 2006]</i>
	Популяционные <i>Метод муравьиных колоний, H-метод [Гуляницкий, 2008]</i>
По количеству используемых систем окрестностей	
Одна	<i>Имитационный отжиг</i>
Много	<i>Поиск в переменных окрестностях</i>
По типу целевой функции	
Статическая	<i>Метод муравьиных колоний, генетический алгоритм</i>
Динамическая	<i>Управляемый локальный поиск</i>
По типу траектории (для итерационных методов)	
Траекторно-непрерывные	<i>G-алгоритм, табу-поиск</i>
Траекторно-разрывные	GRASP

По использованию памяти	
Без использования памяти	<i>Локальный поиск, G-алгоритм, стандартный генетический алгоритм</i>
С кратковременной памятью	<i>Простой табу-поиск</i>
С долговременной памятью	<i>Динамический локальный поиск [Hoos and Stützle, 2005]</i>
С обоими видами памяти	<i>Табу-поиск с использованием критериев стремления</i>
По оценке точности	
С апостериорной оценкой точности	<i>Локальный поиск с известной нижней границей</i>
С априорной оценкой точности	<i>ε-приближенные алгоритмы</i>
Без оценки точности	–
Сходимость по решению	
Сходящиеся с определенной вероятностью	<i>Муравьиные алгоритмы класса $ACO_{bs, \tau_{min}(t)}$ [Dorigo and Blum, 2005]</i>
Не сходящиеся	<i>Стандартный генетический алгоритм</i>
Без оценки сходимости	<i>H-метод</i>
Сходимость по значению	
Сходящиеся с определенной вероятностью	<i>Муравьиные алгоритмы класса $ACO_{\tau_{min}}$ [Dorigo and Blum, 2005], G-алгоритм</i>
Не сходящиеся	<i>Стандартный генетический алгоритм</i>
Без оценки сходимости	<i>H-метод</i>

По принципу принятия решений. В приближенных методах, прежде всего, будем различать два принципа принятия решений – детерминированный и недетерминированный, т.е. отсутствие или наличие псевдослучайных аргументов в функции принятия решения соответственно. На данный момент, подавляющее большинство разработанных и применяемых на практике приближенных методов КО принадлежат именно к недетерминированным, что обусловлено более высокими показателями эффективности таких методов.

По структуре. Классифицирование по структуре, хотя и является важным, но, учитывая отсутствие формального определения метаэвристики, не дает возможности четко различить алгоритмы по принадлежности к простым или метаэвристическим. Например, алгоритм имитационного отжига [Kirkpatrick et al., 1983] разные авторы относят как к метаэвристикам, так и считают простым учитывая то, что его отличие от локального поиска является незначительным, даже по сравнению с такой "примитивной" метаэвристической концепцией, как повторяемый локальный поиск [Lourenço et al., 2002]. Метаэвристические методы можно понимать как комбинацию двух техник: общая схема строится на базовом методе, в которую включается та или иная встроенная процедура. Важным аспектом есть то, что встроенная процедура – это в большинстве случаев самостоятельный алгоритм решения той же задачи, что и метаэвристический метод в целом. Под гибридными алгоритмами понимается такое объединение простых алгоритмов, которое еще не порождает метаэвристичный алгоритм. Например, это может быть

простое последовательное или параллельное выполнение. На практике простые алгоритмы используются все реже, поскольку развитие метаэвристических подходов позволяет получать лучшие результаты, а темпы развития вычислительной техники позволяют применять все более сложные алгоритмы к задачам все большей размерности. Классификацию гибридных метаэвристик детально рассмотрено в работе [Raidl, 2006]. По аналогии с последней, можно предложить классифицировать гибридные алгоритмы по степени гибридизации, порядку выполнения и стратегии управления.

По типу используемых пространств. Далее выделим по типу используемых пространств класс последовательных, или же конструктивных, алгоритмов которые состоят в построении решения из отдельных частей по определенным правилам. Такие алгоритмы работают в некотором расширенном пространстве $S \supseteq X$. Во многих случаях это расширенное пространство можно представить в терминах Определения 1 в таком виде:

$$S = X \cup \bigcup_{i=1, m} X_i, X_i = \{x^i = (\varphi^i, Z, \Omega^i) : \varphi^i : \{y_1, \dots, y_i\} \rightarrow Z\}, i = 1, \dots, m;$$

$$\Omega^i \supseteq \Omega^{i-1}, i = 2, \dots, m; \{y_1, \dots, y_m\} = Y.$$

Простейшим примером является эвристика для задачи коммивояжера, которую часто называют "жадной" или "иди к ближайшему" [Hoos and Stützle, 2005]. На практике конструктивные алгоритмы чаще всего используются либо для задач чрезвычайно большой размерности или задач с затратных вычислением целевой функции, либо для генерирования начальных решений для других методов. Иногда они встроены в состав методов более сложным образом, как, например, в оптимизации муравьиными колониями [Dorigo and Stützle, 2004], конструктивный блок – деятельность искусственных муравьев играет ключевую роль.

Другой класс алгоритмов – итерационные. Такие алгоритмы работают в пространстве полных решений X . Итерационные методы естественно разделяют на два класса по размеру множества решений, которыми алгоритм одновременно оперирует: в случае одноэлементного множества алгоритмы называют траекторными (реже, одноточечными), а в случае, когда множество текущих решений содержит больше одного элемента – популяционными.

В работе [Birattari et al., 2001] предлагается также различать методы комбинаторной оптимизации по количеству различных систем окрестностей которые они используют, типу целевой функции и по типу траектории. Рассмотрим эти характеристики более детально.

По количеству используемых систем окрестностей. По количеству используемых окрестностей можно различить алгоритмы, которые используют в процессе поиска одну или несколько различных систем окрестностей. Формально системы окрестностей обычно определяют так [Blum and Roli, 2003].

Определение 3. Система окрестностей – это такое отображение $N : X \rightarrow 2^{|X|}$, которое каждому решению $x \in X$ ставит в соответствие некоторое множество соседей, которое называется окрестностью решения x .

Большинство приближенных алгоритмов КО используют один тип окрестностей. Это такие алгоритмы как имитационный отжиг, табу поиск [Blum and Roli, 2003]. В повторяемом локальном поиске используется по крайней мере две системы окрестностей. В первой осуществляется локальный поиск до достижения локального минимума и после этого осуществляется возмущение найденного локального минимума. Фактически, возмущение может рассматриваться как шаг в другой системе окрестностей. Эти соображения, в частности, были развиты в алгоритме поиска в изменяемых окрестностях [Mladenovic and Hansen, 1997] – в этом подходе окрестности систематически изменяются в пределах заведомо заданного перечня $\{N_1, \dots, N_l\}$. Процесс конструирования решений в таких подходах как оптимизация

муравьиными колониями или GRASP [Hoos and Stützle, 2005] также можно проинтерпретировать как разновидность поиска в окрестностях, но такая интерпретация не будет отображать базовые алгоритмические идеи этих методов [Birattari et al., 2001].

По типу целевой функции. Некоторые алгоритмы модифицируют оценивание отдельных состояний процесса поиска во время выполнения алгоритма. Одним из примеров является метод "выламывания" (breakout) [Morris, 1993]. Базовой идеей является введение штрафов на включение отдельных компонент в решение, которые изменяют значение целевой функции. Как развитие этого подхода был предложен управляемый локальный поиск [Voudouris and Tsang, 1995]. Табу поиск также можно проинтерпретировать как такой, который использует динамическую целевую функцию, так как некоторые точки пространства поиска являются запрещенными, что отвечает бесконечно большим значениям целевой функции. Однако, остальные среди известных на данный момент методов используют статическую целевую функцию.

По типу траектории. Важным отличием между разными итерационными подходами подходами является то, следуют ли они одной траектории поиска, что отвечает "непрерывным" переходам на графе соседства, или осуществляют большие "прыжки" на графе соседства [Birattari et al., 2001]. Предлагается такое формальное определение траекторно-непрерывного метода.

Определение 4. Метод называется траекторно-непрерывным, если последовательность текущих решений на итерациях x_1, x_2, \dots, x_k владеет таким свойством

$$\forall i = 2, \dots, k \exists N' \in \{N_1, \dots, N_l\} : x_i \in N'(x_{i-1}), \quad (1)$$

где $\{N_1, \dots, N_l\}$ – перечень систем окрестностей которые использует метод. Для популяционных траекторно-непрерывных методов должен существовать такой способ упорядочить решения из множеств текущих решений на итерациях в последовательности, что каждая из последовательностей будет владеть свойством (1). это свойство должно выполняться для каждой последовательности текущих решений, отвечающих отдельным решениям из популяции. Методы не владеющие этим свойством называются траекторно-разрывными.

Имитационный отжиг или табу поиск являются типичными примерами траекторно-непрерывных методов. Алгоритмы, которые осуществляют более сложные переходы, но которые можно представить в виде более простых шагов, также можно проинтерпретировать как траекторно-непрерывных. К таким алгоритмам принадлежат, например, алгоритмы метода поиска переменной глубины [Hoos and Stützle, 2005], в частности, алгоритм Лина-Кернигана предложенный для задачи коммивояжера [Lin and Kernigan, 1973], или алгоритмы, которые базируются на "выбрасывании цепей" (ejection chains) [Glover, 1996]. В повторяемом локальном поиске, GRASP, меметических алгоритмах [Moscato, 1999] генерируются начальные точки для подчиненного локального поиска. Генерация начальных точек отвечает "прыжкам" в процессе поиска, и такие алгоритмы, в общем, выполняют траекторно-разрывными поиск по отношению к графу соседства, который используется в локальном поиске. Отметим, что практически все популяционные подходы являются траекторно-разрывными.

По использованию памяти. Еще одной характеристикой, по которой будем различать приближенные алгоритмы КО является использование памяти. Алгоритмами с краткосрочной памятью называются алгоритмы, которые запоминают определенное количество последних принятых решений, сгенерированных решений, частей решений и т. д. Алгоритмами с долгосрочной памятью называются алгоритмы, которые запоминают информацию обо всем осуществленном процессе поиска в виде набора специальных переменных. Некоторые алгоритмы могут совмещать использование обоих видов памяти, например, табу-поиск с использованием критериев стремления [Glover and Laguna, 1997].

По оценке точности. По оценке точности выделим алгоритмы с апостериорной оценкой точности, априорной оценкой и без соответствующих оценок. Апостериорные оценки вычисляются непосредственно во время процесса решения, учитывая при этом и само полученное решение. Априорная оценка точности – это гарантированная оценка, которая следует из самого алгоритма решения задачи.

По сходимости. Для последнего показателя, по которому предлагается классифицировать приближенные методы комбинаторной оптимизации, – сходимости, в отличие от вышерассмотренных характеристик, принадлежность того или иного метода/алгоритма к классам построенным на основе этой характеристики может измениться в зависимости от получения новых результатов. Сходимость предусматривает два частных случая: сходимость по решению и сходимость по значению. Типы сходимости определяются таким образом [Dorigo and Blum, 2005].

Определение 5. Сходимостью по решению называется достижение с определенной вероятностью алгоритмом состояния, когда одно и то же оптимальное решение будет генерироваться постоянно.

Определение 6. Сходимостью по значению называется достижение с определенной вероятностью алгоритмом состояния, когда алгоритм сгенерирует произвольное оптимальное решение по крайней мере один раз.

Некоторые авторы [Birattari et al., 2001; Blum and Roli, 2003] предлагают разделять приближенные алгоритмы также по тому, были ли они созданы по какому-нибудь природному аналогу, как, например, имитационный отжиг или генетические алгоритмы [Hoos and Stützle, 2005]. Но этот показатель не отражает существенные характеристики алгоритма, поэтому в данную классификацию он включен не был.

Заключение

Предложена классификация приближенных алгоритмов комбинаторной оптимизации. Для отдельных классов также предложена формализация их определения.

Сопоставление разработанной классификации с применяемыми на практике подходами показало, что, ввиду того, что все больше применяемых на практике алгоритмов являются гибридными, совмещая те или иные компоненты разных методов, нередко бывает невозможно четко классифицировать алгоритм в целом. Четкая классификация возможна только при разбиении алгоритма на отдельные процедуры, реализующие те или иные "типичные действия" (например, конструирование решений или возмущение решений). Это в свою очередь ставит вопрос определения перечня различных по функциональности элементарных процедур алгоритмов, который напрямую связан с построением общей (и в тоже время достаточно детальной) схемы приближенных методов комбинаторной оптимизации или, по крайней мере, достаточно широкого их класса. Это может быть одним из направлений будущих исследований.

Список литературы

- [Birattari et al., 2001] Birattari M., Paquete L., Stützle T., Varrentrapp K. (2001) Classification of metaheuristics and design of experiments for the analysis of components. Technical Report AIDA-01-05. Technische Universität Darmstadt. 12 p.
- [Blum, 2003] Blum C., Roli A. (2003) Metaheuristics in Combinatorial Optimization: Overview and Conceptual Comparison. In: ACM Computing Surveys, 35, №3. P. 268–308.
- [Dorigo and Blum, 2005] Dorigo M., Blum C. (2005) Ant colony optimization theory: A survey. In: Theoretical computer science. P. 243–278.
- [Dorigo and Stützle, 2004] Dorigo M., Stützle T. (2004) Ant Colony Optimization. Cambridge: MIT Press, MA. 348 p.
- [Dréo et al., 2007] Dréo J., Aumasson J.-P., Tfaili W., Siarry P. (2007) Adaptive Learning Search, A New Tool To Help Comprehending Metaheuristics. In: International Journal on Artificial Intelligence Tools. World Scientific Publishing Company, vol. 6, №3. P. 483-506.

-
- [Glover, 1996] Glover F. (1996) Ejection Chains, Reference Structures and Alternating Path Methods for Traveling Salesman Problems. In: Discrete Applied Mathematics. №65. P. 223–253.
- [Glover and Laguna, 1997] Glover F., Laguna M. (1997) Tabu Search. Kluwer Academic Publishers. 408 p.
- [Hoos and Stützle, 2005] Hoos H.H., Stützle T. (2005) Stochastic Local Search: Foundations and Applications. San Francisco: Morgan Kaufmann Publ. 658 p.
- [Kirkpatrick et al., 1983] Kirkpatrick S., Gelatt C. D., Vecchi M. P. (1983) Optimization by simulated annealing. In: Science. P. 671–680.
- [Laguna, 2002] Laguna M. (2002) Global optimization and meta-heuristics. In: Encyclopedia of Life Support Systems, Theme 6.5, Topic 2. <http://leeds-faculty.colorado.edu/Laguna/articles/elss.pdf>.
- [Lin and Kernighan, 1973] Lin S., Kernighan B.W. (1973) An effective heuristic algorithm for the travelling salesman problem In: Oper Res J. №21. P. 498–516.
- [Lourenço et al., 2002] Lourenço H. R., Martin O., Stützle T. (2002) Iterated local search. In: Handbook of Metaheuristics: International Series in Operations Research & Management Science, vol. 57 (Eds. F. Glover and G. Kochenberger). Norwell: Kluwer Academic Publishers, MA. P. 321–353.
- [Mladenović and Hansen, 1997] Mladenović N., Hansen P. (1997) Variable Neighbourhood Search. In: Computers & Operations Research. №24. P. 1097–1100.
- [Morris, 1993] Morris P. (1993) The Breakout Method for Escaping from Local Minima. In: Proceeding of the 11th Conference on Artificial Intelligence MIT Press. P. 40–45.
- [Moscato, 1999] Moscato P. (1999) Memetic Algorithms: A Short Introduction In: New Ideas in Optimization (Come F. G. D., Dorigo M., Glover F.). McGraw-Hill. P. 219–234.
- [Mühlenbein and Paaß, 1996] Mühlenbein H., Paaß G. (1996) From recombination of genes to the estimation of distributions I. Binary parameters. In: Lecture Notes in Computer Science: Parallel Problem Solving from Nature, №4. P. 178–187.
- [Raidl, 2006] Raidl G. R. (2006) A Unified View on Hybrid Metaheuristics. In: Lecture Notes in Computer Science. Springer-Verlag. P. 1–12.
- [Voudoris and Tsang, 1995] Voudoris C., Tsang E. (1995) Guided Local Search. Technical Report CSM-247, Department of Computer Science, University of Essex, England.
- [Гуляницький, 2004] Гуляницький Л.Ф. Решение задач комбинаторной оптимизации алгоритмами ускоренного вероятностного моделирования. В: Компьютерная математика. Сб. науч. тр. Киев, Институт кибернетики им. В.М. Глушкова НАН Украины, №1. С. 64–72.
- [Гуляницький, 2006] Гуляницький Л.Ф. Об одном метаэвристическом методе комбинаторной оптимизации. В: Компьютерная математика. № 2. С. 1–6.
- [Гуляницький, 2008] Гуляницький Л.Ф. (2008) До формалізації та класифікації задач комбінаторної оптимізації. В: Теорія оптимальних рішень. №7. С. 2–6.
- [Пападимитриу и Стайглиц, 1985] Пападимитриу Х., Стайглиц К. (1985) Комбинаторная оптимизация. Алгоритмы и сложность. М.: Мир. 512 с.

Об авторе

Сергей Сиренко (Sirenko) – аспирант, Институт кибернетики им. В.М.Глушкова НАН Украины, пр-т Глушкова, 40, Киев, 03680, Украина. E-mail: s.sirenko@gmail.com

Algorithmic and Mathematical
Foundations
of the Artificial Intelligence

АЛГОРИТМЫ РЕШЕНИЯ СИСТЕМ ЛИНЕЙНЫХ ДИОФАНТОВЫХ УРАВНЕНИЙ В ДИСКРЕТНЫХ ОБЛАСТЯХ

Сергей Крывый

Аннотация. Предложены алгоритмы построения минимального порождающего множества решений систем линейных однородных уравнений в множестве натуральных чисел и базиса множества решений системы линейных однородных и неоднородных диофантовых уравнений в кольцах и полях вычетов по модулю некоторого числа.

Abstract. The algorithms for computation of minimal supported set of solutions for systems of linear Diophantine homogeneous equations over set of natural numbers and basis of systems of linear Diophantine homogeneous and inhomogeneous equations in ring and field of remainders on modulo of a number.

Keywords: Systems of linear Diophantine constraints, minimal supported set of solutions, basis of solutions, satisfaction problem of constraints

ACM Classification Keywords: G 2.1 Discrete mathematics: Combinatorics

Conference topics: Algorithmic and Mathematical Foundations of the Artificial Intelligence

Введение

В настоящей работе рассматривается краткий обзор алгоритмов построения минимального порождающего множества решений и базиса множества решений систем линейных диофантовых уравнений в множестве натуральных чисел, в поле и кольце Z_m вычетов по модулю простого и составного числа m . Данная работа является продолжением работ [Крывый, 1999] - [Крывый, 2007]. В основе предлагаемых алгоритмов лежит TSS-метод построения минимального порождающего множества решений систем линейных однородных диофантовых уравнений в множестве натуральных чисел N [Крывый, 1999]. К такого рода системам и методам их решений сводятся задачи математических игр [Донец, 2002], распознавания изображений [Донец, 2005], криптографии [Черемушкин, 2002], распараллеливания циклов [Allen, 1987], построение линейных мозаик [Донец, 2005], унификации в теориях первого порядка [Baader, 1994], арифметики Пресбургера [Comon, 1999], анализ структурных свойств сетей Петри [Miyata, 1990] и многие другие задачи. Описываемые алгоритмы характеризуются оценками временной сложности.

Предварительные сведения

Системой линейных диофантовых констрейнтов (СЛДК) будем называть систему вида $Ax R b$, $A = [a_{ij}]$, $a_{ij}, b \in Z$ (множество целых чисел), $x \in N$, а $R \in \Omega = \{=, \leq, <, \neq, >, \geq\}$, $i = 1, 2, \dots, p$, $j = 1, 2, \dots, q$. Решением СЛДК называется такой вектор $c = (c_1, c_2, \dots, c_q)$, который при подстановке вместо x_j значений c_j в $L_i(x)$ обращает $L_i(x)Rb_i$ в истинное высказывание для всех $i = 1, 2, \dots, p$. СЛДК называется однородной (СЛОДК), если все b_i равны нулю, в противном случае СЛДК называется неоднородной (СЛНДК). Если СЛОДК состоит только из уравнений, то она называется системой линейных однородных диофантовых уравнений (СЛОДУ). Если СЛНДК состоит только из уравнений, то она

называется системой линейных неоднородных диофантовых уравнений (СЛНДУ). Методы решения СЛНДУ сводятся, как известно, к решению соответствующей СЛОДУ и поэтому в дальнейшем основное внимание будет уделяться методам и алгоритмам решения СЛОДУ.

Пусть S - СЛОДУ и $e_1 = (1, 0, \dots, 0, 0)$, $e_2 = (0, 1, \dots, 0, 0)$, ..., $e_q = (0, 0, \dots, 0, 1)$ - единичные векторы из множества N^q , которые называются векторами канонического базиса множества N^q . Введем на множестве N^q отношение порядка \leq , которое определяется таким образом: если $x = (x_1, \dots, x_q)$, $y = (y_1, \dots, y_q) \in N^q$, то $x \leq y$ тогда и только тогда, когда для всех $i=1, \dots, q$, $x_i \leq y_i$. Ясно, что это отношение является частичным порядком и относительно этого порядка можно говорить о минимальных элементах в множестве N^q . Очевидно, что наименьшим элементом в множестве N^q есть нулевой вектор.

Пусть M - множество решений СЛОДУ S . Поскольку система S однородная, то нулевой вектор всегда является ее решением. Это решение будем называть тривиальным, а всякое решение системы S , отличное от тривиального, будем называть нетривиальным решением. СЛОДУ S будем называть несовместной, если множество M состоит только лишь из тривиального решения, в противном случае она будет называться совместной.

Известно, что множество B минимальных элементов множества решений M системы S составляет базис множества M и если $|M| > 1$, то базис B всегда существует, конечен и всякий элемент из M представим в виде неотрицательной линейной комбинации векторов из B . Известно также, что процесс решения СЛНДУ или системы линейных диофантовых неравенств (СЛДН) может быть сведен к решению СЛОДУ, поэтому основное место в исследованиях уделяется СЛОДУ. Следует заметить, что в общем случае такое сведение увеличивает размерность пространства, над которым рассматривается полученная СЛОДУ, что сказывается на эффективности вычислений. Однако, имеются методы сведения, которые не увеличивают размерности пространства [Contejan, 1997].

Критерий совместности СЛОДУ

Критерий совместности СЛОДУ, используемый здесь, и алгоритм его реализации подробно описаны в работах [Крытый, 1999] и [Крытый, 1999,1], поэтому приведем лишь необходимые факты, нужные в дальнейшем, следуя этим работам.

Пусть дана СЛОДУ $S = L_1(x) = 0 \wedge L_2(x) = 0 \wedge \dots \wedge L_p(x) = 0$. Рассмотрим множество векторов канонического базиса $M_0' = \{e_1, \dots, e_q\}$ и первое уравнение $L_1 = a_{11}x_1 + a_{12}x_2 + \dots + a_{1q}x_q = 0$ системы S . С помощью функции $L_1(x)$ разобьем элементы множества M_0' на такие три группы $M_1^0 = \{e^0 \mid L_1(e^0) = 0\}$, $M_1^+ = \{e^+ \mid L_1(e^+) > 0\}$, $M_1^- = \{e^- \mid L_1(e^-) < 0\}$. Ясно, что если одно из множеств $M_1^0 \cup M_1^+$ или $M_1^0 \cup M_1^-$ пусто, то уравнение $L_1(x) = 0$ не имеет нетривиальных решений в множестве натуральных чисел. Допустим, что хотя бы два из множеств M_1^0, M_1^+, M_1^- непусты, тогда рассмотрим множество

$$M_1' = M_1^0 \cup \{e_{ij} \mid -L_1(e_i)e_j + L_1(e_j)e_i\}, e_i \in M_1^-, e_j \in M_1^+.$$

Используя функцию $L_2(x)$, разобьем элементы множества M_1' аналогично предыдущему также на три группы $M_2^0 = \{e^0 \mid L_2(e^0) = 0\}$, $M_2^+ = \{e^+ \mid L_2(e^+) > 0\}$, $M_2^- = \{e^- \mid L_2(e^-) < 0\}$.

Допустим, что хотя бы два из этих множеств непусты, тогда построим множество

$$M_2' = M_2^0 \cup \{e_{ij} \mid -L_2(e_i)e_j + L_2(e_j)e_i\}, e_i \in M_2^-, e_j \in M_2^+.$$

Предположим, что таким способом построено множество M_j' из множеств $M_j^0 = \{e_r^0 \mid L_j(e_r^0) = 0\}$, $M_j^+ = \{e_i^+ \mid L_j(e_i^+) > 0\}$, $M_j^- = \{e_s^+ \mid L_j(e_s^+) < 0\}$ с помощью функции $L_j(x)$ и это множество непусто. Непосредственно из этих построений вытекает такое утверждение [Кривый, 1999].

Теорема 1. Элементы множества M_j' есть решениями системы уравнений $L_1(x) = 0 \wedge L_2(x) = 0 \wedge \dots \wedge L_j(x) = 0$.

Определение 1. Множество M_j' , построенное выше, будем называть усеченным множеством решений системы $S' = L_1(x) = 0 \wedge L_2(x) = 0 \wedge \dots \wedge L_j(x) = 0$.

Пусть $M_j' = \{e_1', \dots, e_k'\}$ - усеченное множество решений системы S' , а M_j -- множество всех ее решений. Тогда имеет место такое утверждение.

Теорема 2. Для всякого вектора $x \in M_j - M_j'$ существует представление в виде неотрицательной линейной комбинации вида $tx = b_1e_1' + b_2e_2' + \dots + b_ke_k'$, где $t, b \in N$, $t \neq 0$, $e_i' \in M_j'$, $i = 1, 2, \dots, k$.

Из этой теоремы следует такой критерий совместности СЛДОДУ в множестве натуральных чисел.

Теорема 3. СЛОДУ $S = L_1(x) = 0 \wedge L_2(x) = 0 \wedge \dots \wedge L_p(x) = 0$ совместна тогда и только тогда, когда её усеченное множество $M_p' \neq \emptyset$.

Допустим, что СЛОДУ S совместна и $M_p' = \{e_1', \dots, e_k'\}$ её усеченное множество решений. Тогда имеют место такие утверждения.

Теорема 4. Векторы из усеченного множества решений являются минимальными решениями СЛОДУ S , т.е. являются ее базисными решениями [Кривый, 1999, 1].

Теорема 5. Пусть $x = (x_1, \dots, x_q)$ минимальное решение СЛОДУ S и $M' = \{e_1' = (\alpha_{11}, \alpha_{11}, \dots, \alpha_{1q}), \dots, e_k' = (\alpha_{k1}, \alpha_{k1}, \dots, \alpha_{kq})\}$ её усеченное множество решений. Тогда верхняя граница x' величины координат произвольного решения СЛОДУ удовлетворяет неравенству $x' = \max x_i \leq k \max \alpha_{ij}$.

Теорема 6. Сложность алгоритма определения совместности СЛОДУ в общем случае имеет экспоненциальную сложность по числу уравнений в системе.

Решение СЛДУ в кольцах и полях вычетов

Кольцом вычетов Z_m по модулю числа m определяется обычным образом: т.е. это алгебра с двумя нульарными операциями 0 и 1, двумя бинарными социативно-коммутативными и дистрибутивными операциями + и -. Заметим, что это кольцо имеет делители нуля. На основании законов для операций в кольце Z_m вытекает справедливость такого тождества: для всех $x, y \in Z_m$ $x + y = 0$ следует $x = -y$.

Из тождеств следует, что в кольце Z_m $x = m - y$, а $-y = x - m$, что дает возможность заменять положительное число x на отрицательное число $-y = x - m$ и наоборот. Такие элементы x и $-y$ будем называть дополнениями (x дополняет $-y$ и наоборот). Кольцо вычетов Z_m называется примарным, если модуль m является степенью простого числа p .

Случай одного линейного однородного диофантового уравнения (ЛОДУ). Пусть дано ЛОДУ

$$L(x) = a_1x_1 + \dots + a_nx_n = 0, \text{ где } a_i, x_i \in Z_m, i = 1, 2, \dots, n.$$

(Условие 1): Допустим, что $a_i \neq 0$ и этот коэффициент взаимно прост с модулем m .

Теорема 7. Множество B решений ЛОДУ $L(x)=0$, построенное комбинированием дополнения первого ненулевого коэффициента, удовлетворяющего условию 1, взятого с отрицательным знаком, с остальными ненулевыми коэффициентами и пополненное векторами канонического базиса, которые соответствуют нулевым коэффициентам ЛОДУ, является базисом множества всех решений этого ЛОДУ. Сложность алгоритма пропорциональна величине l^3 , где $l = \max(s, n)$, $s = \log m$ - число двоичных разрядов числа m , а n - число неизвестных в ЛОДУ.

Следствие 1. Если модуль m является простым числом, то множество B решений ЛОДУ $L(x) = 0$ является базисом множества всех решений этого ЛОДУ.

Случай линейного неоднородного диофантового уравнения (ЛНДУ). Пусть дано ЛНДУ $L(x) = a_1x_1 + \dots + a_nx_n = b$, у которого коэффициент a_k взаимно прост с модулем m . Найдем решение сравнения $a_k y \equiv b \pmod{m}$, которое при данных условиях будет единственным. Пусть этим числом будет s , т. е. вектор $x = (0, \dots, 0, s, 0, \dots, 0)$ будет решением $L(x) = b$. Применяя TSS-метод к этому ЛОДУ, которое соответствует $L(x) = b$, находим базис B множества его решений.

Теорема 8. Множество B с добавленным вектором $x = (0, \dots, 0, s, 0, \dots, 0)$, найденные описанным выше способом, является базисом множества решений ЛНДУ.

Следует заметить, что такого типа СЛОДУ можно решать с использованием алгоритма построения базиса множества решений в множестве натуральных чисел (например, алгоритма Контежан-Деви [Contejan, 1997]). Если применять такого типа алгоритмы, то для приведенной выше СЛОДУ он сгенерирует большое количество решений, в то время как только несколько из них будут составлять базис множества всех решений данной СЛОДУ. Из этого следует, что TSS-алгоритм более предпочтителен, чем традиционные алгоритмы построения базиса множества всех решений СЛОДУ, удовлетворяющих условию 1.

ЛОДУ над примарными кольцами. Рассмотрим ЛОДУ над примарным кольцом Z_m

$$L(x) = a_1x_1 + \dots + a_nx_n = 0, \quad \text{где } a_i, x_i \in Z_m, i = 1, 2, \dots, n, m = p^{t+1}, t > 1, t \in N.$$

Пусть $\text{НОД}(a_1, a_2, \dots, a_n, m) = p^u$, тогда сокращая коэффициенты на p^u , получаем ЛОДУ удовлетворяющее условию 1. Имеет место следующая теорема.

Теорема 9. Множество TSS уравнения $L(x) = 0$, дополненное вектором $s = (p^v, 0, 0, \dots, 0)$, является базисом множества решений ЛОДУ $L(x) = 0$.

Общий случай ЛОДУ. Рассмотрим ЛОДУ, для которых не выполняется условие 1. Предположим, что модуль m имеет разложение на простые множители вида $m = p^c q^d$ (например, $m = 12 = 3^2 \cdot 2$) и дано ЛОДУ $L(x) = a_1x_1 + \dots + a_nx_n = 0$, где $a_i, x_i \in Z_m, i = 1, 2, \dots, n$. Построим по этому ЛОДУ два

ЛОДУ $L_1(x) = a_1'x_1 + \dots + a_n'x_n = 0$, и $L_2(x) = b_1'x_1 + \dots + b_n'x_n = 0$, в которых коэффициенты приведены по модулям p^c и q^d . Построим базисы множеств решений B_1, B_2 для этих ЛОДУ. Имеет место следующая теорема.

Теорема 10. *Множество $B = B_1 \cup B_2$ является базисом множества всех решений ЛОДУ $L(x) = a_1x_1 + \dots + a_nx_n = 0$.*

TSS-метод решения СЛДУ в кольцах вычетов и полях вычетов. Из вышеприведенных теорем следует такая процедура построения базиса множества решений СПОДУ. Она состоит в разбиении СПОДУ S на две подсистемы S' и S'' по модулям p^c и q^d соответственно. Каждая из этих подсистем решается отдельно, находятся вначале базисы B' и B'' соответственно для S' и S'' , а затем базис $B = q^d B' \cup p^c B''$, где $q^d B'$ и $p^c B''$ означает умножение каждого вектора из B' на q^d , а из B'' - на p^c .

В общем случае, если модуль m имеет разложение, содержащее больше двух сомножителей, т.е. $m = p_1^{n_1} p_2^{n_2} \dots p_k^{n_k}$, то получаем k подсистем. Принимая во внимание, что арифметическая сложность выполнения операций сложения и вычитания в кольце Z_m пропорциональна s , где s - максимальная разрядность рассматриваемых чисел; выполнения операций умножения и деления, как и вычисления НОД двух чисел, меньших m , - s^2 , то арифметическая сложность построения базиса множества решений СПОДУ имеет вид $O(l^5)$. Таким образом, арифметическая сложность перехода от предыдущего к последующему ЛОДУ в одной подсистеме пропорциональна величине $O(l^5)$, где $l = \max(n, s, r)$, $s = \log m$. Такая процедура повторяется r раз и в результате имеем $O(l^6)$, где $l = \max(n, s, k, r)$. Иными словами имеет место

Теорема 11. *Множество B , построенное TSS-методом, является базисом множества решений СПОДУ $L(x) = a_1x_1 + \dots + a_nx_n = 0$. Арифметическая сложность построения B пропорциональна величине $O(l^6)$, где $l = \max(n, s, k, r)$.*

TSS-метод решения СЛНДУ. Построение базиса множества решений СЛНДУ сводится к поиску частного решения ЛНДУ и базиса множества решений соответствующего ему ЛОДУ. Характеристику временной сложности дает следующая теорема.

