

Krassimir Markov, Vitalii Velychko, Oleksy Voloshin
(editors)

Natural and Artificial Intelligence

ITHEA

SOFIA

2010

Krassimir Markov, Vitalii Velychko, Oleksy Voloshin (ed.)

Natural and Artificial Intelligence

ITHEA®

Sofia, Bulgaria, 2010

ISBN 978-954-16-0043-9

First edition

Recommended for publication by The Scientific Council of the Institute of Information Theories and Applications FOI ITHEA

This book is engraved in prof. Zinovy Lvovich Rabinovich memory. He was a great Ukrainian scientist, co-founder of ITHEA International Scientific Society (ITHEA ISS). To do homage to the remarkable world-known scientific leader and teacher this book is published in Russian language and is concerned to some of the main areas of interest of Prof. Rabinovich.

The book is opened by the last paper of Prof. Rabinovich specially written for ITHEA ISS. Further the book maintains articles on actual problems of natural and artificial intelligence, information interaction and corresponded intelligent technologies, expert systems, robotics, classification, business intelligence; etc. In more details, the papers are concerned in: conceptual problems of the natural and artificial intelligent systems: structures and functions of the human memory, ontological models of knowledge representation, knowledge extraction from the natural language texts; network technologies; evolution and perspectives of development of the mechatronics and robotics; visual communication by gestures and movements, psychology of vision and information technologies of computer vision, image processing; object classification using qualitative characteristics; methods for comparing of alternatives and their ranging in the procedures of expert knowledge processing; ecology of programming – a new trend in the software engineering; decision support systems for economics and banking; systems for automated support of disaster risk management; and etc.

It is represented that book articles will be interesting for experts in the field of information technologies as well as for practical users.

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

Printed in Bulgaria

Copyright © 2010 All rights reserved

© 2010 ITHEA® – Publisher; Sofia, 1000, P.O.B. 775, Bulgaria. www.ithea.org; e-mail: info@foibg.com

© 2010 Krassimir Markov, Vitalii Velychko, Oleksy Voloshin – Editors

© 2010 Ina Markova – Technical editor

© 2010 For all authors in the book.

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

ISBN 978-954-16-0043-9

C/o Jusautor, Sofia, 2010

PREFACE

ITHEA International Scientific Society (**ITHEA ISS**) is aimed to support growing collaboration between scientists from all over the world. The scope of the books of the ITHEA ISS covers the area of Informatics and Computer Science. ITHEA ISS welcomes scientific papers and books connected with any information theory or its application. ITHEA ISS rules for preparing the manuscripts are compulsory.

ITHEA Publishing House is the official publisher of the works of the members of the ITHEA ISS.

Responsibility for papers and books published by ITHEA belongs to authors.

This book is engraved in prof. Zinovy Lvovich Rabinovich memory. He was a great Ukrainian scientist, co-founder of ITHEA International Scientific Society (ITHEA ISS). To do homage to the remarkable world-known scientific leader and teacher this book is published in Russian language and is concerned to some of the main areas of interest of Prof. Rabinovich.

The book is opened by the last paper of Prof. Rabinovich specially written for ITHEA ISS. Further the book maintains articles on actual problems of natural and artificial intelligence, information interaction and corresponded intelligent technologies, expert systems, robotics, classification, business intelligence; etc.

In more details, the papers are concerned in:

- conceptual problems of the natural and artificial intelligent systems: structures and functions of the human memory, ontological models of knowledge representation, knowledge extraction from the natural language texts;
- network technologies;
- evolution and perspectives of development of the mechatronics and robotics;
- visual communication by gestures and movements, psychology of vision and information technologies of computer vision, image processing;
- object classification using qualitative characteristics;
- methods for comparing of alternatives and their ranging in the procedures of expert knowledge processing;
- ecology of programming – a new trend in the software engineering;
- decision support systems for economics and banking;
- systems for automated support of disaster risk management;

and etc.

It is represented that book articles will be interesting for experts in the field of information technologies as well as for practical users.

The book is recommended for publication by the Scientific Council of the Institute of Information Theories and Applications FOI ITHEA.

Papers in this book are selected from the ITHEA ISS Joint International Events of Informatics "ITA 2010":

| | |
|---------------|---|
| CFDM | Second International Conference "Classification, Forecasting, Data Mining" |
| i-i | International Conference "Information - Interaction – Intellect" |
| i.Tech | Eight International Conference "Information Research and Applications" |
| ISK | V-th International Conference "Informatics in the Scientific Knowledge" |
| MeL | V-th International Conference "Modern (e-) Learning" |
| KDS | XVI-th International Conference "Knowledge - Dialogue – Solution" |
| CML | XII-th International Conference "Cognitive Modeling in Linguistics" |
| INFOS | Thirth International Conference "Intelligent Information and Engineering Systems" |
| NIT | International Conference "Natural Information Technologies" |
| GIT | Eight International Workshop on General Information Theory |
| ISSI | Forth International Summer School on Informatics |

ITA 2010 took place in Bulgaria, Croatia, Poland, Spain and Ukraine. It has been organized by
ITHEA International Scientific Society

in collaboration with:

- ITHEA International Journal "Information Theories and Applications"
- ITHEA International Journal "Information Technologies and Knowledge"
- Institute of Information Theories and Applications FOI ITHEA
- V.M.Glushkov Institute of Cybernetics of National Academy of Sciences of Ukraine
- Taras Shevchenko National University of Kiev (Ukraine)
- Association of Developers and Users of Intelligent Systems (Ukraine)
- Universidad Politecnica de Madrid (Spain)
- University of Calgary (Canada)
- BenGurion University (Israel)
- University of Hasselt (Belgium)
- Dorodnicyn Computing Centre of the Russian Academy of Sciences (Russia)
- Institute of Linguistics, Russian Academy of Sciences (Russia)
- Institute of Mathematics and Informatics, BAS (Bulgaria)
- Institute of Mathematics of SD RAN (Russia)
- New Bulgarian University (Bulgaria)
- The University of Zadar (Croatia)
- Rzeszow University of Technology (Poland)
- Kharkiv National University of Radio Electronics (Ukraine)
- Kazan State University (Russia)
- Alexandru Ioan Cuza University (Romania)
- Moscow State Linguistic University (Russia)
- Astrakhan State University (Russia)

as well as many other scientific organizations. For more information: www.ithea.org .

We express our thanks to all authors, editors and collaborators as well as to the General Sponsor.

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

Sofia – Kiev

June 2010

K. Markov, V. Velychko, O. Voloshin

TABLE OF CONTENTS

| | |
|---|----|
| Preface | 3 |
| Table of Contents..... | 5 |
| Index of Authors | 7 |
| Память и мышление человека в пирамидально-концептуальном представлении | |
| <u>Зиновий Л. Рабинович</u> | 11 |
| Истина или смысл? Два подхода к обработке естественно языковых текстов | |
| <i>Леонид Святогор, Виктор Гладун</i> | 23 |
| Об автоматизации процесса извлечения первичных знаний из естественно-языковых текстов | |
| <i>Александр Палагин, Сергей Крывый, Дмитрий Бибилов</i> | 33 |
| Методы анализа малых возмущений линейных моделей | |
| <i>Алексей Волошин, Григорий Кудин, Владимир Кудин</i> | 41 |
| Многокритериальный метод распределения ограниченных ресурсов | |
| <i>Альберт Воронин</i> | 48 |
| Применение метода «МАСКА» для групповой экспертной классификации научных проектов по многим критериям | |
| <i>Петровский Алексей Борисович, Ройзензон Григорий Владимирович, Тихонов Игорь Петрович, Балышев Александр Владимирович</i> | 56 |
| Сравнение альтернатив в подходе обобщенных интервальных оценок | |
| <i>Михаил Стернин, Геннадий Шепелёв, Алла Рябова</i> | 68 |
| Процедуры нахождения строгого результирующего отношения на множестве альтернатив в задаче линейного упорядочения альтернатив | |
| <i>Павел Антосяк, Алексей Волошин</i> | 78 |
| Методологические принципы распределение квот на выбросы парниковых газов | |
| <i>Алексей Волошин, Ирина Горицына, Сергей Мащенко</i> | 85 |
| Инструментальный комплекс алгоритмов и программ оптимального проектирования сетей с технологией MPLS | |
| <i>Юрий Зайченко, Елена Зайченко</i> | 94 |
| Построение оптимизационной балансовой модели с приростными коэффициентами | |

| | |
|---|-----|
| <i>Елена Единак</i> | 101 |
| Сравнительный анализ методов оценки риска банкротства корпорации | |
| <i>Ольга Прокопенко</i> | 106 |
| Исследование двойственной задачи оптимизации инвестиционного портфеля в нечетких условиях | |
| <i>Юрий Зайченко, Ови Нафас Агаи Аг Гамиш</i> | 115 |
| Мульти-агентные технологии интеллектуального управления в робототехнических, телекоммуникационных и grid-сетях | |
| <i>Адиль Тимофеев</i> | 129 |
| Моделирование реалистических движений и мимики для задач визуализации жестовой информации | |
| <i>Юрий Крак, Юрий Кривонос, Александр Бармак, Антон Тернов, Богдан Троценко</i> | 137 |
| Переменная разрешающая способность как фактор развития средств обработки визуальной информации | |
| <i>Владимир Калмыков</i> | 144 |
| Обобщенная табличная алгебра, обобщенное исчисление строк и их эквивалентность | |
| <i>Дмитрий Буй, Ирина Глушко</i> | 157 |
| Тиражирование данных в динамически настраиваемых распределенных информационных системах | |
| <i>Людмила Лядова, Михаил Стрелков</i> | 165 |
| Мультиагентный подход к решению проблемы равномерного распределения вычислительной нагрузки | |
| <i>Александр Миков, Елена Замятина, Арсений Козлов</i> | 173 |
| Программное обеспечение – экологический подход к исследованиям | |
| <i>Максим Луцкий, Николай Сидоров</i> | 181 |
| Концепция построения и методы реализации автоматизированной системы оперативно-диспетчерского управления газотранспортной системой Украины | |
| <i>Юрий Пономарев, Богдан Клюк, Виктор Борисенко</i> | 189 |
| Моделирование и анализ фактов и связей между ними | |
| <i>Павел Мальцев</i> | 194 |

INDEX OF AUTHORS

| | | | |
|--------------------|------------|-------------------------|---------|
| Александр Бармак | 137 | Елена Зайченко | 94 |
| Адиль Тимофеев | 129 | Елена Замятина | 173 |
| Александр Палагин | 33 | Зиновий Рабинович | 11 |
| Александр Балышев | 56 | Игорь Тихонов | 56 |
| Александр Миков | 173 | Ирина Глушко | 157 |
| Алексей Волошин | 41, 78, 85 | Ирина Горицына | 85 |
| Алексей Петровский | 56 | Леонид Святогор | 23 |
| Алла Рябова | 68 | Людмила Лядова | 165 |
| Альберт Воронин | 48 | Максим Луцкий | 181 |
| Антон Тернов | 137 | Михаил Стернин | 68 |
| Арсений Козлов | 173 | Михаил Стрелков | 165 |
| Богдан Клюк | 189 | Николай Сидоров | 181 |
| Богдан Троценко | 137 | Ови Нафас Агаи Аг Гамиш | 115 |
| Виктор Борисенко | 189 | Ольга Прокопенко | 106 |
| Виктор Гладун | 23 | Павел Антосяк | 78 |
| Владимир Калмыков | 144 | Павел Мальцев | 194 |
| Владимир Кудин | 41 | Сергей Кривый | 33 |
| Геннадий Шепелёв | 68 | Сергей Мащенко | 85 |
| Григорий Кудин | 41 | Юрий Зайченко | 94, 115 |
| Григорий Ройзензон | 56 | Юрий Крак | 137 |
| Дмитрий Бибииков | 33 | Юрий Кривонос | 137 |
| Дмитрий Буй | 157 | Юрий Пономарев, | 189 |
| Елена Единак | 101 | | |



ПАМЯТИ
ЗИНОВИЯ ЛЬВОВИЧА РАБИНОВИЧА
01.08.1918 – 02.11. 2009

2 ноября 2009 года на 92-м году ушёл из жизни выдающийся советский и украинский кибернетик, последовательный сторонник идей академика В.М. Глушкова и основатель научной школы в области создания вычислительных систем, машинной памяти, биокибернетики и искусственного интеллекта, член Ассоциации Создателей и Пользователей Интеллектуальных Систем (АСПИС) и ITNEA International Scientific Society – доктор технических наук, профессор **Рабинович Зиновий Львович**.

Зиновий Львович стоял у истоков отечественной кибернетики и вычислительной техники. Он был генератором научных идей, которые с помощью инженерного таланта стремился воплощать в оригинальных конструкциях. Им была защищена докторская диссертация по совокупности опубликованных работ, что составляло в те годы исключение из правил ВАК. В начале шестидесятых годов прошлого столетия, будучи заведующим отделом Теории цифровых автоматов, Зиновий Львович сплотил вокруг себя уникальный творческий коллектив молодых исследователей, из которого образовалась школа учёных с широким спектром интересов – начиная с теории алгоритмических языков, программирования, интерактивных средств человеко-машинного общения и заканчивая теорией искусственного интеллекта. Именно он первым понял значение растущей пирамидальной сети в качестве модели нейроподобной обучаемой среды и указал на неё как на возможную модель мышления и памяти в мозге человека. В расцвете творческого потенциала З.Л. Рабинович создал концепцию и разработал технический проект ЭВМ нового поколения «Украина» (1965 г.). По своей логико-вычислительной архитектуре, принципам программирования, методам вычислений, возможности решения интеллектуальных задач, элементной базе и вычислительной эффективности эта машина намного превосходила зарубежные аналоги того времени, а по языкам высокого уровня, человеко-машинному диалогу её можно считать аналогом персонального компьютера. «Украина», опередившая идеи своего времени на пятнадцать лет, не пошла в серию только из-за недалёковидности политических руководителей того времени.

З.Л. Рабинович – Лауреат Государственной Премии Украины, Лауреат именных премий Академии Наук Украины – имени С.А. Лебедева и имени В.М. Глушкова; Лауреат Международной награды “ITHEA”, заслуженный деятель науки; входил в состав творческого коллектива авторов «Энциклопедии кибернетики» – первого в мире издания, обобщившего достижения этой науки до 1975 года; был членом Редакционных коллегий Международных журналов ITHEA ISS: International Journal “Information Theories and Applications” и International Journal “Information Technologies and Knowledge”.

Зиновий Львович вёл преподавательскую работу в Киевском государственном университете им. Т.Г. Шевченко и других ВУЗах страны, был отличным популяризатором научных идей; благодаря ему многие студенты выбрали свой путь в кибернетике. Он подготовил 9 докторов и 43 кандидата технических наук. Любил свою работу и свой коллектив. Отличался личной скромностью и интеллигентностью. Будучи организатором и непременным участником конференций «Знание–Диалог–Решение» (KDS) поражал учёных разных стран своей эрудицией, чтением наизусть поэм Пушкина и Лермонтова, для коллег был образцом любознательности и энергичности.

Сохраним светлую память об этом замечательном Учёном и Человеке.

ITHEA International Scientific Society

ПАМЯТЬ И МЫШЛЕНИЕ ЧЕЛОВЕКА В ПИРАМИДАЛЬНО-КОНЦЕПТУАЛЬНОМ ПРЕДСТАВЛЕНИИ

Зиновий Л. Рабинович

Аннотация: В работе рассматриваются концептуальные положения о структурно-функциональном изучении Памяти человека и интеллектуальных сторон мышления и об использовании этих гипотез для построения моделей Памяти и Искусственного интеллекта с обратной корректировкой исходных гипотез. Приводятся основные свойства таких моделей. Результаты многолетних исследований и экспериментов обобщаются в сжатом виде и дополняются ссылками на соответствующие публикации.

Ключевые слова: память, мышление, образ, модель, растущие пирамидальные сети, решение проблем, сознание, подсознание, эмоции.