Теорема 12. *Временная сложность приведенной выше процедуры построения общего решения СЛНДУ выражается величиной $O(l^7)$, где $l = \max(n, s, k, r)$.*

Заметим, что приведенные алгоритмы имеют полиномиальные оценки временной сложности при условии известного разложения модуля на простые множители. Проблема разложения натурального числа на простые множители (которая называется проблемой факторизации) является одной из наиболее важных проблем теории чисел. Имеется несколько алгоритмов ее решения: алгоритмы Полларда, Полларда-Штрассена, решета числового поля [Черемушкин, 2002]. Наиболее эффективным алгоритмом в настоящее время является последний из перечисленных алгоритмов потому, что, в отличие от первых двух алгоритмов, он ищет большие делители заданного числа. Все эти алгоритмы имеют экспоненциальные оценки временной сложности, наилучшая из которых для заданного числа n имеет вид $O(2^{c\sqrt{\ln n \ln \ln n}})$, где c - константа, а n - число неизвестных в СЛДУ.

Очевидно, что описанные алгоритмы применимы к СЛДУ в полях вычетов, поскольку в таких полях условия 1 автоматически выполняются. А это значит, что все приведенные выше факты верны для этих полей и в случае простого модуля нет необходимости в его факторизации. Следовательно, алгоритмы решения СЛДУ в полях вычетов по модулю простого числа имеют полиномиальную оценку временной сложности и эта оценка имеет вид $O(q^2 n^2)$ [Кривый, 2007].

Заключение

В заключение заметим, что приведенные оценки временных сложностей алгоритмов можно уточнять, если проследить все детали процесса вычислений, происходящего в TSS-алгоритме. В данной работе мы ограничиваемся установлением только верхних оценок (т. е. сложность в наихудшем случае) этих алгоритмов. Отметим также, что при малых значениях модуля p сложностью вычисления НОД в полях и кольцах вычетов можно пренебречь и тогда оценка алгоритмов решения систем в таких полях упрощается. Так, например, в поле F_2 , которое часто встречается в приложениях, необходимость вычисления НОД вообще отпадает, поэтому сложность решения СЛОДУ и СЛНДУ в таком поле становится пропорциональна величине qn^2 , где q - число уравнений, а n - число неизвестных в системе.

Библиография

- [Донец, 2002] Донец Г. А. Решение задачи о сейфе на $(0,1)$ -матрицах//Кибернетика и системный анализ. -2002. - N 1. - С. 98--105.
- [Донец, 2005] Донец Г. А., Самер И. М. Альшаламе. Решение задачи о построении линейной мозаики.. Теория оптимальных решений. - К.: Ин-т кибернетики им. В. М. Глушкова НАН Украины. - 2005. - С. 15 -- 24.
- [Кривый, 2006] Кривый С. Л. Алгоритмы решения систем линейных диофантовых уравнений в целочисленных областях// Кибернетика и системный анализ - 2006. - N 2. - С. 3 -- 17.
- [Кривый, 2007] Кривый С. Л. Алгоритмы решения систем линейных диофантовых уравнений в полях вычетов. Там же. - 2007. - N 2. - С. 15 -- 23.
- [Кривый, 2007,1] Кривый С. Л. Алгоритмы решения систем линейных диофантовых уравнений в кольцах вычетов. Там же. - 2007. - N 6. - С. 15 -- 23.
- [Кривый, 1999] Кривый С. Л. О некоторых методах решения и критериях совместности систем линейных диофантовых уравнений в области натуральных чисел. Там же. - 1999. - N 4. - С.12 -- 36.
- [Кривый,1999,1] Кривый С. Л. Критерий совместности систем линейных диофантовых уравнений над множеством натуральных чисел//Доклады НАНУ. - 1999. - N 5. - С.107-112.
- [Черемушкин, 2002] Черемушкин А. В. Лекции по арифметическим алгоритмам в криптографии. - М.: МЦНМО. - 2002. - 103 с.
- [Чугаенко. 2007] Чугаенко А.В. О реализации TSS-алгоритма.. ж. Управляющие системы и машины. -2007. - N 3. - С. 14 -- 26.
- [Baader,1994] Baader F., Ziekmann J. Unification theory. Handbook of Logic in Artificial Intelligence and Logic Programming.-Oxford University Press. -1994. - P. 1--85.
- [Allen,1987] Allen R., Kennedy K. Automatic translation of FORTRAN program to vector form. ACM Transactions on Programming Languages and systems. -1987.- v. 9, N4. - P. 491--542.
- [Contejan, 1997] Contejan E., Ajili F. Avoiding slack variables in the solving of linear diophantine equations and inequations. Theoretical Comp. Science. -1997.- 173. - P.183 -- 208.
- [Pottier, 1991] Pottier L. Minimal solution of linear diophantine systems: bounds and algorithms. In Proc. of the Fourth Intern. Conf. on Rewriting Techniques and Applications. -Como. -Italy. -1991. -P. 162 -- 173.

-
- [Domenjoud, 1991] Domenjoud E. Outils pour la deduction automatique dans les theories associatives-commutatives. Thesis de Doctorat d'Universite: Universite de Nancy I. -1991.
- [Clausen, 1989] Clausen M., Fortenbacher A. Efficient solution of linear diophantine equations. J. Symbolic Computation. - 1989. - 8, N 1,2, -P. 201--216.
- [Romeuf, 1990] Romeuf J. F. A polynomial Algorithm for Solvin systems of two linear Diophantine equations. TCS. - 1990. - 74, N3. - P.329--340.
- [Filgueiras, 1995] Filgueiras M.,Tomas A.P. A Fast Method for Finding the Basis of Non-negative Solutions to a Linear Diophantine Equation. J. Symbolic Computation. -1995.- 19, N2. -P. 507--526.
- [Comon, 1999] Comon H. Constraint solving on terms: Automata techniques (Preliminary lecture notes). Intern. Summer School on Constraints in Computational Logics: Gif-sur-Yvette, France, September 5--8. - 1999. - 22 p.
- [Murata, 1989] Murata T. Petri Nets: Properties, Analysis and Applications. In Proceedings of the IEEE. -1989. - v. 7.- №4. - P. 541-580.
-

Сведения об авторе

Кривый Сергей Лукьянович - Институт кибернетики им. В.М.Глушкова НАН Украины; Украина, Киев.
email: krivoi@i.com.ua

МНОГОКРИТЕРИАЛЬНЫЕ ЗАДАЧИ НА КОМБИНАТОРНОМ МНОЖЕСТВЕ ПОЛИРАЗМЕЩЕНИЙ: СТРУКТУРНЫЕ СВОЙСТВА РЕШЕНИЙ

Людмила Колечкина

Резюме: Рассматривается многокритериальная задача дискретной оптимизации на допустимом комбинаторном множестве полиразмещений. Исследуются структурные свойства допустимой области и различных видов эффективных решений. На основе развития идей евклидовой комбинаторной оптимизации и метода главного критерия предложены и обоснованы возможные подходы для решения многокритериальной комбинаторной задачи на множестве полиразмещений..

Ключевые слова: многокритериальная оптимизация, дискретная оптимизация, Парето-оптимальные, слабо, строго эффективные решения, комбинаторное множество полиразмещений.

ACM Classification Keywords: G2.1 Combinatorics (F2.2), G1.6 Optimization

Введение

Многокритериальные задачи оптимизации на различных множествах продолжают привлекать внимание многих исследователей [1 – 10]. Модели дискретной комбинаторной оптимизации широко применяются при решении важных задач геометрического проектирования, экономики, размещение объектов, управления процессом обработки данных, принятия решений и других. В последнее время в области исследования различных классов комбинаторных моделей, разработки новых методов их решения большое внимание уделяется методам, которые основаны на использовании структурных свойств комбинаторных множеств [2, 8 – 15].

В данной работе формулируется и исследуется качественно новая и актуальная задача, которая объединяет многокритериальность альтернатив и допустимые множества решений, имеющие определенные комбинаторные свойства. Как известно, большинство комбинаторных оптимизационных задач могут быть сведены к задачам целочисленного программирования, но это не всегда оправдано, поскольку при этом теряется возможность учета комбинаторных свойств задач [2].

Систематическое изучение свойств евклидовых комбинаторных множеств и их исследование описаны во многих работах. Рядом с хорошо известными евклидовыми комбинаторными множествами перестановок, размещений, сочетаний, разбиений выделяются более сложные структуры – поликомбинаторные множества. Интерес к таким множествам обусловлен разными прикладными задачами, поскольку значительное их количество хорошо описывается с помощью поликомбинаторных конструкций [12, 14].

Следует отметить, что задачи евклидовой комбинаторной оптимизации на поликомбинаторных множествах неотъемлемо связаны с комбинаторными многогранниками, которые являются выпуклыми оболочками таких множеств, и их свойствами. Повышенный интерес к комбинаторным и поликомбинаторным конфигурациям обусловлен исследованиями последних лет в области компьютерных технологий при создании современных алгоритмов и программ для решения оптимизационных задач. Следовательно, рассмотрение новых задач на поликомбинаторных множествах со многими критериями предопределено потребностями практики.

Данная работа продолжает исследования многокритериальных задач на комбинаторных множествах перестановок, сочетаний, представленные в работах [8, 9].

Постановка задачи.

Рассматриваются многокритериальные задачи вида:

$$Z(\Phi, P_{qk}^{ns}(A, H)): \max \left\{ \Phi(a) \mid a \in P_{qk}^{ns}(A, H) \right\},$$

состоящие в максимизации векторного критерия $\Phi(a)$ на евклидовом комбинаторном множестве полиразмещений, где $\Phi_i: R^n \rightarrow R^1, i \in N_l = \{1, \dots, l\}$.

Для изложения материала используем понятие мультимножества $A = \{a_1, a_2, \dots, a_q\}$, которое определяется основанием $S(A) = \{e_1, e_2, \dots, e_k\}$ т.е. множеством всех его различных элементов и кратностью его элементов $k(e_j) = r_j$ – числом повторений каждого j -го элемента основания, $j \in N_k = \{1, 2, \dots, k\}$, $r_1 + r_2 + \dots + r_k = q$.

Выберем произвольное $n \in N_q$. Упорядоченной n -выборкой из мультимножества A называется набор

$$a = (a_{i_1}, a_{i_2}, \dots, a_{i_n}), \text{ где } a_{i_j} \in A \quad \forall i_j \in N_n, \quad \forall j \in N_n, \quad i_s \neq i_t, \text{ если } s \neq t \quad \forall s \in N_n, \quad \forall t \in N_k.$$

Определение 1. Множество упорядоченных n -выборок из мультимножества A при условии $n < q$ называется множеством размещений с повторениями из n действительных чисел, среди которых k различных, либо общим множеством размещений и обозначается $P_{qk}^n(A)$.

Представим множество N_q в виде упорядоченного разбиения на s , где $s < q$, непустых попарно непересекающихся подмножеств J_1, \dots, J_s , то есть для них выполняются условия: $J_i \cap J_j = \emptyset$, $J_i \neq \emptyset, J_j \neq \emptyset, \forall i, j \in N_s$, а также упорядоченное разбиение числа n на s слагаемых n_1, n_2, \dots, n_s , которое удовлетворяет условию $1 \leq n_i \leq q_i, \quad \forall i \in N_s, |J_i| = q_i$. Очевидно, что $q_1 + q_2 + \dots + q_s = q$, $n_1 + n_2 + \dots + n_s = n$.

Обозначим H – множество элементов вида: $h = (h(1), \dots, h(n)) = (h^1, \dots, h^s)$, где $h(j) \in N_n, j \in N_n$, а h^i – произвольная перестановка элементов множества $J_i \quad \forall i \in N_s$.

Пусть подмультимножество A^i мультимножества A , состоит из тех элементов A , номера которых принадлежат множеству $J_i: A^i = \{a_1^i, \dots, a_{n_i}^i\}, |J_i| = n_i$.

Определение 3. Множество $P_{qk}^{ns}(A, H) = \{(a_{h(1)}, \dots, a_{h(n)}) \mid a_{h(i)} \in A \quad \forall i \in N_n, \forall h \in H\} \subset R^n$

называют общим множеством полиразмещений.

Не теряя общности, упорядочим элементы мультимножества A по неубыванию: $a_1 \leq a_2 \leq \dots \leq a_n$.

Очевидно, что это упорядочение сохраняется и для каждого подмультимножества $A^i, i \in N_s$, из A .

Свойства евклидова множества полиразмещений

Известно, что комбинаторные множества приобретают интересные свойства при погружении в арифметическое евклидово пространство. Будем рассматривать элементы множества полиразмещений как точки арифметического евклидова пространства R^n .

Пусть вектор a – элемент евклидова комбинаторного множества $E(A)$. Отображение $\varphi: E(A) \rightarrow E_\varphi(A) \subset R^n$ называется погружением множества $E(A)$ в арифметическое евклидово пространство, если φ задает взаимно однозначное соответствие $E_\varphi(A) \subset R^n$ по правилу: для $x_j = a_{ij}$ $\forall j \in N_n$.

Известно [12, 14], что выпуклой оболочкой множества полиразмещений $P_{qk}^{ns}(A, H)$ является многогранник полиразмещений $\Pi_{qk}^{ns}(A, H) = \text{conv } P_{qk}^{ns}(A, H)$, множеством вершин которого есть элементы множества полиразмещений: $\text{vert } \Pi_{qk}^{ns}(A, H) = P_{qk}^{ns}(A, H)$.

Теорема 1. Множество $\Pi_{qk}^{ns}(A, H)$ определяется совокупностью всех решений системы

$$\begin{cases} \sum_{j=1}^{n_i} x_j \leq \sum_{j=1}^{n_i} a_j^i, i \in N_s, \end{cases} \quad (1)$$

$$\begin{cases} \sum_{j=1}^{m_i} x_{\alpha_j} \geq \sum_{j=1}^{m_i} a_j^i, m_i \in N_{q_i-1}, \alpha_j \in J_i, \forall i \in N_s \end{cases} \quad (2)$$

$$\alpha_j \neq \alpha_t, \forall j \neq t, \forall j, t \in J_i.$$

Многогранник $\Pi_{qk}^{ns}(A, H)$ будем называть общим многогранником евклидова множества полиразмещений. Рассмотрим некоторые его свойства и связь с общим множеством полиразмещений.

Очевидно, что из системы линейных неравенств (1) – (2) можно выделить s подсистем линейных неравенств, описывающих многогранники размещений $\Pi_{q_i k_i}^{n_i}(A^i)$, являющиеся выпуклой комбинацией множества размещений $a_{h^i}^i, i \in N_s$. Следовательно,

$$\Pi_{q_i k_i}^{n_i}(A^i) = \left\{ x \in R^{n_i} \left| \sum_{j=1}^{n_i} x_j \leq \sum_{j=1}^{n_i} a_{q_i-j}^i, \sum_{j=1}^{m_i} x_{\alpha_j} \geq \sum_{j=1}^{m_i} a_j^i \right. \right\},$$

$$m_i \in N_{q_i-1}, \alpha_j \in J_i, \alpha_j \neq \alpha_t, \forall j \neq t, \forall j, t \in J_i, \forall i \in N_s.$$

Определение 4. Под произведением многогранников M_1, \dots, M_s понимают множество

$$\otimes_{i=1}^s M_i = \left\{ x \in R^{d_1 + \dots + d_s} \mid x = (x_1, \dots, x_s), x_i \in M_i \quad \forall i \in N_s \right\}, \text{ где } M_i - d_i - \text{ мерный многогранник.}$$

Воспользуемся следующей леммой [15].

Лемма. 1) Произведение многогранников является многогранником;

2) $\dim(\otimes_{i=1}^S M_i) = \sum_{i=1}^S \dim M_i$, где $\dim M$ – размерность множества M ;

3) k -мерные грани многогранника $\otimes_{i=1}^S M_i$ образуют множество с элементами вида $\otimes_{i=1}^S F_i$, где F_i – k_i -мерная грань многогранника M_i и $k_1 + \dots + k_s = k$.

Каждый из многогранников $\Pi_{q_i k_i}^{n_i}(A^i)$ представляет собой многогранник размещений. По определению 4

и согласно лемме справедливо равенство

$$\otimes_{i=1}^S \Pi_{q_i k_i}^{n_i}(A^i) = \left\{ x \in R^{d_1 + \dots + d_s} \mid x = (x_1, \dots, x_s), x_i \in \Pi_{q_i k_i}^{n_i}(A^i) \quad \forall i \in N_s \right\}, \quad \text{то есть точка}$$

$x \in \otimes_{i=1}^S \Pi_{q_i k_i}^{n_i}(A^i)$ удовлетворяет каждой из s подсистем системы (1), (2). Следовательно, можно

утверждать, что если a_{h^i} – вершина многогранника $\Pi_{q_i k_i}^{n_i}(A^i)$, то $a(h) = \otimes_{i=1}^S a_{h^i}$. Соответственно

$$a(h) = (a_{h^1}, \dots, a_{h^s}), \quad \text{где } a(h) \in P_{qk}^{ns}(A, H).$$

Справедливы следующие теоремы [12].

Теорема 2. Общий многогранник евклидова множества полиразмещений можно представить как произведение многогранников $\Pi_{q_i k_i}^{n_i}(A^i)$ размещений, т.е. $P_{qk}^{ns}(A, H) = \otimes_{i=1}^S \Pi_{q_i k_i}^{n_i}(A^i)$.

Теорема 3. Множество полиразмещений $P_{qk}^{ns}(A, H)$ совпадает с множеством вершин многогранника $\Pi_{qk}^{ns}(A, H)$.

Теорема 4. Вершина $a(h) \in \text{vert } \Pi_{qk}^{ns}(A, H)$ является смежной с вершиной $a(z) \in \text{vert } \Pi_{qk}^{ns}(A, H)$ тогда и только тогда, когда $a(z)$ образуется из $a(h)$ перестановкой двух неравных друг другу компонент a_i^i и a_j^j , $j \in J_{q_i-1}$, $i \in N_s$.

Следует отметить, что общее число p линейных неравенств, входящих в систему (1), (2), описывающих многогранник полиразмещений $\Pi_{qk}^{ns}(A, H)$ очень велико. Согласно [11] совокупность неравенств подсистемы для некоторого подмножества $J_i, i \in N_s$ системы (1), (2), имеющих одинаковое значение m_i верхнего предела суммирования, будем называть m_i -ой группой неравенств этой подсистемы, где $i \in N_s$. В частности, в каждую m_i -ю группу входит $C_{q_i}^{m_i}$ неравенств. Отсюда имеем общее число

$$\text{неравенств, описывающих многогранник } \Pi_{q_i k_i}^{n_i}(A^i) \text{ равным } p_i = \sum_{i=0}^{q_i} C_{q_i}^{m_i} = 2^{q_i}, \quad i \in N_s.$$

Справедливо следующее утверждение.

Утверждение. Поскольку из q_i координат $a_j^i, j \in J_i$, точки $x \in P_{qk}^{ns}(A, H)$ только k_i различных, то из системы неравенств (1), (2), описывающей общий многогранник полиперестановок $\Pi_{qk}^{ns}(A, H)$, можно исключить некоторые неравенства. Их общее число составляет $N = \sum_{i=1}^s N^i$, где $N^i = 1 + q_i + \sum_{j=i+1}^{q_i} C_{q_i}^j$.

Доказательство. С учетом выполнения условия $a_1 \leq a_2 \leq \dots \leq a_{q_i}$ для любого $j \in N_{m_i-1}$, $m_i \leq q_i, i \leq N_s$, имеет место равенство $a_j^i = a_{j+1}^i$. В этом случае при выполнении неравенств первой группы в подсистеме (1), (2) будут также справедливы неравенства второй, третьей, ..., m_i -ой, $i \in N_s$, групп. Действительно, поскольку $x_j \geq a_1^i, j \in J_i, i \in N_s$, то для любого $m_i \in N_n$ выполняется условие $\sum_{j=1}^{m_i} x_{\alpha_j} \geq m_i a_1^i$. Следовательно, из каждой подсистемы системы (1), (2), описывающей многогранник

полиразмещений $\Pi_{qk}^{ns}(A, H)$, можно исключить неравенства второй, третьей, ..., m_i -ой, $i \in N_s$, групп и общее число неравенств в i -ой подсистеме будет составлять $N^i = 1 + q_i + \sum_{j=i+1}^{q_i} C_{q_i}^j$, а следовательно число неравенств, которое можно исключить из системы (1), (2) будет равно $N = \sum_{i=1}^s N^i$. Если набор

чисел $(a_1^i, a_2^i, \dots, a_n^i)$ обладает свойством $a_j^i = a_{j+1}^i \quad \forall j \in N_{n_i-1} \setminus N_{n_i-m_i}, i \in N_s$, то в подсистеме системы (1), (2) достаточно оставить только неравенства первой, второй, ..., $(m_i - j)$ -ой групп. Доказательство завершено.

При отображении множества полиразмещений $P_{qk}^{ns}(A, H)$ в евклидово пространство R^n сформулируем задачу $Z(F, X)$ максимизации некоторого векторного критерия $F(x)$ на множестве X , причем каждой точке $a \in P_{qk}^{ns}(A, H)$ будет соответствовать точка $x \in X$, такая, что $F(x) = \Phi(a)$.

$$Z(F, X) : \max \{F(x) \mid x \in X\},$$

где $F(x) = (f_1(x), f_2(x), \dots, f_l(x))$ соответствует функционалу $\Phi_i(a)$, $f_i : R^n \rightarrow R^1, i \in N_l$, X - непустое множество, которое определяется следующим образом: $X = \text{vert } \Pi_{qk}^{ns}(A, H)$, где $\Pi_{qk}^{ns}(A, H) = \text{conv } P_{qk}^{ns}(A, H) = \Pi$. Под соответствием векторной функции F вектору функционалов Φ будем понимать соотношение: $\Phi(a) = F(\varphi(a)) \forall a \in P_{qk}^{ns}(A, H)$.

Задача $Z(F, X)$ может содержать также дополнительные линейные ограничения, образующие выпуклое многогранное множество $D \subset R^n$ вида: $D = \{x \in R^n \mid Bx \leq d\}$, где $B \in R^{m \times n}, d \in R^m$. Следовательно допустимое множество имеет вид: $X = \text{vert } \Pi_{qk}^{ns}(A, H) \cap D$.

Разработано множество различных принципов принятия решений в таких задачах. Наиболее традиционные из них связаны с выделением из всего множества $Y = \{y = F(x) \mid x \in \Pi_{qk}^{ns}(A, H) \cap D\}$ множества неулучшаемых или оптимальных по Парето, оптимальных по Слейтеру, оптимальных по Смейлу векторов.

Таким образом, под решением задачи $Z(F, X)$ будем понимать нахождение некоторого подмножества одного из следующих множеств: $P(F, X)$ – множества Парето-оптимальных (эффективных решений), $Sl(F, X)$ – оптимальных по Слейтеру (слабо эффективных) решений, $Sm(F, X)$ – оптимальных по Смейлу (строго эффективных) решений. Напомним [4, 10], что точка $x^* \in X$ называется эффективной (или Парето-оптимальной), если $\exists x \in X : F(x) \geq F(x^*), F(x) \neq F(x^*)$; слабо эффективной (оптимальной по Слейтеру), если $\exists x \in X : F(x) > F(x^*)$ и строго эффективной (оптимальной по Смейлу), если $\exists x \in X : x \neq x^*, F(x) \geq F(x^*)$.

Из приведенных определений этих множеств следует справедливость соотношений между ними: $Sm(F, X) \subset P(F, X) \subset Sl(F, X)$.

Как известно, множество $P(F, X)$ Парето-оптимальных решений не пусто, поскольку допустимая область X ограничена, и внешне устойчиво [10]: $\forall y \in \Pi_{qk}^{ns}(A, H) \exists x \in P(F, X) : F(x) \geq F(y)$.

При построении метода решения многокритериальной задачи $Z(F, X)$ следует учитывать структурные особенности ее допустимой области, т.е. свойства общего многогранника полиразмещений и довольно большое число описывающих его ограничений. Условия различных видов оптимальности решений и общий подход к нахождению слабо эффективных и Парето-оптимальных решений на основе использования представленных структурных свойств множеств эффективных решений, разработаны в работе Н.В. Семеновой в этом номере данного журнала.

Выводы

В статье исследованы сложные комбинаторные многокритериальные задачи на множестве полиразмещений. Рассмотрены некоторые свойства допустимой области комбинаторной многокритериальной задачи, погруженной в арифметическое евклидово пространство, являющейся общим многогранником полиразмещений. Полученные результаты в определенном смысле обобщают и развивают свойства изученного ранее общего многогранника размещений и являются необходимыми и важными для построения различных методов решения указанных классов задач. Дальнейшее развитие данной работы будет направлено на исследование структурных свойств других сложных комбинаторных множеств и на их основе разработку новых методов решения многокритериальных задач комбинаторной оптимизации.

Бібліографія

- [1] Сергиенко И.В. Математические модели и методы решения задач дискретной оптимизации. – К.: Наук. думка, 1988. – 472 с.
- [2] Сергиенко И.В., Каспшицкая М.Ф. Модели и методы решения на ЭВМ комбинаторных задач оптимизации. – Киев – Наук. думка, 1981. – 287 с.
- [3] Сергиенко И.В., Шило В.П. Задачи дискретной оптимизации: проблемы, методы решения, исследования. – К.: Наук. думка, 2003. – 264 с.
- [4] Сергиенко И.В., Козерацкая Л.Н., Лебедева Т.Т. Исследование устойчивости и параметрический анализ дискретных оптимизационных задач. – Киев: Наук. думка, 1995. – 170 с.
- [5] Сергиенко И.В., Лебедева Т.Т., Семенова Н.В. О существовании решений в задачах векторной оптимизации // Кибернетика и системный анализ. – 2000. - №6 – С. 39 – 46.
- [6] Лебедева Т.Т., Семенова Н.В., Сергиенко Т.І., Умови оптимальності та розв'язуваності в задачах лінійної векторної оптимізації з опуклою допустимою множиною // Доповіді НАНУ. – 2003. – №10 – С. 80–85.
- [7] Лебедева Т.Т., Семенова Н.В., Сергиенко Т.І. Устойчивость векторных задач целочисленной оптимизации: взаимосвязь с устойчивостью множеств оптимальных и неоптимальных решений // Кибернетика и системный анализ. – 2005. – №4 – С. 90–100.
- [8] Семенова Н.В., Колечкина Л.Н., Нагорная А.Н. Подход к решению векторных задач дискретной оптимизации на комбинаторном множестве перестановок // Кибернетика и системный анализ – 2008. – №3 – С. 158–172.
- [9] Semenova N.V., Kolechkina L.M., Nagirna A.M. Vector combinatorial problems in a space of combinations with linear fractional functions of criteria // Intern. Journal "Information Theories and Applications", 15. – 2008. – P. 240 – 245.
- [10] Подиновский В.В., Ногин В.Д. Парето-оптимальные решения многокритериальных задач. – М.: Наука, 1982. – 256 с.
- [11] Стоян Ю.Г., Яковлев С.В. Математические модели и оптимизационные методы геометрического проектирования. – Киев: Наук. думка, 1986. – 265 с.
- [12] Стоян Ю.Г., Ємець О.О. Теорія і методи евклідової комбінаторної оптимізації. – К.: Ін-т систем. досліджень освіти, 1993. – 188 с.
- [13] Ємець О. О., Колечкіна Л. М. Задачі комбінаторної оптимізації з дробово-лінійними цільовими функціями.– К.: Наукова думка. – 2005.– 118 с.
- [14] Ємець О.О., Роскладка О.В. Задачі оптимізації на полікомбінаторних множинах: властивості та розв'язання. – Полтава: РВЦ ПУСКУ, 2006. – 130с.
- [15] Емеличев В.А., Ковалев М.М., Кравцов М.К. Многогранники, графы, оптимизация. – М.: Наука, 1981. – 344с.
- [16] Ермольев Ю.М., Ляшко И.И., Михалевич В.С., Тюття В.И. Математические методы исследования операций. – К.: Вища школа, 1979. – 312 с.

Інформація об авторе

Людмила Николаевна Колечкина – Інститут кибернетики ім. В.М. Глушкова НАН України, канд. физ.-мат. наук, доцент, докторант, 03680 МСП Киев 187, проспект академіка Глушкова, 40, Україна; e-mail: ludapl@ukr.net

ВЕКТОРНЫЕ ЗАДАЧИ НА КОМБИНАТОРНОМ МНОЖЕСТВЕ ПОЛИРАЗМЕЩЕНИЙ: УСЛОВИЯ ОПТИМАЛЬНОСТИ И ПОДХОД К РЕШЕНИЮ

Наталия Семенова

Резюме: Рассматривается многокритериальная задача дискретной оптимизации на комбинаторном множестве полиразмещений. Исследуются структурные свойства множеств эффективных решений. Получены необходимые и достаточные условия различных видов оптимальности решений. На основе развития идей евклидовой комбинаторной оптимизации, методов главного критерия, декомпозиции, отсекающих плоскостей Келли, релаксации разработаны и обоснованы возможные подходы для решения многокритериальной комбинаторной задачи на множестве полиразмещений.

Ключевые слова: многокритериальная оптимизация, дискретная оптимизация, Парето-оптимальные, слабо, строго эффективные решения, комбинаторное множество полиразмещений.

ACM Classification Keywords: G2.1 Combinatorics (F2.2), G1.6 Optimization

Введение

Широкое распространение многокритериальных оптимизационных моделей при решении важных задач экономики, проектирования сложных систем, принятия решений в условиях неопределенности и других в последние десятилетия стимулировало внимание многих специалистов к изучению разнообразных аспектов теории векторной, в том числе дискретной, оптимизации [1-10]. Математические модели дискретной оптимизации охватывают широкий круг прикладных задач. Кроме того, ряд теоретических проблем самой математики может быть сформулирован в виде дискретных задач, поэтому все большее возрастает необходимость интенсивного развития теории и методов поиска решений задач дискретной оптимизации как неотъемлемой части научного фундамента для построения современных информационных технологий и систем. В связи с этим изучение некоторых свойств решений векторных дискретных задач, получение необходимых и достаточных условий различных видов оптимальности решений является актуальной проблемой, так как знание таких свойств и условий дает основу для разработки способов проверки оптимальности выбранного решения и построения высокоэффективных методов отыскания различных множеств эффективных решений.

В последнее время в области исследования многих классов комбинаторных моделей, разработки новых методов их решения большое внимание уделяется методам, основанным на использовании структурных свойств комбинаторных множеств [2, 8-14]. Использование информации о структуре выпуклой оболочки допустимых решений, которая является основанием для многих многогранных методов, – один из самых успешных на данный момент подходов к решению задач комбинаторной оптимизации. Но при решении таких задач возникают проблемы, связанные со сложностью математических моделей, большим объемом информации и др. На сегодняшний день в области исследования различных классов комбинаторных моделей и разработки новых методов их решения получены существенные результаты.

В данной работе формулируется и исследуется актуальная и качественно новая задача, сочетающая многокритериальность альтернатив и комбинаторные свойства полиразмещений допустимого множества.

Она продолжает и развивает исследования многокритериальных задач на комбинаторных множествах перестановок, сочетаний, представленные в работах [8, 9] и работе Л.Н. Колечкиной в этом номере журнала. На основании установленной взаимосвязи между многокритериальными задачами на

комбинаторных множествах и оптимизационными задачами на непрерывном допустимом множестве в данной статье изучены некоторые структурные свойства множеств различных видов эффективных решений, а также сформулировано и доказано ряд теорем об условиях оптимальности решений рассмотренных задач. Для векторных задач на комбинаторном множестве полиразмещений предложен один из возможных подходов к их решению, основанный на развитии метода главного критерия для рассматриваемого класса задач, идеях декомпозиции, отсекающих плоскостей Келли, релаксации, и учитывающий довольно большое число ограничений.

Постановка задачи

При отображении множества полиразмещений $P_{qk}^{ns}(A, H)$ в евклидово пространство R^n сформулируем задачу $Z(F, X)$ максимизации некоторого векторного критерия $F(x)$ на множестве X , причем каждой точке $a \in P_{qk}^{ns}(A, H)$ будет соответствовать точка $x \in X$, такая, что $F(x) = \Phi(a)$.

$Z(F, X) : \max \{F(x) \mid x \in X\}$, где $F(x) = (f_1(x), \dots, f_l(x))$, $f_i : R^n \rightarrow R^1, i \in N_l, X = \text{vert } \Pi_{qk}^{ns}(A, H)$,

$X \neq \emptyset, \Pi_{qk}^{ns}(A, H) = \text{conv } P_{qk}^{ns}(A, H), P_{qk}^{ns}(A, H)$ – общий многогранник полиразмещений, построение и свойства которого описаны в работе Л.Н. Колечкиной в этой книге.

Задача $Z(F, X)$ может содержать также дополнительные линейные ограничения, образующие выпуклое многогранное множество $D \subset R^n$ вида: $D = \{x \in R^n \mid Bx \leq d\}$, где $B \in R^{m \times n}, d \in R^m$.

Следовательно ее допустимое множество имеет вид: $X = \text{vert } \Pi_{qk}^{ns}(A, H) \cap D$.

Под решением задачи $Z(F, X)$, как обычно, будем понимать нахождение элементов одного из следующих множеств: $P(F, X)$ - множества Парето-оптимальных (эффективных решений), $Sl(F, X)$ - оптимальных по Слейтеру (слабо эффективных) решений, $Sm(F, X)$ - оптимальных по Смейлу (строго эффективных) решений. Согласно [4, 6, 7] для каждого $x \in X$ справедливы утверждения:

$$x \in Sl(F, X) \Leftrightarrow \{y \in X \mid F(y) > F(x)\} = \emptyset, \quad (1)$$

$$x \in P(F, X) \Leftrightarrow \{y \in X \mid F(y) \geq F(x), F(y) \neq F(x)\} = \emptyset, \quad (2)$$

$$x \in Sm(F, X) \Leftrightarrow \{y \in X \mid y \neq x, F(y) \geq F(x)\} = \emptyset, \quad (3)$$

$$Sm(F, X) \subset P(F, X) \subset Sl(F, X). \quad (4)$$

Поскольку допустимая область X ограничена, то множество $P(F, X)$ не пусто и внешне устойчиво [10]. В случае бесконечного мультимножества A вопрос о существовании элементов множества $P(F, X)$ требует отдельного исследования.

Структурные свойства и условия оптимальности различных множеств эффективных решений

Использование структурных свойств комбинаторных множеств при погружении в арифметическое евклидово пространство позволяет предложить оригинальные подходы к решению соответствующих

оптимизационных задач, а также повысить эффективность традиционных методов комбинаторной оптимизации.

Теорема 1. Элементы множеств $Sm(F, X)$, $P(F, X)$, $Sl(F, X)$ многокритериальной комбинаторной задачи $Z(F, X)$ находятся в вершинах многогранника $\Pi_{qk}^{ns}(A, H)$ полиразмещений.

Доказательство. Учитывая соотношение (4) между введенными множествами эффективных решений и тот факт, что множество допустимых решений X является подмножеством множества вершин общего многогранника $\Pi_{qk}^{ns}(A, H)$ полиразмещений, т. е. $x \in \text{vert } \Pi_{qk}^{ns}(A, H)$, а $\Pi_{qk}^{ns}(A, H) = \text{conv} P_{qk}^{ns}(A, H)$

приходим к справедливости включений: $Sm(F, X) \subset P(F, X) \subset Sl(F, X) \subset \text{vert } \Pi_{qk}^{ns}(A, H)$.

Пусть функции $f_i(x), i \in N_l$, векторного критерия $F(x)$ – линейны, то есть $f_i(x) = \langle c_i, x \rangle, i \in N_l$.

Структурные свойства допустимой области X и множеств разных видов эффективных решений, отмеченные в теореме 1, а также линейность функций векторного критерия позволяют свести решение задачи $Z(F, X)$ к решению задаче $Z(F, G)$ определенной на непрерывном допустимом множестве

$G = \Pi_{qk}^{ns}(A, H) \cap D$. Пусть $C \in R^{n \times l}$ – матрица, где c_j – ее вектор – строка, $i \in N_l$. Многогранник

$\Pi_{qk}^{ns}(A, H)$ представим в виде $\Pi_{qk}^{ns}(A, H) = \{x \in R^n \mid \langle \pi_i, x \rangle \leq \gamma_i, i \in N_q\}$, сведя все неравенства к

одному знаку (\leq). Введем к рассмотрению конус $K = \{x \in R^n \mid Cx \geq 0\}$ перспективных направлений [3]

задачи $Z(F, X)$ и выпуклые замкнутые конусы $0^+ \Pi(y) = \{x \in R^n \mid \pi_i x \leq 0, i \in N(y)\}$, где

$N(y) = \{i \in N_q \mid \pi_i y = \gamma_i\}$, которые могут быть построены для всех точек $y \in \text{vert } \Pi_{qk}^{ns}(A, H)$.

Очевидно, что $N(y) \neq \emptyset$, $X \subseteq y + 0^+ \Pi(y)$. Обозначим $K_0 = \{x \in R^n \mid Cx = 0\}$ – ядро отображения

$C: R^n \rightarrow R^l$, $\text{int } K = \{x \in R^n \mid Cx > 0\}$ – внутренность конуса K . Из формул (1) - (3) следует

справедливость утверждений

$$x \in Sl(C, X) \Leftrightarrow (x + \text{int } K) \cap X = \emptyset, \quad (5)$$

$$x \in P(C, X) \Leftrightarrow x + (K \setminus K_0) \cap X = \emptyset \quad (6)$$

$$x \in Sm(C, X) \Leftrightarrow (x + K) \cap X \setminus \{x\} = \emptyset. \quad (7)$$

Теорема 2. $P(F, G) \cap \text{vert } \Pi_{qk}^{ns}(A, H) \subset P(F, X)$, $Sl(F, G) \cap \text{vert } \Pi_{qk}^{ns}(A, H) \subset Sl(F, X)$,

$Sm(F, G) \cap \text{vert } \Pi_{qk}^{ns}(A, H) \subset Sm(F, X)$.

Доказательство. Поскольку $\text{vert } \Pi_{qk}^{ns}(A, H) \cap D \subset G$, то

$P(F, G) \cap \text{vert } \Pi_{qk}^{ns}(A, H) \cap D \subset P(F, G \cap \text{vert } \Pi_{qk}^{ns}(A, H) \cap D) = P(F, X)$. Аналогично можно доказать

соотношения $Sm(F, X) = Sm(F, D \cap \text{vert } \Pi_{qk}^{ns}(A, H)) \supset Sm(F, G) \cap \text{vert } \Pi_{qk}^{ns}(A, H)$ и

$Sl(F, X) = Sl(F, \text{vert } \Pi_{qk}^{ns}(A, H) \cap D) \supset Sl(F, G) \cap \text{vert } \Pi_{qk}^{ns}(A, H)$.

Теорема 3. Если допустимое множество X задачи $Z(F, X)$ не содержит ограничений, описывающих выпуклое многогранное множество D , или $\Pi_{qk}^{ns}(A, H) \subseteq D$, то есть $X = \text{vert } \Pi_{qk}^{ns}(A, H)$, то $\forall x \in R^n$

справедливы утверждения: $x \in Sl(F, X) \Leftrightarrow x \in Sl(F, \Pi_{qk}^{ns}(A, H)) \cap \text{vert } \Pi_{qk}^{ns}(A, H)$,

$x \in P(F, X) \Leftrightarrow x \in P(F, \Pi_{qk}^{ns}(A, H)) \cap \text{vert } \Pi_{qk}^{ns}(A, H)$,

$x \in Sm(F, X) \Leftrightarrow x \in Sm(F, \Pi_{qk}^{ns}(A, H)) \cap \text{vert } \Pi_{qk}^{ns}(A, H)$.

Доказательство. Учитывая условия данной теоремы и теоремы 2, следует, что $\forall x \in R^n$ справедливы утверждения: $x \in Sl(F, \Pi_{qk}^{ns}(A, H)) \cap \text{vert } \Pi_{qk}^{ns}(A, H) \Rightarrow x \in Sl(F, X)$,

$$x \in P(F, \Pi_{qk}^{ns}(A, H)) \cap \text{vert } \Pi_{qk}^{ns}(A, H) \Rightarrow x \in P(F, X),$$

$$x \in Sm(F, \Pi_{qk}^{ns}(A, H)) \cap \text{vert } \Pi_{qk}^{ns}(A, H) \Rightarrow x \in Sm(F, X).$$

Докажем обратные импликации. Пусть $x \in Sl(F, X)$, откуда согласно теореме 1 следует, что $x \in \text{vert } \Pi_{qk}^{ns}(A, H)$. Предположим, от противного, что $x \notin Sl(F, \Pi_{qk}^{ns}(A, H))$. Учитывая линейность функций $f_i(x), i \in N_l$, векторного критерия $F(x)$ согласно теореме 5 [6] выполняется условие $\text{int } K \cap (\Pi(x) - x) \neq \emptyset$, то есть в конусе $(x + \text{int } K)$ лежат некоторые точки границы многогранника $\Pi_{qk}^{ns}(A, H)$, следовательно существует точка, которая принадлежит этому конусу. Последнее в силу формулы (5) означает, что $x \notin Sl(F, X)$ и приводит к противоречию с условием теоремы. Остальные утверждения данной теоремы доказываются аналогично этому. Доказательство завершено.

Следствие. При условиях теоремы 3 справедливы равенства $P(F, \Pi_{qk}^{ns}(A, H)) \cap \text{vert } \Pi_{qk}^{ns}(A, H) =$
 $= P(F, X), Sl(F, \Pi_{qk}^{ns}(A, H)) \cap \text{vert } \Pi_{qk}^{ns}(A, H) = Sl(F, X),$

$$Sm(F, \Pi_{qk}^{ns}(A, H)) \cap \text{vert } \Pi_{qk}^{ns}(A, H) = Sm(F, X).$$

Если в задаче допустимая область $X = \text{vert } \Pi_{qk}^{ns}(A, H)$, то для любой точки $x \in \text{vert } \Pi_{qk}^{ns}(A, H)$ справедливы необходимые и достаточные условия оптимальности всех указанных выше видов эффективных решений задачи $Z(F, X)$, полученные в работе [6].

Если допустимая область X задачи $Z(F, X)$ имеет вид $X = \text{vert } \Pi_{qk}^{ns}(A, H) \cap D$ и $\Pi_{qk}^{ns}(A, H) \cap D \neq \Pi_{qk}^{ns}(A, H)$, то справедливы лишь достаточные условия оптимальности решений.

Теорема 4. $\forall x \in \text{vert } \Pi_{qk}^{ns}(A, H) : x \in P(F, \Pi_{qk}^{ns}(A, H)) \cap D \Rightarrow x \in P(F, X)$,

$$x \in Sl(F, \Pi_{qk}^{ns}(A, H)) \cap D \Rightarrow x \in Sl(F, X), \quad x \in Sm(F, \Pi_{qk}^{ns}(A, H)) \cap D \Rightarrow x \in Sm(F, X).$$

Доказательство. Поскольку $G = \Pi \cap D$, то $\forall x \in \text{vert } \Pi_{qk}^{ns}(A, H)$ справедливы импликации $x \in P(F, \Pi_{qk}^{ns}(A, H)) \cap D \Rightarrow x \in P(F, \Pi_{qk}^{ns}(A, H) \cap D) = P(F, G) \Rightarrow x \in P(F, X)$,

$x \in Sl(F, \Pi_{qk}^{ns}(A, H)) \cap D \Rightarrow x \in Sl(F, X)$, $x \in Sm(F, \Pi_{qk}^{ns}(A, H)) \cap D \Rightarrow x \in Sm(F, X)$.

Если задача $Z(F, X)$ не содержит линейных ограничений, образующих выпуклое многогранное множество $D \subset R^n$, либо $\Pi \subseteq D$, т.е. $X = \text{vert } \Pi$, то с учетом необходимых и достаточных условий оптимальности (теорема 3) процесс ее решения сводится к поиску эффективных решений задачи $Z(F, G)$ на непрерывном допустимом множестве $G = \Pi_{qk}^{ns}(A, H)$ с последующим выбором из них лишь тех, которые являются вершинами многогранника $\Pi_{qk}^{ns}(A, H)$ полиразмещений.

Анализируя теоремы 2 и 4, приходим к соотношениям между задачами $Z(F, X)$ и $Z(F, G)$: если $x \in R(F, G) \cap \text{vert } \Pi_{qk}^{ns}(A, H)$, то $x \in R(F, X)$, если $x \notin R(F, G) \cap \text{vert } \Pi_{qk}^{ns}(A, H)$, то из этого не следует, что $x \notin R(F, X)$, где $R(F, X)$ обозначает множество $P(F, X)$, $Sm(F, X)$ или $Sl(F, X)$.

Таким образом, теоремы 1 – 4 устанавливают взаимосвязь между задачей $Z(F, X)$ и задачей $Z(F, G)$, определенной на непрерывном допустимом множестве. Это дает возможность применять классические методы непрерывной оптимизации к решению векторных комбинаторных задач на полиразмещениях и на этой основе развивать новые оригинальные методы решения, используя свойства комбинаторных множеств и их выпуклых оболочек.

Общий подход к решению векторных задач на комбинаторном множестве полиразмещений

1. Находим эффективные решения задачи $Z(F, \Pi_{qk}^{ns}(A, H))$.

2. Проверяем их на принадлежность множеству D . Если $x \in P(F, \Pi_{qk}^{ns}(A, H)) \cap D$, то $x \in P(F, X)$.

3. Рассмотрим допустимые решения $x \in X$ задачи $Z(F, X)$, которые являются неэффективными в задаче $Z(F, \Pi_{qk}^{ns}(A, H))$, то есть $x \in X \setminus (P(F, \Pi_{qk}^{ns}(A, H)) \cap D)$ и проверяем их на эффективность.

Для этого воспользуемся необходимыми и достаточными условиями, сформулированными в [10].

Утверждение 1. Допустимое решение x^0 эффективно тогда и только тогда, когда оно является оптимальным решением задачи $Z^1(F, X) : \max \left\{ \sum_{i=1}^m f_i(x) \mid x \in X, f_i(x) \geq f_i(x^0), i \in N_m \right\}$.

Если решение x^0 не эффективно, то в результате решения этой задачи находим эффективное решение x^* , которое более предпочтительно, чем x^0 , т.е. $F(x^*) \geq F(x^0)$.

Продолжая исследования и развивая результаты работ [1, 5, 6, 8 – 13], в данной статье предложен и обоснован подход к решению задачи $Z(F, X)$, основанный на развитии метода главного критерия для рассматриваемого класса векторных задач. Метод решения однокритериальных задач основан на идеях декомпозиции, отсекающих плоскостей Келли, релаксации, при реализации которого учитывается тот

факт, что число ограничений довольно большое. Целесообразным в этом случае является использование процедуры релаксации, т. е. временного отбрасывания некоторых ограничений и решения задачи на более широкой области при оставшихся ограничениях. При построении метода решения многокритериальной задачи $Z(F, X)$ учитываются структурные особенности ее допустимой области.

При разработке метода на начальном этапе необходимо определить исходное решение, что позволит правильно выбрать активные ограничения - неравенства.

Утверждение 2. Если для элементов мультимножества A и коэффициентов $c_j, j \in N_n$, целевой

функции задачи $\text{extr} \left\{ f(x) = \sum_{j=1}^n c_j x_j \mid x \in \text{vert } \Pi_{qk}^{ns}(A, H) \right\}$ выполняются соответственно условия

$a_1 \leq a_2 \leq \dots \leq a_n$ и $c_{i_1} \leq c_{i_2} \leq \dots \leq c_{i_n}, i_n \in N_n$, то максимум функции $f(x)$ на допустимом множестве достигается в точке $z = (z_{i_1}, \dots, z_{i_n}) \in \text{vert } \Pi_{qk}^{ns}(A, H)$, которая определяется как $z_{i_j} = a_j \forall j \in N_n$,

а минимум – соответственно в точке $y = (y_{i_1}, \dots, y_{i_n})$, где $y_{i_{j+1}} = a_{n-j} \forall j \in N_{n-1} \cup \{0\}$.

Справедливость данного утверждения очевидна, так как наибольшее значение суммы попарных произведений достигается при сопоставлении возрастающей последовательности c_i и возрастающей последовательности z_i , а наименьшее значение суммы, соответственно достигается при сопоставлении возрастающей последовательности c_i и убывающей последовательности $y_i, i \in N_n$.

Для решения задачи $Z(F, X)$ необходимо учесть на начальном этапе лишь часть ограничений, которые определяют область X . Поскольку определение эффективных решений задачи $Z(F, X)$ является более важным, чем построение всего множества ограничений, описывающих допустимую область G , поэтому достаточно построить только те ограничения множества G , которые определяют эффективные решения данной задачи. Рассматриваемый здесь метод предназначен для получения таких ограничений.

Для описания и обоснования метода решения задачи введем следующие обозначения. Допустимую область задачи $Z(F, G)$ запишем в виде $G = \{x \in R^n \mid Hx \leq g\}$, $g = (g_1, g_2, \dots, g_u)$, $H \in R^{u \times n}$, H - матрица, которая используется для матрично-векторной формы записи ограничений вида (1), (2) и линейных неравенств, описывающих многогранник D , где все ограничения сведены к одному (\leq) виду неравенств. Обозначим N_u множество, элементы которого определяют номера ограничений системы (1), (2) и дополнительных ограничений, описывающих выпуклое многогранное множество D :

$N_q = \{1, 2, \dots, 2^n + m\}$. Определим множества $G_i = \{x \in R^n \mid \langle h_i, x \rangle \leq g_i\}, i \in N_u; \forall x^s \in R^n$ определим множества $N^a(x^s) = \{i \in N_q \mid \langle h_i, x^s \rangle = g_i\}$ и $N^n(x^s) = \{j \in N_q \mid \langle h_j, x^s \rangle < g_j\}$ – соответственно

активных и неактивных ограничений в точке x^s ; $h_i \in R^n, g_i \in R, i \in N_q$ - соответственно i -я вектор-строка матрицы H и i -я компонента вектора g . Введем в рассмотрение задачу $Z(F, G^v): \max \{F(x) \mid x \in G^v\}$, где $G^v = \{x \in R^n \mid \langle h_i, x \rangle \leq g_i, i \in Q_v \subset N_q\}$, Q_v - множество индексов ограничений, описывающих ее допустимую область на v -м шаге алгоритма, $Q_v = N_q \setminus R_v$, R_v - множество номеров ограничений, которые не включены в эту задачу на v -м шаге.

Определение. Величина $r_i(x) = \langle h_i, x \rangle - g_i$, $i \in N_q$, называется отклонением точки $x \in R^n$ от границы множества G_i , а величина $r(x) = \max \{r_i(x) \mid i \in N_q\}$ называется отклонением точки $x \in R^n$ от границы множества G . Очевидно, что для $i \in N_p$ $r_i(x) = \sum_{j=1}^i x_{\alpha_j} - \sum_{j=1}^i a_j^i$, а для $i \in N_q \setminus N_p$ $r_i(x) = \langle b_i, x \rangle - d_i$, где b_i - i -я вектор-строка матрицы B , $d_i \in R$.

Теорема 5. Эффективное (Парето-оптимальное, слабо, строго эффективное) решение x^0 задачи $Z(F, G^V)$ является эффективным в том же смысле решением задачи $Z(F, G)$ тогда и только тогда, когда выполняется условие $r(x^0) \leq 0$.

Доказательство. Необходимость утверждения очевидна, поскольку допустимое решение x^0 задачи $Z(f, G^V)$ является допустимым решением задачи $Z(F, G)$ тогда и только тогда, когда выполняется условие $r(x^0) \leq 0$. Достаточность утверждения следует из определения задачи $Z(F, G)$ и $r(x)$.

Для рассматриваемого класса векторных задач на комбинаторном множестве полиразмещений предлагается метод главного критерия. Он состоит в том, что исходная многокритериальная задача $Z(F, X)$ сводится к задаче оптимизации по одному критерию $f_r(x)$, $r \in N_l$, который объявляется главным или основным, при условии, что значения всех остальных критериев должны быть не меньше некоторых установленных величин (пороговых значений) t_i , $i \in N_l \setminus \{r\}$. Таким образом, имеем задачу

$$Z(f_r, X(t_i)) : \max \{f_r(x) \mid f_i(x) \geq t_i, i \in N_l \setminus \{r\}, x \in X\}.$$

Оптимальное решение x^0 этой задачи всегда является слабо эффективным, а если оно единственно (с точностью до эквивалентности \square_f), то и эффективным. Если решение x^0 эффективно, то оно является единственным (с точностью до эквивалентности \square_f) решением задачи $Z(f_r, X(t_i))$ при любом фиксированном $r \in N_l$ и $t_i = f_i(x^0)$, $i \in N_l \setminus \{r\}$. Выбор одного из критериев в качестве главного никак не уменьшает свободы выбора оптимального решения. Для определения пороговых значений t_i , $i \in N_l \setminus \{r\}$ воспользуемся утверждением 2, которое дает возможность установления верхних и нижних границ значений критериев $f_i(x)$, $i \in N_l$, на множестве полиразмещений. Предлагается два подхода к решению исходной задачи $Z(F, X)$. Первый состоит в назначении порогов t_i , $i \in N_l \setminus \{r\}$ максимально возможных значений критериев $f_i(x)$, $i \in N_l$, на множестве полиразмещений с последующим расширением допустимого множества задачи $Z(f_r, X)$, если исходная задача окажется недопустимой, а в случае ее допустимости нахождение эффективного либо слабо эффективного решения. Второй подход предполагает поиск оптимального решения задачи $Z(f_r, X(t_i))$ при назначении минимальных значений критериев $f_i(x)$, $i \in N_l$ с последующим сужением допустимой области посредством выбора значений порогов t_i , $i \in N_l \setminus \{r\}$ следующих, упорядоченных по возрастанию, за минимальными значениями критериев. Процедура назначения серии пороговых величин t_i ограничений и в первом и во втором подходе очень проста, так как используя утверждение 2, она сводится после упорядочения коэффициентов критериев к вычислению скалярного произведения двух векторов, т.е. к определению значений линейных критериев. При этом, учитывая структурные особенности множества

полиразмещений, величины t_i можно вычислять более эффективно, используя перестановки элементов каждого i -го, $i \in N_S$, подмножества мультимножества A .

Общая идея предложенного метода решения задачи $Z(F, X)$ состоит в последовательном включении ограничений задачи, описывающих область допустимых решений.

Алгоритм

1. Сводим многокритериальную задачу $Z(F, G)$ к однокритериальной задаче $Z(f, G)$ методом главного критерия. Полагаем $\nu = 0$. Выбираем главный критерий $f_r(x)$, $r \in N_I$. На основании утверждения 2 назначаем величины t_i , $i \in (N_I \setminus \{r\})$, для ограничений на остальные критерии. Выбираем ограничения, определяющие область $G^\nu \subset G$, начальной итемы.

2. Находим оптимальное решение x^ν задачи $Z(f_r, G^\nu(t_i))$ с помощью симплекс-метода.

3. Если полученное оптимальное решение – точка множества полиразмещений, то в найденной точке x^ν проверяем выполнение ограничений, которые не были учтены. Очевидно, ими могут быть лишь те ограничения, описывающие выпуклое многогранное множество D . Если решение x^ν не удовлетворяет этим ограничениям, то добавляем к ограничениям задачи $Z(f_r, G^\nu(t_i))$ наиболее нарушенное, из описывающих многогранник D . Если решение x^ν удовлетворяет указанным ограничениям, то оно является эффективным решением задачи $Z(F, G)$ и, следовательно, задачи $Z(F, X)$.

4. Если полученное решение x^ν не является точкой множества полиразмещений, то строим отсечение, проходящее через смежные вершины и отсекающую его. Прибавляем это отсечение к ограничениям задачи $Z(f_r, G^\nu(t_i))$.

5. Сравниваем значение $f(x^\nu)$ со значением целевой функции, найденным на предыдущем шаге. Если оно уменьшается, то отбрасываем неактивные (несущественные) ограничения в точке x^ν , иначе эти ограничения оставляем. С измененной допустимой областью задачи $Z(f_r, G^\nu(t_i))$ переходим к пункту 2.

То обстоятельство, что в процессе решения задачи ограничения последовательно добавляются и ни одно из ограничений не отбрасывается, если $f(x^\nu)$ остается равным предыдущему значению, гарантирует, что решается только конечное число задач вида $Z(f_r, G^\nu(t_i))$.

Выводы

В статье исследованы сложные комбинаторные многокритериальные задачи на множестве полиразмещений. Получены необходимые и достаточные условия оптимальности различных видов эффективных решений. На основе структурных свойств допустимой области, погруженной в арифметическое евклидово пространство, построен и обоснован метод решения рассмотренного класса задач. Разработанный подход может успешно применяться для решения векторных задач на комбинаторном множестве полиразмещений и предоставляет возможность исследовать и находить элементы множества Парето-оптимальных решений многокритериальных комбинаторных задач с учетом других комбинаторных свойств области допустимых решений. Дальнейшее развитие данной работы будет

направлено на реализацию и адаптацию предложенного алгоритма, а также на разработку новых методов решения векторных задач комбинаторной оптимизации.

Библиография

- [1] Сергиенко И.В. Математические модели и методы решения задач дискретной оптимизации. – К.: Наук. думка, 1988. – 472 с.
- [2] Сергиенко И.В., Каспшицкая М.Ф. Модели и методы решения на ЭВМ комбинаторных задач оптимизации. – Киев – Наук. думка, 1981. – 287 с.
- [3] Сергиенко И.В., Шило В.П. Задачи дискретной оптимизации: проблемы, методы решения, исследования. – К.: Наук. думка, 2003. – 264 с.
- [4] Сергиенко И.В., Козерацкая Л.Н., Лебедева Т.Т. Исследование устойчивости и параметрический анализ дискретных оптимизационных задач. – Киев: Наук. думка, 1995. – 170 с.
- [5] Сергиенко И.В., Лебедева Т.Т., Семенова Н.В. О существовании решений в задачах векторной оптимизации // Кибернетика и системный анализ. – 2000. - №6 – С. 39 – 46.
- [6] Лебедева Т.Т., Семенова Н.В., Сергиенко Т.І., Умови оптимальності та розв'язуваності в задачах лінійної векторної оптимізації з опуклою допустимою множиною // Доповіді НАНУ. – 2003. – №10 – С. 80–85.
- [7] Лебедева Т.Т., Семенова Н.В. Сергиенко Т.І. Устойчивость векторных задач целочисленной оптимизации: взаимосвязь с устойчивостью множеств оптимальных и неоптимальных решений // Кибернетика и системный анализ. – 2005. – №4 – С. 90–100.
- [8] Семенова Н.В., Колечкина Л.Н., Нагорная А.Н. Подход к решению векторных задач дискретной оптимизации на комбинаторном множестве перестановок // Кибернетика и системный анализ – 2008. – №3 – С. 158–172.
- [9] N.V. Semenova, L.M. Kolechkina, A.M. Nagirna. Vector combinatorial problems in a space of combinations with linear fractional functions of criteria // Intern. Journal "Information Theories and Applications", 15. – 2008. – P. 240 – 245.
- [10] Подиновский В.В., Ногин В.Д. Парето–оптимальные решения многокритериальных задач. – М.: Наука, 1982. – 256 с.
- [11] Стоян Ю.Г., Яковлев С.В. Математические модели и оптимизационные методы геометрического проектирования. – Киев: Наук. думка, 1986. – 265 с.
- [12] Стоян Ю.Г., Ємець О.О. Теорія і методи евклідової комбінаторної оптимізації. – К.: Ін-т систем. досліджень освіти, 1993. – 188 с.
- [13] Ємець О. О., Колечкіна Л. М. Задачі комбінаторної оптимізації з дробово-лінійними цільовими функціями.– К.: Наукова думка. – 2005.– 118 с.
- [14] Ємець О.О., Роскладка О.В. Задачі оптимізації на полікомбінаторних множинах: властивості та розв'язання. – Полтава: РВЦ ПУСКУ, 2006. – 130с.

Информация об авторе

Наталія Владимировна Семенова – Інститут кибернетики ім. В.М.Глушкова НАН України, канд. физ.-мат. наук, старший науковий співробітник, 03680 МСП Києв 187, проспект академіка Глушкова, 40, Україна; e-mail: nvsemenova@meta.ua

ПОИСК ОБЩИХ ПРИНЦИПОВ ФОРМАЛИЗАЦИИ В ТЕОРЕТИЧЕСКИХ И ПРИКЛАДНЫХ ИССЛЕДОВАНИЯХ

Петр Василик, Александр Провотар

Аннотация: Обсуждается подход к формализации в различных предметных областях, сопровождающийся использованием качественно отличных логик на разных уровнях дискретизации процессов.

Ключевые слова: множества, противоречия, нейронные сети, обучение нейронных сетей, формализация.

ACM Classification Keywords: F.1.1 Models of Computation – neural networks, C.1.3 Other Architecture Styles – Neural Nets.

Введение

В статье затрагивается вопрос о том, почему формальные модели, используемые при решении различных задач, в одних случаях работают безупречно (получаем ожидаемый результат), в других – приводят к сложным, иногда даже противоречивым выводам. Ответ на этот вопрос (по мнению авторов) нужно искать не в точности (адекватности) моделей, а в ограничениях, которые накладывает дискретный характер происходящих в пространстве и времени процессов (событий). Речь идет о том, что переход от одного дискретного состояния объекта (процесса) к другому зачастую приводит к количественным и качественным изменениям самого объекта (процесса). Это означает, что методы изучения такого процесса на разных уровнях дискретизации должны быть различными. Так как в основе разных методов изучения лежит логический подход, то понятно, что качественно различные дискретные уровни должны обслуживать так называемые качественные логики или логики разных уровней дискретности. По сути, такие логики должны отличаться одна от другой хотя бы тем, что должны существовать утверждения истинные в одной логике и ложные в другой. Исследование объектов при таком подходе осуществляется с помощью логики соответствующего уровня дискретности, которая должна быть построена и верифицирована. Применение же других логик в этом случае может быть причиной появления различных противоречивых утверждений (парадоксов), которые оказывают негативное влияние на процессы познания объектов.

Дальнейшее изложение посвящено аргументации предложенного тезиса. При этом используются на первый взгляд совершенно различные предметные области, которые при более глубоком рассмотрении имеют достаточно общего с точки зрения изложенного выше. Другими словами, возникновение парадоксов и проблем неполноты в рассматриваемых областях (соответственно неадекватности) связано именно с попытками использования логик одного качественного уровня для исследования объектов другого качественного уровня.

Множества и присущие им противоречия

Исследованию парадоксов теории множеств посвящено огромное количество работ. Вместе с тем о причинах их возникновения если и говорится, то очень туманно. Наиболее известные из них – объяснения на основе понятий актуальной и потенциальной бесконечности. Последние понятия сами требуют

аккуратных и точных определений, которые вряд ли можно сформулировать однозначно (или формально).

Почему же они (противоречия) возникают? Рассмотрим простой пример. Ограничимся только конечными множествами. Мощность такого множества, как известно, это количество элементов в нем. Если множества A и B не пересекаются, то справедливо следующее соотношение:

$$|A \cup B| = |A| + |B|.$$

Если одно из этих множеств (например A) счетное, то соотношение будет иметь вид:

$$|A \cup B| = |A|.$$

Следовательно, утверждение "мощность суммы множеств равна сумме мощностей этих множеств" справедливо только в случае конечных множеств. Здесь имеет место сформулированный выше тезис, предполагающий построение новой логики (в данном случае логики операций над множествами) при переходе к бесконечным множествам, которые являются объектами нового качества.

В соответствии с [Василик, 2008] рассмотрим диагональную процедуру Кантора [Кантор, 1985], которая используется при доказательстве некоторых теорем математической логики и теории вычислений. Например, для доказательства несчетности множества всех подмножеств множества натуральных чисел (теорема Кантора) делается предположение о существовании биекции $f: N \rightarrow 2^N$. После этого строится множество A , которое определяется соотношением $x \in A \Leftrightarrow x \notin f(x)$. Так как $f(x)$ – биекция, то $f(m) = A$ для некоторого m . Но в таком случае приходим к противоречию $m \in f(m) \Leftrightarrow m \notin f(m)$, которое (по мнению Кантора) вызвано предположением о существовании биекции, а, следовательно, доказывает теорему.

На самом деле все происходит иначе. Построение множества A будет корректным только в случае определенности множества $f(x)$. Но при $x = m$ множество $f(m) = A$ будет неопределенным. Это означает, что относительно элемента m не можно дать ответ на вопрос $m \in A$ или $m \notin A$. Так как справедливыми будут следующие импликации $m \in A \Rightarrow m \notin f(m) = A$, $m \notin A \Rightarrow m \in A$.

Более глубокий анализ приведенного выше доказательства позволяет сделать вывод о том, что соотношение, определяющее множество A противоречит предположению о биективности отображения f . В самом деле, если $A = f(m)$, то при условии $m \notin A$ соотношение будет определять также и множество $A \cup \{m\}$, а при условии $m \in A$ – также и множество $A \setminus \{m\}$. Поэтому отображение f не только не является биективным, но даже не является функцией.

Направшивается следующий вывод. В случае конечных множеств теорема Кантора справедлива. При переходе к бесконечным множествам предложенное доказательство не представляется корректным. Это значит, что в этом случае (даже если утверждение справедливо) для его доказательства нужно использовать логику другого качества, которая оперирует, соответственно, объектами другого качества.

Еще один пример использования диагональной процедуры Кантора доказательства существования всюду определенной одноместной функции, которая не принадлежит множеству \mathcal{F} всюду определенных одноместных функций (при условии существования универсальной функции для \mathcal{F}) [Мальцев, 1965], тоже приводит к противоречию. А именно, введение новой функции $g(x) = F(x, x) + 1$, где F – универсальная для класса \mathcal{F} , означает, что значение функции g в произвольной точке x можно вычислить как значение выражения $F(x, x) + 1$. Предположение о том, что g принадлежит классу \mathcal{F} означает, что $g(x) = F(i, x)$ для некоторого i . В этом случае получаем противоречие $F(i, i) = F(i, i) + 1$, которое (как утверждается) доказывает, что функция g не принадлежит классу \mathcal{F} .

На самом деле противоречие возникает из-за некорректного (противоречивого) определения функции g . Кроме того, возникшее противоречие вовсе не доказывает, что функция g не принадлежит классу \mathcal{F} .

Действительно, мы имеем два определения функции g . В соответствии с первым $g(i)=F(i,i)$, в соответствии с другим – $g(i)=F(i,i)+1$. Таким образом, функция g в точке i имеет два значения, следовательно, не является функцией. Более того, возникшее противоречие означает, что мы не можем определять функцию g соотношением $g(x)=F(x,x)+1$ при условии, что F – универсальная для класса \mathcal{F} функция. Но только не то, что функция g не принадлежит классу \mathcal{F} . В этом случае причина некорректности доказательства лежит снова в плоскости произвольного использования качественно разных логик.

Аналогичные ситуации возникают и случае рассмотрения алгоритмических систем и даже знаменитой теоремы Геделя о неполноте [Антипенко, 1986].