ACM Classification Keywords: I.2 Artificial Intelligence, I.2.4 Knowledge Representation Formalisms and Methods, I.2.8 Problem Solving, I.5.1 Pattern recognitions – Models.

Введение

Результатом информационного подхода к возможному объяснению мыслительных процессов, которые отражают интеллектуальную сторону деятельности Человека, явилась разработка нейрокибернетической модели Памяти.

Среда Памяти предстаёт как единая совокупность нейроподобных пирамидальных структур, охваченных обратными связями. С функциональной точки зрения память выступает как хранитель, переработчик и генератор знаний и как процессор мышления. Однако наш подход заключается в неразрывном рассмотрении функциональных и структурных свойств памяти. Такой общий взгляд на проблему сформулирован в названии статьи; он представляется полезным в контексте стремления передать (некоторые) функции естественного интеллекта искусственному – машинному интеллекту. Соответственно, в методах математического моделирования мышления выступают, в их взаимосвязи, понятия естественного и искусственного интеллектов, отражая в своём материале познание первого и его моделирование – как реализация второго – в их концептуальных представлениях.

Некоторые понятия, использованные в названии работы, нуждаются в уточнении, поскольку они играют ключевую роль в понимании проблемы.

Итак, **Память** воспринимается как нейросеть, в которой и осуществляются процессы мышления, происходит спонтанное или целенаправленное возбуждение пространственно-распределённых компонент (кластеров) нейросреды памяти, воплощающих информационные «смыслы». Запоминание информации – это лишь часть функций Памяти, которая в самом широком смысле понимается как система высшей нервной и информационной деятельности Человека; она находится в активном состоянии именно при решении сложных и разнообразных задач, которое мы соотносим с процессами мышления [7, 11, 14, 18, 23, 25, 26, 27, 31, 32, 35, 36, 37, 40, 41, 42].

Входы в Память и выходы из неё осуществляются прямыми и обратными информационными связями с внешней, по отношению к ней, средой организма, и вообще – внешней по отношению к организму средой [7, 11, 36].

Далее, название работы определяет всё вышесказанное как относящееся именно к Человеку. А это уже означает, что Память представляется как состоящая из двух информационных систем – так называемой

чувственно-образной и языковой системы. Языковая система занимает в среде Памяти часть в несколько раз меньшую, чем остальная, чувственно-образная система. Однако она имеет определяющее значение с позиций превращения человека в Человека [3, 13, 15, 18, 19, 23, 42].

В названии статьи также указывается *концептуальность* Памяти в сочетании с её особым отличительным свойством – *пирамидальностью* как характерным признаком информационных структур в нейросреде.

Концептуальность (как термин) подразумевается в том смысле, что он определяет главное и фундаментальное свойство *знания*. Знание рассматривается не само по себе, не изолированно от Памяти, не как абсолют, а как продукт мышления и, следовательно, тесно зависит от Памяти. Концептуальность базового понятия «знание» принципиально означает его «неустранимость» из некоторой системы понятий. Знание концептуально и структурировано: например, если данное конкретное знание уподобить дереву, то концептуальность представляется его стволом, а различные подробные проявления этого знания – ветвями дерева. То есть, все эти ветви имеют одну общую, одинаковую часть, а каждая из них обладает ещё и своими свойствами [7, 18, 20, 23, 31, 32, 33, 36, 40, 41, 42].

Изучением структуры, свойств и информационных характеристик Памяти занимаются представители разных наук, используя методы биологии, психологии и кибернетики. Для целей информационного моделирования мы выделяем три, наиболее характерные, черты Памяти: *ассоциативность*, *иерархичность* и насыщенность *обратными связями*.

Ассоциативность Памяти. Одним из фундаментальных свойств Памяти, которое определяет процессы нашего мышления, является ассоциативность процессов возбуждения, возникающих в разных функциональных областях памяти при формировании понятий. Ассоциации могут возникать под влиянием внешних и внутренних факторов и приводить к непредсказуемым результатам. Если в формировании понятия принимают участие определённые участки Памяти, которые реализуют структуру образа, то, вероятно, наличие общих компонент структур разных образов и определяет эти образы как ассоциативно связанные между собой. Следовательно, данное свойство характеризует Память как семантическую сеть взаимосвязанных образов, реализованную в нейронной среде. Именно в ассоциативной нейронной среде выполняются процессы восприятия, переработки и сохранения знаний, используемых в процессах мышления и взаимодействия с внешней (по отношению к Памяти) средой. Свойство ассоциативности образов – как уже имеющихся, так и создаваемых в нейросреде – непосредственно влияет на интеллект Человека, который должен отображать бесконечное разнообразие внешнего мира [7, 23, 25, 27, 29, 31, 32, 33, 40, 41, 42].

Иерархичность Памяти. Рассмотрим важнейшее, в рамках предлагаемой концепции, свойство Памяти, называемое *пирамидальностью*. По сути, данное свойство задаёт иерархическую структуру сети, состоящей из нейроподобных элементов – такую, в которой по мере продвижения «наверх» производится «свёртывание» информации. Уровни пирамиды соответствуют степени обобщения фактов и знаний в памяти Человека. Уровни пирамиды насыщены внутренними связями, а «этажи» пирамиды охвачены межуровневыми связями. Связи носят многонаправленный характер, передавая нейровозбуждение как «по вертикали» – в высшие отделы коры мозга, так и распространяя его «по горизонтали» и на более низкие уровни. Иерархия нейроструктур Памяти, закреплённая эволюцией «живого вещества» (В.И. Вернадский), также обусловлена необходимостью структурировать и упорядочить в сознании бесконечное разнообразие внешнего мира [1, 4, 7, 12, 16, 19, 21, 26, 27, 31, 32].

Обратные связи в Памяти. Процесс временного возбуждения иерархической сети под влиянием внешних стимулов или внутренних побуждений, который направлен на поиск решения некоторой проблемы, назовём «*внутренним взором*». Этот процесс носит внутренний (когнитивный) характер. Мы не можем себе представить процесс достижения определённого решения как некий однонаправленный

процесс. Даже когда мы производим простой перебор в памяти, мы возвращаемся на предыдущие стадии обработки, т.е. – зацикливаем некоторые процедуры вместе с участвующими в них локальными структурами памяти. При этом «внутренний взор» управляет перебором и контролирует достижение цели. Следовательно, вся пирамидальная семантическая сеть пронизана обратными связями, нисходящими до самого «дна» пирамид. Примером обратного распространения возбуждения может служить наша реакция на образ (или даже на слово) «лимон», которая распространяется от сложившегося образа до рецепторов. Очевидно, данная реакция также служит результатом *внутреннего взора* Памяти. Прямые и обратные связи формируют Память как единую среду и как целостную информационную систему. Известно, что наличие обратных связей позволяет приводить сложную систему либо к самовозбуждению, либо к устойчивому состоянию; этот же механизм служит достижению информационных целей [7, 11, 25, 27, 31, 32, 33, 36, 40, 41, 42].

Распространение возбуждения в нейросреде, которое сопровождает любой мыслительный акт, не носит хаотический характер (за исключением стрессовых ситуаций). Например, если наблюдается некоторая ситуация, то для её понимания и выработки решения в памяти привлекаются образы подобных ситуаций. При этом происходит поиск соответствующей модели ситуации, то есть – её образ. Как указывалось, образ представлен совокупностью понятий (конкретных или абстрактных), а также виртуальных (внеязыковых, чувственных) представлений, которые реализованы в структуре нейросети. Чтобы найти образ, адекватный или близкий по смыслу заданному, требуется вести поиск, двигаясь при этом по структурам памяти и направляя его к решению задачи. В зависимости от сложности задачи поиск может свестись к многократному повторению «метода проб и ошибок», то есть – к возврату на предыдущие этапы решений. В процессе такого движения к цели задействуются как прямые, так и обратные связи; образно говоря, некоторая «программа поиска решения» высвечивает *внутренним взором* цель, средства и этапы достижения цели [3, 4, 10, 14, 16, 19, 26, 29, 31, 32, 33].

Вот, собственно, те основные понятия, которые положены в основу пирамидально-концептуального представления структур Памяти и процессов Мышления. Ниже мы остановимся на некоторых узловых этапах становления и развития данной концепции.

Растущие пирамидальные сети

Математическим аппаратом моделирования иерархического (пирамидального) способа размещения информации в Памяти, а также вытекающего из данного размещения процесса иерархической обработки знаний, могут служить **Растущие пирамидальные сети (РПС)** [1, 2, 4, 5, 6, 19, 21]. Здесь уместно предположить, что иерархичность самого процесса обработки знаний в Памяти может означать, что всякое целенаправленное Мышление человека вовлекает в этот процесс иерархические подструктуры в контексте поставленной цели.

Растущие пирамидальные сети – как один из видов семантических сетей, использовались эффективно и широко: например, для решения задач обнаружения закономерностей функционирования технических систем и прогнозирования их состояния; в медицинской диагностике; в химической отрасли; в научных, экономических и многих других приложениях [24, 34]. В то же время, являясь особым классом нейроподобных сетей, в которых возникновение новых понятий и забывание старых сопровождается «рождением» и «отмиранием» элементов сети, *Растущие пирамидальные сети* прочно вошли в теоретический фундамент Искусственного интеллекта. Воздействие внешней среды на входные элементы сети анализируется и приводит к генерированию *новых понятий*, которые закрепляются в ней *структурно* [4].

РПС представляет собой семантическую сеть, или направленный граф, состоящий из узлов и связей между ними [19]. Количество входных связей превышает количество выходных, что обуславливает *пирамидальную архитектуру* сети. На нижнем уровне пирамиды находятся узлы – рецепторы, на верхнем формируются несколько конечных узлов. Рецепторы являются входными элементами, которые возбуждаются, если значение признака (информационного сигнала или атрибута) входного объекта соответствует диапазону чувствительности данного рецептора.

В процессе обучения (или «анализа выборки») на вход РПС последовательно поступают объекты *обучающей выборки*, и рецепторы «снимают» показания соответствующих признаков. При этом комбинации возбуждённых рецепторов запоминаются путём появления в сети *новых узлов*, которые генерируются по формальным правилам и связываются с ранее созданной структурой. Каждый узел интерпретируется логической функцией. Сложность *логических понятий* по мере их продвижения к верхним уровням сети – возрастает. После окончания просмотра выборки окончательно формируется *финальная сетевая структура*, в которой зафиксированы комбинационные свойства множества объектов. Поскольку предполагается, что в выборке нет тождественных объектов, то количество конечных узлов сети равно длине выборки.

Финальная структура РПС представляет собой «информационный портрет выборки», или некоторый «образ» среды; этот образ проявил себя в специфических комбинациях первичных признаков объектов, что формально отражено системой логических выражений. Формулы верхнего уровня, обобщающие внутренние логические комбинации, задают *модель* той ситуации, которая отражена в выборке; иначе говоря, мы вправе говорить о *закономерности*, которая сформировала данную выборку и отображена в структуре сети. Имея модель, мы также можем говорить об *отличии* одной ситуации от другой. Таким образом, РПС позволяет решать задачи прогнозирования и распознавания.

Таким образом, динамическую модель РПС можно считать концептуальным, *образным представлением* внешней среды или объекта, если его свойства объективно отражаются в выборке, проявляются через рецепторы и закрепляются в финальной архитектуре сети. Важнейшим преимуществом данной модели является тот факт, что процесс обучения фиксируется протоколом формирования новых понятий – от начальной позиции до финальной структуры, причём результаты обучения выражены в формально-логическом виде.

Парадигма познания

Из концепции *Растущей пирамидальной сети* вытекают, по крайней мере, два равноценных методологических подхода к её изучению и развитию. Первый подход – *бионический* – означает применение аппарата РПС для моделирования структур Памяти и механизмов мышления в Мозге – в целях расширения и углубления представлений о них. Второй подход – *кибернетический* – обуславливает реализацию аппарата РПС не только как программную, на обычных компьютерах, но и как архитектурно-структурную [1, 2, 4, 5, 6, 7, 11, 16, 19, 25, 27, 29, 31, 32, 33, 36, 40, 41, 42]. Такой кибернетический подход, собственно говоря, образует *научную парадигму изучения «функции через структуру»*. В идейном плане данный подход восходит к работам Мак-Каллоха и Питтса по созданию сетевых структур, но особенно большую популярность он приобрёл после разработки Ф. Розеблаттом обучаемой модели – перцептрона. Однако работами В.М. Глушкова и М. Минского было показано, что возможности перцептрона по решению проблем, связанных с интеллектом, достаточно ограничены и что необходимо продолжать поиски. В последние годы структурный подход получил значительный импульс благодаря *Дж. Хокинсу*, который, остроумно критикуя традиционную методологию изучения систем

методом «чёрного ящика», прорицает создание «думающих машин» на пути структурно-морфологического изучения Мозга человека [38].

Историческая справка. Структурный подход к изучению и воспроизведению интеллекта активно развивался в Институте кибернетики им. В.М. Глушкова Украинской Академии Наук, начиная с шестидесятых годов прошлого века. В практическом плане была поставлена задача разработки новых методов организации памяти ЭВМ будущих поколений. В результате углубления исследований и после предложения В.П. Гладуном идеи Растущей сети была сформулирована более общая концепция, которая заключалась в соединении двух подходов: «синтез структуры по заданной функции» и «синтез функции через структуру». Далее, на протяжении шестидесятых и семидесятых годов эта синтезированная методология получила научно-теоретическое и программное подкрепление. Тогда были разработаны программные модели Растущих пирамидальных сетей, Семантических балансных сетей и другие, а сам подход получил название «структурно-функциональное моделирование». Изучались алгоритмы обучения, распознавания, выявления закономерностей, прогнозирования состояний сложных систем и другие. Впоследствии, как уже упоминалось выше, на основе моделей был создан ряд аппаратурно-программных комплексов, ориентированных на решение конкретных задач [19, 34]. Тем самым был реализован тезис В.М. Глушкова «о единстве ближних и дальних целей в науке».

В теоретическом плане был сформулирован методологический **принциппознания естественного и искусственного интеллекта в их неразрывной связи**. Из данного принципа следует, что биокрибернетические модели памяти и мышления должны с самого начала учитывать главные черты устройства и работы человеческого мозга.

В результате развития этого гносеологического принципа (которое проходило в обстановке острых научных дискуссий при участии автора данной статьи) в области Искусственного интеллекта удалось сформулировать важнейший теоретический результат: *«Инструмент познания естественного интеллекта, воспроизводящий его функционально, должен быть подобен объекту познания структурно»*. (Разумеется, данное утверждение нельзя механически переносить на инструменты изучения неинтеллектуальных систем – объектов и явлений Природы).

Таким образом, ещё почти полвека назад в парадигме Искусственного интеллекта был совершён прорыв. Его следует соотносить с отечественной кибернетической школой, возглавляемой академиком В.М. Глушковым. При этом оформление гипотез относительно информационного моделирования на ЭВМ структур Памяти и процессов Мышления происходило в тесном контакте с биокрибернетической школой академика Н.М. Амосова.

Плодотворная идея затем была закреплена серией монографий (в том числе – изданных за рубежом), а также созданием в СССР большого количества программных систем, решающих целый класс задач **problemsolving**. Что касается резонансной книги **Дж. Хокинса** «Об интеллекте» (в соавторстве с С. Блейкли), то она вполне лежит в русле нашей парадигмы.

Механизмы мышления

Какое значение имеют Растущие пирамидальные сети для компьютерного моделирования важнейших **мыслительных функций Мозга и понимания мыслительных процессов?**

Прежде всего, их применение оказывается весьма эффективным для выработки в структуре Памяти и понимания механизмов Мышления при выполнении таких важнейших функций жизнедеятельности как опознание образов и решение проблем, которые являются опорными понятиями в Искусственном интеллекте [2, 14, 16, 26, 28, 29, 31, 32, 33].

Первое из них – **опознание объекта** (образа, ситуации) – означает, что вводимая в Память информация об объекте средствами РПС обобщается и уже в структурированном (пирамидальном) виде сопоставляется со знаниями, запечатлёнными в Памяти, в результате чего и происходит либо распознавание объекта, либо констатация отсутствия достаточной информации о нём в Памяти. В последнем случае могут быть определены наличные совпадения его некоторых свойств со свойствами других, запечатлённых в Памяти объектов, что может восприниматься как некоторое подобие анализируемого объекта его образному представлению.