Общий вывод таков: рассмотренные выше примеры показывают, что уровень формализации логик бесконечных объектов не достиг уровня формализации логик конечных объектов. Использование же не достаточно обоснованных формализаций приводит, как мы видели, к противоречивым построениям. Поэтому проблема в этом случае состоит в построение логики бесконечных объектов, а, следовательно, сводится к проблеме неполноты теоретических построений [Василик, 2006].

Системы искусственного интеллекта и присущая им неполнота

Здесь будут рассматриваться экспертные диагностические системы ГОМЕОПАТ разных версий, в создании которых авторы принимали непосредственное участие [Проватар, 2000]. Одна из особенностей таких систем состоит в том, что они при постановке диагноза явно приближаются к процессу постановки диагноза человеком.

Процесс имеет две стадии:

- 1) ограничение пространства диагностирования путем выбора правдоподобных гипотез заболевания (в медицинской литературе это известно как "дифференциальная модель диагностики");
- 2) применение стратегии диагностики для идентификации заболеваний, которые наиболее полно соответствуют симптомам.

Одна из версий таких систем сначала идентифицирует набор заболеваний, которые характеризуются некоторыми или всеми симптомами, которые есть у пациента. После этого определяет в этом наборе заболевания, которые наилучшим образом соответствуют обнаруженным симптомам. Далее осуществляется поиск заболеваний, которым отвечают нерассмотренные симптомы и процесс повторяется до тех пор, пока не останется ни одного нерассмотренного симптома.

На первой фазе консультации в систему вводятся как позитивные, так и возможные негативные показания. Каждый позитивный симптом вызывает те узлы дерева заболеваний, которые имеют с ним связи, включая не только терминальные сущности заболеваний, а и узлы более высокого уровня, которые задают категории болезней.

В результате введения начальных симптомов и данных относительно

- а) симптомов, которые не имеют отношения к данному заболеванию
- б) симптомов, связанных с данным заболеванием
- в) симптомов, которые должны быть обнаруженными, при условии правильности диагноза, но не были зарегистрированными
- г) симптомов, не обнаруженных у пациента порождается "модель заболевания" для каждого вызванного узла.

В соответствии с особенностями этих моделей система начинает с того, что формирует свой первый дифференциальный диагноз. Набор альтернативных моделей исследуется на другой стадии - принятия заключительного решения.

Процесс принятия решения зависит от числа альтернативных заболеваний. Если количество болезней-кандидатов превышает некоторую фиксированную границу, то применяется стратегия ИСКЛЮЧЕНИЯ, в соответствии с которой задаются вопросы с четко определенным заболеванием. Если нужный симптом отсутствует, то такое заболевание исключается из списка. Для выбора одной болезни из болезней-кандидатов применяется стратегия ДИСКРИМИНАЦИИ.

Как только появляется новая информация, все модели заболеваний переоцениваются и конструируется новая дифференциальная модель диагностики. Такой процесс проходит ряд итераций до тех пор, пока одно или несколько заболеваний не подтвердятся. В этом случае все симптомы помечаются как "объясненные", а все заболевания, которые имеют каузальные отношения с данными заболеваниями, начинают анализироваться.

В результате этой процедуры система может у одного пациента найти несколько заболеваний. Это существенный шаг вперед, поскольку традиционные диагностические системы выходят из предположения о наличии одной болезни.

Таким образом, система действует в соответствии со стратегией конструирования, которую применяет врач-клиницист и последующей проверкой дифференциального диагноза. Такая модель диагноза включает двухступенчатую процедуру, которая сводится к тому, что сначала выдвигаются гипотезы заболевания на основе введения данных о пациенте (процесс "снизу вверх"), а потом осуществляется их оценка с помощью дополнительных симптомов обнаруженных заболеваний (процесс "сверху вниз").

В другой версии системы диагностирование сводится к задаче выбора "наиболее подходящих" объектов из некоторого множества по заданным характеристикам этих объектов с применением элементов теории нечетких множеств.

Как известно, понятие нечеткого множества было введено Л. Заде в работе [Zadeh, 1965]. Принадлежность некоторого элемента такому множеству определяется любым числом единичного интервала $[0;1]$, в отличие от обычных множеств, где элемент либо принадлежит множеству, либо нет. Теория нечетких множеств нашла применение в различных областях компьютерных наук, в частности, в теории представления знаний, методах идентификации и распознавания образов, теории принятия решений (в связи с тем, что не всегда можно точно определить принадлежность объекта некоторому классу или точно сформулировать его свойство), теории оптимизации и др. Теория нечетких множеств дала возможность формализовать неточные и интуитивные утверждения, что оказалось очень полезным при разработке фундаментальных принципов искусственного интеллекта.

Задача выбора "наиболее подходящего" объекта (или объектов) из некоторого множества объектов по заданным характеристикам этих объектов путем построения уровневых множеств и степеней принадлежности для используемых первичных нечетких подмножеств формально описывается следующим образом.

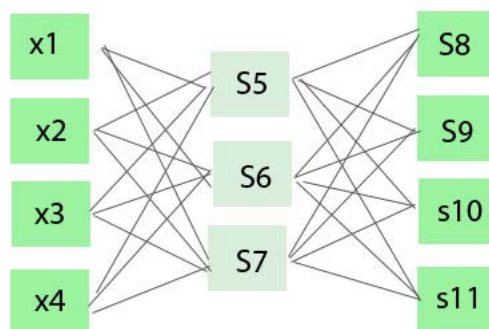
Пусть имеется некоторое множество объектов $X = \{X_1, X_2, \dots, X_k\}$ и некоторое множество их характеристик $C = \{C_1, C_2, \dots, C_n\}$, причем между этими множествами существует отношение $(\infty : \infty)$, т.е. каждый элемент множества X может иметь несколько характеристик, а так же один и тот же элемент множества C может являться характеристикой сразу нескольких объектов. При этом если элементу X_i соответствует элемент C_j то говорим, что объект X_i имеет характеристику C_j .

Задача заключается в том, чтобы по наперед заданным характеристикам определить наиболее подходящий объект (или несколько объектов) с максимальной степенью принадлежности множеству X . Для определения "оптимального" множества объектов используются нечеткие множества [Дудка, 2001].

Удобным инструментом для представления информационных моделей являются нейронные сети (НС). В общем случае, сеть принимает некоторый входной сигнал из внешнего мира, и пропускает его через себя с преобразованиями в каждом нейроне. Таким образом, в процессе прохождения сигнала по связям сети происходит его обработка, результатом которой является определенный выходной сигнал.

Для проектирования нейронной сети в одной из версий системы ГОМЕОПАТ была выбрана наиболее распространенная структура нейронных сетей - многослойная. Эта структура подразумевает что каждый нейрон произвольного слоя связан со всеми выходами (аксонами) нейронов предыдущего слоя или со всеми входами НС в случае первого слоя. Другими словами сеть имеет следующую структуру слоев: входной, промежуточный (скрытый) и выходной. Такие нейронные сети также называют полносвязными [Катеринич, 2007].

Для решения задачи диагностирования использовалась НС следующей архитектуры.



Задача обучения МНС в классическом виде представлялась так. Пусть задана некоторая последовательность x^* входных данных. Необходимо найти такое решение x , при котором можно классифицировать вновь представленные входные данные. Критерий $R(x, x^*)$ определяет качество решения. Множество решений x определяется выбором алгоритма настройки весовых коэффициентов $w_i^{(a)}$. При такой постановки задачи, процесс обучения сводится к получению наилучшего решения из множества возможных. Другими словами, обучение МНС – это процесс накопление информации x^* и параллельно процесс выбора решения x .

НС системы использует алгоритм обратного распространения суть которого заключается в распространении сигналов ошибки от выходов НС к ее входам, в направлении, обратном прямому распространению сигналов в обычном режиме работы (режим распознавания). Другими словами используя технологии последовательной настройки нейронов, начиная с последнего, выходного слоя, и заканчивая настройкой элементов первого слоя. Обучение НС может быть проведено необходимое число раз. Для обучения используется так называемое δ -правило, которое заключается в реализации стратегии обучения «с учителем». Ошибка обучения вычисляется по следующей формуле $\delta = y^* - y$ в алгоритме градиентного спуска (весовой коэффициент) $w_i(k+1) = w_i(k) - \gamma \delta x_i$, $\gamma > 0$, где γ - коэффициент «усиления алгоритма», x_i - i -й вход синаптической связи нейрона, y^* - необходимый выход нейрона, y – реальный выход.

Рассмотренные выше примеры разных версий системы ГОМЕОПАТ, использующие различные модели диагностирования, в большинстве случаев дают достаточно приближенную картину диагноза пациента.

По-видимому, проблема состоит в том, что используемые информационные модели не учитывают каких-то важных показателей (факторов), которые влияют на целостную картину постановки дифференциального диагноза и, следовательно, используемая формальная (информационная) модель не полная. Скорее всего, и это подтверждается многочисленными исследованиями последних лет, необходимо учитывать так называемые "внешние факторы" или внешние воздействия на биологический объект полевых структур разной природы.

Формализуемое и не формализуемое в вычислительных задачах

Общеизвестно, что при решении задач из различных областей человеческой деятельности используется (как правило) некоторый язык, позволяющий формализовать исходную постановку задачи (построить формальную модель) с целью возможного применения формальных методов (если таковые существуют) самого языка для решения поставленной задачи. Таким образом, процесс формализации в данном случае сводится к описанию проблемы (задачи) на некотором формальном языке. Хорошо если этот формальный язык – язык математики, а, следовательно, обладающий развитыми методами решения формализованных задач. В этом случае, обычно, говорят о полной формализации проблемы. Если формальный язык не является языком математики – говорят о неполной или частичной формализации проблемы. При этом не исключается существование другого формального языка, с помощью которого достигается полная или частичная формализация проблемы. Впрочем, возможны ситуации, когда не удается получить формальную модель проблемы. Если такие модели в принципе существуют, то проблема является полностью или частично формализованной. В противном случае говорят о неформализованных предметных областях, т. е. о таких областях, где существуют неформализованные проблемы.

В [Проватар, 2008] предлагается так называемый алгоритмический подход, позволяющий в некотором смысле упорядочить наше представление о формализации. Следствием этого подхода является возможность классифицировать проблемы на допускающие формализацию, слабо формализуемые и неформализуемые.

Актуальность затрагиваемых вопросов состоит в том, что используя предлагаемый подход в каждом конкретном случае, можно дать ответы на вопросы о возможности точного решения задачи на компьютере, приближенного решения задачи с использованием компьютера или невозможности получения ни точного, ни приближенного решения. Они тесным образом связаны с проблемами формальной проверки правильности доказательств в математике и корректности (следовательно, надежности) программного обеспечения, основанных на формальных методах проверки непротиворечивости математических же алгоритмов, лежащих в основе функционирования программ.

Особенности формализуемых моделей согласно [Проватар, 2008] состоят в следующем. Во-первых, в каждой из них можно решить проблему разрешимости. Во-вторых, решение этой проблемы фактически осуществляется построением алгоритмов. В-третьих, каждый из этих алгоритмов обеспечивает точное решение проблемы разрешимости с использованием компьютера.

Для слабо формализованных задач характерно следующее. Во-первых, их решение на соответствующих формальных моделях достигается. Во-вторых, решение проблем фактически осуществляется построением алгоритмов. В-третьих, каждый из алгоритмов обеспечивает приближенное, но приемлемое решение соответствующей проблемы с использованием компьютера.

Наконец, отметим особенности неформализуемых проблем. Во-первых, решение поставленных задач на имеющихся формальных моделях получить не удастся. Во-вторых, попытка решения проблем сводится фактически к построению алгоритмов и формальных моделей входных и выходных данных.

Из вышеизложенного следует, что в качестве языка формализации проблем можно использовать произвольный язык (например, язык математики). Такой язык должен иметь развитые средства для описания дескриптивных (входные и выходные данные) и императивных данных (алгоритм решения

проблемы). Такая характеристика языка очень напоминает описание языка программирования (разница только в том, что современные языки программирования не содержат необходимых средств описания дескриптивных данных). Поэтому, формализация проблемы означает использование точных формальных моделей для дескриптивной и императивной частей проблемы. Другими словами, это программа на некотором языке с формализованным синтаксисом и, желательно, семантикой. В этой связи напомним, что логика высказываний является полностью формализованной. Для логики предикатов [Мендельсон, 1976] с помощью современных подходов формализуется только множество тождественно истинных предложений. Для арифметики множество истинных предложений, так же как и его дополнение – не формализуются.

Заключение

Рассмотренные выше примеры позволяют сделать предположение о том, что всякие количественные усложнения в постановках проблем могут привести к новым задачам, но методы (логика) решения этих задач будут оставаться прежними. Поэтому решение задач такого рода позволяет лишь усовершенствовать существующие подходы и методы. Переход к качественно новым задачам требует в первую очередь разработки совершенно новых подходов и методов. Использование же существующих не может быть приемлемым по причинам, которые обсуждались выше.

Библиография

- [Василик, 2008] Василик П.В., Проватар А.И. О паралогичности некоторых логических построений //КИСА. 2008. - №3. - С. 180-186.
- [Кантор, 1985] Кантор Г. Труды по теории множеств. - М.: Наука, 1985. - 430с.
- [Мальцев, 1965] Мальцев А.И. Алгоритмы и рекурсивные функции. - М.: Наука, 1965. - 392с.
- [Антипенко, 1986] Антипенко Л.Г. Проблема неполноты теории и ее гносеологическое значение. - М.: Наука, 1986. - 224 с.
- [Василик, 2006] Василик П.В., Проватар А.И. К проблеме неполноты теоретических построений (на примере формализации некоторых задач естествознания). // КИСА. - 2006. - №4. - С. 145-150.
- [Проватар, 2000] Проватар А.И., Дудка Т.Н., Гошко Б.М. Применение метода резолюций на семантических сетях //Проблемы программирования. –2000. –№1–2.
- [Zadeh, 1965] Zadeh L. A. Fuzzy sets. Information and Control. 8(3):338–353, 1965.
- [Катеринич, 2007] Л.Катеринич, А.Проватар. Диагностирование на нейрон-ных сетях в системе Гомеопат // XIII-th International Conference: Knowledge Dialogue Solution. - Sofia, 2007. - V1. - P.64-68.
- [Проватар, 2008] А.И. Проватар. Формализация: алгоритмический подход. //Проблемы программирования. – 2008.
- [Мендельсон, 1976] Мендельсон Э. Введение в математическую логику. – М.: Наука, 1976. –320 с.
- [Дудка, 2001] ДудкаТ.Н., Проватар А.И. Применение нечетких множеств и уровневых чисел для решения задачи выбора //Проблемы программирования. – 2001.– №1. – С.21-26.

Информация об авторах

Василик Петр Васильевич – доктор биологических наук, ведущий научный сотрудник Института Кибернетики имени В.М. Глушкова, aprowata@unicyb.kiev.ua

Проватар Александр Иванович – доктор физико-математических наук, профессор, заведующий кафедрой информационных систем факультета кибернетики Киевского национального университета имени Тараса Шевченка, aprowata@unicyb.kiev.ua

USING SAT FOR COMBINATIONAL IMPLEMENTATION CHECKING

Liudmila Cheremisinova, Dmitry Novikov

Abstract. *The problem of checking whether a system of incompletely specified Boolean functions is implemented by the given combinational circuit is considered. The task is reduced to testing out if two given logical descriptions are equivalent on the domain of one of them having functional indeterminacy. We present a novel SAT-based verification method that is used for testing whether the given circuit satisfies all the conditions represented by the system of incompletely specified Boolean functions.*

Keywords: *design automation, verification, simulation, SAT.*

ACM Classification Keywords: *B.6.2 [Logic Design]: Reliability and Testing; G.4 [Mathematical Software]: Verification; B.6.2 [Reliability and Testing]: Error-checking.*

Introduction

Verification takes place at all design levels: from the conceptual design down to the design of combinational circuits. In this process, the objective of verification is to ensure that implemented and specified behaviors are the same; at this point, the design is error-free. Validating the functionality of digital circuits and systems is an increasingly difficult task. Multiple chip design projects are reporting that approximately 70% of their design time is spent in verification. This is due to the growing complexity of the designs that has not been accompanied by improvements in functional verification techniques.

While design tools have been made, and they, at least partially, support the complexity of highly integrated designs, there is a lack of verification support. Widely used tools for verification are logic simulators. At present, logic simulation is the most widely used technique for ensuring the correctness of digital integrated circuits in industry because of its scalability and predictable run-time behavior. This technique is based on verifying a digital system by stimulating inputs of the circuit with binary signals values that propagate in the circuit leading to a corresponding activation of the outputs, whose values must be consistent with the expected ones.

The past ten years have seen efforts in developing commercial formal verification tools. Instead of testing all input combinations explicitly (as simulators do), they formally prove a circuit's functionality to be consistent with its specification. Formal verification techniques have the potential of providing more general results than traditional simulation methods: it is possible to guarantee that a specific property holds for a design under all possible input stimuli. Now combinational equivalence checking (CEC) plays an important role in VLSI design; its usual application is verifying functional equivalence of combinational circuits after multi-level logic synthesis. In a typical scenario, there are two structurally similar implementations of the same design, and the problem is to prove their functional equivalence. This problem was addressed in numerous research publications, some of them are referenced in [Drechsler, 2000; Kropf, 1999; Mishchenko, 2006; Ganai, 2002; Kuehlmann, 2002; Kunz, 2002]. In a modern CEC flow based on formal verification approach, both circuits to be verified are transformed into a single circuit called a miter. It is derived by combining the pairs of inputs with the same names and feeding the pairs of outputs with the same names into EXOR gates, which are ORed to produce the single output of the miter. The miter is a combinational circuit with the same inputs as the original circuit and there is constant 0 on its output if and only if the two original circuits produce identical output values under all possible input assignments.

In the paper, the verification task is examined for a case, when desired functionality of the system under design is incompletely specified, i.e. intended behavior of implemented design allows functional indeterminacy. Such a

case usually occurs on early stages of designing when assignments to primary inputs of designed device exist which will never arise during normal mode of the device usage. Thus when hardware implementing this device, its outputs in response of these inputs may be arbitrary defined. In this case verification methodologies must consider only possible input scenarios to the design under verification and verify that every possible output signal of implemented behavior has its intended (described in initial specification) value.

The considered case could be thought as solvable by means of simulation-based tools of verification. But we propose a Boolean satisfiability-approach (SAT-approach) of checking whether a given combinational circuit implements a particular design specification. In this paper, we propose the following contributions to the problem of combinational verification: 1) the behaviour of the combinational circuit to be designed is specified with functional indeterminacy; 2) we show how it is possible to use SAT tools [Goldberg, 2002] for the considered case.

So in this paper we consider the verification problem for the case, when: 1) desired incompletely specified functionality is given in the form of a system of incompletely specified Boolean functions; 2) functions of the system are specified on intervals (cubes) of values of Boolean input variables and these intervals are large enough; 3) the system is implemented in the form of a combinational circuit in the basis of the elementary gates AND, OR and NOT.

We will discuss a novel SAT based method for testing whether the given circuit implements all multiple-output cubes representing the system of incompletely specified Boolean functions.

Background

Vectors and assignments. Let us consider a Boolean vector $\mathbf{x} = (x_1, x_2, \dots, x_n)$ of input variables. Let us call a set of equalities of type $x_i = \sigma_i$ (where $\sigma_i \in \{0, 1\}$, $i = \{1, 2, \dots, n\}$) as a variable value assignment \mathbf{a} for the vector \mathbf{x} of input variables. A variable value assignment \mathbf{a} for the vector \mathbf{x} can be a complete if all x_i are assigned or a partial otherwise. In the last case some of variables may be don't-care, meaning that any assignment to these variables is permissible. A complete variable value assignment represents a minterm and partial assignment represents a cube in n -dimensional Boolean space $E^n = \{0, 1\}^n$. A cube represents a product of literals (from now on, literal is a Boolean variable or its negation). A cube of the size k fixes values of exactly k variables and covers 2^{n-k} minterms. In general case a cube c_k (and the appropriate product) covers another cube c_l (and the appropriate product) if the literals of c_k are a subset of the literals in c_l .

Boolean functions. A completely specified Boolean function (CSF) $f(\mathbf{x}) = f(x_1, x_2, \dots, x_n)$ is a many-to-one mapping from n -dimensional ($n \geq 0$) Boolean space into a single-dimensional one: $E^n \rightarrow E$. A don't-care for a logic function allows it to have either 0 or 1 as a possible value. If, for some input combinations (minterm), the output of the function is a don't-care, this Boolean function is called as incompletely specified one (ISF). ISF is a mapping $E^n \rightarrow \{0, 1, -\}^m$, where the symbol “-” denotes don't-care condition. A CSF has only care minterms, which correspond to the assignments, for which it takes values 0 or 1. An ISF additionally has don't-care minterms, which correspond to the assignments, for which the function is flexible and can be either 0 or 1.

CSF $f(\mathbf{x})$ is specified by a pair of sets U_f^1 and U_f^0 of cubes (or minterms) that represent its on-set and off-set that divide Boolean space E^n into two parts. In the case of ISF there exists a don't-care set $E^n \setminus (U_f^1 \cup U_f^0)$ of cubes (that is not empty). A CSF $g(\mathbf{x})$ implements an ISF $f(\mathbf{x})$, if the CSF can be derived from the ISF by assigning either 0 or 1 to each don't-care minterm or, that is the same, if

$$U_f^1 \subseteq U_g^1, \quad U_f^0 \subseteq U_g^0 \quad (1)$$

A system of Boolean functions $F = \{f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x})\}$ (or $\mathbf{f}(\mathbf{x})$ in the vector notation, where \mathbf{x} and \mathbf{f} are vectors of input and output variables) of completely specified Boolean functions (CSF) is a mapping between

n -dimensional and m -dimensional Boolean spaces. In the case of ISF don't-care minterms may differ for different functions.

Let us specify a system $f(x)$ of ISFs as a set of multiple-output cubes. A multiple-output cube (u, t) is a pair of ternary vectors (products) of dimensions n and m that are called as its input and output parts correspondingly. The input part u represents a cube in E^n or a product of some literals $x_i \in X$. The output part t is a ternary vector of values of functions for the cube u or a product of some literals $f_i \in F$. For each $f_i \in F$ the j -th entry t^j of t is 1 or 0 ($t^j = 0$) if all the minterms covered by the cube u are in the on-set $U_{f_i}^1$ or in the off-set $U_{f_i}^0$ correspondingly; otherwise t^j is don't-care. For example, for the case $(u, t) = (-010-; -0)$ (or $(u, t) = (\bar{x}_2 x_3 \bar{x}_4; \bar{f}_2)$) we may state that all four minterms (belonging to the interval $-010-00100, 00101, 10100, 10101 \in U_{f_2}^0$ and $f_2(x_1, 0, 1, 0, x_5) = 0$ but we can say nothing about the value of the function f_1 on these minterms: some of them can be care some don't-care ones.

A Conjunctive normal form (CNF) represents a Boolean function as conjunction of one or more clauses, each being in its turn a disjunction of literals. From now on, we consider only clauses that do not simultaneously contain a literal and its negation.

A CNF denotes a unique completely specified Boolean function $f(x)$ and each of its clauses corresponds to an implicate of the function. A Boolean variable can be assigned a truth value (0 or 1). Also, clauses and CNF may assume values depending respectively on the values of the corresponding literals and clauses. CNF representation is popular among SAT algorithms because each clause must be satisfied (evaluate to 1) for the overall CNF to be satisfied. The SAT problem is concerned with finding an assignment $X' \rightarrow \{0,1\}$ to the variables of some subset $X' \subseteq X$ that makes CNF equal to 1 or proving that it is equal to the constant 0. If the first outcome takes place they say that the CNF is satisfied and refer to $X' \rightarrow \{0,1\}$ as a satisfying assignment.

Matrix Models of Boolean functions and CNF. Matrix representation of CNF formula C containing k clauses and n distinct variables is a ternary matrix C having a row for each clause and a column for each variable. The entry c_j^i of the matrix in the i -th row and the j -th column is 1, 0 or "-" depending on in what a form (x_j or \bar{x}_j) the variable x_j appears or does not appear in i -th clause of C . The same manner, matrix representation of ISF $f(x)$ is a pair of ternary matrixes U_f^1 and U_f^0 having rows for all cubes from U_f^1 and U_f^0 , correspondingly.

The system $f(x)$ of ISFs given by the set S of multiple-output cubes (u_i, t_i) can be represented by a pair of ternary matrices U and T of the same cardinality (Fig. 1). The matrix U contains as its rows all input parts of multiple-output cubes from S ; similarly matrix T specifies as its rows all output parts.

	x_1	x_2	x_3	x_4	x_5		f_1	f_2	
	-	-	1	1	1		1	-	1
	1	1	-	-	-		1	0	2
$U =$	-	0	0	0	-		0	1	3
	0	1	-	1	0		0	0	4
	-	0	1	0	-		-	0	5
	-	1	-	1	1		-	1	6
						$T =$			

Figure 1: An example of ISF system

Representation of a function in multiple-output cubes form has the following distinctive features. Cubes $u_i, u_j \in U$ can intersect each other (in contrast to a representation in the minterm form). Don't-care value of an element t_i^j of the matrix T means that either the function f_j is don't-care on the whole cube u_i or f_j does not take the same value (1, 0 or "-") on the whole interval u_i , i.e. there exist at least two minterms covered by the cube u_i on which f_j has different values. For example, the cubes $u_1 = --111$ and $u_2 = 11---$ intersect on the minterm 11111 so we may say that $f_2(1,1,1,1,1) = 0$ and f_2 do not take the same value on the whole interval u_1 .

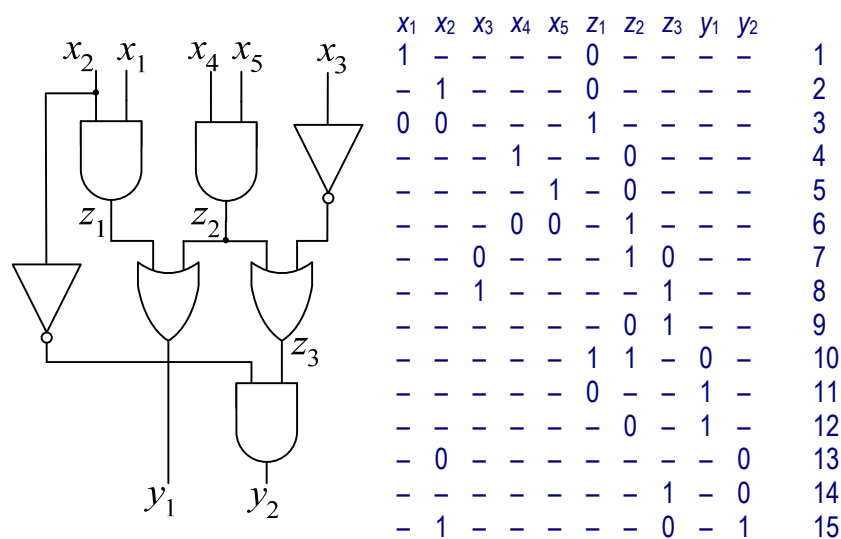


Figure 2: An example of a combinational circuit: a) the circuit; b) the corresponding conventional CNF

Combinational circuit. A combinational circuit under consideration refers to a gate-level network where primary inputs are connected to primary outputs through an interconnection of basic gates that implement elementary Boolean functions such as AND, OR, NOT, NAND etc. As usual we consider further only acyclic circuits.

The topological description of an acyclic combinational circuit can be represented using a directed acyclic graph, where nodes correspond to the gates, primary inputs and outputs of the circuit; edges correspond to circuit wires connecting the nodes. Incoming edges of a node are called its fanins and outgoing edges are called fanouts. A node in the circuit is multiple fanout if its output is a fanin to different gates. The node and its output signal are named the same. Nodes without fanins are the sources of the graph, called as primary inputs of the circuit; nodes without fanouts are the sinks, called as the primary outputs. Internal nodes of the graph correspond to logical gates implementing elementary Boolean functions. An example of a circuit (that will be tested later whether it implements the system of two ISFs depicted in Fig. 1) with five inputs, two outputs and seven gates is shown in Fig. 2,a. Here AND, OR, NOT gates are used as the basic ones.

Let us call the functionality of a circuit node in terms of its immediate fanins as the local function of the node, and the functionality of a circuit node in terms of the circuit primary inputs as the global function of the node. Thus the functionality of the circuit in terms of its primary inputs is the system of global functions implemented on the circuit primary outputs. For example, the local function of the node connected with primary output y_1 is $y_1 = z_1 \vee z_2$ and the corresponding global function is $y_1 = x_1 x_2 \vee x_4 x_5$.

Simulation-based verification

At present, logic simulation is the most widely used technique for ensuring the correctness of digital integrated circuits in industry. This technique computes the values of the internal signals and primary outputs of a circuit, given the values of its primary inputs. One round of simulation begins with stimulating primary inputs of the simulated circuit with binary signals values simulation and then this one is propagated through the circuit leading to a corresponding activation of the circuit primary outputs, whose values must be consistent with the expected ones. The complexity of simulating a particular set of input values is linear in the simulated circuit size (let us remember we consider only combinational circuit, so there are no internal state variables). But the overall Boolean space of values of n primary input variables contains up to 2^n combinations. Due to the complexity of

constructing all these combinations and verifying the compatibility between implementation and specification, simulation is infeasible for state-of-the-art designs.

A special type of simulation is of the most interest: guided simulation, when inputs are assigned based on certain information, provided by the design specification. In our case inputs could be assigned minterms covered by input parts u_i of multiple-output cubes of ISF-system. So the first step on the way of ISF-system verification consists in representing all the multiple-output cubes as multiple-output minterms. Then parallel binary simulation [Cheremisinova, 2008] of the combinational circuit can be performed under all input assignments corresponding to the minterms simultaneously.

Such a method could reduce in some cases the search space of simulators but only in the case when the input parts u_i of multiple-output cubes are "small enough" covering a small number of minterms. However though the specification of the designed circuit with n inputs would be specified with a small number of multi-output implicants, the overall size of Boolean space covered by them could be near to 2^n . So for the case when ISF system contains "big" multiple-output cubes covering a great number of minterms we propose a novel SAT-based method of testing whether the circuit implements such a multiple-output cubes.

SAT-based verification

SAT-solvers can be circuit-based or CNF-based. The former represent the SAT problem as a circuit composed of simple gates, while the latter use conjunctive-normal-form. To tackle the problems of circuit verification using the second type of SAT-solvers, they usually require their input to be in CNF because instances of SAT are usually represented as CNF formulas. This type of solvers is more general and can also be applied to circuits, by converting them into CNF form.

CNF encoding of combinational circuit. Majority of SAT applications derived from circuit representation rely on some a version of the Tseitin transformation for producing conventional CNF of the circuit. A circuit-to-CNF conversion uses as many variables as there are primary inputs and gates in the circuit.

When the conventional transformation is applied to a combinational circuit, for output of each gate (except output ones) its own internal Boolean variable is introduced and only local functions of the gates are considered. Then CNF formula is associated with each gate, and captures the consistent assignments between gate inputs and output. These all the gate local CNFs are joined then in the overall circuit CNF by using the conjunction operation. Both the size of the resulting CNF and the complexity of the conventional translation procedure are linear in the gate number of the original combinational circuit.

The derivation of CNF for a gate representing a local function $y = f(z_1, z_2, \dots, z_k)$ is based on defining a new Boolean function $\varphi(y, f) = y \sim f(z_1, z_2, \dots, z_k)$ (as in [Kunz, 2002]), that is true in the only case when both functions y and $f(z_1, z_2, \dots, z_k)$ assume the same value. Next, the function φ should be represented as a CNF form. As an example consider 2-input AND and OR gates, the formulas φ for them can be transformed the following way:

$$\begin{aligned}\varphi^{\wedge}(y, f) &= \varphi^{\wedge}(y, z_1, z_2) = y \sim f(z_1, z_2) = y \sim z_1 z_2 = (\bar{y} \vee z_1 z_2)(y \vee \bar{z}_1 \vee \bar{z}_2) = (\bar{y} \vee z_1)(\bar{y} \vee z_2)(y \vee \bar{z}_1 \vee \bar{z}_2); \\ \varphi^{\vee}(y, f) &= \varphi^{\vee}(y, z_1, z_2) = y \sim f(z_1, z_2) = y \sim z_1 \vee z_2 = (\bar{y} \vee z_1 \vee z_2)(y \vee \bar{z}_1 \vee \bar{z}_2) = (\bar{y} \vee z_1 \vee z_2)(y \vee \bar{z}_1)(y \vee \bar{z}_2).\end{aligned}$$

The above CNF formulas of 2-input AND and OR gates could be obtained reasoning from the truth table for the relational representations of AND and OR gates, identifying what combinations of the inputs and the gate output are admissible or possible (Fig. 3). In other words, the rightmost column of the truth tables is true iff the characteristic function of $f(z_1, z_2)$ is equal to y [Kropf, 1999.]. Then we construct a clause for each row of the truth table where the final column has 0 by formulating a disjunction of literals z_1, z_2, y in the negated form relative to

the values of the considered truth table. The disjunction of all these clauses results in the CNF formula. So for 2-input AND and OR gates we will obtain the following representations (the same as ones shown above):

$$\begin{aligned}\varphi^{\wedge}(y, z_1, z_2) &= (\bar{y} \vee z_1 \vee z_2)(\bar{y} \vee z_1 \vee \bar{z}_2)(\bar{y} \vee \bar{z}_1 \vee z_2)(y \vee \bar{z}_1 \vee \bar{z}_2) = (\bar{y} \vee z_1)(\bar{y} \vee z_2)(y \vee \bar{z}_1 \vee \bar{z}_2); \\ \varphi^{\vee}(y, z_1, z_2) &= (\bar{y} \vee z_1 \vee z_2)(y \vee z_1 \vee \bar{z}_2)(y \vee \bar{z}_1 \vee z_2)(y \vee \bar{z}_1 \vee \bar{z}_2) = (\bar{y} \vee z_1 \vee z_2)(y \vee \bar{z}_1)(y \vee \bar{z}_2).\end{aligned}$$

z_1	z_2	y^{\wedge}	φ^{\wedge}	z_1	z_2	y^{\vee}	φ^{\vee}
0	0	0	1	0	0	0	1
0	1	0	1	0	1	1	1
1	0	0	1	1	0	1	1
1	1	1	1	1	1	1	1
0	0	1	0	0	0	1	0
0	1	1	0	0	1	0	0
1	0	1	0	1	0	0	0
1	1	0	0	1	1	0	0

Figure 3: Truth tables of the 2-input AND and OR functions

For the general case here are the conventional CNF representations of NOT, k -input AND and OR gates comprising the example circuit in Fig. 2, a:

$$\begin{aligned}\varphi^{-}(y, z) &= (z \vee y)(\bar{z} \vee \bar{y}); \\ \varphi^{\wedge}(y, z_1, z_2, \dots, z_k) &= (z_1 \vee \bar{y})(z_2 \vee \bar{y}) \dots (z_k \vee \bar{y})(\bar{z}_1 \vee \bar{z}_2 \vee \dots \vee \bar{z}_k \vee y); \\ \varphi^{\vee}(y, z_1, z_2, \dots, z_k) &= (\bar{z}_1 \vee y)(\bar{z}_2 \vee y) \dots (\bar{z}_k \vee y)(z_1 \vee z_2 \vee \dots \vee z_k \vee \bar{y}).\end{aligned}$$

It is possible to eliminate the output variable y of NOT $y = \bar{z}$ gate and two appropriate clauses of circuit conventional CNF if to subsume y in fanout gates of the NOT gate replacing all instances of y with the negated input variable z of this gate. Fig. 2, b shows a circuit and its conventional CNF.

Conventional CNF analysis. SAT problem for a CNF formula is formulated as follows [Kunz, 2002]: given a CNF formula representing a Boolean function $y = f(x_1, x_2, \dots, x_n)$, the problem consists of identifying a set of assignments to the formula variables, $\{x_1 = a_1, x_2 = a_2, \dots, x_n = a_n\}$, such that all CNF clauses are satisfied (taking into account that a clause is satisfied if at least one its literal is equal to 1, i.e., $f(a_1, a_2, \dots, a_n) = 1$, or proving that no such assignment exists. Recall that a CNF formula is satisfiable if there exists an assignment $\{x_1 = a_1, x_2 = a_2, \dots, x_n = a_n\}$ providing $f(a_1, a_2, \dots, a_n) = 1$. This assignment is known as a satisfying assignment.

When we have a multiple-output circuit we can state a problem of finding out primary input assignments making some primary output to be one. To test whether the output be 1 the unit clause (a clause consisting of the only literal) corresponding to the tested output y_i is added to the circuit CNF. Once the overall problem is formulated in CNF, a SAT solver can be used to solve it. The resulting satisfying assignments of the circuit CNF and only they form the on-set of the global function y_i and furthermore only on-set minterms satisfy the CNF. For example, after adding to the CNF of the circuit (Fig. 2) the unit clause y_1 and testing CNF for existence partial assignments satisfying every clause one could find CNF partial assignments (for the subspace restricted by the function y_1) given in Fig. 5.

In general case let y^{σ} (where $\sigma \in \{0, 1\}$) be a literal of the variable y , precisely, $y^1 = y$ and $y^0 = \bar{y}$. Then the unit clause y^{σ} represents the assignment $y = \sigma$. Let we have to check whether an output y_i of a circuit to be constantly σ . We cannot directly use SAT solver to show that this statement is true but we could without problems to prove using SAT solver that y_i equals (in the case when \bar{y}_i is not the constant σ) or does not equal to $\bar{\sigma}$ for some argument assignments. If we cannot prove y_i equals $\bar{\sigma}$ for some arguments assignments thereby we prove that y_i is the constant σ .

x_1	x_2	x_3	x_4	x_5	z_1	z_2	y_1
1	1	-	1	1	1	1	1
-	0	-	1	1	0	1	1
0	-	-	1	1	0	1	1
1	1	-	0	-	1	0	1
1	1	-	-	0	1	0	1

Figure 4: CNF partial assignments of the circuit CNF (Fig. 2,b) appended by the unit clause y_1

Such proving is known as proving by contradiction, often it is very convenient when using SAT solver. Thus to show that circuit primary output y_i is the constant σ using prove by contradiction we suppose it is not. At the first step we add to the circuit CNF the unit clause $y_i \bar{\sigma}$ and search using SAT solver for a satisfying assignment \mathbf{a} to CNF formula for which $y_i(\mathbf{a}) = \bar{\sigma}$ is true. Such an assignment $\mathbf{a} \cup \{y_i = \bar{\sigma}\}$ is called as a counter-example. If there exists no counter-example then the circuit implements the constant σ on the output y_i .

SAT-Based Model of Testing of Multiple-Output Cubes of ISF System

Assume we have some CNF formula C describing a circuit and a system $f(x)$ of ISFs the arguments and the functions of which correspond to primary inputs and outputs of the circuit. Hereinafter let us consider as an example the system $f(x)$ and the circuit (and CNF specifying it) shown in Figures 1, 2.

A problem under discussion is verifying if the given circuit implements the ISF system $f(x)$. It is right if it takes place for each pair of ISF $f_i(x) \in f(x)$ and the appropriate circuit output $y_i(x)$. The global CSF $y_i(x)$, realized by the i -th circuit output, implements a function $f_i(x) \in f(x)$ iff (1) takes place.

A multiple-output cube $(\mathbf{u}_i, \mathbf{t}_i)$ of ISF system imposes conditions on values of some functions f_j : $f_j(\mathbf{u}_i) = t_j$ for all j for which $t_j \in \{1, 0\}$. Further, we are interested in only those components of \mathbf{t}_i which are not don't-care. The truth of the conditions (1) guarantees the circuit has the same functionality as the ISFs system: for every input stimulus implied by input part of any multiple-output cube $(\mathbf{u}_i, \mathbf{t}_i)$ the Boolean vector of values of the circuit outputs is covered by the output part \mathbf{t}_i . In terms of circuit CNF the conditions (1) could be reformulated as follows. For every multiple-output cube $(\mathbf{u}_i, \mathbf{t}_i) \in f(x)$ a partial value assignment $\mathbf{u}_i \cup \mathbf{t}_i$ of input and output variables should be satisfying assignment for circuit CNF. Below, cubes $(\mathbf{u}_i, \mathbf{t}_i)$ will be checked if they are implemented by CNF one by one, with no particular order.

Let us consider in more details the procedure of SAT-solving for an elementary cube $(\mathbf{u}_i, t_i) \in (\mathbf{u}_i, \mathbf{t}_i)$, where $t_i = f_i(\mathbf{u}_i) = \sigma$ and $\sigma \in \{1, 0\}$. At the first step, keeping in mind the prove by contradiction, we assign $f_i(\mathbf{u}_i)$ to be $\bar{\sigma}$ (where $\sigma = t_i$) i.e. we suppose $(\mathbf{u}_i, y_j \bar{\sigma})$ takes place and reduce the circuit CNF C making a set of assignments setting all literals of $\mathbf{u}_i \cup y_j \bar{\sigma}$ to 1 obtaining $C(\mathbf{u}_i \cup y_j \bar{\sigma})$.

For instance, the set of initial assignments for (\mathbf{u}_6, t_6^2) (Fig. 1) will be $x_2 = x_4 = x_5 = 1, y_2 = 0$. This means that all clauses of C having at least one of the literals x_2, x_4, x_5, \bar{y}_2 are discarded and literals $\bar{x}_2, \bar{x}_4, \bar{x}_5$ and y_2 are removed from all clauses having them. After making these assignments we search for a satisfying assignment \mathbf{a} to the obtained CNF $C(\mathbf{u}_i \cup y_j \bar{\sigma})$. In our case for CNF $C(x_2 = 1, x_4 = 1, x_5 = 1, y_2 = 0)$ there exists such an assignment, for instance, $x_1 = z_1 = z_2 = z_3 = y_1 = 1$ (1-----1111-). That proves that there exists a counter-example for (\mathbf{u}_6, t_6^2) , proving that the global function $y_2(x)$ of the circuit does not implement $f_2(x)$ for some input pattern from \mathbf{u}_6 : $f_2(\mathbf{u}_6) \neq y_2(\mathbf{u}_6)$. This conflicting input pattern in our case is 1 1 - 1 1.

The above procedure is applicable for the case when the output part \mathbf{t}_i of a multiple-output cube $(\mathbf{u}_i, \mathbf{t}_i)$ consists of the only component having definite value (0 or 1). But in general case the output part $\mathbf{t}_i = y_1^{\sigma_1} y_2^{\sigma_2} \dots y_k^{\sigma_k}$ of a multiple-output cube $(\mathbf{u}_i, \mathbf{t}_i)$ consists of more than one component, for instance k , having definite values. Proof by contradiction which tries to find a counter-example forces to test the following assignment:

$$\mathbf{u}_i \cup \bar{\mathbf{t}}_i = \mathbf{u}_i \cup \neg(y_{i1}^{\sigma_1} y_{i2}^{\sigma_2} \dots y_{ik}^{\sigma_k}) = \mathbf{u}_i \cup (y_{i1}^{\bar{\sigma}_1} \vee y_{i2}^{\bar{\sigma}_2} \vee \dots \vee y_{ik}^{\bar{\sigma}_k}). \quad (2)$$

So in this case we make initial assignments setting only literals of \mathbf{u}_i to 1, then add to CNF $C(\mathbf{u}_i)$ the clause $y_{i1}^{\bar{\sigma}_1} \vee y_{i2}^{\bar{\sigma}_2} \vee \dots \vee y_{ik}^{\bar{\sigma}_k}$ ($y_{ij}^{\bar{\sigma}_j} = 0, 1$). Or, it is the same, only add to CNF $l+1$ clauses (where l is the number of literals in \mathbf{u}_i): l unit clauses of the type x_j ($x_j \in \mathbf{u}_i$) and a clause $y_{i1}^{\bar{\sigma}_1} \vee y_{i2}^{\bar{\sigma}_2} \vee \dots \vee y_{ik}^{\bar{\sigma}_k}$ of size k .

For example the cube $(\mathbf{u}_2, \mathbf{t}_2)$ implies three clauses: $x_1, x_2, \bar{y}_1 \vee \bar{y}_2$ to be added to CNF. For the extended CNF there exists no satisfying assignment, that fact proves the values of the global functions $y_1(\mathbf{x})$ and $y_2(\mathbf{x})$ are equal correspondingly to $f_1(\mathbf{u}_2)$ and $f_2(\mathbf{u}_2)$ for all input patterns from \mathbf{u}_2 : $f_1(\mathbf{u}_2) = y_1(\mathbf{u}_2), f_2(\mathbf{u}_2) = y_2(\mathbf{u}_2)$.

Conclusion

In this paper, we propose the following contributions to the problem of combinational verification.

1. We consider a case when one of the compared descriptions is incompletely specified.
2. We show how it is possible to use SAT tools for the considered case.
3. We suppose the method of checking whether multiple-output cubes of ISF are implemented by circuit conventional CNF.

Bibliography

- [Drechsler, 2000] R. Drechsler. Formal Verification of Circuits. Kluwer Academic Publishers, 2000.
- [Kropf, 1999] T. Kropf. Introduction to Formal Hardware Verification. Springer, 1999.
- [Mishchenko, 2006]. A. Mishchenko, S. Chatterjee, R. Brayton, N. Eén. Improvements to Combinational Equivalence Checking. In: Proc. ICCAD'06, Nov. 5–9, 2006, San Jose, CA, 2006.
- [Ganai, 2002]. M.K. Ganai, L. Zhang, P. Ashar, A. Gupta, Malik S. Combining strengths of circuit-based and CNF-based algorithms for a high-performance SAT solver. In: Proc. ACM/IEEE Design Automation Conference, 2002, pp. 747–750.
- [Kuehlmann, 2002] A. Kuehlmann, A.J. van Eijk Cornelis: Combinational and Sequential Equivalence Checking. In: Logic synthesis and Verification. Ed. S.Hassoun, T.Sasao and R.K.Brayton. Kluwer Academic Publishers, 2002, pp. 343–372.
- [Kunz, 2002] W. Kunz, J. Marques-Silva, S. Malik. SAT and ATPG: Algorithms for Boolean Decision Problems. In: Logic synthesis and Verification. Ed. S.Hassoun, T.Sasao and R.K.Brayton. Kluwer Academic Publishers, 2002, pp. 309–341.
- [Goldberg, 2002] E. Goldberg, E. Novikov. BerkMin: A fast and robust SAT-Solver. In: Proc. European Design and Test Conference, 2002, pp. 142–149.
- [Cheremisinova, 2008] L. Cheremisinova, D. Novikov. Simulation-based approach to verification of logical descriptions with functional indeterminacy. In: Information Theories & Applications (IJ ITA), 2008, Vol. 15, No. 3, pp. 218–224.

Authors' Information

Liudmila Cheremisinova – Principal Researcher, The United Institute of Informatics Problems of National Academy of Sciences of Belarus, Surganov str., 6, Minsk, 220012, Belarus, e-mail: cld@newman.bas-net.by

Dmitry Novikov – Post graduate student, The United Institute of Informatics Problems of National Academy of Sciences of Belarus, Surganov str., 6, Minsk, 220012, Belarus, e-mail: yakov_nov@tut.by

Philosophy and Methodology
of Informatics

BUSINESS INFORMATICS

Krassimir Markov, Krassimira Ivanova, Ilija Mitov, Valeria Mitova

Abstract: A definition of the concept "business informatics" based on the General Information Theory is discussed in the paper.

Keywords: Business Informatics, General Information Theory

ACM Classification Keywords: K.4 Computers and Society; Automation; J. Computer Applications, J.0 General, J.1 Administrative Data Processing, Business

Conference topic: Philosophy and methodology of informatics.

1. Introduction

The concept "Business informatics" became popular about fifteen years ago when the usual concepts had lost their generality, abstractness and actuality. Concepts such as "automation of the company's management", "information service of the decision making", "information modeling of the micro-economical processes", and so on, had caused theoretical rationalizing of the whole information basics and searching of a new unified concept.

Till now there is no common definition of the concept "Business informatics". In the paper we propose our understanding of it based on the General Information Theory [Markov et al, 2006].

1.1. Business

We use the law oriented understanding of the concept "business". We assume only formally registered companies as main business structures – it is correctly to hold on law-determined meaning and to connect the concept "business" with more concrete definition as doing some activities as juridical-registered organization and the name (firm) with defined in advance purposes, with adhering to some rules, norms and laws, received by the community as a regulator of these activities.

Let's remember that like every social formation the company is a social organism, which, like other organisms, generates, develops and dies. The life cycle of the company strongly depends on harmonic development of each their elements, which is subordinated of the main goal of existing of every organism – **to develop itself, searching its place in the environment**. From here follows that not "making money" but "development and entering in the environment" are main goals of each company. Every deviation of this goal leads to isolating of the company and step-by-step disappearing.

There exist different ways to represent the structure of the company. Here we will represent two main approaches from the point of view of:

- the elements, which construct the company – so cold "horizontal" view;
- the management structure and levels of decision making – so cold "vertical" view.

The main elements of every company (**horizontal view**) are:

- **Human resources** (H) - this concept usually is connected with the firm staff. Here we expand the examining and we will mention that each company is connected with sizeable number of "non-staff" physical or juridical persons, which in one or a different way contribute to its development. No one company can develop without a contingent of "partners". The market presence of the company is

unthinkable without "clients". The activity of "competitors" is one of the main stimuli for the development of the company. So, the external relationships determine the place of the company in the business environment.

- **Material resources** (M) include all objects which are processed by the "human resources". Material resources include the means of production and needed prime and raw materials as well as the output production which is still under the control of the company.
- **Financial resources** (F) are indication of the company capabilities.

Some resources may belong to the company (i.e. its own) as well as some may belong to the other companies or individuals but for the given period are on disposition of the company (black and gray circles around the letters) (Fig. 1).

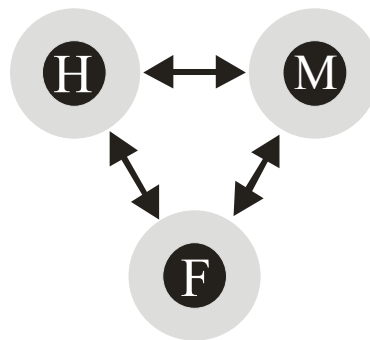


Fig.1. The main elements of a company

In general, the company's development is quantity and quality changing of its main resources to achieve the basic goal – adaptation to the neighboring environment and consolidation the relationships with the business partners.

The companies are functioning as cybernetic systems. So this is another (**vertical**) view to the company (Fig. 2).

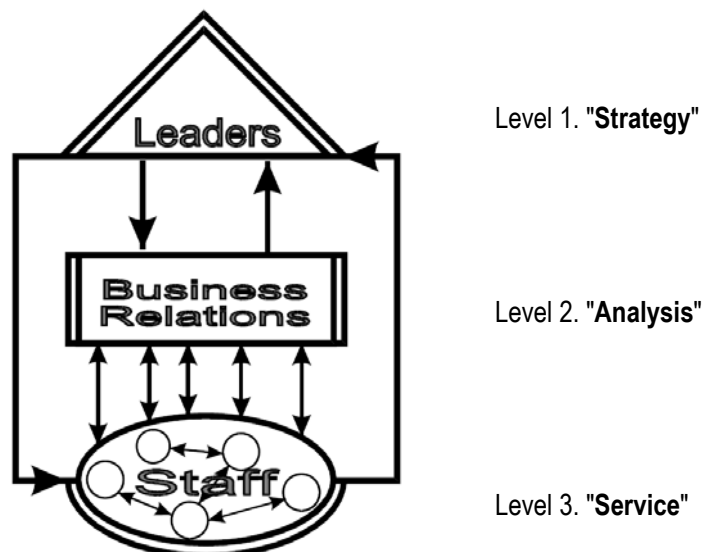


Fig. 2. The cybernetic structure of a company

We may separate at least three levels depending of types of management – strategic, tactical and operative management:

- Level 1. "**Strategy**" – The Leaders make long-term **strategic** decisions;
- Level 2. "**Analysis**" – The internal business relations are concentrated at this level. As a rule, the accountancy department serves as mediator between leaders and the staff of the company and provides the **tactical** management and the short-term decision making;
- Level 3. "**Service**" – The realization of the leaders' strategy is provided at this level by the company's staff. Staff members provide **operative** management and decision making.

When the company is a small one, the contacts between the leaders and the staff are direct and may be provided by the leaders. It is clear, the great companies and corporations need a middle level for support the communication between high and low management levels. The good practice advices to have direct contacts, too. But they are episodic and have more informative character.

1.2. Informatics

Our understanding of the "informatics" follows the General Information Theory (GIT) [Markov et al, 2006]. The concept information is defined as special kind of reflection of one entity in another for which there exists reflection evidence.

The real world contains unlimited number of entities. When an entity contacts another, there exists a great possibility to join third entity in this process. It is clear; the third entity may contact and reflect each of others as well as the process of realization of the interaction between them – the process of realization of the contact is a specific (temporal) forming relationship between entities and during the process of establishing the contact the entities form new (temporal) entity which in the same moment may be reflected by the third entity. So, the third entity may reflect any traces of this interaction from both first and second entities.

In the special case when the third entity contains reflections of the first entity received by both two different ways:

1. by transitive impact of the first entity on the third one through the second entity,
2. by impact of the first entity on the third one which is different from the transitive one, i.e. it can be direct impact or transitive impact through another entity (-ies)

then the third entity became as an external relationship between first entity and its reflection in the second entity – it became as "**reflection evidence**" of this relationship.

The first entity is called **reflection source**; the second entity is called **reflection recipient**; and the third entity is called **reflection evidence**.

In this special case, when there exist the triple

"(source, recipient: evidence)",

the reflection of the first entity in the second is called **information** in the second for the first entity.

In the nature there exist countless entities, interactions and reflections. The "Theory of reverberation" is a good step to understanding the reflection phenomena [Павлов, 1987]. From point of view of the General information theory it is important that only the subject may use reflections and reflection evidences in accordance with any goals. For instance, in the story "Cinderella" the crystal shoe is only a beautiful entity till the moment of proving that this is the shoe which reflects exactly the foot of Cinderella and it became an element of information triad presenting in a given way the Cinderella, i.e. any information about her.

For the Subject:
a reflection
of an entity (*source*)
in another entity (*recipient*)
is **information** about the first entity,
iff there exists corresponded
evidence (*reflection evidence*).

In other words, every reflection may be assumed by the Subject as information provided that there exists evidence what really it reflects. I.e. the information is a kind of reflection, but not every reflection is information. Information is only these reflections for which the subject has evidences what they reflect.

This understanding is in the foundation of the General Information Theory (GIT). The basic structure of the GIT presented in [Markov et al, 2006] is reproduced in the Table 1.

GIT is completed by three theories: "Theory of the *information*", "Theory of the *information subject* (INFOS)" and "Theory of the *information interaction* (inforaction)". We have no place to present GIT in details.

Table 1. **Basic Structure of the General Information Theory**

Occurrence	Specificity	Subject	Theory
Reflection	Information Relationship	Evidence	Theory of the Information
Activity	Information Expectation	Witness	Theory of the Infos
Modeling	Information Modeling	Information Subject (Infos)	
Interaction	Information Interaction	Society	Theory of the Inforaction

The main result that follows from the Table 1 is the definition of the informatics in more general view as science about the types and properties of:

- information and the information structures;
- information subjects;
- information interaction.

2. Business Informatics

The traditional understanding of the concept "Business Informatics" is as an interdisciplinary discipline, which is aimed to study information structures, operations and processes that are inherent to the business and support their automation.

Now we can try to extend this definition taking into account the concepts presented in GIT as well as the views of the company structure presented above.

At the first place, the concept of "Business Informatics" we may define as a science about the unity of:

- Business Information,
- Business Information Subjects,
- Business Information Interaction.

Integrating the **horizontal** and **vertical** views of the company we may build a specific "**pyramidal**" model [Марков и др., 1994] separated on three levels (Fig. 3):

- "Strategy" (the highest level of the model);
- "Analysis";
- "Service" (the lowest level).



Fig.3. The "pyramidal" model of a company

This way we have possibility to make more detailed description of the main scientific areas of the Business Informatics.

- **Business Information:** The variety of reflections in the business does not allow us to present and discuss all of them in the frame of this paper. Important task of the Business Informatics is to propose relatively complete investigation and classification of the business-reflections. In the same time we need to remark that the dynamics of the development of the different companies as well as of the changing of the business environment, on principle does not allow complete description. As a rule the scientists are contented by the systematizing of the formally defined reflections which are used by the community structures - taxation authorities, statistics, etc. Usually out of consideration are types of reflections created by the company in the environment and vice versa – by the environment in the company;
- **Business Information Subjects:** One business reflection can be perceived in a different manner from different business subjects. The classification of the business subjects has a crucial role because of the subjective character of perception of business reflections. Often even formally defined types of reflection have perceived differently and this leads to serious practical problems. As we mentioned above, in general the business information subjects may be divided in two main groups – which belong to the company and the rest, which belong to the environment: partners, clients, rivals and neutrals, which in given moment may change their type;
- **Business Information Interaction:** The investigation of the business information subjects is determined by the necessity of provide qualitative automated service of theirs interaction. The important information need to be collected, stored, processed and distributed to corresponded decision makers in appropriate mode suitable for quick and non-vague perceiving, in one hand, and ensuring correct information interaction in the frame of the company as well as in the environment, from other hand. That is the main goal of the information service.

3. Conclusion

Building of adequate information models of business activities, based on common theoretical foundation, leads to more clear and deep understanding of these activities and thence to theirs optimizing and automating.

Let's underline that main subject of the Business Informatics is to examine as a whole both forming elements of the business – "organization" and "information" [Марков и др., 1994].

In conclusion we need to point that main goal of the Business Informatics is to ensure the automation of business activities in accordance with the pyramidal information model [Марков и др., 1994]. For instance there exists variety of software systems for information service, which may be classified to separate levels of the pyramid:

- the lowest level "**Service**" is formed by components for the operative information service of activities of and with the main company resources – human, material and financial. In concrete implementations here we may see program systems for information service of processing the personal information, the salaries and other working contracts, the logistic activity in one or more storehouses, the fixed assets amortization, etc.;
- the level "**Analysis**" contains the systems for information service of accounting activities. These are the systems from new quality where the accent is made on the serious analysis and preparation of prognoses;
- the highest level "**Strategy**" includes systems for information service of the decision making in the conditions of collective information interaction and multi-variant decisions. Usually, the OLAP and Data Warehouses are classified at this level.

The leading information service systems render an account to main areas of the Business Informatics – Business Information, Business Information Subjects and Business Information Interaction in accordance with different levels of management.

The approach, presented in this paper, had been implemented in the Complex FOI – a Bulgarian program product for information service of company management and accounting.

Bibliography

[Markov et al, 2006] Kr. Markov, Kr. Ivanova, I. Mitov. Basic Structure of the General Information Theory. IJ ITA, Vol.14, No.: 1, pp.5-19.

[Марков и др., 1994] К.Марков, К.Иванова, И.Митов, Й.Икономов, В.Василев. Информационно моделиране на фирмени дейности. Списание "Известия на Икономическия Университет - Варна"; бр. 3-4, 1994; стр.119-127.

[Павлов, 1987] Т.Павлов. Избрани философски съчинения в три тома, Том 2, Теория на отражението. Изд. "Наука и изкуство", София, 1987.

Authors' Information

Krassimir Markov - Institute of Mathematics and Informatics - BAS, Sofia, Bulgaria,
e-mail: kmarkov@math.bas.bg

Krassimira Ivanova - Institute of Mathematics and Informatics - BAS, Sofia, Bulgaria,
e-mail: kivanova@math.bas.bg

Iliia Mitov - Institute of Information Theories and Applications FOI ITHEA, Sofia, Bulgaria,
e-mail: mitov@foibg.com

Valeria Mitova - Institute of Information Theories and Applications FOI ITHEA, Sofia, Bulgaria,
e-mail: mitova@foibg.com

BASIC PRINCIPLES OF ORGANIZATION OF THE MEDIUM AND THINKING PROCESS OF THE HUMAN IN ITS CONCEPTUAL PRESENTATION

Yuriy Belov, Zinoviy Rabinovich

Abstract: *Given in the report conceptual presentation of the main principles of fractal-complexity Ration of the media and thinking processes of the human was formulated on the bases of the cybernetic interpretation of scientific information (basically from neurophysiology and neuropsychology, containing the interpretation giving the best fit to the authors point of view) and plausible hypothesis's, filling the lack of knowledge.[1-4]*

Keywords: *organization memory, pattern recognition, problem solving, emotional factor, bionic model.*

ACM Classification Keywords: *1.2.4 Knowledge Representation Formalisms and Methods, 1.5.1 Pattern recognition - Models.*

Conference topic: *Problems of Computer Intellectualization.*

Introduction

Main author's activity anticipating this paper (and being served the bases of it) may be found in [5-13] in historical order. One of active co-authors of these papers was the neurophysiologist Voronkov G.S. (Moscow Lomonosov State University). The listed papers and this report as well have on their base a concept considering the mind as the psychological process being in the memory of the brain, particularly on its deliberate and intuitive (underdeliberate and overdeliberate) levels.

With respect to the whole neuro network the memory is only a single part as it operates the data (signals) coming into it from sensory organs as well as from other outside components of the organism. The data comes into the memory in the appropriate form being constitute from the outside influence signals.

We find it is reasonable for the sake of comprehension to construct the following association with the control and complex computation system. Let the central computer of the system be the memory and the special sensors delivering to it input information be the sensory organs. In contrast to a computer data processing in the memory takes place immediately inside instead of being realized by a special processor interacting with the memory (present day high-performance computer development is attracted by distributed data processing, that is mentioned at the end of the report).

All the process of semantic processing in the memory is characterized by its realization in the different aspects:

- Inductive data generation (as the transition from the particular to the general, i.e. from "micro" to "macro" [4]).
- Deductive specialization of the generalized information (as the transition from the general to the particular, i.e. from "macro" to "micro" [4]).
- Abductive transformation (as the transition from the particular to the particular, over the same significance information units).
- First two types of operations we will call vertical - ascending and descending, respectively. So, the vertical operations perform fractal-complexity and complexity-fractal transitions correspondingly whereas the horizontal ones save belonging of the rank to input or output information. Uncovering the mechanizes of the operations, their combination and sequence is the sense of determining principles of structural

memory and the processes taking place in it that are caused by the interaction with the environment and providing the mental thinking operation.

Particularly important aspect of the human vital functions being of dominant interest of artificial intelligence in sense of modeling is the function of deliberating decision formation as the goal reaching. That is why we are going to consider conceptual memory model and the model of processes in the memory orienting right on that functions [5-13]. Conceptual model here is an implementation of fundamental functional properties common for different forms and signal nature. However first of all is necessary to imagine the physical substance of "memory" and how the limits of it are defined in the neuro network.

Memory Concept and its Global Organization

The memory functions include such concepts as "recognizing", "remembering", "imagining" etc. Unlike perceiving the information from environment, it is suitable to represent all these actions as demonstrating the so-called "mental sight", i.e. a look which is initialized within the memory and acted as exciting the certain sensory memorized structures in the memory network, their combination etc.

So, what is this structure?

To answer this question and understand the global defining principle of the memory organization (that is required of the conceptual model) it is necessary to start with the basic hypothetical prerequisite that ensues, as it were, from "common sense" (but coordinate with experimental data).

In addition, such prerequisite as the main basic hypothesis is the following one:

"The pattern reconstruction in the memory (imagination, mental sight) is determined by exciting all its elementary components which took part in the pattern perception" [11].

It would be considered as a law of nature, as an immutable but inexplicable fact. Really, how does it happen that fluctuations of potential of the neural net components are transformed to visible (also audible, tangible, palpable) patterns? However, it happens, isn't it? So, this fact should be the basis for the further constructions.

These constructions should already result in capability to form the signals inside the memory. These signals are able to excite the components of the pattern perceived and fixed in the memory when is already no information about this pattern at the memory input. Concerning this external information, the internal exciting signals are the information flow, which to this it reverses.

So, we conclude the aforementioned fact that the "memory" begins where the reverse bonds finish. They finish just at the level of "cc" conceptors, which repeat "r" receptors. Reverse bonds shouldn't extend over "r" receptors since the illusions occurs during mental sight, i.e. one sees, hears and touches that it isn't sensed at present time. So, the memory limit is just positioned between "r" receptors and "c" conceptors: therefore, duplicating the first ones by the last ones is necessary for it [6].

The last ones ("c" conceptors) are actually the smallest sensor, components of the perceived patterns and the most elementary sub-patterns regarding to the Pattern. These sub-patterns hierarchically and naturally group together into larger sub-patterns, then the last ones group in even larger sub-patterns etc. up to the sensor concentration of the whole Pattern in the single structure unit which denotes only its symbol.

Thus, when constructed information signals enter the memory from outside, the nonlinear pyramidal model of the pattern expressed by this information will be constructed in the memory. The down-going reverse bonds from the top of the pattern pyramid up to its base (consisting of conceptors which repeat input receptors) are necessary for memorizing the pattern, i.e. for making possible the pattern reconstruction by "mental sight" according to the above-mentioned main hypothesis. So, the model of the concrete pattern, both as an object and as a structure in

the memory, is the pyramid hierarchical construction with up-going convergent inductive (from partial to general) bonds and down-going divergent deductive (from general to partial) bonds.

The memory as a whole consists of the set of such models as the memory loops associatively bound by the general components. The memory is a structural realization of the semantic network as a system with knowledge fixed in it.

The above-mentioned fact can be illustrated by the network which is enough simple for visualization (Fig.1). The network consists of fully constructed (i.e. fixed in the memory) ABCD pattern and CDEF pattern entered into the memory but not fixed in it.

It is caused by that the pyramid (i.e. the model) of the ABCD pattern already has reverse bonds and pyramid of the CDEF pattern isn't still constructed in full.

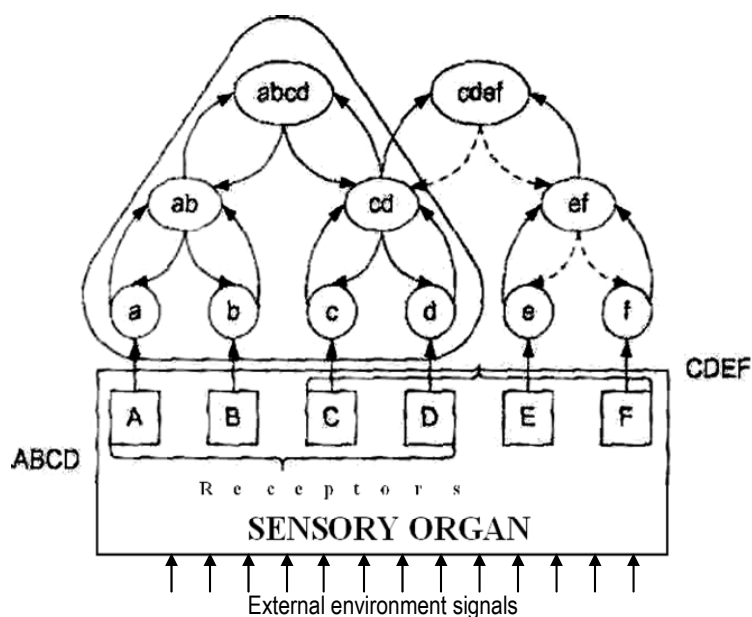


Fig.1. Elementary structures of pattern perception, memorizing and recognition

These patterns have one general CD sub-pattern, which is the associative element of both pyramids. Such using of the general parts of the pattern provides: first, economy of constructing the semantic structures in the memory; second, spontaneous (i.e. self-originated) parallelism during processing the pattern information in it that is essentially favorable to effectiveness of this processing (see below).