Разумеется, что с использованием средств РПС может, причём эффективно, осуществляться *обучение* Памяти как формирование и фиксация в ней достаточной информации о новых объектах путём хотя бы повторного ввода первичной информации о них, либо по указаниям учителя.

Однако наиболее важной и, по-видимому, основной функцией механизма мышления, имея в виду его (мышление) как *целенаправленное*, безусловно, является **решение проблем**. Решаемая проблема формально представляема исходной и целевой ситуациями. Процесс её решения может быть интерпретирован как нахождение цепи преобразований *исходной ситуации в целевую*. И при этом подразумевается самое широкое толкование данных терминов. То есть, проблемой, например, может являться обычная математическая задача, или задание исходных данных и целевой функции. Другой пример проблемы – оказаться в заданном месте, двигаясь из исходного пункта. Алгоритм решения проблемы состоит из цепочки рассуждений и сравнения достигнутого результата с желаемым на каждом шаге.

Ясно, что в нашей жизнедеятельности проблем бесчисленное множество, они непрерывно возникают, и именно их решение, как осмысленное, и составляет сущность человеческого мышления. Шаг навстречу пониманию процессов рассуждений и вывода может быть сделан с помощью модели РПС. В аппарате РПС возможно планирование решений по преобразованию исходной ситуации в целевую, причём сам процесс реализуется сменой состояния (структуры или архитектуры) сети во времени. При этом составленный план может уже восприниматься как собственно решение, если он уже является окончательным и его выполнение требует лишь тривиальных действий [1, 2, 4, 5, 6, 16, 19, 21].

Таким образом, математический аппарат Растущих пирамидальных сетей оказывается весьма полезным в целях системного изучения проблемы, в которой «увязаны» Мышление, Память и её пирамидальная структура. Но и с противоположной стороны может оказаться полезным использование соответствующих особенностей естественных механизмов мышления для совершенствования аппарата РПС в целях структурно-аппаратной поддержки контроля текущих промежуточных результатов, например – в процессе получения целевого результата, в реализации семантических связей временного совмещения операций и т.п. А это уже, собственно, означает ранее упомянутое взаимное использование знаний в области естественного и искусственного интеллекта.

Этот, далеко не полный перечень направлений изучения Памяти человека и процессов Мышления, оказывается актуальным и востребованным для создания компьютеров нового поколения и кластерных систем. В частности, трудно себе представить, чтобы в интеллектуальном диалоге с человеком «искусственный мозг» не обращался к системе знаний, выработанных человечеством и не владел бы способами обработки этих знаний – вплоть до открытия новых. Работы в данном направлении уже ведутся давно [3, 7, 12, 13, 14, 15, 16, 17, 20, 22, 28, 39].

Решение проблем с помощью аппарата Растущих пирамидальных сетей, в том числе – при использовании его для моделирования структур Памяти и процессов Мышления, характеризуется совершенно очевидной **целеустремлённостью** действий в целях получения искомого результата [26].

Эти результаты могут использоваться для пополнения знаний – как в программной памяти компьютера, так и в компьютерных структурных моделях Мозга.

Такой вид мышления, осуществляемый в Памяти Мозга, который обладает *целеустремлённостью*, приводит к результату и осознаётся индивидуумом, принято считать *правильным*. Правильное мышление осуществляется при совместном действии *Сенсориума* (чувственного восприятия) и обязательно *Языковой системы* [3, 7, 11, 13, 15, 18, 19, 23, 36, 42]. Следовательно, как явление *сознания* (в широком смысле этого термина), такое мышление является индивидуальным и специфичным для каждого человека. В этом плане его изучение выходит за пределы нейрокибернетики, в поле которой выполнена данная работа, и относится к компетенции психологии личности.

В то же время следует подчеркнуть, что вышеуказанные концептуальные характеристики Памяти далеко не исчерпывают всех её особенностей. Следует обязательно принять во внимание тот факт, что человеческая Память, помимо смысловых реакций, обладает ещё эмоциональным выходом. **Эмоциональные факторы** проявляются как текущие изменения характеристик Памяти – глобальных и локальных, долговременных и мгновенных – и наблюдаются в возбуждённых или заторможенных реакциях субъекта, в глубине и концентрации внимания, в неповторяемости поведения и его непредсказуемости. Проявлению эмоций сопутствует лавинообразное нарастание возбуждения в нейронной сети. Так же, как *сознание* и *подсознание*, эмоциональные факторы для каждого человека сугубо индивидуальны. Именно сочетание *мировоззрения, знания и культуры* совместно с эмоциональными факторами определяет, собственно, то, что принято считать *душой* Человека. В материалистическом понимании данная категория является внутренним продуктом индивидуального (*филогенетического*) развития Личности в социальном и информационном Мировом, сложным «социально-эмоциональным фактором» его внутреннего состояния и внешнего поведения. Дальнейшие исследования в этом направлении, по нашему мнению, представляют для воссоздания искусственного интеллекта большой интерес [3, 10, 13, 15, 18, 19, 23, 35, 36, 37, 42].

Однако необходимо отметить, что к понятию «правильное мышление» эмоциональный фактор иногда никакого отношения не имеет. Именно эмоции зачастую заставляют нас поступать вопреки логике, совершать нецелесообразные и даже необъяснимые поступки. В то же время эмоции являются необходимым участником *творчества* (но не формального сочинительства), когда в сознании человека они отражают радости жизни или её невзгоды. Воспроизведение эмоций в ходе творчества и влияние их на результат, по-видимому, являются не столь отдалённой целью Искусственного интеллекта.

В связи с обозначенными здесь перспективами и, тем более – для эффективности новых технологических решений собственно задач Искусственного интеллекта, бесценное значение имеет Международная Программа **CDS** (*Cognition, Dialogue, Solution*), которая реализуется совместно Болгарским Институтом информационных теорий и применений **FOI ITHEA** и Ассоциацией создателей и пользователей интеллектуальных систем – **АСПИС**, созданной при Институте кибернетики им. В.М. Глушкова НАН Украины. Название **CDS** Программы является производным от наименования Международной конференции **KDS** (*Knowledge–Dialogue–Solution*). Она проводится регулярно в разных странах, пользуется популярностью и издаёт свои Труды. Именно сотрудничество учёных в рамках этой конференции привело к объединению их усилий в научной Программе, одним из разделов которой является создание новых интеллектуальных информационных технологий на основе Растущих пирамидальных сетей.

Здесь будет уместно подчеркнуть, что Программа **CDS** находится в поле одного из важнейших международных научно-исследовательских направлений, обозначенного термином «Мозг» (поскольку включает в содержание структурные и функциональные элементы нейроподобных сетей, которые

рассматриваются большинством исследователей в качестве моделей Мозга). В науковедческой литературе определены ещё два «стратегических» направления: «Общество» и «Вселенная», приоритетность которых для познания Мира – очевидна.

Модельное изучение интеллектуальных процессов: обобщение результатов.

Публикации последних лет, периода 2008 – 2009 годов [39, 40, 41, 42], в которых получили своё оформление и дальнейшее развитие изложенные выше идеи, вызвали значительный интерес учёных разных стран, были поддержаны в дискуссиях, а отдельные положения приняты научным сообществом и могут рассматриваться как **фундаментальные** в области Искусственного интеллекта. Поэтому в данной ретроспективной статье представляется необходимым обобщить совокупность предшествующих работ и сделать некоторые **выводы** относительно свойств Памяти. Итак:

1. **Концептуальность.** Главным и фундаментальным результатом, который служит отправным моментом для последующих исследований, является концептуальный подход к представлению образных структур в Памяти, в её нейросети, и взаимосвязь этих структур с информационными процессами, в них происходящими. Понятие «концептуальности» означает **неразрывную и взаимообусловленную связь** между структурной средой Памяти и процессами интеллектуальной обработки информации в этой среде. Исключение любого компонента из рассмотрения будет означать разрушение системного подхода.

2. **Иерархичность.** Структурная модель Памяти должна представлять собой *сеть* и быть организованной по принципу усложнения и углубления понятий. Процессы интеллектуальной обработки информации («машинного мышления») в модели Памяти, которая изучается на системном уровне, поддерживаются как **прямыми связями** между носителями образов («кластерами понятий»), так и **обратными связями** между уровнями иерархии. Поиск решения в Памяти представляется в виде **потока направленного возбуждения** узлов локальных нейронных структур (пирамидальных структур), ответственных за формирование образов, который развивается во времени и пространстве. На рисунке 1 показана гипотетическая структура памяти, которая отобразит процессы интеллектуальной обработки информации.

3. При воспроизведении некоторых **психических и нейробиологических функций** Памяти в нейросетевых моделях получают объяснение феномены, присущие естественным процессам и трактуемые как *образное мышление*. С позиций интеллектуального моделирования, где легко воспроизводятся процедуры лавинообразной активизации, пространственно-временной локализации, горизонтального и вертикального распространения возбуждения, могут быть интерпретированы или выдвинуты правдоподобные гипотезы относительно некоторых важных психологических процессов и явлений, таких, например, как: **внутренний самоконтроль цели, пути решения и достижения результата, спонтанность возбуждения, случайность, непредсказуемость поиска**, а также относительно таких проявлений Мозга как **сознание, подсознание, эмоции и душа**.

4. При воспроизведении **творческой основы** естественного интеллекта, которая в общих чертах характеризуется: разумным поведением, адекватностью реакций, согласованностью с внешней средой и её преобразованием, усвоением, накоплением, развитием и забыванием знаний, решением задач и проблем, обобщением понятий, распознаванием образов, разнообразной творческой деятельностью и так далее, необходимо опираться на предметные научные понятия: **семантическая сеть, растущая (динамическая) сеть, образ, «правильное мышление», планирование решений, обнаружение закономерностей, прогнозирование, представление, поиск и генерирование знаний** и другие. Для этого разработаны способы наглядного графического и модельного представления образов в Памяти и некоторых процессов Мышления.

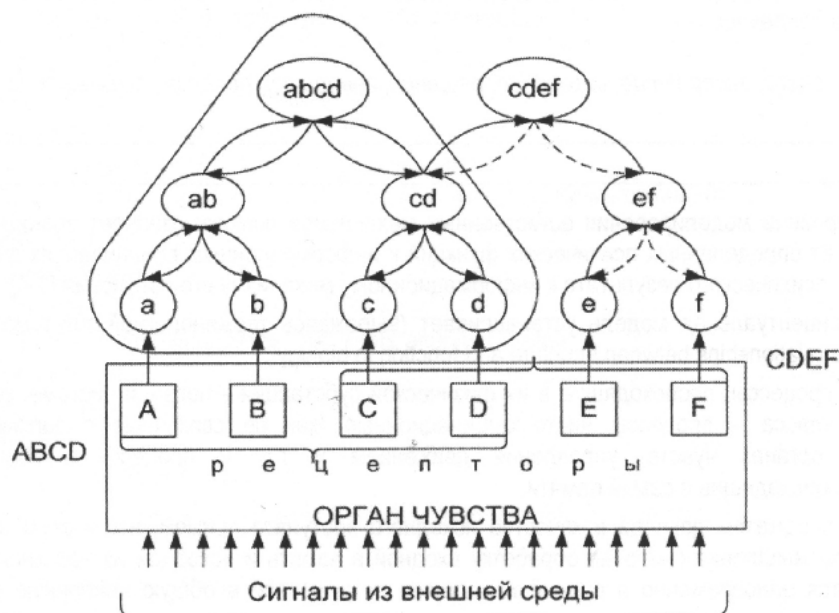


Рис. 1. Элементарные структуры восприятия образа, запоминания и распознавания.

5. Определён спектр характеристик Памяти и информационных действий в ней, которые целесообразно вместе со структурным подкреплением перенести в компьютер и суперЭВМ, что позволило бы улучшить не только их вычислительные ресурсы, но и резко повысить эффективность взаимодействия с Человеком за счёт **интеллектуального партнёрства**. Особенно перспективным видится результат такого перенесения для мультипроцессорных суперЭВМ с кластерной архитектурой. С другой стороны, моделирование Памяти (которое, кстати, уже привело к созданию *ассоциативной памяти* ЭВМ) открывает новые пути создания принципиально **новых компьютерных архитектур**.

Заключение

Методология. Исходя из выдвинутой парадигмы познания Интеллекта – «функция через структуру», которая является фундаментом нейрокибернетики и в русле которой и реализуется принцип структурного подобия Мозга и его кибернетической модели, мы приходим к научной методологии. Именно эта методология предполагает изучение углублённого в познание Естественного интеллекта Человека с позиций разных наук и моделирование его (исходя из возможностей современных технологий) компьютерными средствами. Она же (методология) предполагает использование этого познания в создании Искусственного интеллекта – в том числе и с использованием второго в познании первого. В этом проявляется **диалектика познания** Человеком Вселенной и самого себя. Но познание должно обязательно идти рядом с **осознанием**, то есть – пониманием ведущей духовной роли Человека в этом процессе.

От редакции. Коллеги и ученики З.Л. Рабиновича сочли необходимым расширить список публикаций, упомянутых автором в его статье, и отразить многолетний плодотворный труд этого выдающегося Кибернетика в области Памяти, Мышления и Искусственного интеллекта. Публикации расположены в хронологическом порядке.

Благодарности

Статья напечатана при финансовой поддержке проекта **ITHEA XXI** Института Информационных теорий и Приложений FOI ITHEA Болгария (www.ithea.org).

Литература

1. Гладун В.П., Рабинович З.Л. Формирование понятий с помощью обучаемой сети нейроподобных элементов // Материалы симпозиума «Основные подходы к моделированию психики и эвристическому программированию». – Изд. ИК АН ГрССР, Тбилиси, 1968.
2. Гладун В.П., Рабинович З.Л., Попурий Е.П. Обучаемый классификатор // Конференция по теории автоматов и искусственному мышлению. Аннотации докладов. – Изд. ФАИ УзССР, Ташкент, 1969. – С. 33 – 34.
3. Рабинович З.Л. Некоторые вопросы автоматизации творческих процессов // Вторая Всесоюзная конференция по философским проблемам кибернетики. – Изд. Института философии АН СССР, 1970.
4. Гладун В.П., Попурий Е.П., Рабинович З.Л. Экспериментальная классификация и формирование понятий В сб. Кибернетика и вычислительная техника. – «Наукова думка». – К.: 1970, вып.7. – С. 28 – 32.
4. Гладун В.П., Ефименко О.Н., Рабинович З.Л. Выделение характерных сочетаний признаков объектов на ЦВМ // Управляющие системы и машины – 1972 – № 2 – С. 34 – 37.
6. Гладун В.П., Ефименко С.М., Рабинович З.Л. Программа выделения характерных сочетаний признаков. – РФАП АН УССР. – К.: 1974.
7. Рабинович З.Л. Подход к исследованию механизмов высших функций мозга путём их моделирования на основе обучающихся ассоциативных структур // В сб.: Проблемы адаптивного управления. – Ростов-на-Дону. – Изд. РГУ. – 1974. – С. 211 – 225.
8. Глушков В.М., Рабинович З.Л. Проблемы автоматизации дедуктивных построений // Управление, информация, интеллект / Под ред. Берга А.Н., Бирюкова Б.В., Геллера Е.С., Поварова Т.Н. – М.: Мысль, 1976. – ч. 4, гл. 2. – С. 300–326.
9. В.П. Гладун, З.Л. Рабинович. Формирование модели мира в системах искусственного интеллекта (Предварительная публикация). – М.: 1977. – 19 с.
10. Гладун В.П. Эвристический поиск в сложных средах. // Киев: Наукова думка, 1977. – 166 с.
11. Рабинович З.Л. Некоторый бионический подход к структурному моделированию целенаправленного мышления // Кибернетика 1979, №2, с. 116–124.
12. Gladun V., Rabinovich Z. Formation of the World Model in Artificial Intelligence Systems // Machine Intelligence, 9. – Ellis Horwood Ltd., Chichester. – 1980. – P. 299 – 309.
13. Глушков В.М., Стогний А.А., Биба И.Г., Ващенко Н.Д., Галаган Н.И., Гладун В.П., Рабинович З.Л., Сакунов И.А., Хоменко Л.В. // Системы автоматизации творческих процессов в научных исследованиях, проектировании и задачах управления роботами. – Кибернетика – 1981. – № 6. – С. 110 – 115.
14. Гладун В.П., Рабинович З.Л. Решение задач планирования действий и формирования понятий на основе дедуктивных и индуктивных построений. // В кн.: IX Всесоюзный симпозиум по кибернетике (г. Сухуми, 10 – 15 ноября 1981 г.). – Тезисы докладов. – К.: 1981. – С.94 – 96.
15. Глушков В.М., Гладун В.П., Рабинович З.Л. Автоматизация творческих процессов в научных исследованиях // В кн.: IX Всесоюзный симпозиум по кибернетике (г. Сухуми, 10 – 15 ноября 1981 г.). – Тезисы докладов. – К.: 1981. – С. 110 – 111.
16. Гладун. Планирование решений – Киев: Наукова думка. – 1990. – 168 с.
17. Коваль В.Н., Рабинович З.Л., Авербух А.Б. Поддержка систем искусственного интеллекта на основе интеллектуальных решающих машин – нового класса ЭВМ // III Конференция по искусственному интеллекту “КИИ–92. Сборник научных трудов, том 2. – Тверь, 1992. С. 169–172.