To finish the construction of the CDEF pattern means creating the reverse bonds in it (shown via dotted line). Finishing the construction, i.e. changing the model of demonstrated pattern into the memorized one can be performed by several successive demonstrations of the same pattern that results in beating the genetically innate reverse bonds or even in their appearance.

The creation of the pattern structures in the memory (models) is scientifically valid. The mentioned above conception on constructing these structures nevertheless requires the following plausible hypothesis:

constructing the direct up-going bonds for perceiving the concrete patterns by the memory precedes an appearance of the reverse down-going bonds, which ensure their memorization (see the main hypothesis) by learning.

Thus, a brain memory medium of a new-born child (or other individuals with enough developed nervous system) is already saturated by bonds needed to obtain the information from sensors organs. The memory itself is

originated in full by constructing the already reverse bonds in metabolic process begun from the birthday (and while preparing the relevant capabilities even before it). Therefore, the brain as the thinking mechanism creates itself but under the environment impact. Let rather look aside and notice that the so-called innate abilities are apparently caused by peculiarities of the genetically innate direct bonds. The demonstration of these abilities (as the construction of the relevant full pattern structures in the memory) already occurs as a result of learning; i.e. phenomenal talents appear as a result of coincidence of two factors; the genetically constructed relevant structures in the memory Medium and its subsequent learning as an impact of external factors.

As a consequence, the brain's memory in its conceptual representation is the hierarchical semantic structure restricted from above with ending of - connections (straight lines coming from sensory organs) and from bottom - with ending of backward connections to structure units, which directly perceive information entering the memory (formed by specific organs of sense from specific signals). And whole thinking from simple recognition of the pattern information to its analysis, synthesis and the subsequent actions on sets of patterns, connected with creation and transformation of different situations, are occurs in the closed restricted space, connected with dual-side informational links with the external (toward it) environment. The first link is intended for information perceiving from the environment, the second - for information (directing, informative and other, created directly in memory) production into the environment.

Pattern Recognition

Now we consider recognition as the simplest psychical function, which determines the presence of the model of the recognized pattern in the memory as its definite structure.

If such structure exists (e.g. abed model in the described above Fig. 1) then the secondary spike of exciting this structure already passed by the reverse bonds will take place after presenting the prototype of the structure and its switching-off. It just means implementing the "mental sight", i.e. recognizing the presented pattern in accordance with the main hypothesis. If there aren't the fully reverse bonds are absent, i.e. the pattern isn't filled in (such as cdef in Fig. 1) the secondary spike will not take place.

Process modeling already at this stage has needed the special tools taking into account temporary dependences of spreading epyexcitements in neuron networks, and the computer model is constructed using this tools.

Thus, it is hypothetically claimed that recognition of the demonstrated pattern is determined by the secondary spike of exciting the relevant structure as its remembrance [13].

This hypothesis is supported by modeling the components of the olfactory sensory system [12] (performed by neurophysiologist G.S. Voronkov). The results are showed in Fig.2 (the secondary spike doesn't occur if the existing reverse bonds in the olfactory sensory system are interrupted and it obviously occurs if the reverse bonds exist, i.e. there is the full model of the demonstrated pattern in the memory).

Since exciting processes can be both as probabilistic and as possibilistic. The results not defined in full and misoperations (e.g. due to associative bonds (cdef) between the models of two patterns) are certainly possible in a recognizing process. However, it is that it takes place in fact.

"Recognizing" in the presented interpretation is the function implemented quite automatically in any living organism, which has the relevant nervous system.

The implementation of this function in the broad sense of this term already including "comprehension" should relate to the thinking process in which described actions are implemented only as its first stage. During the construction of mathematical models of memory patterns, the apparatus of growing pyramidal networks [14] can be effectively used, as a type of semantic networks.

Pattern Perception and Problem Solving

The introduced conception of pattern structures in the memory as their models is general for all thinking functions since it is a basis for organizing all the memory.

It is very important for modeling just the human thinking processes to consider the memory as a whole system consisting of two subsystems: sensory one, lingual one [6] and higher associative subsystem, which store the patterns and their lingual notations and notions. The structures of these subsystems are interrelated by direct and reverse bonds, which define a correspondence between them and their mutual effect by transferring the excitation. Note that language variety is realized by the additional lingual subsystems whose structures wouldn't bind with the sensor system structures [7] but interact with it only by the structure of subsystem of single language. The type of bonds between sensor and one or other lingual subsystems just defines capability and the level of recognized thinking in one or another language.

The process of human thinking, as specified above is determined by interaction between the sensory and lingual subsystems of the Medium at the cognitive and intuitive (as uncognitive) levels.

The cognitive thinking is organically bound with the lingual expression of thoughts, i.e. individual talks, as it were, with himself or herself. Thus, the cognitive thinking is called verbal although the language wouldn't be a speech. For example, the relevant signals even enter the organs of speech in the first dominating case.

The principally sequential character of cognitive thinking results from it, because it is impossible to pronounce more than one thought at the same time.

In general, the cognitive thoughts are represented by the so-called "complete" dynamic structures, which integrate the relevant excited structures of the sensor and lingual subsystems at the different levels of their hierarchy.

Excitement the incomplete structures is related to uncognitive thinking which isn't restricted by strong interaction of structures of the sensor and lingual subsystems.

Therefore, such dynamic structures can arise at once at the different levels of these subsystems not resulted in "pronouncing" the excited "sense".

So, the amount of information processed (even spontaneously) can be much greater here than during cognitive thinking.

Furthermore, such combinations can arise at the intuitive thinking level, which hasn't lingual equivalents, and so they don't rise to the consciousness level (e.g. savage thinking).

Therefore, the component of an uncognitive intuitive thinking is of quite great importance besides the cognitive component of the purposeful process of solving the problems (man does think!).

The following hypothesis will be quite natural in accordance with above-mentioned facts [5]:

The Problem to be solved is specified in the memory by the models of original and goal situations, and its solution is an activated chain of cause-effect bonds that results in transforming the first one to the second one. The process of constructing the chain consists of two interrelated processes operating at the same time: sequential cognitive process (as reasoning) and spontaneous activating the structures in the memory by their associative bonds with the models of the original and goal situations (we shall ignore below the "model" term and technology that equate the structure in the memory to the situation itself).

Realizing the problem to be solved (the Problem situation) creates some "tension" in the memory, and so the special "problem generator" (PG) [5] term is very useful for obvious illustrating and considering the process specified in the hypothesis, which poles are the original and goal situations and its "tension" maintains the existence of the problem situation.

Creating the activated chain, which locks these poles and means solving the Problem liquidates this "tension", i.e. terminates the PG existence. The links of the specified chain are intermediate situations between the original and goal situations (Fig.2) and they can be defined not only by single-sided transformation but also by cross transformation of these situations.

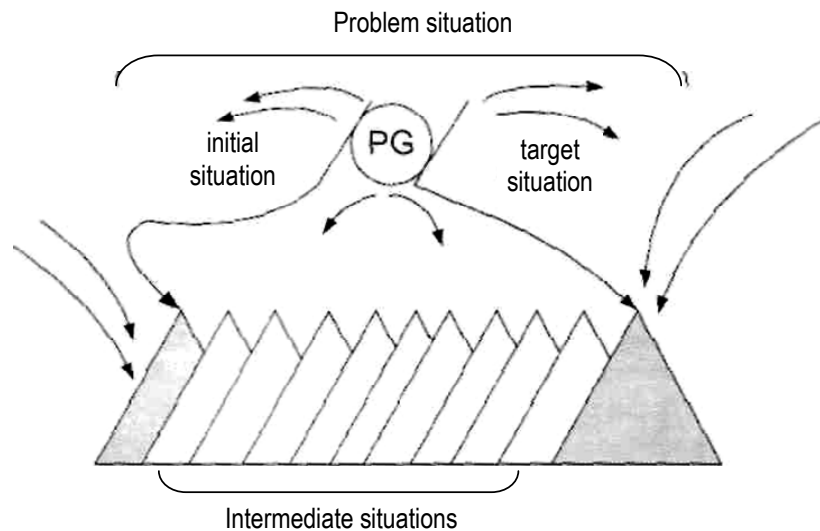


Fig. 2. Chain of the problem solving as a pattern situations transmission

However, the locking chain wouldn't be created in the continuous process (e.g. if there isn't enough knowledge in the memory) that promotes creating the new intermediate PG which defines the break in the constructed chain of its polesjocking, i.e. a new pair of original and goal situations.

The Problem solution can result in creating the new structure (i.e. the new knowledge) in the memory due to beating the new bonds between its components. If there are enough intension and lifetime of the dynamic structure, this process is similar to conversion of the information dynamically stored (i.e. as a short-term memorizing in the statistically fixed memory).

As the "distance" between the original and goal situations decreases due to creating the links of the required chain, which locks them, increase of the activity of the second process and the process as a whole can result in avalanche sudden locking of the PG poles, i.e. solving the Problem as a result of enlightenment. Furthermore, it can occur quite unexpectedly and accidentally just as a result of the second process, only if the first process which is the cognitive reasoning doesn't exist. The second process occurs nevertheless because PG is already excited.

In general, enlightenment is a property of creative processes, which can be schematically considered as a sequence with step-by-step domination of cognitive and intuitive thinking [15].

In the first case, the obtained result is comprehended and the new intermediate chain (sub-chain) is proposed. In the second case (i.e. at the next stage), this sub-chain is already reached and changed, if possible, up to obtaining the final result. Thus, the total process of solving the problem is probabilistic (or possibilistic) with a wide range of its quantitative characteristics. As such its speed and the time depends on the complexity of the problem being solved and the level of excitement of PG. The first factor is stipulated with length of the chain of interconnected structures, which communicate the initial and target situations (i.e. "the distance" between them). The second factor is stipulated mainly with how strongly a human is dedicated to the problem in concern, i.e. how the emotion of interest is involved into its solution.

Emotional Factor in Thinking

As long as the procedure of information processing in memory (in general in neural system) is characterized not only with sense but also with emotional part, let us according to the subject of the report define the major features of interaction between these contributors.

Division of a whole neuron network into separate fields and preparation of informational signal-receptors and memory environment, which accepts these signals, hold also in case, when emotional components are presented. Such components perform emotional influence by creation and transmission of specialized mediators.

These influences, which are established and transmitted in the whole hierarchy of memory, define its particular emotional state (e.g. depression or alternatively inspiration). In turn this state represents a background for mental processes taking place inside the memory. In other words, it guides the activation and interaction of memory structures.

Consequently, the emotional part actively participates in mental process and influences its functional performance. Nevertheless such performance, embodied with corresponding neuro-physiological indicators, naturally effects the emotional condition. For instance, when mental function is involved for solution of a complex problem, as it has been already stated, the emotion of interest (or even inspiration) plays the major role. Moreover, when theoretical problem is solved (the poles of PG are connected) with sudden conjecture, the emotion of extreme joy can appear.

As such, information processing inside the memory can be defined as emotional and sense simultaneously. In addition, this processing in its conceptual representation can be easily represented in terms of described principles of memory organization.

Naturally, major emotional organs are located on the memory level, which accepts 1:1 receptor signals. It means, that during perceiving a pattern by memory from outside they will be organically included into its formed structure.

During the pattern recognition according to the main hypothesis all components of memory will be excited, i.e. the mental pattern obtains the "emotional coloring".

Consequently, memory as a system with structured organization during the information processing creates two independent results in thinking - sense and emotional (correspondingly - idea and humor, which stipulates the state of memory).

Emotional organs are located in neural system and outside memory. They participate in interaction with memory similarly to guided fulfilling mechanisms (like, moving apparatus).

I.e. emotional mediators are presented not only in internal loops of memory structures, but also in reflective loops of its connections with outside (in respect to memory) environment (see above).

In addition, mutual emotional influence appears between them in such interaction. In other words, memory does not only manage, but also is managed by external environment in respect of its adjustment to a particular condition.

Concerning the emotional part of information processing it is necessary to stress, that it especially reflects in subconscious intuitive human thinking and at the same time has a significant value in his/her conscious thinking (for example, in speed-up of searching for particular information in his/her internal database).

As a whole, the introduced conceptual model explains a lot of psychological phenomena, which are related to processes of thinking, as well as clarifies the material substance of its mechanisms (including human's capability erudition, smartness, inspiration etc). It supports our sureness in its plausibility.

It is worth to pay special attention to the fact, that though perceptive meaning is introduced in the model, the model itself can serve as a sufficient basis for its further advanced concretizations.

Possibilities of Bionic Use of Conceptual Model

Conceptual model (CM) has already essential bionic value [5]. But it makes no sense to apply nature as a whole to machines (e.g. legs and wheels). For example, when implementing the analogues of CM characteristics in computer architecture, it is necessary to remember its destination first of all.

For developing the universal high-performance and high-intellectual computers (i.e. computers that have enough advanced internal intelligence), it is quite reasonable to reflect the following characteristics in their architectures (as some analogues of the main CM principles) [16]:

- Distributed processing the information and its operational storage (i.e. processing part of the machine is to be some memory-processor medium).
- The two-component machine computational process: first, the successive one which perceives the user's tasks, initiates, organizes and controls the process of their performing; and second, parallel one which is the component of the total computational process and is responsible for performing the tasks in each own branch.
- Machine tasks and knowledge representation as semantic and associative networks realized by graphs and their hierarchical processing; knowledge is represented by complex data structures at its upper level and in details at the lower level.
- Possibility to adapt to tasks and to organize the total computational process step-by-step with its dynamic scheduling and controlling the results obtained at each step.

The universal computer with the architecture built based on specified principles must promote the effective realizing of different information technologies of alternative classes: a symbolism and connectivism including neural-computer ones. However, in the second case, the technologies are realized at the program models, which can also have sufficiently high characteristics (e.g. a great number of neuron-like elements) and can be structurally realized in part by paralleling processing in memory-processor medium. Furthermore, such computer must promote the effective technology realizing and integrating different technology processes.

A new class of multimicroprocessor cluster computers with specified properties (the so-called intellectual solving machines (ISM)) and, in particular, its model for broad using was developed in the Institute of Cybernetics of NAS of Ukraine) (this project was supported by the grant from the USA chief of the project - Prof. Koval V. beside him the main authors Prof. Bulavenko O. and Prof. Rabinovich Z.) [17, 18].

ISMs in accordance with above-mentioned fact combine the distributed information processing with the internal higher-level language (which has the developed means of knowledge presenting and processing) and dynamic centralized-decentralized (successive and parallel respectively) control. Exactly this set of characteristics stipulates the belonging of ISM to the new class

The high level of machine intellect structurally realized in their cluster architectures contribute the essential increasing the efficiency of user-machine interaction and possibility to perform, at that, complex computations which need the deep fractality of processing data. It provided by using in programming the high and very high level languages and automatic hierarchical organization of program transformation and performance. Such organization, as a fact, provide for the consequence of information processing based on fractal-complexity and complexity transfers.

Bibliography

1. Behtereva N.P.: Healthy and ill human brain. // Nauka, Moscow - 1988 (in Russian).
2. Bloom F., Lazerson A., Hofstadter L.: Brain, mind and behavior. // Mir. Moscow - 1988 (in Russian).
3. Dowling J.E.: Creating mind: how the brain works. // W.W. Norton & Company Inc. - N.Y., London. - 212 p.
4. Haken H., Haken-Krell M: Mystery of perception. // Mir. Moscow - 2002 (in Russian).
5. Rabinovich Z.L.: A particular bionic approach to structural modeling of purposeful thinking. // Cybernetics - 1979. - No 2. - p. 115-118 (in Russian).
6. Voronkov G.S., Rabinovich Z.L.: Sensor and lingual systems - two form of knowledge representation. // News of artificial intelligence. 1993. p. 116-124 (in Russian).
7. Rabinovich Z.L.: About thinking mechanisms and intelligent computers. // Cybernetics and system analysis - 1993. - No 3. - p. 69-78 (in Russian).
8. Rabinovich Z.L., Voronkov G.S.: // Knowledge representation and processing in interaction of sensor and language human neuro-system. // Cybernetics and system analysis - 1998. - No 2. - p. 3-11 (in Russian).
9. Voronkov G.S., Rabinovich Z.L.: // Cogniting model of memory and thinking. // International Journal of Information theories and applications - 2000. - Volume VII. - No 4. - p. 164-169.
10. Rabinovich Z.L.: On natural thinking mechanisms and computer intelligence. // Cybernetics and system analysis - 2003. - No 5. - p. 82-88 (in Russian).
11. Rabinovich Z.L.: Conceptual representation about pattern recognition and problem solving in human memory and possibility of its using in artificial intelligence. // KDS 2005 Proceedings. - Volume I. - XI Conf. - June 20-30. 2005, Varna, Bulgaria, -p. 1-8.
12. Belov Yu.A., Tkachuk S.V., Iamborak R.V.: Mathematical and computer modeling and research of cognitive processes in human brain. // KDS 2005 Proceedings. - Volume I. - XI Conf. – June 20-30. 2005. Varna, Bulgaria, -p. 23-36.
13. Rabinovich Z.L., Belov Yu.A.: Conceptual idea of nature mechanisms of recognition, purposeful thinking and potential of its technical applications. // Lecture Notes in computer science. - Springer- Verlag GmbH. - 3561/2005.
14. Gladun V.P.: Decision planning. // Naukova Dumka, Kyiv - 1987 - 167 p. (in Russian).
15. Glushkov V.M., Rabinovich Z.L.: Problems of deductive constructions automation. // Information management and intelligence. - Moscow - 1976 - part 4. chapter 2. - p. 300-326 (in Russian).
16. Rabinovich Z.L.: On concept of machine Intelligence and its development. // Cybernetics and system analysis - 1995. - No 2. - p. 163-173 (in Russian).
17. Koval V.N., Bulavenko O.N., Rabinovich Z.L.: Intellectual solving machines as basic highly productive computation systems. // Control systems and machines - 1998. No 36. - p. 43-52 (in Russian).
18. Koval V.N., Bulavenko O.N., Rabinovich Z.L.: Parallel architectures and their development on the basis of intelligent solving machines. // Proc. Of the Intern. Conf. of parallel computing in electrical engineering. - Warsaw, Poland, September 22-25 - 2002 - p. 21-26.

Authors' Information

Belov Yuriy Anatoliyevich – Dr., Prof., Head of Department, Theoretical Cybernetics Department, Faculty of Cybernetics, Kyiv National Taras Shevchenko University, Glushkov av. 2, build. 6, Kyiv-03127, Ukraine

Rabinovich Zinoviiv Lvovich – Dr., Prof., V.M. Glushkov Institute of Cybernetics of NAS of Ukraine, Prospekt akad. Glushkova 40, 03680 Kiev, Ukraine; e-mail: aduis@rambler.ru

ОБЯЗАТЕЛЬНО ЛИ ОЩУЩЕНИЯ ЯВЛЯЮТСЯ ИЗОМОРФНЫМИ «ОБРАЗАМИ» МИРА: АНАЛИЗ С НЕЙРОФИЗИОЛОГИЧЕСКИХ ПОЗИЦИЙ НЕКОТОРЫХ АСПЕКТОВ ТЕОРИИ ОТРАЖЕНИЯ¹

Геннадий Воронков

***Аннотация.** Работа направлена на выяснение причин застоя в развитии теории отражения, являющейся до сих пор основной философской концепцией в проблеме сознания и мышления, проблеме, тесно связанной в настоящее время с проблемой «сознания» в искусственном интеллекте. В критическом плане рассматриваются понятие «образ», используемое в теории отражения, и критика теории символов, содержащаяся в теории отражения. Кратко обсуждается возможный путь развития теории отражения.*

***Ключевые слова:** теория отражения, теория символов, ощущение, «образ», изоморфизм, модель, соответствие.*

Введение

Конкретные научные представления и научная картина мира, которую дает философия, находятся в отношениях взаимного обусловливания. Основной философской концепцией в проблеме сознания и мышления, проблеме, тесно связанной теперь с проблемой «сознания» в искусственном интеллекте, остается до сих пор теория отражения. Однако на современном этапе развития наук о мозге и наук, центральным понятием в которых является понятие «информация», явно ощущается недостаток в обобщениях философского характера, а теорию отражения, никак нельзя, по крайней мере, назвать «локомотивом» для этих наук. Стремление выяснить причину такого «состояния дел» в теории отражения послужило иницилирующим началом для данной работы. Работа посвящена анализу некоторых аспектов теории отражения с точки зрения нейрофизиологической концепции, развиваемой автором (и в соавторстве), – взгляда на мозг как нейронную модель.

Теория отражения не является завершенной, некоторые её положения, в частности в отношении ощущений, требуют развития и экспериментальных исследований. Об этом писали и сами основатели теории отражения: «... на деле остаётся ещё исследовать и исследовать, каким образом связывается материя, якобы не ощущающая вовсе, с материей ..., обладающей ясно выраженной способностью ощущения. Материализм ясно ставит нерешенный ещё вопрос и тем самым толкает его к разрешению, толкает к дальнейшим экспериментальным исследованиям.»[1]. Понимание незавершенности этой теории обусловило, видимо, особую осторожность и тщательность у авторов в формулировании её основных положений. Тем не менее, в текстах, излагающих и интерпретирующих её (например, в [1]) имеются высказывания, требующие, как нам представляется, критического рассмотрения, поскольку, либо они допускают неоднозначное толкование ключевого в этой теории термина «образ», либо раскрывающийся в них в явном виде смысл этого термина не согласуется, на наш взгляд, с рядом известных фактов² в отношении работы сенсорных систем, а также с результатами анализа понятий «модель» и

¹ В основу статьи положен внепрограммный доклад, прочитанный на 2-й конференции «Философия сознания: история и современность» (секция 4 «Междисциплинарные проблемы философии сознания»). Философский факультет, МГУ. 17-18 ноября 2006.

² Некоторые из них приведены в трудах (в частности, в [2]) Гельмгольца, одного из авторов теории символов.

«соответствие» [4, 5]¹, проводимого автором как необходимое условие развития упомянутого выше представления [3], кратко формулируемого как «мозг есть нейронная модель». Естественно, критического рассмотрения, по этой причине, требуют и следствия, вытекающие из теории отражения в её исходном варианте, в частности, ленинская критика теории символов.

Конкретно, в настоящей работе рассматриваются в критическом плане (1) понятие «образ», используемое в теории отражения, и (2) критика теории символов, содержащаяся в теории отражения, а также (3) кратко обсуждается возможный путь развития теории отражения.

Основной текст

1. Понятие «образ». Приведем несколько цитат из работы [1] в отношении понятия «образ» (ощущение): «... вне нас существуют вещи. Наши восприятия и представления – образы их»; «Ощущения, то есть образы внешнего мира, существуют в нас, порождаемые действием вещей на наши органы чувств». В этих высказываниях слово, понятие «образ» (ощущение) дается без комментариев, без определения. Поэтому читатель вправе понимать его наиболее привычным образом – как сходное (похожее), верное отображение действительности. Другими словами, под «образом» (ощущением) понимается нечто, которое в каком-то аспекте свойств изоморфно объекту, его вызывающему. Это понимание может быть прочитано и в «осторожных» словах Энгельса, цитируемых в этой же работе [1]: «Нет ни единого случая, насколько нам известно до сих пор, когда бы мы вынуждены были заключить, что наши научно-проверенные чувственные восприятия производят в нашем мозгу такие представления о внешнем мире, которые по своей природе отклоняются от действительности² или что между внешним миром и нашими чувственными восприятиями его существует природенная несогласованность». В явном виде понимание «образа» как обязательно изоморфного, сходного (похожее), хотя бы в каком-то аспекте свойств, с объектом, его вызывающим, содержится в следующих двух цитатах из [1]: 1) на стр. 218, где критикуется теория символов, излагаемая Гельмгольцем: «Если ощущения не суть образы вещей, а только знаки или символы, не имеющие «никакого сходства» с ними, то исходная материалистическая посылка Гельмгольца подрывается, подвергается некоторому сомнению существование внешних предметов, ...» и 2) на стр. 26, где упоминается критика Энгельсом Дюринга: «... материализм последовательный должен ставить здесь «образы», картины или отображения на место «символа»».

Такое понимание «образа» - как обязательно изоморфного (похожее) хотя бы в каком-то аспекте физических свойств, объекту, его вызывающему, - переносится в явном или в неявном виде и в другие варианты теории отражения, например, в современные информационные теории, оперирующие представлением о моделировании в мозге реальной действительности. В этих теориях обязательная изоморфность моделей в отношении их оригиналов принимается, как правило, «по умолчанию», как само собой разумеющееся.

Однако, представление об обязательной изоморфности «образов» (ощущений), то есть об обязательной схожести, похожести их «по своей природе» с объектами их вызывающими, входит в противоречие с известными фактами в отношении работы и организации сенсорных (анализаторных) систем. Так, обонятельное ощущение (запах розы, к примеру) само по себе ни в коей мере ничего не говорит нам о природе, о физических свойствах воздействующего запахового объекта: на основании только испытываемого обонятельного ощущения мы ничего не можем узнать о физических свойствах воздействующего начала, которым вызывается это ощущение, - ни о структуре запаховых молекул, ни о физических характеристиках предметов, источающих запах. То же, в принципе, можно сказать в отношении вкусовых «образов» (ощущений). Ничего также не говорит нам о физических свойствах света (о его сложном спектральном составе) восприятие (ощущение) белого света; более того, смесь двух

¹ В настоящей работе предпринята попытка использовать определения понятий «модель» и «соответствие», данные в [4, 5], как теоретические инструменты для анализа философской концепции - «теории отражения».

² Выделения подчеркиванием здесь и далее сделаны автором настоящей работы.

разных линий видимого спектра, может вызывать такое же ощущение цвета, которое вызывается одной другой (третьей) линией спектра. Эти примеры свидетельствуют о наличии целого ряда сенсорных «восприятий», в которых отсутствует какая бы то ни было изоморфность ощущения («образа») по отношению к качественным свойствам объекта (оригинала), вызывающего данное ощущение. В то же время, в этих примерах, между воздействующим объектом и ощущением, которое им вызывается, как правило, имеет место соответствие: определенному воздействию соответствует определенное, специфическое (но, не обязательно изоморфное) ощущение ¹ («образ»).

Такое «состояние дел» в сенсорных «восприятиях» вполне объяснимо с нейрофизиологических позиций. Например, в концепции «мозг есть нейронная модель» такой (неизоморфный) характер «восприятия» находит объяснение в известном нейрофизиологическом принципе «все или ничего», лежащем в основе нейронных механизмов, реализующих соответствие: с точки зрения этой концепции, нейрон сенсорного пути, принимающий воздействия на разветвления апикального дендрита, подобен воронке; расширенная часть её выполняет роль замка, открываемого только определенным воздействием (своего рода ключом) и запускающего по типу спускового курка активность в узкой части «воронки»; узкая часть работает по типу «все или ничего» и поэтому её активность не несет никакой специфики входного воздействия. Кроме того, предполагаемые в ряде теорий пространственные или пространственно-временные «узоры-паттерны» (в качестве кодов), а также специфика, заключающаяся в сочетаниях (или комбинациях) стимулов, утрачиваются в результате «схождения» в «воронке» нейрона. ²

Показанная выше для целого ряда сенсорных «образов» (ощущений) неизоморфность в отношении объектов, их вызывающих, не означает отсутствие изоморфности у всех сенсорных «образов» (ощущений). Так, изоморфно представлены пространственные отношения между элементами («точками»), представляющими зрительное поле в мозгу - эти отношения являются такими же, как отношения между точками в реальном поле зрения. Таким образом, ощущение предметного видения изоморфно реальным объектам в аспекте пространственных отношений составляющих их элементов. То же, в принципе, по-видимому, можно сказать в отношении звуковых последовательностей и их «образов», представленных на самых низких мозговых уровнях. Сходно (изоморфно) представлены в мозгу также другие (не только пространственные и временные) отношения между объектами. Например, изоморфно по отношению к внешнему миру представлен такой его аспект как «избирательные» связи-отношения между объектами. Эти связи-отношения реализуются специальными нейронными механизмами - избирательными аксонными связями между нейронами, представляющими объекты в мозге. ³ В настоящее время этот аспект мозговой организации интенсивно изучается. Здесь же нам важно было обратить внимание на существование «образов», не изоморфных объектам, их вызывающим, и объяснимость этого обстоятельства с нейрофизиологических позиций, а также на то, что это обстоятельство не было отмечено и учтено в теории отражения.

2. Теория символов. Отмеченное выше отсутствие изоморфности «образов» (по отношению к вызывающим их оригиналам) в целом ряду сенсорных «восприятий» является принципиально важным моментом, например, для выработки отношения к «критике теории символов», данной в работе [1]. «Теория символов» есть «теория, по которой ощущения и представления человека представляют из себя не копии действительных вещей и процессов природы, не изображения их, а условные знаки, символы, иероглифы ...» ([1], стр. 216). Ниже Ленин приводит слова «крупнейшего представителя» этой теории - Гельмгольца: «Идея и объект, представленный ею, суть две вещи, принадлежащие, очевидно, к двум

¹ Оно может соответствовать и нескольким разным, в принципе, объектам.

² Это общепризнаваемое в нейрофизиологии обстоятельство не служит препятствием для создания нейрофизиологических представлений, объясняющих сложную работу мозга. Так, в частности, в концепции «мозг есть нейронная модель» рассматриваются, к примеру, иерархическая организация сенсорного пути именно из таких нейронов, а также реализация ими функций конъюнкции и дизъюнкции, организация памяти и некоторые возможные нейрофизиологические механизмы решения задач [3, 4, 6, 7].

³ В концепции «мозг есть нейронная модель» эти связи-отношения именуется «сигнификативными связями»

совершенно различным мирам» - и критикует их, называя их «вопиющей неправдой», и комментируя: «... так разрывают идею и действительность, сознание и природу только кантианцы». В этих ленинских высказываниях не учитывается факт существования «образов», не изоморфных оригиналам. Ибо, такие «образы» являются (по определению)¹ ничем иным как знаками, символами этих оригиналов. В настоящее время в физиологическом эксперименте регистрируют нейроны, поставленные² в соответствие разным, простым и сложным стимулам - линиям определенного наклона, направлениям движения, скоростям движения, рукам, лицам и другим объектам и свойствам. Такие нейроны являются тоже ничем иным как символами – ибо, они тоже не изоморфны оригиналам. В концепции «мозг есть нейронная модель» (в частности) сенсорный мир представлен в мозгу сложной иерархически организованной моделью из подобных, выше описанных нейронов. Это не приводит к «несогласованности восприятий с миром», к «разрыву идеи и действительности» ни в теоретическом, ни в экспериментальном аспектах: согласованность обеспечивается наличием соответствия между миром объектов и представляющими их нейронами. Это соответствие осуществляют физиологические механизмы, уже в достаточной степени изученные экспериментально и теоретически. Эти механизмы содержат, как отмечалось выше, «элемент», работающий по типу «все или ничего». Активностью такого механизма не передаётся качественная специфика³ (свойства) входного воздействия (ибо соответствующая специфика ощущения априорна, - она, как и механизм соответствия, формируется в филогенезе и онтогенезе и при обучении). Роль этого механизма состоит только в реализации имеющегося соответствия - «когда одна определенная сущность актуализирует определенную другую, поставленную ей в соответствие» (*qui pro quo*). Поэтому такой механизм одновременно выступает и как связующий, и как разграничивающий механизм (как своего рода граница) между внешним миром и его нейронной моделью. В таком представлении о взаимодействии среды и мозга цитированные выше слова Гельмгольца прочитываются как адекватно описывающие и трактующие реальное «состояние дел» в физиологии органов чувств, а критика «теории символов»⁴ Гельмгольца, данная в [1], представляется недостаточно корректной.

3. Возможный путь развития теории отражения. Теория отражения базируется на понятиях «образ» и «отражение». Под отражением понимается один из типов взаимодействия материальных тел (сущностей), под «образом» понимается результат этого взаимодействия – изоморфное отображение (ощущение). Однако эта теория в своем исходном варианте сталкивается, как следует из проведенного выше анализа, с трудностью объяснить факт существования неизоморфных образов-ощущений.

Эта трудность в теории отражения преодолевается, с нашей точки зрения, если в ней понятие «отражение» заменить на понятие «соответствие», а понятие «образ» на понятие «модель»⁵. С наиболее общих позиций, суть этой замены заключается в расширении предмета исследования в теории отражения, а, следовательно, и в расширении самой теории. Ибо, по сути, предлагается заменить предмет её исследования, именно, взаимодействие материальных тел по типу отражения заменить на более широкий - взаимодействие по типу соответствия, которое охватывает и взаимодействие по типу отражения.

¹ Неизоморфным оригиналу может быть только поставленный ему в соответствие знак или символ.

² Постановка в соответствие и формирование механизма, реализующего соответствие, – процессы, происходящие в филогенезе и онтогенезе и при обучении.

³ Как было отмечено выше, специфика входного воздействия играет лишь роль «конфигурации ключа к замку» в механизме соответствия.

⁴ В настоящей статье не обсуждается «эмпиристическая теория ощущений» (противостоящая «нативистической»), которую защищал Гельмгольц [2].

⁵ В смысле определений этих понятий здесь и в [4] и [5].