18. Воронков Г.С., Рабинович З.Л. Сенсорная и языковая системы – две формы представления знаний // *Новости искусственного интеллекта*. – 1993. – № 2. – С. 116–124.
19. Виктор Гладун. Процессы формирования новых знаний. София: СД «Педагог 6», 1994. 192 с.
20. Рабинович З.Л. О концепции машинного интеллекта и его развития // *Кибернетика и системный анализ*. – 1995. – № 2. – С. 163–173.
21. Рабинович З.Л., Яценко В.А. Подход к моделированию мыслительных процессов на основе нейроподобных растущих сетей // *Кибернетика и системный анализ*, 1996, № 5, с. 3–20.
22. Коваль В.Н., Булавенко О.Н., Рабинович З.Л. Интеллектуальные решающие машины как основа высокопроизводительных вычислительных машин. // *Управляющие системы и машины*. № 36, 1998. – С. 43–52.
23. Рабинович З.Л., Воронков Г.С. Представление и обработка знаний во взаимодействии сенсорной и языковой нейросистем человека // *Кибернетика и системный анализ*, 1998, № 2. – С. 3–11.
24. Киселёва Н.Н., Ващенко Н.Д., Гладун В.П., Леклер С.Р., Джексон А.Г. Прогнозирование неорганических соединений, перспективных для поиска новых электрооптических материалов // *Перспективные материалы*. – 1998. – № 3. – С. 28 – 32.
25. G. Voronkov, Z. Rabinovich. Cognitive model of memory and thinking // *International Journal on Information theories and applications*, 2000, vol. 7, No 4, pp. 164–169.
26. В.П. Гладун. Партнёрство с компьютером. Человеко-машинные целеустремлённые системы. "Port Royal", Киев – 2000. – 118 с.
27. Воронков Г.С., Рабинович З.Л. Естественная среда памяти и мышления: модельное представление // *Труды Международной научно-практической конференции KDS–2001, том 2 – Санкт-Петербург, июнь 2001. – «Лань», Сев-Зап. государственный заочный технический университет. – С. 110–115.*
28. Koval V., Bulavenko O., Rabinovich Z. 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).– pp. 21–26.
29. G. Voronkov, Z. Rabinovich. On neuron mechanisms used to resolve mental problems of identification and learning in sensorium // *International Journal on Information theories and applications*, 2003, vol. 10, No 1, pp. 23–28.
30. Рабинович З.Л. О естественных механизмах мышления и интеллектуальных ЭВМ // *Кибернетика и системный анализ*. – 2003. – № 5. – С. 82–88.
31. З.Л. Рабинович. Концептуальное представление об опознании образов и решении проблем в памяти человека и возможностях его использования в искусственном интеллекте // *XI-th International Conference «Knowledge – Dialogue – Solution». Proceedings. Vol. 1. FOI-Commerce, Sofia, 2005, pp. 1–8.*
32. Zinovi Rabinovich. Conceptual idea of identification of patterns and problem solving in the human's memory and the possibilities to use it in artificial intelligent // *International Journal "Information Theories @ Applications"*, 2005, vol. 12, No 1, pp. 41–48.
33. Zinovi L. Rabinovich, Yuriy A. Belov. Conceptual Idea of Natural Mechanism of Recognition, Purposeful Thinking and Potential of Its Technical Application. Mechanisms, Symbols and Models. First Intern. Work-Conference on the Interplay, Spain, June, 2005. Proceedings, Part 1.
34. Л. Святогор. К вопросу о развитии интерфейса «Разработчик – Заказчик». // *XI-th International Conference «Knowledge –Dialogue – Solution». Proceedings. Vol. 1. FOI-Commerce, Sofia, 2005, pp. 371– 374.*
35. З. Рабинович, Ю. Белов. Сознание, подсознание и эмоции, душа и KDS // *XI1-th International Conference «Knowledge –Dialogue – Solution». Proceedings. FOI-Commerce, Sofia, 2006, pp. 11–18.*
36. З.Л. Рабинович. Аспекты нейролингвистики в свете концептуального представления о структуре памяти и процессах мышления и их бионическое использование // *Труды конференции «Горизонты прикладной лингвистики» – Симферополь, 2006. – TIP, 2006.*

37. Zinovy Rabinovich, Yuriy Belov. Consciousness, subconsciousness and emotions, soul and KDS // International Journal "Information Technologies and Knowledge, 2007, vol. 1, No 2, pp. 152–158.
38. Хокинс Дж., Блейкли С. Об интеллекте: Пер. с англ. – М.: ООО «И.Д. Вильямс», 2007. – 240 с.
39. Рабинович З.Л. Об интеллектуализации ЭВМ (история и перспективы) // Міжнародна конференція "50 років Інституту кібернетики імені В.М. Глушкова НАН України". – Україна, Київ, 24–26 грудня 2007 року. Праці конференції. – Київ, 2008. – С. 174–176.
40. Рабинович З.Л., Белов Ю.А. Главные принципы организации памяти и мыслительных процессов в мозге человека (в их концептуальном представлении) // IV Міжнародна школа-семинар «Теорія прийняття рішень» – Ужгород, 29.09 – 04.10.2008. Праці школи-семінару – Ужгород, 2008. – С. 32–33.
41. Yuriy Belov, Zinovy Rabinovich. Basic Principles of Organisation of the Medium and Thinking Processes of the Human in Its Conceptual Presentation // Book Series "Information Science and Computing", 2008, No 7. "Artificial Intelligence and Decision Making", vol. 2/2008. Pp. 219 – 227.
42. Ю.А. Белов, З.Л. Рабинович. Память человека и мышление – образное и символьное. Концептуальное модельное представление // Доклады Академии Наук (РФ), 2009, том 427, № 6, Секция «Информатика». – С. 761–764.

Информация об авторе



Зиновий Львович Рабинович

1918 – 2009,

д.т.н., проф.

Институт кибернетики им. В.М. Глушкова НАН Украины.

ИСТИНА ИЛИ СМЫСЛ? ДВА ПОДХОДА К ОБРАБОТКЕ ЕСТЕСТВЕННО ЯЗЫКОВЫХ ТЕКСТОВ

Леонид Святогор, Виктор Гладун

Аннотация: *Выполнен сравнительный анализ двух принципиально разных подходов к семантической обработке текстов и дискурсов: логическая интерпретация и онтологический анализ. В зависимости от семантической сложности текста устанавливается сфера применимости каждого направления. Подробно рассматривается метод онтологического анализа, который опирается на систему знаний и формализованную категорию «смысл». Уточняется понятие «машинное понимание текстов ЕЯ».*

Ключевые слова: *логический анализ, онтология, смысл, текст, машинное понимание*

ACM Classification Keywords: *I.2.4 Knowledge Representation Formalisms and Methods; I.2.7 Natural Language Processing – Text analysis*

Введение

Логический и содержательный анализ текстов естественного языка (ЕЯ) не могут быть противопоставлены друг другу: каждый соотносится со своей областью применения. Если логическая обработка, в принципе, устанавливает непротиворечивость высказываний «логическим знаниям» о предметной области (ПрО) и их выводимость из системы аксиом, то в задаче семантического анализа слабо-формализованных и разнообразных по тематике текстов ЕЯ на первый план выдвигается понимание авторского (коммуникативного) замысла, *извлечение* его из текстового материала и *лексическое оформление*.

С позиций искусственного интеллекта Д.А. Поспелов выделил пять уровней машинного понимания текстов различной степени семантической сложности. На первом и втором уровне востребуются только те термины, которые содержатся в тексте непосредственно или могут быть получены из текста путём его логического расширения. На третьем уровне для понимания текста нужно привлечь некоторую дополнительную информацию, которую нельзя вывести логическими методами. Четвёртый и пятый уровни требуют, чтобы к интерпретации текста были привлечены такие понятия, которые в тексте отсутствуют, но накоплены в базах знаний о мире, социуме и используются в предметных областях человеческой деятельности.

В последних двух вариантах толкование текста требует привлечения широкого контекстного материала: знаний о целях коммуникации, о социальной среде, её моральных и культурных ценностях и других системных характеристик.

Цель и содержание данной работы состоят в том, чтобы проанализировать и сравнить между собой два различных подхода к семантической обработке текстов – логическую интерпретацию и онтологический анализ. На основании сравнения проявляется сфера приложения каждого направления. Подробно рассматривается метод онтологического анализа, который опирается на систему знаний и формализованную категорию «смысл». Уточняется понятие «машинное понимание текстов ЕЯ».

ТЕКСТ и ДИСКУРС

Объектом семантического разбора является текст или дискурс. Необходимо обозначить рамки этих понятий следующими вопросами: Что такое текст? Каково его функциональное назначение? Что мы хотим из него извлечь? Последний вопрос является кардинальным с прагматической точки зрения.

Для выяснения этих вопросов дадим следующее **определение**. *Текст есть лексическая, грамматически и тематически оформленная, конечная форма передачи замысла автора партнёру по коммуникации – в общей знаковой, языковой и знаниевой среде.*

Современная теоретическая лингвистика опирается на семь критериев текста, сформулированных в 1981 г. Бограндом и Дресслером: 1) связность, 2) целостность, 3) интенциональность, 4) приемлемость, 5) информативность, 6) ситуационность и 7) интертекстуальность [Тичер и др., 2009]. В русле данной работы нас будут интересовать следующие параметры текста: связность, целостность (тематическое единство) и интенциональность (авторский замысел).

В данном исследовании никаких ограничений на тематическое содержание текстов не накладывается. Большинство исследователей считают *функциональным назначением текста* фиксацию и передачу рациональных и культурных знаний в процессе коммуникации людей и сообществ. Текст рассматривают как вербализованную и закодированную средствами языка мысль автора (или его замысел).

На практике сложилось нечёткое разделение текстов на два вида: те, которые функционируют в узкой специализированной предметной области (ПрО) или в документообороте, и другие, неформатные тексты с *широким спектром* семантических отношений, например – дискурсы. Первый вид характеризуется небольшим количеством семантических отношений и достаточно высокой структурируемостью данных. *Дискурс* определяется как «*связный текст в контексте многих конституирующих и фоновых факторов – социокультурных, психологических и т.д.*» [Штерн, 1998, с. 87]. Он характерен локальными (микро) и глобальными (макро) семантическими структурами, причём последние характеризуются сложными, нелинейными схемами их *смыслового взаимодействия*, которые зависят от *авторского замысла*.

Существо дискурса, его суть определяются не столько количеством включённых в ситуацию объектов, сколько переплетением связей между ними. Связи эти имеют преимущественно ассоциативный характер. В многообразии непредсказуемых отношений между семантически оформленными элементами заключена сложность текста для его восприятия.

Под *сложным текстом* мы будем понимать такой, который нельзя представить конечным числом предикативных отношений в силу существования сюжетных, кореферентных и имплицитных связей между отдалёнными частями (смысловыми блоками) текста. Именно референциональные связи «*отражают коммуникативную установку автора, его интенции, фонд знаний партнёра, коммуникативную организацию высказывания, отношение к контексту*» [Штерн, 1998, с. 272]. Связи многообразны и трудно-перечислимы, не всегда могут быть установлены логическими методами, а их совокупность вместе с объектами воссоздаёт фрагмент реальной или ментальной действительности.

Отсюда вытекают трудности принципиального характера, когда мы хотим извлечь из текста его главное содержание, то есть – выявить ту коммуникативную цель, которую преследует автор. Предполагая, что эта цель выражена в тексте некоторым «*СМЫСЛОМ*», и что вербализованный смысл отделён от автора, мы тем самым признаём *объективность* данной категории. «*Смысл существует сам по себе как инвариант, который выясняется (устанавливается) во время перевода или пересказа текста*» [Штерн, 1998, с. 286]. Более того, чётко утверждается, что «*сам текст на естественном языке является*

лишь проводником смысла, а замысел и знания автора лежат во вторичной структуре (смысловой структуре или макроструктуре текста) [Гаврилова, Червинская, 1992, с. 104].

Если считать, что текст служит, образно говоря, «контейнером» мысли и смысла, то необходимо подчеркнуть, что данный контейнер «погружён в море знаний». Это означает, что толкование текста невозможно без привлечения терминов и понятий, которые в самом тексте не упоминаются, но существуют в обиходе коммуникации; они называются *контекстным знанием*. Сложный текст может быть осмыслен, то есть – понят только в контексте широких языковых знаний об окружающей его среде.

ЛОГИКА

Отношение логики к смыслу и смысла – к логике

Будем исходить из предпосылки, что в классической логике значение истинности любого сложного высказывания, которое построено из простых (аксиоматических) высказываний и логических связок *и*, *или*, *не* – есть функция истинности простых высказываний, причём аргументы и функции заданы на множестве значений (*истина*, *ложь*). Центральной проблемой теории является проверка *выводимости* утверждений и их логической *непротиворечивости*. Хотя сами по себе логические высказывания инвариантны к содержательной (смысловой) стороне терминов, предполагается, что от бессмысленных высказываний должны уберечь «правильные» аксиомы, поскольку на стадии их формирования закладывается семантическая связь с внешним миром, а также «корректные» формальные правила вывода (причём обязаны также соблюдаться «правила применения правил»). Всё это, однако, при конечной системе аксиом и при достаточно длинном выводе не убергает – в момент перехода с логического языка на естественный – от появления *«смысловых парадоксов»*.

Несмотря на это, в логике встречается мнение, что в процессе формального вывода сохраняется *«связь по смыслу»* [Одинцов, 1996; Рубашкин, 1989]. Рассмотрим пример. Возьмём две аксиомы, не вызывающие никаких возражений: *«В огороде – бузина»* и *«В Киеве – дядько»*. Объединив их импликацией *если – то*, получим безукоризненно истинное высказывание. Однако порождённая здесь логическая истина демонстрирует как раз факт отсутствия смысла. Разумеется, здесь недостаёт аксиомы, связывающей объекты двух посылок, но мы это обнаружили уже за рамками логики («слишком поздно»). В данном примере «связь по смыслу» вывела нас за пределы смыслового поля.

Проблема, однако, заключается не в том, что между *истиной* и *смыслом* нет *однозначного соответствия*, а в том, что такого соответствия нельзя требовать. *Смысл вообще не является компетенцией логических исчислений*, и в логике это понятие не имеет определения. Смысл, понимаемый как совокупность неформальных связей между частями дискурса, *не измеряется двоичной переменной*. Считать смыслом *имя предиката* не корректно – это имя получает смысловую интерпретацию только в естественно-языковом выражении. В свою очередь, *логические связки и, или, не* также не имеют никакой другой интерпретации, кроме «математической» – это просто табличные числовые функции. На пути превращения логики Аристотеля в формальную систему, как отмечает Д.А. Поспелов, *«исчезла логика и победила математика»* [Поспелов, 1982, с. 31].

Не всё благополучно обстоит также и с аксиоматикой. Попытка придать семантику даже простому аксиоматическому утверждению обходится дорого: приходится уточнять понятия, затем – уточнять вторичные понятия, и этот процесс лавинообразно нарастает, вовлекая в себя всё новые пласты знания [Рубашкин, 1989]. Именно поэтому *«Логик пока не идёт дальше рассмотрения узкоспециализированных понятийных систем»* [там же, с. 185].

Однако мы вправе требовать связи истинности со смыслом в ином – этическом плане. Как отмечают психологи, человек *«выделяет особый смысл, который связан с ориентацией на истину как на*

бытийную ценность и является нормативным по отношению к профессиональной деятельности учёного» [Балл, 2006, с. 110].

Сформулируем кратко итог. **Истинность есть мера выводимости высказываний, а смысл есть мера знания о мире.** С позиций искусственного интеллекта можно считать, что категория смысла универсальнее истины.

Что мы хотим извлечь из текста?

В конечном итоге, все методы семантического анализа направлены на формализацию и обработку знаний. Однако при решении задач семантической обработки текстов вырисовывается принципиальная разница не столько в методах, сколько в целевой установке обработки знаний: если логический аппарат ориентирован на вывод и получение *новых* знаний, то семантический анализ текстов нацелен на извлечение *готовых* знаний и их перекрёстный анализ. При этом, как указывалось, *истинность или ложность обнаруженных в тексте конкретных знаний не являются целью семантической обработки*: приоритетная цель заключается в установлении смысловых отношений между лексическими элементами, кодирующими знание.

В зависимости от степени формализации текста и сложности мы можем извлечь из него либо ту информацию, которая в нём содержится – в терминах непосредственно составляющих, либо другую информацию, которая раскрывает содержимое терминами контекстного знания. В последнем случае необходимо построить систему контекстного знания – онтологию. Как будет показано ниже, целью семантической обработки является извлечение из текста и дискурса онтологического смысла. Тем самым категория «смысл», извлечённая из ментальной сферы, приобретает определённость и конструктивность, что позволяет вести обработку машинными средствами. Однако, как справедливо указывает *И.Б. Штерн*, «*рассматривая понимание естественного языка как оперирование смыслом, прежде всего надо разрешить принципиальные проблемы, связанные с определением сущности и структуры смысла*» [Штерн, 1998, с. 286]. Один из возможных способов решения этих проблем предложен в следующих разделах.

Краткий вывод. В тех текстах, где данные могут быть структурированы и ограничены по множеству объектов и отношений (например, в рамках ПрО), целью диалога или поиска служит извлечение данных. В сложных текстах требуется определить смысл как цель коммуникации.

В следующем разделе рассмотрен пример логико-предикативной модели представления знаний, используемой для машинного диалога. Он необходим для сопоставления с онтологической моделью.

Принципы логической обработки знаний

Типичные черты предикатно-логического подхода присутствуют в поисковой системе, подробно и ясно описанной в работе [Одинцов, 1996]. Предметной областью служит организация производства, снабжения и сбыта в металлургической отрасли Украины (фрагмент отрасли).

Информация о ПрО представлена в *базе знаний* в виде семантических структур **фактов**, причём данные структуры явно определяют стандартные связи между объектами. Фактом называют *локализованную минимальную совокупность понятий естественного языка, отражающих связь между классами объектов, классами и их элементами, а также между элементами различных классов*. Факты называются также суждениями или высказываниями. Классы объектов перечислены в *базе данных*. Семантическая структура факта представлена деревом, в узлах которого находятся понятия, а дуги отражают семантические отношения между ними. Одновременно такая графическая структура транслируется в предикатную форму.

Запрос, который формулируется на естественном языке с незначительными ограничениями на синтаксис, преобразуется через предикатную форму в семантическую структуру, подобную структурам базы знаний. После этого производится **распознавание** структуры запроса в базе знаний. При совпадении структура базы знаний преобразуется в предикат, а последний транслируется в ответ на естественном языке.

Поисковая система поддерживает составные и сложные вопросы. Она может не только продублировать вопрос с выдачей искомым данных, но и представить прогноз относительно ресурсов, выполнения плана и другие знания.

Как утверждает автор рассматриваемой работы, предикат типа: **«Находится(предприятие (А), город(В))»** «связывает понятия «предприятие» и «город» в **единое, имеющее смысл, целое**» [Одинцов, 1996, с. 19]. С этим можно согласиться, если рассматривать данную ситуацию как *«функциональное отношение между предикатом и актантами, которое отражает семантическую структуру данной ситуации»* [Штерн, 1998, с. 265]. И при этом добавить, что *смысл целого* проверяется за пределами логики.

В описанной системе логика предикатов, дополненная структурами знаний, служит естественным аппаратом для проверки истинности выводимых утверждений: предметная область в понятийном плане ограничена, данные структурированы, предикатные отношения перечислимы (сведены в таблицу), синтаксис упрощён. Однако, *«возникают огромные технические трудности, которые перерастают в принципиальные препятствия, как только модель мира становится достаточно объёмной, а число тем, по которым человек хочет вести диалог с ЭВМ – велико»* [Поспелов, 1982, с. 192].

В новейших работах логико-семантического направления наблюдается стремление больше *сблизить логику с семантикой*. Для этого, например, в работе [Никитченко, 2008] в теорию вводятся *«модели миров»*. Успех в данном направлении, по нашему мнению, может быть достигнут, если номинативные структуры логики будут связаны с системой репрезентации знаний, а выводимость будет означать обнаружение формальных смысловых связей. Тогда семантическая нагрузка может быть возложена в моделях миров на интенциональные и экстенциональные связи, а прерогативой логики останется вывод и синтез сложных семантических понятий из простых.

ОНТОЛОГИЯ И СМЫСЛ

Онтологическая картина мира

Наша задача заключается в том, чтобы **перевести понятие «смысл» из ментальной или подразумеваемой категории – в конструктивную**. Это возможно сделать лишь в том случае, если смысл будет включён в формальную систему репрезентации знаний – семантическую сеть; наиболее пригодной для наших целей является **онтология** [Святогор, Гладун, 2009].

Известны онтологии двух типов – абстрактные и предметные онтологии. *Абстрактные онтологии* (Дж. Совы, СУМО, Микрокосмос) создаются с философских позиций, чтобы ответить на вопрос: какие категории познания создаёт человек для понимания универсума – природы, космоса, материи. Однако на уровнях межличностного и социального общения человек, стремясь быть понятным, предпочитает структурировать мир иначе. В этом случае для передачи коммуникативных знаний подходят *предметные онтологии*, которые лучше передают ситуации и события в окружающей среде. Предметная онтология способна сочетать в себе, в терминологии Ю.Д. Апресяна, «научную» и «наивную» картины мира.

Предметная лексическая онтология представляет собой семантическую сеть, в которой узлам соответствуют языковые понятия, описывающие объекты или состояния – *концепты*, а связи отображают различные *отношения* между ними. В соответствии с законами познания внешнего мира сеть

имеет *иерархическую* архитектуру, где чем больше уровень общности понятия, тем более высокое место в структуре оно занимает. «На стадии концептуализации эта иерархическая структура превращается в пирамиду знаний» [Гаврилова, Червинская, 1992, с. 44]. Формальные концепты и отношения сети имеют под собой мощную *систему интерпретации* на естественном языке.

Для репрезентации глобальной системы всеобщих знаний, ориентированной на анализ текстов с произвольной тематикой, разработана предметная онтология **ИО*3**; она подробно и вместе с примерами описана в работе [Гладун и др., 2008].

Трёхуровневая иерархическая онтология ИО*3

Основными принципами построения трёхуровневой иерархической онтологии – **ИО*3** являются:

- структурирование мира на основе материалистической парадигмы академика *В.И. Вернадского*;
- введение *трёх уровней* иерархии («слоёв знаний»), внутри которых развиваются многочисленные подуровни связанных концептов;
- выбор в качестве *концептов*: имён, именующих понятия, объекты, ситуации, состояния и атрибуты, и сложившихся в языковой среде общего знания и общения;
- выбор в качестве *концептуальных связей*: формальных, атрибутивных и ассоциативных отношений;
- задание *интерпретирующей системы* на множестве декларативных знаний, хранящихся в учебниках, энциклопедиях и толковых словарях;
- замыкание концептов на *словарь* естественного языка.

Верхний уровень иерархии образует дерево, вершиной которого служит глобальный концепт *Материя*. Далее он раскрывается через понятия: *Косное вещество*, *Живое вещество*; последнее разделяется на *Биосферу* и *Ноосферу*, и далее деление продолжается, переходя к концептам среднего уровня. Здесь находятся понятия, которые сложились в науке для изучения природы, человека и общества по отраслям знаний. Наконец, общенаучные понятия раскрываются на нижнем уровне через конкретные характеристики и отношения бытового уровня, которые могут быть спрогнозированы в корпусе текстов. Кроме того, на третьем уровне формулируются и подключаются к среднему уровню концепты предметных областей, которые могут, в принципе, представлять любые области знаний и деятельности человека.

Построение пирамиды онтологии подчинено чёткому принципу репрезентации знаний через *интенционалы* и *экстенционалы*. Каждое понятие в семантике рассматривается в единстве его этих двух категорий. Это значит, что каждый интенционал должен быть раскрыт соответствующим ему по семантическому отношению экстенционалом. Практически это означает, что каждый концепт, выступающий в роли интенционала или экстенционала, имеет *значение* из множества **имён** объектов или ситуаций, которые имеют *денотат* в реальном мире и лексический *смысл* в естественном языке [Шевченко, 2003]. В онтологии **ИО*3** в роли экстенционала могут выступать: *части понятия* (при классификации), его *признаки* (при атрибуции), *значения* (при уточнении понятия), *примеры* (в качестве представителей класса), а также *ассоциативные понятия* (по усмотрению эксперта). Очевидно, что при движении по вертикали пирамиды (вниз) экстенциональные компоненты становятся интенциональными по отношению к собственным определителям. Экстенционалы (имена концептов) взаимодействуют со словарём естественного языка.

Для того, чтобы придать пирамиде знаний конструктивные, процедурные свойства, воспользуемся понятием «графа концептуальных зависимостей», введенного *Р. Шенком*. Уникальным свойством данной конструкции является тот факт, что «**граф концептуальных зависимостей служит метаязыком для внутренней смысловой репрезентации текстов, который отображает смысловую структуру ситуации**» [Штерн, 1998, с. 197].

Иерархическую трёхуровневую онтологию – **ИО*3** будем считать ориентированным *концептуальным графом*, все пути на котором ведут из любого концепта к глобальной вершине *Материя*.

Относительно количества концептов и внутренних связей онтологии никаких ограничений не делается: её структура является *расширяемой* за счет добавления новых знаний, в том числе – разных ПрО.

Концептуальный онтологический граф **ИО*3** имеет следующие преимущества, которые выделяют его среди других систем репрезентации знаний:

- граф служит инструментом *формализации* понятия «смысл текста естественного языка»;
- он позволяет свести процедуру семантического анализа текста к *извлечению смысла* и реализует данную процедуру *стандартными графическими операциями*;
- допускает *машинное представление* и машинную обработку текстовых знаний.

Семантические процедуры извлечения из текста его глубинного содержания будем в дальнейшем называть *онтологическим анализом*. Результатом онтологического анализа служит *онтологический смысл*. В следующем разделе показано, каким образом онтологический анализ (выделение смысла) может быть реализован алгоритмически.

Онтологический смысл: формальное представление

В искусственном интеллекте задача извлечения знаний из текстов конкретизируется как «*понимание и выделение смысла текста*» [Гаврилова, Червинская, 1992, с. 104]. Как было показано выше, фундаментальной основой для выделения смысла может служить онтология. Однако без формального определения смысл остаётся объектом ментального рассмотрения в психологии и литературоведении. Попытка построить лингвистическую модель «*Смысл – Текст – Смысл*» [Мельчук, 1999] не привела к конструктивным результатам, поскольку, используя формальные семантические, синтаксические и морфологические структуры, не опиралась на систему знаний. Поэтому нашей первой задачей было дать формальное определение этой категории мышления и сделать её *вычислимой*.

Полное рассмотрение дано в работе [Святогор, Гладун, 2009]; здесь приводятся только основные определения.

Определение 1. *Конструктом смысла* является пара концептов, связанных между собой формальным, атрибутивным или ассоциативным отношением.

Определение 2. Онтологический смысл строится из конструктов смысла.

Онтологическим смыслом некоторого концепта является подграф концептуального графа онтологии ИО*3, определяемый начальным и финальным концептами. Указанный подграф называется *смысловой траекторией* начального концепта.

Определение 3. Значением начального концепта смысловой траектории служит *ключевое слово* текста; значением финального – вершина онтологии: *Материя*.

Определение 4. Совокупность смысловых траекторий всех ключевых слов текста формально определяет *онтологический смысл текста*.

Таким образом, смысловой анализ начинается с извлечения из текста очередного ключевого слова. Основанием для такой операции служит определение, предложенное в работе [Гаврилова, Червинская, 1992, с. 109]: «*Набор ключевых слов – это набор опорных точек, по которым развёртывается текст при кодировании в память и осознаётся при декодировании, это семантическое ядро цельности*». Со своей стороны, мы опираемся на гипотезу, что ключевые («значимые») слова являются основным семантическим инструментом автора текста для кодирования своего замысла.

Таким образом, **формальный онтологический смысл** получается в результате проекции дискретизированного текста на онтологию **ИО*3**. Результатом отображения в сети каждого ключевого слова служит **смысловая траектория** – дискретная упорядоченная цепочка взаимосвязанных концептов, которая «считана» с подграфа онтологического графа. Задание онтологического смысла в таком формате пригодно для компьютерного представления и «машинного понимания».

Онтологический смысл отличается от других известных определений, а тем более – от интуитивного понимания смысла, тем, что он *задан в многомерной координатной системе представления знаний о внешнем мире и имеет – в фиксированной онтологии – однозначное графическое представление.*

МАШИННОЕ ПОНИМАНИЕ

Понимание и связность

Как отмечалось выше, многие исследователи в области искусственного интеллекта определяют **понимание реалии** как «*постижение её смысла через конституирующие её контексты*» [Штерн, 1997, с. 92]. «*Процесс понимания текста заключается в извлечении смысла, заложенного в текст, и именно текст является носителем этого смысла ... Рассматривая понимание естественного языка как оперирование смыслом, прежде всего необходимо разрешить принципиальные проблемы, связанные с определением сущности и структуры смысла*» [Штерн, 1998, с. 285].

На пути разрешения этих принципиальных проблем понимания естественного языка нами предложено понятие онтологического смысла. Подчеркнём ещё раз, что онтологический смысл не есть отображение высказывания на множество *ноль* и *единица*. Это есть *отображение высказывания на связанное множество именованных концептов, репрезентующее систему знаний* – онтологию.

В самом общем случае **понимание** является целью коммуникации и прерогативой двух интеллектов. На ментальном уровне процесс понимания означает возбуждение в памяти человека моделей реальности (языковых и образных), нахождение аналогий или закрепление новых связей; на машинном уровне факт «компьютерного понимания» означает возбуждение на графе онтологии соответствующего подграфа смысла и связывание его с другими подграфами. В том и другом случае реконструируется некоторая языковая ситуация (реалия).

Соединение онтологии со смыслом позволяет соединить *семантику и прагматику текста в единый процедурный комплекс*. Действительно, онтология знаний отвечает за связь текста с внеязыковой действительностью, а смысловая траектория отвечает за кодирование онтологического смысла в машинной памяти. Такой синтез семантики и прагматики позволяет нам дать следующее определение:

машинным пониманием текста называется процесс построения семантических траекторий, возбуждаемых текстом на графе онтологии, которая во внутренней памяти машины выполняет роль системы репрезентации знаний о внешнем мире.