Вместо заключения

Более чем за столетнее существование теория отражения в её исходном варианте не привела к существенному прогрессу в решении проблемы ощущения (и сознания). Кратко, об этом может свидетельствовать следующее, в значительной степени общепринятое в настоящее время представление. Принято считать на основе нейрофизиологических данных, что ощущения коррелируют с активностью определенных нейронов. Например, есть основания полагать, что ощущение предметного видения связано с активностью нейронов стриарной зрительной коры, но не с активностью, к примеру, непосредственно зрительных нейронов - рецепторов сетчатки, что мотивационные ощущения (к примеру, ощущение жажды) сопровождают активность нейронов в гипоталамусе, и что не сопровождается какими-либо ощущениями активность моторных нейронов. С другой стороны, из нейрофизиологических данных не следует (пока, по крайней мере) обратное, именно, что активность нейронов является коррелятом ощущений, то есть, что активность нейронов является следствием ощущений. Но, в таком случае, остается непонятной роль ощущений в принципе, и возникают вопросы, почему, зачем существуют ощущения, сопровождающие активность определенных афферентных нейронов, и нельзя ли сконструировать (в принципе) автомат («зомби»), не обладающий ощущениями, но обладающий поведением (в самом широком смысле этого слова), ничем не отличающимся от поведения человека. Эти вопросы реально поставлены соответственно в нейрофизиологической и философской современной литературе. Можно видеть, что общее представление об отношении мозг/ощущение, лежащее в основе этих вопросов, отличается от декартовского дуализма только по форме, но не по сути. Другими словами, прогресс современной науки по проблеме ощущений почти не заметен с очень давнего времени. Данная статья не дает видимого решения этой проблемы, она – лишь конкретная попытка движения в этом направлении.

Библиография

1. Ленин В.И. Материализм и эмпириокритицизм. Госполитиздат, 1953.
2. Гельмгольц Г. О восприятиях вообще. // Перевод главы из «H. von Helmholtz. Handbuch der Physiologischen Optik, 1910» в «Хрестоматия по ощущению и восприятию». (Ред. Ю.Б. Гиппенрейтер и М.Б. Михалевская). Изд. московского университета. 1975, с. 61-87.
3. Воронков Г.С. Сенсорная система как нейронная семиотическая модель адекватной среды.// В сб.: Сравнительная физиология высшей нервной деятельности человека и животных. М.: Наука, 1990, с. 9-21
4. Воронков Г.С. Новое содержание в старых понятиях: к пониманию механизмов мышления и сознания.// XI International Conference "Knowledge-Dialogue-Solution", Varna, Bulgaria. KDS-2005, Proceedings. Vol. 1, Fo-commerce, Sofia, 2005, pp 17-23.
5. Воронков Г.С. Понятия «модель» и «соответствие»: нейрофизиологический и общий аспекты.// Тезисы докладов и выступлений IV Российского философского конгресса «Философия и будущее цивилизации». Москва, 2005. Том 1. с. 707-708.
6. Воронков Г.С. Механизмы решения задач в элементарном сенсориуме: нейронные механизмы опознания и сенсорного обучения.// Нейрокомпьютеры: разработка, применение, 2004, № 2-3, с. 92-100.
7. Voronkov G.S., Rabinovich Z.L. On neuron mechanisms used to resolve mental problems of identification and learning in sensorium.// Int. J. "Information theories and Applications", 2003, v. 10, N 1. pp. 23-28.

Информация об авторе

Геннадий С. Воронков – Московский государственный университет им. М.В. Ломоносова, Россия, Москва, 119992, Ленинские Горы 1/12, Биологический факультет. E-mail: av13675@yandex.ru

ГИПЕРСЛУЧАЙНЫЕ МАРКОВСКИЕ МОДЕЛИ

Игорь Горбань

Аннотация: Для описания физических процессов в непредсказуемо меняющихся статистических условиях предложены гиперслучайные марковские модели. Введены новые определения понятий сходимости в среднеквадратическом последовательности гиперслучайных величин и последовательности гиперслучайных функций, позволившие ввести понятия непрерывности, дифференцируемости и интегрируемости гиперслучайных функций. Обобщено понятие марковского процесса на случай гиперслучайных процессов. Получены прямое и обратное уравнения Колмогорова, описывающие диффузионные гиперслучайные процессы. Исследованы винеровский и гауссовский марковский гиперслучайные процессы.

Keywords: теория гиперслучайных явлений, марковский процесс, сходимость в среднеквадратическом.

ACM Classification Keywords: G.3 Probability and Statistics

Введение

Для описания различных физических эффектов широко используются вероятностные подходы. Теория вероятности создавалась, а затем развивалась в первую очередь для описания физических событий, величин, процессов и полей, обладающих свойством статистической устойчивости. Под статистической устойчивостью или стабильностью понимается существование определенного предела частоты соответствующего математического объекта (события, величины, процесса или поля) при устремлении объема выборки к бесконечности. Статистически неустойчивые физические явления (а они составляют подавляющее большинство реальных явлений) с трудом поддаются описанию известными вероятностными методами.

Для описания и изучения статистически нестабильных явлений недавно [1] на базе классической теории вероятностей и математической статистики была предложена новая физико-математическая теория гиперслучайных явлений, рассматривающая как математические, так и физические аспекты проблемы учета статистической неустойчивости физических явлений.

Математическими моделями гиперслучайных физических явлений (событий, величин, процессов, полей) являются соответствующие математические объекты, под которыми понимается множество случайных событий, величин или функций, каждый элемент которого ассоциируется с определенными статистическими условиями $g \in G$.

В настоящее время сформированы математические основы новой теории, однако, остаются еще не изученными многие вопросы, представляющие прикладной интерес, в частности вопросы ее применения для решения статистических задач физики.

Целью настоящей статьи является разработка основ построения гиперслучайных марковских моделей для решения таких задач.

Сходимость в среднеквадратическом, непрерывность, дифференцируемость и интегрируемость гиперслучайных функций

В работах [1, 2] были введены понятия четырех типов сходимости последовательностей гиперслучайных величин и гиперслучайных функций в предположении, что для всех элементов последовательности статистические условия одинаковы. Это предположение существенно ограничивает класс решаемых

задач. В этой связи представляется целесообразным переопределить понятия сходимости, в частности понятие сходимости в среднеквадратическом.

Пусть имеется последовательность гиперслучайных величин $X = \{X_1, \dots, X_N\}$ и гиперслучайная величина X . Для всех X_1, \dots, X_N и X определены условия $g \in G$ и условные функции распределения $F_1(x_1/g), \dots, F_1(x_N/g)$ и $F(x/g)$.

Тогда последовательность X сходится в среднеквадратическом к X , если при $N \rightarrow \infty$ и $g_N \rightarrow g$

$$M[|X_N/g_N - X/g|^2] \rightarrow 0,$$

т.е. случайная последовательность $X_1/g_1, \dots, X_N/g_N$ сходится в среднеквадратическом к случайной величине X/g . При этом будем писать

$$\text{l.i.m.}_{\substack{N \rightarrow \infty \\ g_N \rightarrow g}} X_N/g_N = X/g$$

или $\text{l.i.m.}_{N \rightarrow \infty} X_N = X$.

Аналогично, пусть имеется последовательность гиперслучайных функций $X(t) = \{X_1(t), \dots, X_N(t)\}$ ($t \in T$) и гиперслучайная функция $X(t)$, для которой определены условные функции распределения $F_1(x_1; t/g), \dots, F_1(x_N; t/g)$ и $F(x; t/g)$.

Тогда последовательность $X(t)$ сходится к $X(t)$ в среднеквадратическом, если для всех $t \in T$ при $N \rightarrow \infty$ и $g_N \rightarrow g$

$$M[|X_N(t)/g_N - X(t)/g|^2] \rightarrow 0,$$

т.е. $\text{l.i.m.}_{\substack{N \rightarrow \infty \\ g_N \rightarrow g}} X_N(t)/g_N = X(t)/g$.

Заметим, что в приведенных определениях сходимости последовательности гиперслучайных величин и последовательности гиперслучайных функций в отличие от предложенных в работах [1, 2] определениях этих же понятий не обязательно совпадение статистических условий последовательности и объекта, к которому эта последовательность стремиться (g_1, \dots, g_N, g могут быть в общем случае разными).

Гиперслучайную функцию $X(t)$ ($t \in T$) будем называть гиперслучайной функцией второго порядка, если математическое ожидание нижней границы квадрата этой функции ограничено для всех $t \in T$: $M_t[X^2(t)] < \infty$.

Новое понятие сходимости последовательностей гиперслучайных функций позволяет ввести понятия непрерывности, дифференцируемости и интегрируемости.

Гиперслучайную функцию второго порядка $X(t) = \{X(t)/g_t \in G\}$ будем называть непрерывной в среднеквадратическом в точке t , если

$$\text{l.i.m.}_{\Delta t \rightarrow 0} X(t + \Delta t) = X(t),$$

т. е. для всех условий $g_t, g_{t+\Delta t} \in G$

$$\lim_{\substack{\Delta t \rightarrow 0 \\ g_{t+\Delta t} \rightarrow g_t}} M[|X(t + \Delta t)/g_{t+\Delta t} - X(t)/g_t|^2] = 0.$$

Гиперслучайную функцию $X(t)$ второго порядка назовем дифференцируемой в среднеквадратическом в точке t , если существует функция $X'(t)$ (производная), описываемая следующим выражением:

$$X'(t) = \text{l.i.m.}_{\Delta t \rightarrow 0} \frac{X(t + \Delta t) - X(t)}{\Delta t},$$

т. е. для всех условий $g_t, g_{t+\Delta t} \in G$

$$\lim_{\substack{\Delta t \rightarrow 0 \\ g_{t+\Delta t} \rightarrow g_t}} \mathbb{M} \left[\left| \frac{X(t + \Delta t)/g_{t+\Delta t} - X(t)/g_t}{\Delta t} - X'(t)/g_t \right|^2 \right] = 0.$$

Гиперслучайную функцию $X(t)$ второго порядка будем называть интегрируемой на интервале $T(\tau)$, если при произвольном разбиении интервала $T(\tau)$ на N интервалов $\Delta t_n = t_n - t_{n-1}$ независимо от выбора точек t_n существует функция $Y(\tau)$ (интеграл гиперслучайной функции $X(t)$), определяемая выражением

$$Y(\tau) = \lim_{\substack{\max \Delta t_n \rightarrow 0 \\ g_{t_n} \rightarrow g_t}} \sum_n X(t_n) \Delta t_n = \int_{T(\tau)} X(t) dt,$$

т. е. для всех $g_{t_n}, g_t \in G (n = \overline{1, N})$

$$\lim_{\max \Delta t_n \rightarrow 0} \mathbb{M} \left[\left| \sum_n X(t_n)/g_{t_n} \Delta t_n - \int_{T(\tau)} X(t)/g_t dt \right|^2 \right] = 0.$$

Иначе, гиперслучайная функция $X(t)$ второго порядка непрерывна, дифференцируема или интегрируема, если непрерывны условия и соответственно непрерывны, дифференцируемы или интегрируемы составляющие случайные функции $X(t)/g_t$ для всех $g_t \in G$.

Заметим, что на основании известных теорем для случайных функций справедливы следующие утверждения.

1) Гиперслучайная функция $X(t)$ второго порядка непрерывна в среднеквадратическом в точке t тогда и только тогда, когда для всех $g_t \in G$ математические ожидания $m_{x/g_t}(t)$ случайных функций $X(t)/g_t$ непрерывны в точке t , а ковариационные функции $R_{x/g_1 g_2}(t_1, t_2)$ этих случайных функций непрерывны в точке $t = t_1 = t_2$.

2) Гиперслучайная функция $X(t)$ второго порядка дифференцируема в точке t тогда и только тогда, когда для всех $g_t \in G$ математические ожидания $m_{x/g_t}(t)$ случайных функций $X(t)/g_t$ дифференцируемы в точке t и в точке $t_1 = t_2$ существуют смешанные производные второго порядка

$$\frac{\partial^2}{\partial t_1 \partial t_2} R_{x/g_1 g_2}(t_1, t_2) \text{ от ковариационных функций } R_{x/g_1 g_2}(t_1, t_2).$$

3) Гиперслучайная функция $X(t)$ второго порядка с математическими ожиданиями $m_{x/g_t}(t)$ и ковариационными функциями $R_{x/g_1 g_2}(t_1, t_2)$ интегрируема, если существуют интегралы

$$\int_{T(\tau)} m_{x/g_t}(t) dt, \int_{T(\tau)} \int_{T(\tau)} R_{x/g_1 g_2}(t_1, t_2) dt_1 dt_2. \text{ При этом}$$

$$\begin{aligned} \mathbb{M} \left[\int_{T(\tau)} X(t) / g_t dt \right] &= \int_{T(\tau)} m_{x/g_t}(t) dt, \\ \mathbb{M} \left[\int_{T(\tau)} \int_{T(\tau)} (X(t_1) / g_{t_1})(X(t_2) / g_{t_2}) dt_1 dt_2 \right] &= \\ &= \int_{T(\tau)} \int_{T(\tau)} R_{x/g_{t_1}g_{t_2}}(t_1, t_2) dt_1 dt_2 + \int_{T(\tau)} m_{x/g_{t_1}}(t) dt \int_{T(\tau)} m_{x/g_{t_2}}(t) dt. \end{aligned}$$

Непрерывный марковский гиперслучайный процесс

Пусть $X_0 = X(t_0), \dots, X_N = X(t_N)$ — значения непрерывного гиперслучайного процесса $X(t)$ в произвольные моменты времени $t_0 < t_1 < \dots < t_N$. Гиперслучайный процесс $X(t)$ назовем марковским, если для любого момента времени t_N и любого условия $g_{t_N} \in G$ одномерная условная плотность вероятности

$$\begin{aligned} f_1(x_N; t_N; g_{t_N} / x_0, \dots, x_{N-1}; t_0, \dots, t_{N-1}; g_{t_0}, \dots, g_{t_{N-1}}) &= \\ = f_1(x_N; t_N; g_{t_N} / x_{N-1}; t_{N-1}; g_{t_{N-1}}). \end{aligned} \quad (1)$$

Отсюда следует, что многомерная плотность вероятности марковского гиперслучайного процесса $X(t)$ может быть представлена следующим образом:

$$\begin{aligned} f_N(x_0, \dots, x_N; t_0, \dots, t_N; g_{t_0}, \dots, g_{t_N}) &= \\ = f_1(x_0; t_0; g_{t_0}) \prod_{n=1}^N \Pi(x_n; t_n; g_{t_n} / x_{n-1}; t_{n-1}; g_{t_{n-1}}), \end{aligned} \quad (2)$$

где $\Pi(x_n; t_n; g_{t_n} / x_{n-1}; t_{n-1}; g_{t_{n-1}}) = f_1(x_n; t_n; g_{t_n} / x_{n-1}; t_{n-1}; g_{t_{n-1}})$ — плотность вероятности перехода случайной величины $X(t) / g_t$, находящейся в состоянии x_{n-1} в момент времени t_{n-1} в условиях $g_{t_{n-1}}$ в состояние x_n в момент времени t_n в условиях g_{t_n} .

Из выражения (1) следует, что если значения гиперслучайного процесса в любые несовпадающие моменты времени независимы при всех условиях g_t [1], то процесс — марковский. Обратное утверждение неверно.

Плотность вероятности перехода $\Pi(x; t; g_t / x'; t'; g_{t'})$ является неотрицательной величиной, нормированной к единице:

$$\int_{-\infty}^{\infty} \Pi(x; t; g_t / x'; t'; g_{t'}) dx = 1.$$

Кроме того, эта плотность вероятности обладает свойством сингулярности:

$$\lim_{\substack{t \rightarrow t' \\ g_t \rightarrow g_{t'}}} \Pi(x; t; g_t / x'; t'; g_{t'}) = \delta(x - x')$$

и удовлетворяет обобщенному уравнению Маркова (уравнению Смолуховского)

$$\Pi(x; t; g_t / x_0; t_0; g_{t_0}) = \int_{-\infty}^{\infty} \Pi(x; t; g_t / x'; t'; g_{t'}) \Pi(x'; t'; g_{t'} / x_0; t_0; g_{t_0}) dx'.$$

Для марковского гиперслучайного процесса $X(t)$ плотность вероятности перехода $\Pi(x; t; g_t / x_0; t_0; g_{t_0})$ из состояния x_0 в момент времени t_0 в условиях g_{t_0} в состояние x в момент времени t в условиях g_t определяется уравнением

$$-\frac{\partial}{\partial t_0} \Pi(x; t; g_t / x_0; t_0; g_{t_0}) = \sum_{v=1}^{\infty} \frac{A_v(x_0; t_0; g_{t_0})}{v!} \frac{\partial^v \Pi(x; t; g_t / x_0; t_0; g_{t_0})}{\partial x_0^v}, \quad (3)$$

где

$$\begin{aligned} A_v(x_0; t_0; g_{t_0}) &= \\ &= \lim_{\substack{\Delta t \rightarrow 0 \\ g_{t_0+\Delta t} \rightarrow g_{t_0}}} \frac{1}{\Delta t} M \left[(X(t_0 + \Delta t; g_{t_0+\Delta t} / x_0; t_0; g_{t_0}) - X(t_0; g_{t_0} / x_0; t_0; g_{t_0}))^v \right]. \end{aligned}$$

Это выражение прямо следует из известной теоремы для случайного марковского процесса [3, 4].

Случайная величина $X(t_0 + \Delta t; g_{t_0+\Delta t} / x_0; t_0; g_{t_0}) - X(t_0; g_{t_0} / x_0; t_0; g_{t_0})$ представляет собой приращение состояния, происходящее за время Δt . Поэтому коэффициенты $A_v(x_0; t_0; g_{t_0})$ можно трактовать как локальные скорости изменения начальных моментов v -го порядка приращения состояния. По аналогии с уравнениями, описывающими диффузионные случайные процессы, гиперслучайные процессы, описываемые уравнением (3) с коэффициентами, равными нулю для всех $v \geq 3$, будем называть диффузионным или первым (обратным) уравнением Колмогорова:

$$\begin{aligned} -\frac{\partial}{\partial t_0} \Pi(x; t; g_t / x_0; t_0; g_{t_0}) &= \\ &= a(x_0; t_0; g_{t_0}) \frac{\partial}{\partial x_0} \Pi(x; t; g_t / x_0; t_0; g_{t_0}) + \frac{1}{2} b(x_0; t_0; g_{t_0}) \frac{\partial^2}{\partial x_0^2} \Pi(x; t; g_t / x_0; t_0; g_{t_0}), \end{aligned} \quad (4)$$

а уравнение

$$\begin{aligned} \frac{\partial}{\partial t} \Pi(x; t; g_t / x_0; t_0; g_{t_0}) &= -\frac{\partial}{\partial x} \left[a(x; t; g_t) \Pi(x; t; g_t / x_0; t_0; g_{t_0}) \right] + \\ &+ \frac{1}{2} \frac{\partial^2}{\partial x^2} \left[b(x; t; g_t) \Pi(x; t; g_t / x_0; t_0; g_{t_0}) \right] \end{aligned} \quad (5)$$

– уравнением Фоккера – Планка – Колмогорова или прямым уравнением Колмогорова, где $a(x_0; t_0; g_{t_0}) = A_1(x_0; t_0; g_{t_0})$ – коэффициент сноса, а $b(x_0; t_0; g_{t_0}) = A_2(x_0; t_0; g_{t_0})$ – коэффициент диффузии.

Гиперслучайные марковские процессы, описываемые уравнениями (4), (5), будем называть диффузионными.

Уравнения (4) и (5) – зависимые.

Из уравнения (5) следует уравнение для одномерной плотности вероятности

$$\frac{\partial}{\partial t} f_1(x; t; g_t) = -\frac{\partial}{\partial x} \left[a(x; t; g_t) f_1(x; t; g_t) \right] + \frac{1}{2} \frac{\partial^2}{\partial x^2} \left[b(x; t; g_t) f_1(x; t; g_t) \right]. \quad (6)$$

Гиперслучайный диффузионный марковский процесс будем называть однородным во времени, если плотность вероятности перехода $\Pi(x; t; g_t / x_0; t_0; g_{t_0})$ не зависит прямо от времени t , t_0 и условий g_t , g_{t_0} , а определяется лишь их разностями $\tau = t - t_0$, $g_\tau = g_t - g_{t_0}$:

$\Pi(x; t; g_t / x_0; t_0; g_{t_0}) = \Pi(x / x_0; \tau; g_\tau)$. Одномерная плотность вероятности такого процесса $f_1(x)$, а также коэффициенты сноса $a(x)$ и диффузии $b(x)$ не зависят от времени и условий.

Если непрерывный гиперслучайный марковский процесс стационарен в узком смысле при всех условиях [2], то он однороден. Это следует из известной теоремы для случайных марковских процессов [4]. Обратное утверждение неверно.

Из соотношения (6) следует, что для стационарного в узком смысле при всех условиях диффузионного однородного гиперслучайного марковского процесса

$$\frac{d}{dx}[b(x)f_1(x)] = 2a(x)f_1(x) + C,$$

где C – константа, определяемая из условия нормировки.

Стохастическим дифференциальным уравнением, описывающим гиперслучайный процесс, будем называть уравнение вида

$$\frac{dx}{dt} = h(x, t, g_t) + g(x, t, g_t)n(t; g_t),$$

где $h(x, t, g_t)$ и $g(x, t, g_t)$ – детерминированные функции, удовлетворяющие условиям Липшица

$$|h(x, t, g_t) - h(y, t, g_t)| + |g(x, t, g_t) - g(y, t, g_t)| \leq L|x - y| \quad (L = \text{const} > 0),$$

$n(t) = \{n(t; g_t), g_t \in G\}$ – гиперслучайный гауссовский белый шум с нулевым математическим ожиданием и спектральной плотностью мощности $N_0(g_t)/2$, зависящей от условий g_t в момент времени t .

Заметим, что в данном случае гиперслучайный белый шум отличается от описанного в работе [1] гиперслучайного белого шума при всех условиях тем, что статистические условия и спектральная плотность мощности могут быть разными для разных моментов времени t .

Рассмотрим два примера диффузионных гиперслучайных марковских процессов: винеровский и гауссовский марковский.

Винеровский гиперслучайный процесс

Разберем следующую задачу статистической механики. Пусть в газе или жидкости находится микрочастица единичной массы. Температура среды T непредсказуемо меняется в пределах $[T_1, T_2]$.

При фиксированной температуре T скорость теплового движения молекулы в фиксированном направлении представляет собой случайную величину, описываемую в приближении Максвелла гауссовским законом распределения с нулевым математическим ожиданием и дисперсией kT/m [5], где k – постоянная Больцмана, m – масса молекулы.

Из-за непредсказуемого изменения температуры среды скорость теплового движения молекулы статистически нестабильна и может быть описана гиперслучайной величиной, границы дисперсии которой равны kT_1/m , kT_2/m . Молекулы, сталкиваясь с частицей, вызывают ее перемещение. В каждый фиксированный момент времени происходит большое число таких столкновений. Силу удара $N(t; g_t)$, вызывающего движение частицы вдоль заданного направления, можно рассматривать как гиперслучайную функцию гауссовского типа (при фиксированных статистических условиях g_t (температуре среды) сила удара описывается гауссовской случайной функцией в виде гауссовского белого шума с нулевым математическим ожиданием и спектральной плотностью мощности $N_0(g_t)/2$).

При этом скорость движения частицы $V(t; g_t)$ также оказывается гиперслучайной функцией.

На основании второго закона Ньютона эта скорость описывается следующим стохастическим дифференциальным уравнением:

$$\frac{dv(t; g_t)}{dt} = n(t; g_t). \quad (7)$$

Решением уравнения при нулевом начальном условии ($v(0; g_t) = 0$) является чисто диффузионный (винеровский) гиперслучайный процесс

$$v(t; g_t) = \int_0^t n(t_1; g_{t_1}) dt_1. \quad (8)$$

Из выражения (8) видно, что значение процесса в текущий момент времени t в условиях g_t определяется множеством статистических условий в момент времени t и предшествующие ему моменты времени $t_1 < t$. Это значение зависит от частоты встречаемости в реализации тех или иных условий.

Гиперслучайный винеровский процесс обладает свойствами, похожими (но не идентичными) свойствам случайного винеровского процесса:

- гиперслучайный винеровский процесс является центрированным ($m_v(t) = M[V(t; g_t)] = 0$),
- дисперсия этого процесса описывается интегралом:

$$\sigma_v^2(t) = \int_0^t \int_0^t M[n(t_1; g_{t_1})n(t_2; g_{t_2})] dt_1 dt_2 = \int_0^t \frac{N_0(g_{t_1})}{2} dt_1$$

- (в частном случае, когда условия g не зависят от времени t дисперсия содержит линейный множитель: $\sigma_v^2(t) = \frac{N_0(g)}{2} t$),
- процесс – гауссовский. Его плотность вероятности описывается выражением

$$f_1(v; t; g_t) = \frac{1}{\sqrt{2\pi\sigma_v(t)}} \exp\left(-\frac{v^2}{2\sigma_v^2(t)}\right), \quad (9)$$

- процесс – марковский (т.к. описывается выражением $v(t_3) = v(t_2; g_{t_2}) + \int_{t_2}^{t_3} n(t; g_t) dt$),
- процесс имеет нулевой коэффициент сноса ($a(v; t; g_t) = 0$) и коэффициент диффузии $b(v; t; g_t) = \frac{N_0(g_t)}{2}$, в общем случае зависящий от времени.

Границы дисперсии $\sigma_{ix}^2(t)$, $\sigma_{sx}^2(t)$ винеровского гиперслучайного процесса описываются выражениями

$$\sigma_{iv}^2(t) = \frac{N_{i0}}{2} t, \quad \sigma_{sv}^2(t) = \frac{N_{s0}}{2} t,$$

где N_{i0} и N_{s0} – соответственно нижняя и верхняя границы спектральной плотности мощности $N_0(g_t)$ гиперслучайного белого шума.

Отсюда следует, что диапазон изменения дисперсии расширяется пропорционально t .

Для гиперслучайного винеровского процесса с учетом приведенных свойств прямое уравнение Колмогорова имеет вид

$$\frac{\partial f_1(v; t; g_t)}{\partial t} = \frac{1}{4} N_0(g_t) \frac{\partial^2 f_1(v; t; g_t)}{\partial v^2},$$

а его решение описывается выражением (9).

Гауссовский марковский гиперслучайный процесс

Обобщением рассмотренного винеровского гиперслучайного процесса является гиперслучайный гауссовский марковский процесс $X(t)$, удовлетворяющий стохастическому дифференциальному уравнению

$$\frac{dx(t; g_t)}{dt} + \alpha x(t; g_t) = \gamma n(t; g_t), \quad (10)$$

где α, γ — постоянные коэффициенты.

К такому уравнению можно прийти, в частности, рассматривая предыдущую задачу с учетом вязкости среды.

Тот факт, что рассматриваемый гиперслучайный процесс $X(t)$ является гауссовским следует из того, что $N(t; g_t)$ — гауссовский белый шум, а уравнение (10) — линейное. То, что гиперслучайный процесс $X(t)$ является марковским, следует из того, что случайный процесс $X(t; g_t)$ — марковский [4].

Общим решением однородного уравнения, соответствующего уравнению (10), является $x(t; g_t) = Ce^{-\alpha t}$, где C — константа. Частным решением этого уравнения является

$x(t; g_t) = \gamma e^{-\alpha t} \int_0^t e^{\alpha t_1} n(t_1, g_{t_1}) dt_1$, а его общим решением —

$$x(t; g_t) = x(0, g_0) e^{-\alpha t} + \gamma e^{-\alpha t} \int_0^t e^{\alpha t_1} n(t_1, g_{t_1}) dt_1, \quad (11)$$

где $x(0, g_0)$ — начальные условия в нулевой момент времени в условиях g_0 .

Из выражения (11) видно, что значение процесса в текущий момент времени t в условиях g_t также как и для винеровского гиперслучайного процесса определяется множеством статистических условий g_{t_1} в предшествующие моменты времени $t_1 < t$ и в момент времени t . Но в отличие от винеровского процесса это значение зависит не от частоты встречаемости в реализации тех или иных условий g_{t_1} , а от последовательности следования этих условий.

Гиперслучайный гауссовский марковский процесс обладает свойствами, похожими (но не идентичными) свойствам случайного гауссовского марковского процесса:

- математическое ожидание гиперслучайного гауссовского марковского процесса не зависит от изменения во времени условий и определяется условиями g_0 в первоначальный момент времени:

$$m_x(t) = x(0; g_0) e^{-\alpha t},$$

- дисперсия этого процесса $\sigma_x^2(t) = \frac{\gamma^2}{2} \int_0^t N_0(g_{t_1}) e^{2\alpha(t-t_1)} dt_1$,

- ковариационная функция процесса $R_x(t_1, t_2) = \sigma_x^2(t) e^{-\alpha|\tau|}$, где $\tau = t_2 - t_1$; $t = \min(t_1, t_2) \geq 0$,

- коэффициент сноса $a(x; t; g_t) = -\alpha x(t; g_t)$, а коэффициент диффузии $b(x; t; g_t) = \gamma^2 \frac{N_0(g_t)}{2}$.

Нетрудно убедиться, что при $\alpha > 0$ границы дисперсии, $\sigma_{ix}^2(t)$, $\sigma_{sx}^2(t)$ и границы ковариационной функции $R_{ix}(t_1, t_2)$, $R_{sx}(t_1, t_2)$ гауссовского марковского гиперслучайного процесса описываются выражениями

$$\sigma_{ix}^2(t) = \frac{N_{i0}\gamma^2}{4\alpha}(1 - e^{-2\alpha t}), \quad \sigma_{sx}^2(t) = \frac{N_{s0}\gamma^2}{4\alpha}(1 - e^{-2\alpha t}),$$

$$R_{ix}(t_1, t_2) = \sigma_{ix}^2(t)e^{-\alpha|t|}, \quad R_{sx}(t_1, t_2) = \sigma_{sx}^2(t)e^{-\alpha|t|}.$$

Отсюда следует, что с ростом t диапазон изменения дисперсии гиперслучайного процесса постепенно возрастает, но при $t \rightarrow \infty$ стремится к независящему от времени интервалу $\left[\frac{N_{i0}\gamma^2}{4\alpha}, \frac{N_{s0}\gamma^2}{4\alpha} \right]$

(рис. 1,а). Дисперсия процесса в момент времени t определяется в первую очередь статистическими условиями в этот момент времени и непосредственно предшествующие ему моменты времени.

При увеличении величины τ (интервала между отсчетами) диапазон изменения ковариационной функции уменьшается и при $\tau \rightarrow \infty$ стремится к нулю (рис. 1,б). При этом коэффициент корреляции процесса $r_x(t_1, t_2) = \frac{R_x(t_1, t_2)}{\sigma_x^2(t)} = e^{-\alpha|t|}$ не зависит от изменения статистических условий во времени.

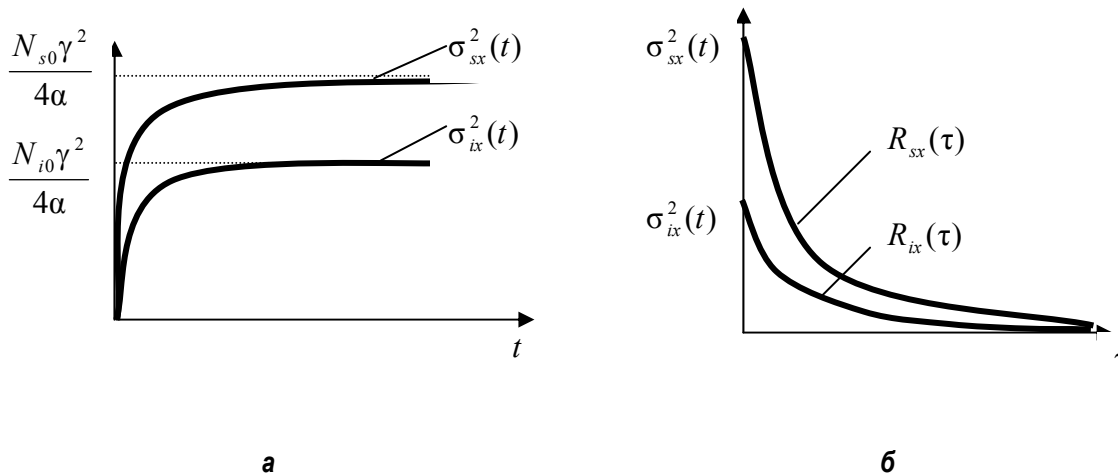


Рис. 1.

Прямое уравнение Колмогорова для плотности вероятности гауссовского гиперслучайного процесса имеет вид

$$\frac{\partial f_1(x; t; g_t)}{\partial t} = \alpha \frac{\partial}{\partial x} [x f_1(x; t; g_t)] + \frac{\gamma^2 N_0(g_t)}{2} \frac{\partial^2 f_1(x; t; g_t)}{\partial x^2}.$$

Решение этого уравнения описывается гауссовской функцией

$$f_1(x; t; g_t) = \frac{1}{\sqrt{2\pi\sigma_x(t)}} \exp\left(-\frac{(x - m_x(t))^2}{2\sigma_x^2(t)}\right).$$

Выводы

1. Для описания физических процессов в непредсказуемо меняющихся статистических условиях предложены гиперслучайные марковские модели.
2. Предложены новые определения понятия сходимости в среднеквадратическом последовательности гиперслучайных величин и последовательности гиперслучайных функций, позволившие ввести понятия непрерывности, дифференцируемости и интегрируемости гиперслучайных функций.
3. Обобщено понятие марковского процесса на случай гиперслучайных процессов. Получены прямое и обратное уравнения Колмогорова, описывающие диффузионные гиперслучайные процессы.
4. Исследованы винеровский и гауссовский марковский гиперслучайные процессы. Установлено, что значения этих процессов определяются статистическими условиями в текущий момент времени и моменты времени, предшествующие ему. Дисперсия винеровского гиперслучайного процесса определяется частотой повторяемости статистических условий в моменты времени, предшествующие рассматриваемому. Дисперсия гауссовского марковского гиперслучайного процесса определяется в основном статистическими условиями в текущий и непосредственно предшествующие моменты времени. Математическое ожидание и коэффициент корреляции гиперслучайного гауссовского марковского процесса не зависят от изменения условий во времени.