Таким мы видим результат онтологического анализа ЕЯ тестов.

Ещё один полезный результат от введения онтологического смысла состоит в уточнении термина **связность текста**. Связный текст есть функция связанной мысли. Однако, будучи отделён от мысли, текст, особенно дискурс, имеет дискретную тематическую или сюжетную структуру. В литературоведении под связностью (например, повести) понимают возможность наблюдать за развитием выделенного сюжетного фрагмента в контексте полного сюжета, отслеживая определённую тему на фоне других. Связность дискурса реализуется авторскими приёмами, чаще всего через имена действующих лиц, место и время действия. На уровне формальной логики связность текста полностью достигается переносом

слова из предыдущего предложения в последующее, т.е. связность умозаключений является, в некотором смысле, синонимом выводимости.

Благодаря формализации смысла в системе знаний можно ввести конструктивное понятие **связность текста по онтологическому смыслу**. Два ключевых слова в тексте порождают две смысловые траектории, которые либо *сходятся* в одну точку – на вершине пирамиды, либо *пересекаются* на одном из трёх уровней онтологии **ИО*3**. Первый общий концепт этих траекторий говорит о том, **что** связывает два ключевых слова на концептуальном уровне. Именно этот общий концепт *именует ту ситуацию*, в которой ключи вступают в первое взаимодействие. Чем выше в пирамиде происходит замыкание траекторий, тем менее связным является изложение.

Краткие выводы. На уровне графических структур мы соединили семантику с прагматикой, «заставив» **машину понимать смысл**. «Понимание» и «смысл» соотносятся как синонимы. Связность текста устанавливается через пересечение траекторий смысла.

Заключение. ЗАЧЕМ НАМ НУЖЕН ОНТОЛОГИЧЕСКИЙ СМЫСЛ ?

Не отрицая ментальное происхождение смысла, которое следует из утверждения, что *«смысловым содержанием фраза наполняется в психическом процессе»* [Шевченко, 2003, с. 25], мы можем дополнить: **«смысловое содержание фразы выявляется в онтологическом процессе машинного понимания текста»**.

Какие последствия может иметь введение «онтологического смысла» для теории и практики?

В теоретическом плане – открыта научная методология глубинного анализа естественно-языковых текстов сложной семантики через синтез онтологического смысла. **Смысл становится вычисляемой семантической функцией текста и дискурса**. В поле знаний об искусственном интеллекте разработана цельная концепция: «текст – знания – смысл», а в инженерии знаний сделан шаг по **«превращению искусства в ремесло»**.

В практическом плане – извлечение из текста онтологического смысла позволяет разгрузить пользователя сети Интернет от рутинной работы по анализу потоков ненужной (нерелевантной запросу) информации. «Смысловой фильтр» не только «отсеет мусор», но представит текст в сжатом виде, сохранив только его смысл. Благодаря этому в системах поддержки принятия решений руководитель или менеджер получает возможность быстро вникнуть в суть проблемы – **не читая сообщение полностью, а просмотрев его «смысловой портрет»**.

В перспективе пользователь сможет построить собственную онтологическую сферу его интересов и в рамках **«персональной онтологии» знаний** оперировать с актуальной информацией на смысловом уровне.

Благодарности

Статья напечатана при финансовой поддержке проекта **ITHEA XXI** Института Информационных теорий и Приложений FOI ITHEA Болгария (www.ithea.org).

Литература

- [Тичер и др., 2009] С. Тичер, М. Мейер, Р. Водак, Е. Веттер. Методы анализа текста и дискурса * Пер. с англ. – Х.: Изд-во Гуманитарный Центр, 2009. – 356 с.
- [Штерн, 1998] Штерн І.Б. Вибрані топіки та лексикон сучасної лінгвістики. Енциклопедичний словник. – К.: «АртЕк», 1998. – 336 с.
- [Гаврилова, Червинская, 1992] Гаврилова Т.А., Червинская К.Р. Извлечение и структурирование знаний для экспертных систем. – М.: Радио и связь, 1992. – 200 с.

- [Одинцов, 1996] Одинцов Б.Е. Пресуппозиция в ассоциативных сетях и машинное понимание текстов. – Львов, 1996. – Препринт. Львовский госуниверситет им. И. Франко. – 60 с.
- [Рубашкин, 1989] Рубашкин В.Ш. Представление и анализ смысла в интеллектуальных информационных системах. – М.: Наука, 1989. – 192 с.
- [Поспелов, 1982] Поспелов Д.А. Фантазия или наука: на пути к искусственному интеллекту. – М.: Наука, 1982 – 224 с.
- [Балл, 2006] Балл Г.А. Психология в рациогуманистической перспективе: Избранные работы. – К.: Изд-во Основа, 2006. – 408 с.
- [Никитченко, Шкільняк, 2008] Нікітченко М.С., Шкільняк С.С. Математична логіка та теорія алгоритмів: підручник. – К.: Видавничо-поліграфічний центр «Київський університет», 2008. – 528 с.
- [Святогор, Гладун, 2009] Л. Святогор, В. Гладун. Определение понятия «Смысл» через онтологию. Семантический анализ текстов естественного языка. – International BookSeries, Number 9. IntelligentProcessing. Supplement to the International Journal "Information Technologies & Knowledge" Volume 3 / 2009. – ITHEA, Sofia, 2009. – p. 53.
- [Гладун и др., 2008] В. Гладун, В. Величко, Л. Святогор. Структурирование онтологии ассоциаций для конспектирования естественно-языковых текстов. – International BookSeries, Number 2. Advanced Research in Artificial Intelligence. Supplement to the International Journal "Information Technologies & Knowledge" Volume 2 / 2008. – ITHEA, Sofia, 2008. – p. 153.
- [Шевченко, 2003] Шевченко А.И. Актуальные проблемы теории искусственного интеллекта. – «Наука і освіта», 2003.
- [Мельчук, 1999] Мельчук И.А. Опыт теории лингвистических моделей «Смысл – Текст». – М.: Школа «Языки русской культуры», 1999. – 346 с.
- [Штерн, 1997] И. Штерн. Интродуктивные модели гуманитарных знаний: концептуальне гештальты versus понятия \ KDS–97. Шестая Международная конференция «Знания – Диалог – Решение». Сборник научных трудов в двух томах. Том 1. – Ялта, 1997. – Стр. 89.

Информация об авторах

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

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

ОБ АВТОМАТИЗАЦИИ ПРОЦЕССА ИЗВЛЕЧЕНИЯ ПЕРВИЧНЫХ ЗНАНИЙ ИЗ ЕСТЕСТВЕННО-ЯЗЫКОВЫХ ТЕКСТОВ

Александр Палагин, Сергей Крывый, Дмитрий Бибиков

Аннотация. *Предлагается формализация процесса извлечения знаний из естественно-языковых текстов, на основании которой строится автоматизированная система анализа определений, временных и событийных отношений, имеющих в текстах.*

Ключевые слова: *обработка естественно языковых текстов, онтологии, формализация*

ACM Classification Keywords: *H4m. Miscellaneous*

Введение

Бурное развитие науки и техники на протяжении последних десятилетий 20-го столетия и начала 21 столетия привела к огромному накоплению количества научно-технической информации. Информационный бум и лавинообразный рост информации приводит к сильному «зашумлению» несущественной информацией интересующей нас предметной области. В связи с этим одному человеку, даже очень высокой квалификации, не под силу освоить, понять и воспользоваться этой информацией с целью проведения научных исследований. Единственным видимым выходом из сложившейся ситуации является автоматизация процесса поиска и обработки необходимой информации. Для решения этой проблемы создаются технологии, ориентированные на смысловые структуры. Ведущей парадигмой структурирования информации на сегодняшний день являются онтологии или иерархические концептуальные структуры, представляющие собой модель предметной области, состоящей из иерархии понятий (концептов) предметной области, связей между ними и законов, действующих в рамках этой модели. Успешное решение проблемы автоматизации процесса построения такого типа структур зависит от успешного решения следующих проблем:

- проблема анализа естественно-языковой текстовой информации с целью извлечения знаний [1,5,6,7];
- проблема построения автоматизированной системы поиска и извлечения знаний, ее архитектуры и инструментария пользователя [1,2,6,7];
- проблема интеграции знаний из нескольких предметных областей с целью обеспечения эффективности проведения исследований междисциплинарного характера [1].

Третья из вышеприведенных проблем, в частности, тесно соприкасается с проблемой эффективного использования систем автоматического поиска доказательств теорем в формальных логических теориях и с подобными ей проблемами. Основу систем автоматического поиска доказательств составляют пруверы – программы, выполняющие доказательство. Успешное применение прувера будет только тогда, когда в его распоряжении имеется вся необходимая информация для успешного проведения доказательства. Например, если пруверу нужно доказать теорему Лагранжа о делимости порядка конечной группы на индекс ее подгруппы, то пруверу недостаточно иметь только аксиоматику теории групп. Ему также понадобится аксиоматика теории делимости, а может и аксиоматика Пеано вместе с некоторыми дополнительными фактами из других областей. Решение этой проблемы облегчается, если имеется интегрированная система, включающая в себя необходимые сведения из других областей знаний. Тогда прувер сам находит нужную ему информацию и использует ее для успешного проведения доказательства.

В данной работе предлагается некоторая формализация процесса анализа естественно-языковых текстов (ЕЯТ) и представления результатов этого анализа.

Формальная постановка проблемы

Процесс автоматизации какой-либо деятельности, как правило, требует формализованной постановки задачи, которая дает возможность выполнения анализа данной задачи с целью выработки метода ее решения. Когда речь идет об автоматизации процесса извлечения знаний из ЕЯТ и построения соответствующей онтологии, то необходимо определить понятия «знание» и «извлечение знания». С целью формализации вышеуказанных понятий, введем следующие определения, пользуясь нотацией констрейнтного программирования [3].

Пусть дано некоторое множество D , на котором определена конечная совокупность $R = \{R_1, \dots, R_k\}$ отношений $R_i \subseteq D^n, i = 1, 2, \dots, k$, конечной арности. Языком ограничений L на D называется непустое множество $L \subseteq R$. Проблема выполнимости ограничений из L формулируется следующим образом.

Для произвольного множества D и языка ограничений L на D проблемой выполнимости ограничений $CSP(L)$ является решение такой комбинаторной задачи:

дана тройка $P = (V, D, C)$, где

- $V = \{v_1, \dots, v_m\}$ - конечное множество переменных;

- $C = \{c_1, \dots, c_q\}$ - конечное множество ограничений, где ограничение c_i из C является парой (s_i, R_i) ,

где $s_i = (v_{i1}, \dots, v_{ij})$ - кортеж, состоящий из переменных, $R_i \in L - n_j$ -арное отношение на D ;

найти функцию $\varphi: V \rightarrow D$ такую, что $\forall (s_i, R_i) \in C$ кортеж $(\varphi(v_{i1}), \dots, \varphi(v_{ij})) \in R_i$ либо убедится в том, что её не существует, $i = 1, 2, \dots, k$. Множество D в этом случае называется областью проблемы, а функция φ называется интерпретацией $CSP(L)$.

В случае анализа ЕЯТ с целью извлечения знаний множество D , как область проблемы, интерпретируется как множество объектов, извлечённых из входного текста T , которое факторизовано по некоторому отношению эквивалентности R (это отношение будем называть отношением синонимии) и в котором «закодированы» отношения $R_i, i = 1, 2, \dots, k$. Множество переменных $V = \{v_1, v_2, \dots, v_m\}$ принимает значения в этом факторизованном множестве объектов, фигурирующих в тексте T (это могут быть лексико-грамматические разряды, конкретные объекты (люди, даты, предметы и т.п.)).

Проблемой извлечения знаний из ЕЯТ называется проблема поиска интерпретации $\varphi: V \rightarrow D$ с явным построением отношений R_i совокупности $L \subseteq R$. При этом, отношения $R_i \in L, i = 1, 2, \dots, k$, извлеченные из текста T , будем называть **знаниями**.

Приведенное определение достаточно общее и его необходимо конкретизировать. Конкретизация интерпретации и отношений в таком случае определяется целями, которые преследуются при анализе данного текста T . В качестве примеров конкретизации можно привести следующие.

А) Лексико-грамматический анализ приводит к конкретизации интерпретации $\varphi: V \rightarrow T$ и отношений $R_i \in L$. Интерпретация φ в данном случае представляется в виде суперпозиции двух функций φ_1 и φ_2 ,

т.е. $\varphi(V) = \varphi_2(\varphi_1(V)) = \varphi_1 * \varphi_2(V)$, где $*$ означает суперпозицию функций. Функции φ_1 и φ_2 реализуют процесс синтаксического и семантического анализа предложений текста T , а отношения R_1 и R_2 - это синтаксические ограничения (синтаксические правила языка, в котором представлен текст T) и семантические ограничения.

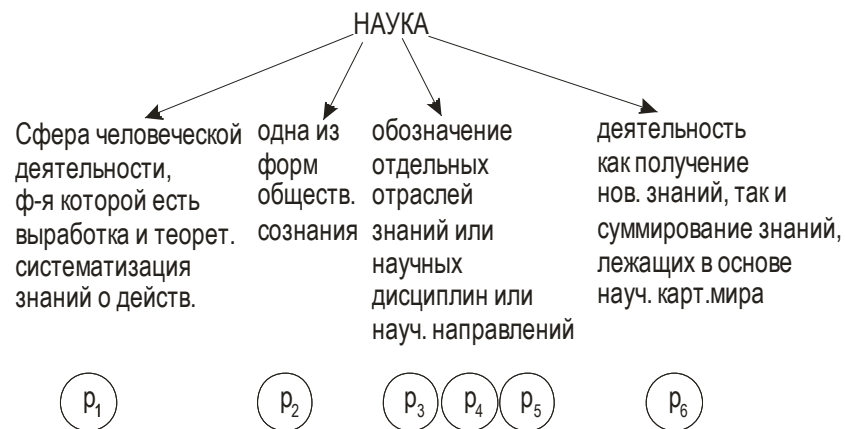
Функцию φ_1 тоже можно рассматривать как суперпозицию отображений φ_{11} и φ_{12} , которые реализуют соответственно морфологический и синтаксический анализ предложений ЕЯТ T и которые вместе с отображением φ_2 составляют классическую систему лексико-грамматического анализа [4].

Б) *Силлогистика Аристотеля* является другим примером уточнения интерпретации φ и отношений $R_i \in L$. В этом случае интерпретация φ носит теоретико-множественный характер, а отношения $R_i \in L$ - это отношение включения для множеств и его свойства. Более полное описание этого уточнения можно найти в работах [5,6].

В) *Текст библиографического характера* является примером хорошо структурированного текста. Это значит, что проблема извлечения знаний из такого текста решается относительно просто. Детальное описание этого процесса приведено в работе [5].

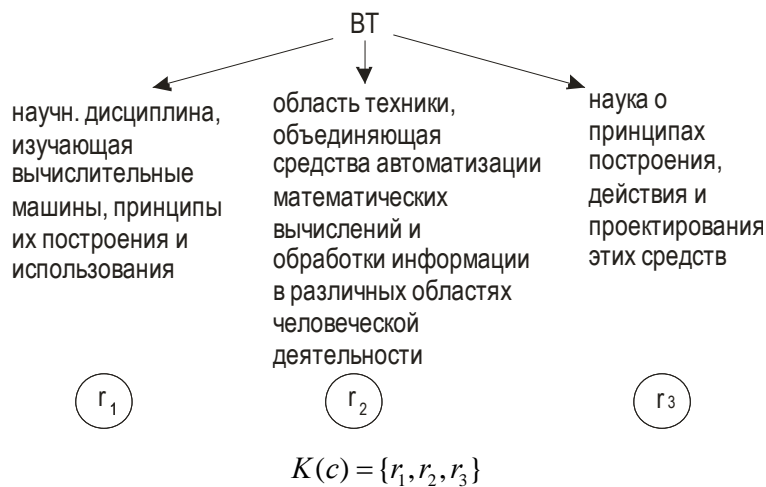
Обработка определений в ЕЯТ и их формальное представление

При анализе ЕЯТ первоочередным является обнаружение двух фундаментальных отношений, которые присутствуют практически в любом ЕЯТ. Это отношения эквивалентности и частичного порядка. Первое из этих отношений определяет классы синонимичных объектов, а второе отношение - иерархию подчиненности классов эквивалентности. Оба эти отношения составляют основу построения онтологий, а знания, полученные на этом этапе - будем называть первичными. В отношении частичного порядка может вкладываться различный семантический смысл: это может быть *отношение таксономии* («принадлежать» множеству, классу, группе и т. п.), *отношение партономии* («состоит из»), *отношение генеалогии* («отец-сын»), *причинно-следственное отношение* («если - то»), *атрибутивное отношение* и т. д. [8]. В качестве примера рассмотрим множество определений. В энциклопедическом словаре находим такое определение понятия «НАУКА».



$K(a) = \{p_1, p_2, p_3, p_4, p_5, p_6\}$ - класс эквивалентности.

Последующие понятия также взяты из энциклопедического словаря, где они определяются.



Приведенные примеры показывают, что построение классов эквивалентности не составляет особых трудностей. В результате построения классов появляются объекты:

$$K(a) = \{p_1, p_2, p_3, p_4, p_5, p_6\}, K(b) = \{q_1, q_2, q_3, q_4, q_5\}, K(c) = \{r_1, r_2, r_3\}.$$

Проблема появляется при вычислении второго отношения, которое определяет отношение подчиненности (иерархии) между имеющимися классами эквивалентности. Однако при таком абстрактном

представлении классов $K_i(x)$ это отношение определить нельзя. Для этого необходимо знать структурные характеристики элементов из классов $K_i(x)$. Поэтому естественным образом появляется необходимость в структуре элементов из классов эквивалентности. Например, если вернуться к вышерассмотренным примерам, то каждый элемент из класса $K(a)$ принимает вид:

$$p_1 = (p_{11}, p_{12}, p_{13}), p_2 = (p_{21}), p_3 = (p_{31}, p_{32}, p_{33}), p_4 = (p_{41}, p_{42}), p_5 = (p_{51}, p_{52}),$$

$$p_6 = (p_{61}, p_{62}), \text{ где}$$

p_{11} = «сфера человеческой деятельности», p_{12} = «выработка знаний о объективной действительности»,

p_{13} = «система знаний о объективной действительности», p_{21} = «форма общественного сознания»,

p_{31} = «отрасль знаний», p_{32} = «научная дисциплина», p_{33} = «научное направление»

p_{41} = «деятельность по получению новых знаний» p_{42} = «суммирование знаний о НКМ».

Аналогично структурируются и остальные элементы в классах эквивалентности.

$$r_1 = (r_{11}, r_{12}, r_{13}, \dots), r_2 = (r_{21}, r_{22}, r_{23}, \dots), r_3 = (r_{31}, r_{32}, r_{33}, \dots),$$

$$q_1 = (q_{11}, q_{12}, q_{13}, \dots), q_2 = (q_{21}, q_{22}, q_{23}, \dots), q_3 = (q_{31}, q_{32}, q_{33}, \dots), q_4 = (q_{41}, q_{42}, q_{43}, \dots)$$

$$q_5 = (q_{51}, q_{52}, q_{53}, \dots)$$

Из приведенной структуризации вытекает следующая формализация. Если класс эквивалентности относится к объекту a , то его формальное определение выглядит как дизъюнкция элементов, составляющих этот класс. Каждый элемент, входящий в тот или иной класс эквивалентности описывается соответствующим предикатом, т.е. если $K(a) = \{p_1, p_2, p_3, p_4, p_5, p_6\}$, то $p(a) \Leftrightarrow p_1(a) \vee \dots \vee p_6(a)$, где p_i -предикаты, характеризующие элементы из класса $K(a)$, а их дизъюнкция характеризует весь класс понятия a .

Далее, если $q_i \in K(a)$ и $q_i = (q_{i1}, q_{i2}, \dots, q_{ik})$, то элемент p_i (или объект p_i), характеризующийся атрибутами p_{ij} , представляется в виде конъюнкции вида

$$p_i(a) \Leftrightarrow p_{i1}(a) \wedge \dots \wedge p_{ik}(a),$$

где $p_{ij}(a)$ - предикат, характеризующий отдельный атрибут понятия a , $i = 1, \dots, l; j = 1, \dots, k$.

Следовательно, каждый класс $K(a)$ описывается дизъюнктивной формой вида

$$p(a) \Leftrightarrow (p_{11}(a) \wedge \dots \wedge p_{1m_1}(a)) \vee \dots \vee (p_{l1}(a) \wedge \dots \wedge p_{lm_l}(a)).$$

Введенная формализация определяет отношение частичного порядка, которое вводится следующим образом:

$$K(a) \leq K(b) \Leftrightarrow (\exists p_i(a))(\exists q_j(b))(q_j(b) \leq p_i(a)),$$

где $q_j(b) \leq p_i(a)$ означает, что $q_j(b)$ входит в виде конъюнктивного члена в $p_i(a)$.

Введенное таким образом отношения частичного порядка естественным образом требует предикатно-реляционного представления объектов из классов эквивалентности и самих этих классов [7]. Для

иллюстрации сказанного, вернемся к вышеприведенному примеру. Выясним, каким образом определяется тот факт, что классу «НАУКА» подчиняется класс «ИНФОРМАТИКА».

Класс «НАУКА», обозначенный как $K(a)$, описывается формулой

$$p(a) \Leftrightarrow p_1(a) \vee p_2(a) \vee p_3(a) \vee \dots \vee p_6(a), \text{ где}$$

$$p_1(a) \Leftrightarrow \text{СФЕРА-ЧЕЛОВЕЧ-ДЕЯТ}(a), \quad p_2(a) \Leftrightarrow \text{ОТРАСЛЬ ЗНАНИЙ}(a),$$

$$p_3(a) \Leftrightarrow \text{НАУЧНАЯ-ДИСЦИП}(a), \quad p_4(a) \Leftrightarrow \text{НАУЧН-НАПРАВЛ}(a),$$

$$p_5(a) \Leftrightarrow \text{ДЕЯТ-ПОЛУЧ-НОВЫХ-ЗН}(a), \quad p_6(a) \Leftrightarrow \text{ДЕЯТ-СУММИР-ЗН-НКТ}(a),$$

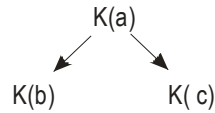
Класс «ИНФОРМАТИКА», обозначенный $K(b)$, описывается формулой

$$q(b) \Leftrightarrow q_1(b) \vee q_2(b) \vee q_3(b) \vee \dots \vee q_l(b),$$

где $q_1(b) \Leftrightarrow p_3(a) \wedge \text{ОБЩ-СВОЙ-НАУЧ-ИНФ}(b) \wedge \text{ЗАКОН-ПРОЦ-НАУЧ-ДЕЯТ}(b) \wedge \dots$

Используя определенное выше отношение частичного порядка, находим, что $p_3(a)$ входит в определение класса $K(a)$, причем $q_1(b) \vee p_3(b) = p_3(b)$ (в силу закона поглощения), а это значит, что $K(b) \leq K(a)$.

Аналогично определяется подчинение класса $K(c)$ классу $K(a)$, в результате чего получаем граф



Полученную таким образом иерархию можно изменять или модифицировать путем диалога с пользователем с целью достижения более правильного представления.

Подводя итог всему сказанному, вышеописанный процесс обработки текстов определений формализуется следующим образом.

Пусть T – множество текстов определений. По этому множеству T строится множество классов, которое определяется отношением эквивалентности R и составляет фактор множество D . На полученном таким образом множестве D определяются специальные отношения R_2, \dots, R_k , которые описывают характеристические свойства элементов из D , т.е. элементов из классов эквивалентности. Эти отношения представляются в виде предикатов, которые определяют отношение частичного порядка R_1 . Это отношение является вторым отношением, по которому строится онтология. Более точно, онтология строится по транзитивному замыканию R_1^* отношения R_1 , которое согласовано с отношением R .

Определение 1. Отношение R и R_1^* будем называть согласованными, если $\forall a, b \in D$ имеет место включение $(a, b) \in R^* R_1^*$, где $R^* R_1^*$ – суперпозиция отношений R и R_1^* .

Из этого определения естественным образом следует первичная онтология: $O = (D = T / R, \mathfrak{R} = \{R_1, R_2, \dots, R_k\}, \varphi, A)$, где $\varphi: D \rightarrow T$ – интерпретация, A – множество аксиом, которое определяется предикатами, описывающими характеристические свойства элементов из D , R_2, \dots, R_k – соответствующие им отношения, а R_1 – отношение частичного порядка.

Обработка ЕЯТ относительно временных и объектных отношений

Рассмотрим еще один пример конкретизации отношений R и R_1 , которые будем называть временными и объектными.

Отношение эквивалентности R определяется конкретным объектом (отсюда и название этого отношения), фигурирующим в тексте T (например, личность, театр, институт, вуз, кафедра и т.п.), а все объекты с ним связанные в тексте T , составляют класс эквивалентности по этому отношению. Отношение R_1 определяет связь объектов из классов эквивалентности отношения R либо в хронологическом порядке (временные зависимости), либо генеалогическом, либо каком-нибудь ином подобном порядке.

Отношение R и R_1 являются источником построения новых отношений. Например если R_1 описывает временные зависимости, то можно определить отношение, которое связывает объекты, относящиеся к данному конкретному моменту времени (год, месяц, день и т.п.). Поясним это примером.

Пусть $D = \{a, b, c, d\}$ - объекты, фигурирующие в данном тексте, который обрабатывается. Тогда $D/R = \{K(a), K(b), K(c), K(d)\}$ состоит из классов, элементами которых являются моменты времени t_{x_i} , связанные с объектом $x \in D$, $i = 1, \dots, j_x$. Допустим, что нас интересуют объекты, которые фигурируют в моменты времени t . Тогда момент времени t определяет отношение R_t , состоящее из объектов, связанных этим моментом времени. Арность этого отношения определяется мощностью множества D . Если $D = \{K(a), \dots, K(b)\}$, то $R_t = \{(a', \dots, b') \mid a' \in K(a) \wedge \dots \wedge b' \in K(b)\}$.

Далее, на отношениях R_t определяется отношения линейного порядка: $R_t \leq R_{t'} \Leftrightarrow t \leq t'$. Пользуясь этим отношением порядка, можно ввести следующее понятие.

Определение 2. Элементарным временным сценарием для временного интервала $[t_1, t_k]$ называется цепь $R_{t_1} \leq R_{t_2} \leq \dots \leq R_{t_k}$, где R_{t_i} - отношения, определенные выше.

Это определение можно модифицировать в зависимости от семантики отношений R и R_1 . Действительно, цепь отношений $R_{t_1} \leq R_{t_2} \leq \dots \leq R_{t_k}$ описывает хронологию некоторых событий во времени, связанных с конкретными людьми, объектами и т.п. Такого типа цепь в действительности может служить основой некоторого реального сценария.

Заключение

Описанные в данной работе способы автоматизации обработки ЕЯТ составляет основу как теоретического, так практического анализа извлечения знаний из ЕЯТ. Используя эту основу и прежде всего ее реализацию, предполагается наращивание ее мощности за счет построения новых мета отношений над построенными отношениями, являющимися отдельными частями знаний, имеющихся в исследуемом тексте.

Литература

1. Палагин А.В., Кривий С.Л., Петренко Н.Г., Знание ориентированные информационные системы с обработкой естественно-языковых объектов: основы методологии и архитектурно-структурная организация. – ж. УСИМ. – 2009. - №3. – С.42-55
2. Палагин А.В., Петренко Н.Г. Системно-онтологический анализ предметной области. – ж.УСИМ – 2009.№4.–С.3-14.

3. Cohen D. Jeavons P. The Complexity of Constraint Languages. In "Handbook of Constraint Programming . - Edited by F. Rossi, P. van Beek and T. Walsh. -2006. – P. 245 - 280.
4. Апресян Ю.Д. Лингвистический процессор для сложных информационных систем. -М.: Наука.-1992.- 324с.
5. Палагін О.В., Кривий С.Л., Петренко М.Г., Бібіков Д.С. Алгебро-логічний підхід до аналізу та обробки текстової інформації.-ж. «Проблемы программирования». - 2010. -№ 2.- (в печати).
6. Кулик Б.А. Логика естественных рассуждений.- С.-Петербург: Невский диалект.- 2001.- 127 с.
7. Рубашкин В.Ш. Представление и анализ смысла в информационных системах. – М.: Наука.- 1989.-188 с.
8. Gavrilova T., Laird D. Practical Design of Business Enterprise Ontologies. In Industrial Applications of Semantic Web. Eds. Bramer m., Terzyn V. - 2005. – Springer. – P. 61-81.

Информация об авторах

Александр Васильевич Палагин – академик НАН Украины, заместитель директора Института кибернетики им. В.М. Глушкова НАН Украины, Украина, Киев; проспект акад. Глушкова, 40
e-mail: palagin_a@ukr.net

Область научных интересов: Интеллектуальные информационные системы

Сергей Лукьянович Кривый – профессор Киевского национального университета имени Тараса Шевченко; Украина, Киев; ул. Владимирская, 40 e-mail: krivoi@i.com.ua

Область научных интересов: Дискретная математика, обработка знаний, анализ, верификация, оптимизация и проектирование программного обеспечения

Дмитрий Сергеевич Бибииков – аспирант Института кибернетики им. В.М. Глушкова НАН Украины; e-mail: bb_coff@mail.ru

Область научных интересов: Искусственный интеллект, автоматизация поиска доказательств в формальных логических языках

МЕТОДЫ АНАЛИЗА МАЛЫХ ВОЗМУЩЕНИЙ ЛИНЕЙНЫХ МОДЕЛЕЙ

Алексей Волошин, Григорий Кудин, Владимир Кудин

Аннотация. Предложено применение методов базисных матриц (МБМ), теории возмущения псевдообратных и проекционных матриц (ТПМ) для анализа влияний малых возмущений в линейных моделях.

Ключевые слова: линейные модели, количественный и качественный анализ, базисная матрица, псевдообращение матриц

ACM Classification Keywords: : I. Computing Methodologies – I.6. Simulation and modeling

Введение

подавляющее большинство технических, экономических процессов описываются в классе линейных моделей (ЛМ). Например, основная макроэкономическая модель, модель Леонтьева (МЛ) [Волошин, 2004] - в виде системы линейных алгебраических уравнений (СЛАУ) при ограничениях на переменные в виде гиперпараллелепипеда. Известно, что:

1. Наличие нечеткостей, неопределенностей значений параметров ЛМ [Кудин, 2008] предопределяет наличие в контуре принятия решения экспертов (ЛПР), которые должны качественно проанализировать структуру модели и указать механизм (процедуру) устранения неопределенностей при ее формировании.
2. Учет неточности представления ЛМ (так называемая проблема адекватности математической и машинной модели) предопределяет разработку механизма согласования результатов проведения вычислений при разной точности представления модели.
3. Некорректность модели при проведении вычислений в рамках ЛМ может существенно повлиять на качественные характеристики, например, величину ранга [Кудин, 2002].
4. Структурные связи в классе ЛМ с прямоугольной матрицей ограничений используются при построении современных методов [Кириченко 2007].

Цель исследования. Предложить новые методы исследования влияния малых изменений (возмущений) ЛМ, на точность решения, величину невязок на широком классе ЛМ.

Концепция анализа ЛМ предполагается двухстадийной. Первая стадия содержит анализ модели - "порождающей" (эталонной, математической, точной), а вторая стадия - анализ "возмущенной" (неточной, нечеткой, машинной).

Постановка задачи

Предметом исследования будет ЛМ:

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

где $A = \{a_{ij}\}_{i,j=1,m}$ матрица размерности $(n \times m)$, $a_j = (a_{j1}, a_{j2}, \dots, a_{jm})$, $j \in J = \{1, 2, \dots, n\}$, $I = \{1, 2, \dots, m\}$ - строки матрицы A , $u = (u_1, u_2, \dots, u_m)^T$, - вектор переменных, $C = (c_1, c_2, \dots, c_n)^T$, - вектор градиента ограничений модели.