Литература

- [Горбань, 2007] Горбань И.И. Теория гиперслучайных явлений. К.: ИПММС НАН Украины, 2007. – 184 с.
- [Горбань, 2006] Горбань И.И. Стационарные и эргодические гиперслучайные функции. // Известия вузов. Радиоэлектроника. – 2006. – № 6. – С. 54 – 70.
- [Горбань, 2003] Горбань И.И. Теорія ймовірностей і математична статистика для наукових працівників та інженерів. – К.: Інститут проблем математичних машин і систем НАН України, 2003. – 244 с.
- [Королюк и др., 1985] Королюк В.С. и др. Справочник по теории вероятностей и математической статистике. – М.: Наука, 1985. – 640 с.
- [Яворский, Детлаф, 1968] Яворский Б.М., Детлаф А.А. Справочник по физике для инженеров и студентов ВУЗов. М.: Наука, 1968. – 940 с.

Сведения об авторе

Горбань Игорь Ильич – заместитель генерального директора по научной работе ГП „УкрНИУЦ”, доктор технических наук, профессор, Украина, Киев, ул. Святошинская, 2;
e-mail: gorban@ukrmdnc.org.ua.

СТРУКТУРЫ, НЕОПРЕДЕЛЁННОСТЬ: МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

Владимир Донченко

Аннотация: Рассмотрены общие проблемы, связанные с математическим моделированием структуры объекта и неопределённости в информации о нём, включая природу, источники и математические методы их описания - моделирования. Проведена систематизация методов описания неопределённости.

Ключевые слова: Структурность, неопределённость, обратные задачи, нечёткие множества, преобразование Хока, псевдообращение по Муру – Пенроузу.

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

Вступление

Прикладная математика является основой универсального метода верифицированного описания объектов внешнего мира с целью их эффективного использования. Собственно, все физические теории представляют собой математические модели для специальных предметных областей (физических явлений), реализующие ту или иную степень абстракции в описании исследуемой области. В связи с задачей моделирования вообще и математического в частности принципиальными являются два момента, которые хотелось бы отметить. Прежде всего – это представление о структуре объекта: обычным является использование оборотов «структура объекта», «описание структуры объекта», и т.д. Объект выступает для исследователя в виде того, что интуитивно понимается как «структура». Таким образом, методы передачи структуры в математическом моделировании являются принципиальными в математическом описании (представлением в виде математической модели) исследуемого объекта. Другим принципиальным моментом, характеризующим исследуемый с целью математического моделирования объект, является то, что часто передаётся классическим выражением «неопределённость в поведении» исследуемого объекта. Классическим математическим средством описания неопределённости в её статистическом, «случайном», проявлении является теория вероятностей и математическая статистика (ТВиМС). Вторая половина XX столетия в математике характеризуется интенсивными усилиями по созданию математических средств описания и оперирования с неопределённостью, альтернативных ТВиМС, к которым можно отнести теорию построения оценок с гарантированной точностью (теорию минимаксного оценивания), теорию нечётких множеств, а также преобразование Хока (ПХ). В то же время, многие надежды, связывавшиеся с появившимися теориями, не оправдались. В значительной мере это относится к теории нечётких множеств. Как представляется, многообразие методов математического описания неопределённости, к которым можно отнести: 1) детерминированный, в том числе обратные задачи; 2) статистический; 3) метод получения оценок с гарантированной точностью; 4) метод нечётких множеств; 5) ПХ, – порождает необходимость приведения их к общей основе. Это означает осмысление природы неопределённости, создание общего методологического подхода, который позволил с единой точки зрения рассматривать разнообразные математические методы её описания.

Предлагаемая работа посвящена обсуждению двух, как представляются автору принципиальных, автору концепций «структуры» и «неопределённости». В ней классифицированы основные математические методы «структурирования» исследуемого объекта, и изложена концепция «множественных моделей неопределённости», связывающая представление о неопределённости в поведении исследуемого объекта с экспериментом. Платформа «множественных моделей неопределённости» позволяет с единых позиций рассматривать математические методы описания неопределённости, к которым отнесены: детерминированный с проблемой скрытых параметров; обратные задачи: статистический; метод оценки параметров с гарантированной точностью (минимаксный); преобразование Хока (Hough Transform); нечёткий. Для последнего вида описания неопределённости обсуждаются внутренние и внешние проблемы развития, в широких предположениях доказано утверждение, позволяющее говорить о нечёткости как о статистическом феномене.

«Структура объекта»: «связь» и «взаимная зависимость» основных частей объекта

В представлении о структуре объекта реализуется представление о том, что объект состоит из взаимосвязанных частей. Таким образом термин «структура» является эквивалентом термина «связи» между частями объекта, рассматриваемого как единое целое. Причем оба эти термина используется на интуитивном, не определяемом строго уровне. Важно только отметить, что в таком контексте математические средства описания того, что называют «структурой» объекта, оказываются средствами описания «связей», взаимной зависимости, которые существуют между выделенными частями исследуемого объекта.

«Структура объекта» – математические средства передачи

Если посмотреть на математику в части её прикладных возможностей: прежде всего с точки зрения возможностей описания структуры исследуемого объекта, то следует отметить, что, для перечисления основных средств описания структуры хватит пальцев одной руки: это отношения, операции, функции, наборы подмножеств. Все основные структуры объектов в тех или иных областях, включая и саму математику, являются комбинацией четырёх основных, упомянутых выше. Чёткое выделение основных способов описания структур - связей в математике делает, с одной стороны, абсолютно прозрачными выражения вида: «структура группы», которой снабжено то или иное множество; «структура линейного пространства», характерная для исследуемого объекта; «структура топологического пространства», «структура поля» и т.д. С другой – лишней раз демонстрирует плодотворность реализации программы Георга Кантора: все математические объекты – это множества, структурированные тем или иным способом: с тем или иным вариантом связей между элементами исследуемого множества.

Заметим, что, хотя функции, как и операции, можно рассматривать как частный случай отношений, исторически они используются специальным, а не частным образом. Не обращаясь к строгим определениям, отметим следующие принципиальные примеры перечисленных выше основных или базовых математических «структур». Важным примером отношений являются отношение частичного порядка на множестве действительных чисел; примером операций – операции сложения и умножения на множестве действительных чисел, а также – покомпонентного умножения или покомпонентного сложения для конечных наборов чисел (числовых векторов). Важным примером функций являются: норма, скалярное произведение, расстояние(метрика). Топология, в том числе и на множестве действительных

чисел, задаётся подходящими наборами множеств: тех, которые интерпретируются как открытые с соответствующими ограничениями-требованиями. Могут задаваться наборы замкнутых множеств, или – наборы окрестностей. Как и топология, измеримое пространство задаётся парой: множество и алгебра или σ - алгебра его подмножеств. Напомним, что по своему предназначению измеримое пространство используется как область определения меры или как носитель информации о событиях в теории случайных процессов.

Выражение «множество с заданной на нём структурой абелевой группы» означает что на множестве задана коммутативная бинарная (обозначаемая обычно «+») и нульарная операции (выделен элемент, обозначаемый через 0) с определёнными свойствами (например, ассоциативность) и связанные между собой определённым образом (например, $a+0=a$). Точно таким же образом, «структура линейного пространства» («векторного») для множества означает задание на нём структуры абелевой группы («+») и континуального набора операций умножения на скаляр, связанных между собой (например, $\lambda(a+b) = \lambda a + \lambda b$). Собственно, «векторами» называют элементы множества, являющегося линейным (векторным) пространством. Выражение «задана структура топологического пространства» означает, что для последовательностей элементов множества имеет смысл предельный переход, точнее – заданы наборы множеств, которые интерпретируются как открытые, а отсюда появляются замкнутые, окрестности и т.д.

Становится понятным выражения «топологическое векторное пространство» или «заданная структура топологического векторного пространства» (для абстрактного множества). Это означает, что в линейном (векторном) пространстве задаются открытые (замкнутые) множества : определён предельный переход для последовательностей элементов (векторов), причём, такой предельный переход согласован с основными линейными операциями (сложение, умножение на скаляр).

Вводя такие классические алгебраические объекты, как «алгебра» и «модель», академик А.И.Мальцев [Мальцев, 1970] определял «алгебру» как множество с тем или иным набором операций, а «модель» – как множество с фиксированным набором операций и отношений.

Что понимают под наблюдением (экспериментом, опытом, испытанием)?

Понятие неопределённости в поведении исследуемого явления или системы тесно связано с понятием «опыта», «эксперимента», «наблюдения», «испытания», которые рассматриваются в рамках категории «опыта». Перечисленные понятия имеют общенаучное содержание и часто употребляются как эквивалентные. Кроме того, они употребляются как эквивалентные между собой и в теории вероятностей и математической статистике. В своём общенаучном смысле эти понятия предназначены для описания деятельности, связанной с непосредственной фиксацией фактов на уровне явления: в процессе непосредственного взаимодействия людей с внешним миром

Для выяснения конкретики общенаучного содержания определения «эксперимент», «опыт», «наблюдение» обратимся к нейтральному, обезличенному источнику, каковым является, к примеру, БСЭ (Большая советская энциклопедия).

В т. 18 БСЭ на стр.463-464 отмечается, что категория «опыта» совпадает по своей сути с категорией «эксперимента» и «наблюдения». Что касается «эксперимента» и «наблюдения», то в том же издании БСЭ, но в томе 30 на стр.6 в статье, посвящённой понятию «эксперимент» отмечается, что термин

происходит от латинского *experimentum*: проба, опыт, – и означает «метод познания, при помощи которого в контролируемых и управляемых условиях исследуются явления действительности». В той же статье отмечается, что «эксперимент» отличается от «наблюдения» тем, что в первом осуществляется активное оперирование с объектом исследования. Таким образом, в цитируемых источниках понятия «эксперимент» и «наблюдение» различаются активностью или пассивностью в оперировании с исследуемым объектом. В то же время в статье, посвященной «наблюдению» в том же издании БСЭ на стр. 186 в т. 17 отмечается, что наблюдение, обычно, является частью «эксперимента». В той же статье в связи с «наблюдением» появляется сочетание «регистрация наблюдений». Таким образом, можно сделать вывод о том, что опыт или эксперимент является методом познания, который заключается в воссоздании стандартных условий наблюдения исследуемого явления и фиксации соответствующих результатов: того что при созданных условиях появляется. Конечно же, вне поля зрения сознательно оставляется обсуждение вопросов о том, как и каким именно образом обеспечивается создание тех или иных условий, а также вопрос о том, как формируется представление о том, что же именно считать результатом эксперимента.

Эмпирический аспект: наблюдение (эксперимент, опыт, испытание)

Таким образом, из приведённых выше цитирований можно сделать вывод, что принципиальными составляющими «эксперимента» являются:

- воссоздание условий наблюдения для явления, которое исследуется;
- фиксация результатов эксперимента: того, что появляется в результате воспроизведения условий эксперимента..

Воспроизведение условий может носить активный или пассивный характер. В первом случае говорят об «эксперименте», для характеристики второго – употребляют термин «наблюдение». Хотя – подчеркнём ещё раз – оба термина могут употребляться как эквивалентные.

Отметим также, что термин «наблюдение» может употребляться для обозначения части «эксперимента», которая заключается в фиксации (регистрации) результатов «эксперимента».

Анализ статистических подходов к определению эксперимента с небольшими вариациями повторяет выделение приведенных выше основных составляющих эксперименту, выдвигая специфические дополнительные условия в том, что касается «стохастического эксперимента»

«Опыт», «эксперимент», «наблюдение», «испытание»: теоретико-вероятностный контент

Понятия «опыта», «эксперимента», «наблюдения», «испытания» являются также специальными понятиями теории вероятностей. Они имеют, в основном, общий эмпирический контент. И это является естественным, поскольку статистические (теоретико-вероятностные) методы являются признанным математическим средством моделирования неопределенности в изучении явлений, где она проявляется в виде случайности. Детальнее о случайности, как виде неопределённости, ниже.

В теоретико-вероятностных источниках в определениях понятий «опыт», «эксперимент», «наблюдение» отсутствует единство и единственность в определении понятий. В разных источниках употребляются разные варианты в применении к одним тех же, как можно понять из модельных примеров, объектам. Кроме того, в некоторых источниках обсуждаемые термины употребляются с эпитетами «стохастический», «вероятностный», «случайный». Еще раз заметим, что все они употребляются как эквивалентные. Ниже

приводится анализ употребления соответствующих понятий в тех или иных источниках и у тех или иных авторов.

В основополагающей книге А.Н. Колмогорова «Основные понятия теории вероятностей» на стр.12 отмечается, что «применение теории вероятностей к реальному миру опыта происходит в соответствии со следующей схемой.

1. Считается, что имеется определенный комплекс условий \mathcal{C} , который может воспроизводиться неограниченное количество раз.
2. Изучается определенный круг событий, которые могут происходить при воссоздании условий \mathcal{C} .

В таком почтенном источнике, как учебник Б.В. Гнеденко «Курс теории вероятностей», понятие наблюдения в виде «испытания» также связывается с комплексом условий \mathcal{C} . Но – дополнительно – и со связанным с этим комплексом условий набором событий (там же, стр.21). Испытание понимается как воссоздание упомянутого комплекса условий и проверке того, выполняется ли при этом воссоздании условий то или иное событие, выбранное из набора событий (там же, стр. 26). Таким образом «испытания» – это „наблюдение“, результаты которого используются для проверки того, выполняется ли исследуемое событие.

У Г. Крамера в классическом издании «Математические методы статистики» на стр. 157-158 понятие „эксперимента“ не определяется явно, но выделяются такие, которые могут быть повторенными многократно при одних и тех же условиях. Среди этого типа экспериментов дополнительно выделяются те, для которых в серии экспериментов «результат ... может изменяться от одного наблюдения к другому самым неправильным образом». Автор также замечает, что «...в этих случаях мы будем говорить, что имеем дело с последовательностью случайных экспериментов» (там же, стр.158).

В учебнике А.В. Скорохода «Элементы теории вероятностей и случайных процессов» уже в начале книги на стр.5 отмечается, что «одним из основных понятий теории вероятностей является понятие стохастического эксперимента». И дальше указывается, что «так называются эксперименты, результаты которых нельзя предусмотреть». Опять же понятие стохастического эксперимента объясняется примерами (там же, стр.5). А на стр.9 (там же) подчеркивается такая черта стохастических экспериментов, как «возможность повторять их большое число раз». Кроме того, в качестве важной черты стохастического эксперимента, отмечается наличие определенного, не состоящего из одного элемента «числа событий» с ним связанных, среди которых выделяются элементарные события. Отметим, что хотя понятие события не определяется, понятие элементарного события определяется строго (там же, стор.6,7).

В фундаментальном издании, которым является «Справочник по теории вероятностей и математической статистике», изданного авторским коллективом в составе В.С. Королюка, М.И. Портенко, А.В. Скорохода, А.Ф. Турбина, авторы на стр.5 отмечают, что «эксперимент определяется определенным комплексом условий, которые или воспроизводятся искусственно, или осуществляются независимо от воли экспериментатора. Кроме того, они отмечают, что эксперимент определяется также результатами эксперимента, то есть определенными событиями, которые наблюдаются как результат осуществления этого комплекса условий». Авторы также различают «детерминированные» и «случайные» или «вероятностные» эксперименты. К первым относят те, «в которых условия эксперимента однозначно определяют наступление (или не наступление) событий, которые ожидаются». Что же касается «случайных» или «вероятностных экспериментов», то они определяются как такие, в которых «при одних

и тех же условиях возможно появление событий, которые исключают друг друга». Отметим, – о чём ниже, – что отмеченных свойств эксперимента недостаточно, чтобы его можно было назвать стохастическим.

В классическом издании «Теория вероятностей» М. Лоева на стр. 13 отмечается, что «...наука имеет дело с закономерностями в испытаниях, которые повторяются», а также что «... долгое время Homo sapiens изучал только детерминированные испытания, в которых условия (причины) полностью определяют результаты (последствия)». Определяются также «случайные испытания» (там же, стр. 13) как такие, в которых при воссоздании их многократно, наблюдаемая частота любого из возможных результатов группируется вокруг определенных чисел. Таким образом, и в этом издании используются понятия, которые связаны с определенным комплексом условий: «испытание» и «случайные испытания», причём последние связываются с теми, в которых частоты появления разных результатов из числа возможных группируются вокруг определенных чисел. Конечно, обязательной является многообразие «испытаний» и наличие разных, вообще говоря, результатов в разных испытаниях.

В энциклопедическом по широте охвата издании, каковым является двухтомник В.Феллера «Введение в теорию вероятностей и ее приложения», в первом томе, понятие эксперимента считается интуитивно ясным и объясняется большим количеством примеров, которым посвященный §2 первого раздела. Собственно, речь идет о формализованном варианте виртуального опыта в общенаучном понимании, который характеризуется фиксированным набором возможных результатов.

Приведённые выше варианты определения «эксперимента» в теории вероятностей являются типичными и для других изданий, среди которых отметим учебник И.И. Гихмана, А.В.Скорехода, М.И. Ядренко «Теория вероятностей и математическая статистика», монографию «Теория вероятностей. Основные понятия, предельные теоремы, случайные процессы» Ю.В. Прохорова и Ю.А. Розанова, обстоятельный учебник А.Н. Ширяева «Вероятность», учебник И.Н. Коваленко и А.А. Филипповой «Теория вероятностей и математическая статистика», надежный учебник А.А. Боровкова «Курс теории вероятностей». В последнем автор отождествляет случайность с неопределенностью, связывая эту случайность – неопределенность, с незнанием (там же, стр.9), хотя на следующей странице, ссылаясь на принцип неопределенности в физике, отмечает, что неопределенность может быть принципиально свойственной исследуемому явлению. Кроме того, определяя объекты изучения в теории вероятностей (там же, стр.1), автор отмечает, что ими являются явления, для которых, с одной стороны «... те или другие эксперименты или наблюдения могут быть воспроизведены многократно при одинаковых условиях». С другой стороны, на той же странице отмечается, что «теорию вероятностей интересуют те эксперименты, результат которых, выраженный каким-то образом, может изменяться от опыта к опыту». Автор также связывает с результатами определённые события, которые могут в связи с этими результатами рассматриваться, и отмечает, что в этом случае события называют «случайными». Таким образом, в определении «эксперимента» по Боровкову общенаучный контент этого понятия (комплекс условий и, возможность многократного его воссоздания) связывается с изменчивостью результатов от наблюдения к наблюдению. Б.А.Севастьянов в «Курсе теории вероятностей и математической статистики» называет возможность многократного воссоздания условий в числе ключевых моментов «эксперимента», поскольку именно возможность многократного воссоздания позволяет делать вывод о наступлении или не наступлении тех или иных событий (понятие, которое автором не определяется). Автор называет такие события массовыми, хотя естественнее было бы отнести это название к явлению, по наблюдением которого можно сделать вывод о наступлении или нет исследуемых событий. У У. Гренадера и В. Фрайбергера в «Кратком курсе вычислительной вероятности и статистики» на стр. 10, отмечается возможность повторения эксперимента при одинаковых условиях, а М. Де Гроот в монографии «Оптимальные статистические решения» заявляет, что «статистика как наука занимается теориями и

методами, которые используются для принятия решений в условиях неопределенности и неполной информации» (там же, стр.11). А на стр. 14 (там же) отмечается, что «эксперимент употребляется здесь (в работе – *примечание автора*) в самом широком понимании для обозначения, в сущности, любого процесса, все возможные результаты которого могут быть указаны заранее и действительный результат которого является одним из указанных». Автор вводит специальное обозначение S для множества результатов и называет его выборочным пространством эксперимента.

Подводя итоги в определении и использовании понятий «эксперимент», «опыт», «наблюдение», «испытание» – иногда с эпитетами «случайный» или «стохастический» – в теории вероятностей отметим, что они употребляются, как эквивалентные в общем русле общенаучного понимания эксперимента. Специально подчеркиваются и выделяются следующие важные отличительные черты того, что в теории вероятностей и математической статистике понимают под экспериментом.

1. Наличие фиксированного комплекса условий, при котором наблюдается исследуемое явление. Указанный комплекс условий может воспроизводиться активно или пассивно.
2. Возможность многократного воспроизведения комплекса условий (массовость).
3. Возможность описания всех возможных результатов. Это множество результатов называют пространством элементарных событий, выборочным пространством эксперимента и т. п. Вместо множества возможных результатов может рассматриваться множество возможных событий, которые могут иметь или не иметь место в связи с наблюдающимися результатами.
4. Случайность, которая связывается с изменчивостью результатов от эксперимента к эксперименту или с наступлением в разных экспериментах событий, которые исключают друг друга: с непредсказуемостью результатов ли, событий ли от эксперимента к эксперименту.

Основные элементы формализации наблюдения: условия, результат и его регистрация

Как следует из вышеприведенного анализа, двумя основными составляющими эксперимента и в общем эмпирическом контексте и в статистическом понимании являются условия эксперимента (будем называть их также комплексом условий) и его результаты: того, что при этих условиях появляется.

1. Что касается условий эксперимента, то они должны допускать возможность их многократного воссоздания (массовость). Каждый из возможных комплексов условий, при котором можно проводить эксперимент будем обозначать k , соответственно – через K будем обозначать совокупность разных вариантов возможных комплексов условий, при которых можно проводить эксперимент.
2. Результат эксперимента, будем обозначать его y , – это то, что может появиться при воспроизведении комплекса условий $k \in K$. Через Y_k будем обозначать, вообще говоря, – множество всех возможных результатов, которые могут появиться при воспроизведении условий k , поскольку фиксация комплекса условий, вообще говоря, не гарантирует однозначного результата эксперимента.

Важно отметить, что термин «результат эксперимента» для обозначения того, что может появиться при воспроизведении комплекса условий (то, что выше обозначается через y), часто употребляют с другим содержанием: в смысле фиксации результатов экспериментов, что, как отмечалось выше, иногда понимают как часть эксперимента и обозначают также термином «наблюдение», понимаемом в узком смысле.

В дальнейшем «регистрация результатов» будет рассматриваться как составляющая эксперимента.

Определение 1. Регистрацией результата эксперимента будем называть фиксацию того, что определяет две составляющие эксперимента: условия κ и результат y – т.е. пару $s = (\kappa, y)$ «условие-результат». Соответственно, под регистрацией серии из N экспериментов (выборкой) будет пониматься последовательность пар

$$s_1, \dots, s_N = (\kappa_1, y_1), \dots, (\kappa_N, y_N). \quad (1)$$

Замечание 1. Как показывает анализ, результат эксперимента, как значение y , часто не отличают от «регистрации результата эксперимента» как того, что фиксируется в связи с проведённым экспериментом и обозначают одним и тем же термином «результат эксперимента».

Основные составляющие наблюдения: детализация условий

Необходимость учёта в одной серии экспериментов с разными условиями привела к необходимости структуризации условий наблюдения. Такая структуризация обеспечивает контролируемое изменение условий наблюдения от одного эксперимента серии к другому. При таком изменении часть комплекса условий остается, вообще говоря, неизменной по умолчанию, а часть изменяется контролируемым образом. Собственно: это означает, что любой из возможных комплексов $\kappa \in K$ условий проведения эксперимента представляется парой $\kappa = (x, f)$, в которой $x \in X$ обозначает вариативную, изменяемую от эксперимента к эксперименту часть, а f – неизменную по умолчанию для серии экспериментов часть комплексов условий наблюдения.

Определение 2. Экспериментом с управляемыми условиями (УпрУЭкс) будем называть такой, в котором условия представляются в виде $\kappa = (x, f)$, $x \in X, f \in \mathfrak{F}$, $K = X \times \mathfrak{F}$, x будем называть вариативной частью условий, f – частью условий по умолчанию, κ – полными условиями эксперимента.

Замечание 2. Отметим, что при детерминированном подходе неизменная часть условий ассоциируется с однозначностью связи результата наблюдения с вариативной частью условий, т.е. – с функцией (функцией отклика) от вариативной, изменяемой части условий. Само же исследование в рамках УпрУЭкс экспериментов в литературе называют моделью «вход-выход» системы с очевидным делением на то, что называют входом, выходом и функцией отклика системы.

Регистрация наблюдений: практика

Следует отметить, в экспериментальной практике регистрация эксперимента в смысле определения 1 подменяется другими вариантами того, что называют фиксацией s_1, \dots, s_N . Такими вариантами в рамках выделенных выше составляющих эксперимента могут быть следующие:

$$s_1, \dots, s_N = \begin{cases} y_1, \dots, y_N \\ (x_1, y_1), \dots, (x_N, y_N) \\ (\kappa_1, y_1), \dots, (\kappa_N, y_N) \end{cases} \quad (2)$$

Множественные модели неопределённости (МнМоН)

Определение 3. В рамках введённых выше понятий множественными моделями неопределённости для исследуемого явления будем называть такое описание неопределённости, которое базируется на множественности значений Y, Y_x, Y_κ : того, что появляется или может появиться в результате серии экспериментов в (2). В зависимости от того, как понимается регистрация эксперимента, это может быть:

$$\begin{aligned}
 Y &= \bigcup_{i=1}^N \{y_i\}, \\
 Y_x &= \bigcup_{i:x_i=x} \{y_i\}, x \in \bigcup_{i=1}^N \{x_i\}, \\
 Y_\kappa &= \bigcup_{i:\kappa_i=x} \{y_i\}, \kappa \in \bigcup_{i=1}^N \{\kappa_i\}
 \end{aligned} \tag{3}$$

Собственно, (3) фиксирует множественность значений того, что может появиться при фиксированном комплексе условий. В первом случае из (3) комплекс условий по умолчанию является одинаковым для всех экспериментов серии, во втором – в серии экспериментов условия варьируются. Множество возможных вариантов условий в серии экспериментов определяется множеством $\bigcup_{i=1}^N \{x_i\}$, а Y_x определяет множество значений y в тех экспериментах серии, в которой вариативная часть одна и та же и определяется вариативной частью $x \in \bigcup_{i=1}^N \{x_i\}$. В третьем варианте то же касается $\bigcup_{i=1}^N \{\kappa_i\}$, которое описывает множества всех возможных вариантов условий серии.

МнМоН: неопределённость в детерминированных наблюдениях

Для детерминированности характерна однозначная связь «полные условия – результат»:

$$\kappa \rightarrow y_\kappa, Y_\kappa = \{y_\kappa\}, \kappa \in K. \tag{4}$$

Однако, если в серии экспериментов условия изменчивы, а регистрация проводится в виде y_1, \dots, y_N вместо $(x_1, y_1), \dots, (x_N, y_N)$ или $(\kappa_1, y_1), \dots, (\kappa_N, y_N)$, то возникает множественность (не одноэлементность) Y , которая, собственно, и является неопределённостью. Такая неопределённость в детерминированном эксперименте связана с проблемой скрытых параметров: дополнительных условий, которые нужно учитывать при регистрации эксперимента, чтобы явление стало детерминированным, т.е., чтобы наблюдения могли быть охарактеризованными в соответствии с (4).

МнМоН: случайность

Случайность в исследуемом явлении с одной стороны характеризуется тем, что связь возможных результатов с полными условиями эксперимента в (3) неоднозначна: для каждого из фиксированных условий в разных экспериментах могут появляться разные результаты:

$$\kappa \rightarrow Y_\kappa \neq \text{"одноэлементное множество"}, \kappa \in K.$$

С другой стороны, для явления, которое называют случайным (а сам эксперимент – стохастическим), должен выполняться закон устойчивости частот. Этим термином обозначается предположение о том, что частоты тех или иных групп возможных результатов должны сходиться к предельному значению, которое на должно зависеть от серии экспериментов, по которому оно получено, но характеризовать само исследуемое явление: быть одинаковым для разных серий экспериментов.

МнМоН: гарантированные оценки (минимакс)

Этот подход связан с дальнейшей априорной структуризацией вариативной или функциональной части условий в рамках детерминированного описания явления: $x = (x^{(1)}, x_V^{(2)})$ или $f = (f^{(1)}, f^{(2)})$ и предположением о том, что в эксперименте фиксируется только одна из частей, например $x^{(1)}$ или $f^{(1)}$ (наблюдаемая компонента), а про вторую – известно, что она принадлежит множеству $E_{x^{(1)}}$ или $E_{f^{(1)}}$ соответственно, которое определяется наблюдаемой компонентой.

МнМоН: интервальный подход

В модель наблюдений со структурированной вариативной частью очевидным образом вкладывается интервальная модель неопределённости. Действительно, достаточно предположить, что в обозначениях предыдущего пункта

$$y = f(x^{(1)}) + x^{(2)}, x^{(2)} \in (-\Delta_{x^{(1)}}, \Delta_{x^{(1)}}) = E_{x^{(1)}}.$$

МнМоН: нечёткие множества

Место нечёткости [Zadeh, 1962] во множественных моделях неопределённости может быть определено в рамках статистической интерпретации нечетких множеств [Donchenko, 1998, а) б)]. Эта интерпретация определяется следующей теоремой.

Теорема [Donchenko, 1998 а), б)]. Для нечёткого множества, задаваемого парой (E, μ) носитель-функция принадлежности, в случае, когда E - пространство с мерой, а μ - измерима, можно построить вероятностное пространство (Ω, B_Ω, P) , событие $A \in B_\Omega$, полную группу событий $H_e = \{\eta = e\}, e \in E, \eta \in E$ - значная случайная величина, так, что

$$\mu(e) = P(A | H_e), e \in E.$$

МнМоН: обратные задачи

Важным классом неопределённостей в детерминированных задачах являются обратные задачи, т.е. задачи в которых необходимо определить множество возможных вариативных частей условий (входов), которые обеспечивают заданное значение результата (выхода). Отметим важную роль псевдообращения по Муру - Пенроузу [Алберт, 1977] и его развитию в работе [Кириченко, 1997] для линейных задач и для применения в задачах кластеризации и распознавания образов [Кириченко, Донченко., 2007].

МнМоН: преобразование Хока

Специальным случаем неопределённости является ПХ [Hough, 1962]. Этот вид неопределённости порождён множественностью возможных вариантов функций отклика в наблюдениях: $\kappa_i = (x_i, f_i), i = \overline{1, N}$. Простейшей моделью наблюдений такого рода может служить бинаризованное изображение, на котором представлено несколько прямых. Выборка представляет собой координаты точек изображения с единичным значением яркости. Детальнее с ПХ и его математической формализацией можно познакомиться в [Donchenko, 2003].

Заклучение

В работе рассмотрены проблемы математического описания структуры(связей составляющих) исследуемого объекта, рассмотрены истоки неопределённости в поведении таких объектов. Перечислены основные математические средства описания структур - связей между составляющими элементами описываемого объекта. Приведены примеры основных структур и важных их комбинаций. Предложенная в работе концепция «множественных моделей неопределённости» связывает неопределённость с экспериментом и позволяет на единой основе рассматривать многообразие математических методов описания неопределённости. В работе также определено место каждого из математических методов в рамках предложенной концепции.

Литература

- [Алберт, 1977] Алберт А. Регрессия, псевдоинверсия, рекуррентное оценивание. – М.: Наука. – 1977.– 305 с.
- [Donchenko, 2003] Donchenko V.S. Hough Transform and Uncertainty// Proceedings X International Conference "Knowledge – Dialog – Solution". – June 16-23, 2003. – Varna (Bulgaria). – P.391-395.
- [Донченко,1968,а)] Донченко В.С. Умовні розподіли та нечіткі множини // Вісник Київського університету. – 1998.–Вип. №3 – С. 175-179.
- [Донченко,1968,б)] Донченко В.С. Імовірність та нечіткі множини .// Вісник Київського університету. Серія фізико-математичні науки. – 1998. – Вип. №4. – С. 141-144
- [Hough] Hough P.V.C. Method and Means for Recognizing Complex Patterns. - U.S. Patent 3069354. – December 1962.
- [Кириченко, 1997] Кириченко Н.Ф. Аналитическое представление псевдообратных матриц //Киб. и СА.- №2. –1997.– С.98-122.
- [Кириченко, Донченко, 2007] Кириченко Н.Ф., Донченко. В.С. Псевдообращение в задачах кластеризации// Киб. и СА.- №4, 2007– С.98-122.
- [Мальцев,1970] Мальцев А.И. Алгебраические системы. – М.: Наука.–1970. – 392 с.
- [Zadeh, 1965] Zadeh, Lotfi. Fuzzy Sets// Information and Control. – June, 1965. – 8(3).–P. 338-353.

Информация об авторе

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



Капитонова Юлия Владимировна

Институт кибернетики им. В.М. Глушкова Национальной академии наук Украины, ассоциация искусственного интеллекта с глубоким прискорбием сообщают о предвременной кончине Капитоновой Юлии Владимировны.

Капитонова Ю.В. работала в Институте кибернетики им. Глушкова НАН Украины с 1963 года и до последних дней своей жизни после окончания Киевского государственного университета. Свою трудовую и научную деятельность Капитонова Ю.В. начала под руководством академика В.М. Глушкова, выполнив ряд научных работ теоретического и практического характера. В Институте кибернетики она прошла путь от младшего научного сотрудника до заведующего отделом, доктора физико-математических наук, профессора, заслуженного деятеля науки и техники. На протяжении всей трудовой деятельности Капитонова Ю.В. занималась разработкой теории проектирования и внедрением практических средств автоматизации проектирования (системы "Проект", Макроконвейер, САД, АПС и др.). Под её руководством было воспитано множество кандидатов и докторов наук, где проявился незаурядный талант требовательного и внимательного наставника научной молодежи.

Светлая память о Капитоновой Юлии Владимировне останется в наших сердцах, в памяти её учеников и коллег.