Предполагаются заданными модели вида (1) с известными (или установленными) свойствами и слабо возмущенные (реальные вычисления с разным уровнем точности и учетом нечеткости задания

значений или малых возмущений). Это ставит, как начальную задачу, анализ непротиворечивости структурных элементов ЛМ (1), установление ее невырожденности (величины ранга) матрицы ограничений, направленной коррекции величины ранга матрицы ограничений изменением отдельных элементов модели, исследование свойств разрешимости, единственности или неединственности решений, а также дальнейшие задачи - анализа влияния малых возмущений на свойства ЛМ..

Основные положения метода базисных матриц (МБМ)

В МБМ введены в рассмотрение строчные базисные матрицы. Базисные матрицы в ходе итераций последовательно изменяются замещением строк вспомогательной СЛАУ строками (нормальными ограничениями) основной СЛАУ. В общем случае в исследуемой модели количество ограничений превышает количество переменных. В частности, для МЛ предполагается $m = n$ (для анализа вводится в рассмотрение вспомогательная СЛАУ с известными свойствами соответствующей размерности).

Определение 1. Квадратная матрица A_b , составленная с m линейно независимых нормалей ограничений (вспомогательной СЛАУ), будем называть базисной, а решение соответствующей ей системы уравнений $A_b u_0^T = c^0$ базисным. Две базисные матрицы с отличной одной строкой будем называть смежными.

Пусть $\beta_{ij}, i, j = 1, 2, \dots, m$ элементы базисной матрицы A_b , e_{ri} - элементы матрицы A_b^{-1} обратной к A_b ; $e_k = (A_b^{-1})_k$ - столбец обратной матрицы. Решение $u_0 = (u_{01}, u_{02}, \dots, u_{0m})^T$, системы уравнений $A_b u = c^0$, где c^0 - подвектор C , - компоненты которого состоят из правых частей ограничений, нормали, которых образуют базисную матрицу A_b ; $\alpha_r = (\alpha_{r1}, \alpha_{r2}, \dots, \alpha_{rm})$ - вектор развития нормали ограничения $a_r u_1 \leq c_r$ по строкам базисной матрицы A_b , $\Delta_r = a_r u_0^T - c_r$ - невязка r -го ограничения в вершине u_0 ; $J_b, J_H, J = J_b \cup J_H$ - множества индексов базисных и небазисных ограничений. Установлены формулы связи базисного решения, коэффициентов развития нормалей ограничений, коэффициентов обратной матрицы, невязок ограничений при переходе к базисной матрице \bar{A}_b (смежной), которая образовывается из матрицы A_b заменой ее строки a_k на a_l , что не входит в базисную матрицу A_b [Кудин, 2002]. При нахождении формул и основных соотношений между элементами метода при переходе от одной базисной матрицы к следующей считаем, что $a_{i1}, a_{i2}, \dots, a_{im}$ - нормали ограничений, $a_j u^T \leq c_j, j \in J_b$, где $J_b = \{i_1, i_2, \dots, i_m\}$ - индексы ограничений, нормали которых образуют строки базисной матрицы A_b , 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_b .

Теорема 1. Между элементами МБМ в смежных базисных матрицах имеют место соотношения

$$\bar{\alpha}_{rk} = \frac{\alpha_{rk}}{\alpha_{lk}}, \bar{\alpha}_{ri} = \alpha_{ri} - \frac{\alpha_{rk}}{\alpha_{lk}} \alpha_{li}, \quad r = \overline{0, n}; i = \overline{1, m}; i \neq k; \quad (2)$$

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

$$\bar{u}_{oj} = u_{oj} - \frac{e_{jk}}{\alpha_{1k}} \Delta_l, \quad j = \overline{1, m}, \quad (4)$$

$$\bar{\Delta}_k = -\frac{\Delta_l}{\alpha_{1k}}, \quad \bar{\Delta}_r = \Delta_r - \frac{\alpha_{rk}}{\alpha_{1k}} \Delta_l, \quad r = \overline{1, n}; \quad r \neq k; \quad (5)$$

причем условием невырожденности является $\alpha_{1k} \neq 0$, допустимости опорного базисного решения

$$\alpha_{1k} < 0 \quad (6)$$

Теорема 2. Для существования единственного решения (1) необходимо и достаточно, чтобы $\alpha_{1k}^{(i)} \neq 0$, $i = \overline{1, m}$, где $\alpha_{1k}^{(i)}$ ведущие элементы симплексной итерации МБМ по замещению строк базисной матрицы нормальями ограничений (1).

Следствие. Ранг матрицы ограничений системы (1) определяется количеством корректных замещений строк матрицы ограничений вспомогательной системы векторами нормальями (1), по формулам (2)-(6).

Теорема 3. Необходимым условием невырожденности новой \bar{A}_b образованной замещением нормали a_l , которая занимает k -ю строку в базисной матрице возмущенной нормалью $(a_l + a'_l)$ являются выполнения условия $\exists i e_{ik} \neq 0$, где e_{ik} элемент A_b^{-1} , такого, что $\alpha'_{li} \neq 0$. Решение будет

определяться соотношением $\bar{u}_{oj} = u_{oj} - \frac{e_{jk}}{\alpha_{1k}} \bar{\Delta}_l$, причем если $\bar{\Delta}_l = 0$, то $\bar{u}_{oj} = u_{oj}$, $j = \overline{1, m}$.

Следствие. Если ведущий элемент итерации $\alpha_{1k} \neq 0$, то при возмущении сохраняющем невырожденность должно выполняться $\alpha'_{li} + \alpha_{1k} \neq 0$.

Теорема дает условия направленного восстановления свойства невырожденности базисной матрицы (1) и анализировать влияние возмущений на решение.

Основные положения теории псевдообратных матриц

Определение 1. Для матрицы $A \in R^{m \times n}$ псевдообратную матрицу $A^+ \in R^{n \times m}$ можно определить согласно условию:

$$\forall b \in R^m, \quad A^+ b = \arg \min_{x \in \Omega_A(b)} \|x\|^2, \quad \text{где } \Omega_A(b) = \text{Arg} \min_{x \in R^n} \|Ax - b\|^2. \quad (7)$$

В практике применения псевдообращения важными являются проекционные матрицы, которые определяются и вычисляются с использованием матриц A и A^+ : проекционные матрицы

$$Z(A) = I_n - A^+ A, \quad Z(A^T) = I_m - A A^+ \quad (8)$$

- это проекторы на подпространство, ортогональное подпространству вектор - строк матрицы A и на подпространство, ортогональное подпространству вектор - столбцов матрицы A соответственно; взвешенные проекционные матрицы

$$R(A) = A^+ (A^+)^T, \quad R(A^T) = (A^+)^T A^+. \quad (9)$$

Если предположить, что расширение матрицы A происходит добавлением новой строки $a^T \in R^n$ после $(i-1)$ -ой строки ($i = \overline{2, m+1}$), т.е. образуется матрица

$$A_{i,a} = (a_{(1)} : \dots : a_{(i-1)} : a : a_{(i)} : \dots : a_{(m)})^T \in R^{(m+1) \times n}. \quad (10)$$

то при известной псевдообратной матрице $A^+ \in R^{n \times m}$ для рекуррентного вычисления псевдообратной матрицы $A_{i,a}^+ \in R^{n \times (m+1)}$ имеют место соотношения - прямые формулы Гревия.

Теорема 3. Если неизвестную псевдообратную матрицу $A_{i,a}^+$ представить в виде

$$A_{i,a}^+ = (p(1) : \dots : p(i-1) : p(i) : p(i+1) : \dots : p(m+1)) \in R^{n \times (m+1)}, \quad (11)$$

где

$$P_i = (p(1) : \dots : p(i-1) : p(i+1) : \dots : p(m+1)), \quad (12)$$

т.е. считать, что матрица $A_{i,a}^+$ может быть получена из матрицы $P_i \in R^{n \times m}$ добавлением после $(i-1)$ - го столбца вектора $p(i)$, то для матрицы P_i имеет место формула:

$$P_i = (1 - p(i)a^T)A^+. \quad (13)$$

При этом неизвестный вектора $p(i)$ определяется в зависимости от свойств линейной зависимости вектора a от векторов подпространства вектор – строк матрицы A :

1) если вектор a линейно независим от векторов подпространства вектор – строк матрицы A , т.е.

$$a^T Z(A)a > 0, \text{ то } p(i) = Z(A)a \|Z(A)a\|^{-2}; \quad (14)$$

2) если вектор a линейно зависим от векторов подпространства вектор – строк матрицы A , т.е.

$$a^T Z(A)a = 0, \text{ то } p(i) = R(A)a(1 + a^T R(A)a)^{-1} \quad (15)$$

Теорема 4. Если для матрицы $A_{i,a} \in R^{(m+1) \times n}$ (после $(i-1)$ - ой строки ($i = \overline{2, m+1}$) матрицы A расположена вектор – строка a^T), известна её псевдообратная матрица $A_{i,a}^+ \in R^{n \times (m+1)}$, то для определения псевдообратной матрицы $A^+ \in R^{n \times m}$ (строка a^T из матрицы $A_{i,a}$ удаляется) имеют место обратные формулы Гревия.

1) в случае линейной зависимости вектор – строки a^T , которая удаляется, от векторов подпространства вектор – строк матрицы $A_{i,a}$, а это определяется выполнением условия

$$a^T p(i) < 1 \quad (16)$$

псевдообратная матрица $A^+ \in R^{n \times m}$ имеет вид

$$A^+ = (I_n + p(i)a^T(1 - a^T p(i))^{-1})P_i \quad (17)$$

При этом ранг псевдообратной матрицы не изменяется, т.е. $rank A = rank(A^T : a)^T$.

2) в случае линейной независимости вектор – строки a^T , которая удаляется, от векторов подпространства вектор – строк матрицы A , а это определяется выполнением условия

$$a^T p(i) = 1, \quad (18)$$

матрица $A^+ \in R^{n \times m}$ определяется формулой

$$A^+ = (I_n - p(i)p^T(i) \|p(i)\|^{-2})P_i, \quad (19)$$

При этом ранг псевдообратной матрицы падает, то есть: $rank A = rank(A^T : a)^T - 1$.

Имеют место соотношения

$$\|p(i)\|^2 = R_{ii}(A_{i,a}^T), \quad (20)$$

$$1 - p^T(i)a = Z_{ii}(A_{i,a}^T). \quad (21)$$

Замечание. Прямые и обратные формулы Гревилля получены для проекционных, а также и для взвешенных проекционных матриц.

Следствие. Операция псевдообращения коммутативна с транспонированием, поэтому формулы Гревилля аналогично выписываются для варианта добавления (удаления) столбца матрицы.

Псевдообращение при возмущении элементов исходной матрицы. Пусть для матрицы

$$A = (a(1) \dots a(n)) \equiv (a_{(1)} \dots a_{(n)})^T \equiv (a_{ij})_{\substack{i=1, \dots, m \\ j=1, \dots, n}} \in R^{m \times n}. \quad (22)$$

где $a(j) = (\alpha_{1j}, \dots, \alpha_{ij}, \dots, \alpha_{mj})^T \in R^m$, $a_{(i)} = (\alpha_{i1}, \dots, \alpha_{ij}, \dots, \alpha_{in}) \in R^n$

известна её псевдообратная матрица

$$A^+ = (p(1) \dots p(m)) = (p_{(1)} \dots p_{(m)})^T \equiv (p_{ij})_{\substack{i=1, \dots, n \\ j=1, \dots, m}} \in R^{n \times m}, \quad (23)$$

$$p(j) = (p_{1j}, \dots, p_{ij}, \dots, p_{mj})^T \in R^n, \quad j = \overline{1, m}, \quad (24)$$

$$p_{(i)} = (p_{i1}, \dots, p_{ij}, \dots, p_{im})^T \in R^m, \quad i = \overline{1, n}. \quad (25)$$

Если элемент α_{ij} матрицы $A \in R^{m \times n}$ изменяется на некоторую величину $\delta_{ij} \in R$, т.е. она приобретает возмущённый вид

$$\tilde{A} = A + \Delta A(\delta_{ij}). \quad (26)$$

Возмущение $\Delta A(\delta_{ij})$ можно представить в матричном виде

$$\Delta A(\delta_{ij}) = \delta_{ij} e_m(i) e_n^T(j), \quad (27)$$

где векторы $e_n(j) \in R^n$, $e_m(i) \in R^m$ определены в соответствии обозначения

$$e_m(i) = (0 \dots 1 \dots 0)^T \in R^m. \quad (28)$$

Возникает задача - определить как влияет возмущение такого вида на элементы псевдообратной матрицы \tilde{A}^+

$$\Delta A^+(\delta_{ij}) = \tilde{A}^+ - A^+, \quad (29)$$

Теорема.5 Если для матрицы $A \in R^{m \times n}$ (формула (22)) известна её псевдообратная матрица $A^+ \in R^{n \times m}$ (формулы (23), (24)) и некоторый её элемент α_{ij} изменяется на величину $\delta_{ij} \in R$, т.е. имеют место формулы (25), (26), то формулы возмущений для псевдообратной матрицы имеют следующий вид:

1. если

$$a_{(i)}^T p(i) < 1, \quad p_{(j)}^T a(j) < 1, \quad (30)$$

т.е. когда векторы $e_i(m)$ и $e_j^T(n)$ линейно независимы от векторов – столбцов и

векторов - строк матрицы A соответственно:

$$\Delta A^+(\delta_{ij}) = -p(i)e_i^T(m)Z(A^T) - Z(A)e_j(n)p_{(j)}^T + Z(A)e_j(n)e_i^T(m)Z(A^T)\left(p_{ji} + \frac{1}{\delta_{ij}}\right), \quad (31)$$

2. если

$$a_{(i)}^T p(i) = 1, \quad p_{(j)}^T a(j) < 1, \quad (32)$$

т.е. когда вектор $e_i(m)$ линейно зависим от векторов - столбцов, а вектор $e_j^T(n)$ линейно независим от векторов - строк матрицы A :

$$\Delta A^+(\delta_{ij}) = -K(A^+, \delta_{ij}) A^+, \quad (33)$$

$$K(A^+, \delta_{ij}) = \frac{kk^T}{\|p(i)\|^2 \delta_{ij}^2 - 2p_{ij} \delta_{ij} + 1}, \quad (34)$$

$$k = p(i) \delta_{ij} - e_j(n); \quad (35)$$

3. если

$$a_{(i)}^T p(i) = 1, \quad p_{(j)}^T a(j) = 1 \quad (36)$$

и при этом

$$p_{ji} \delta_{ij} = -1, \quad (37)$$

т.е. когда векторы $e_i(m)$ и $e_j^T(n)$ линейно зависимы от векторов - столбцов и векторов - строк матрицы A соответственно, а ранг возмущённой матрицы $A + \delta_{ij} e_i(m) e_j^T(n)$ падает:

$$\text{rank}(A + \delta_{ij} e_i(m) e_j^T(n)) = \text{rank}(A) - 1. \quad (38)$$

Для этого случая

$$(A + \Delta A(\delta_{ij}))^+ = A^+ - \frac{p(i) p^T(i)}{\|p(i)\|^2} A^+ - A^+ \frac{p_{(j)} p_{(j)}^T}{\|p_{(j)}\|^2} + \alpha \frac{p(i)}{\|p(i)\|} \frac{p_{(j)}^T}{\|p_{(j)}\|}, \quad (39)$$

$$\alpha = \left(\frac{p_{(j)}^T}{\|p_{(j)}\|} A^{+T} \frac{p(i)}{\|p(i)\|} \right), \quad (40)$$

4. если

$$a_{(i)}^T p(i) = 1, \quad p_{(j)}^T a(j) = 1 \quad (41)$$

и при этом

$$p_{ji} \delta_{ij} \neq -1, \quad (42)$$

т.е. когда векторы $e_i(m)$ и $e_j^T(n)$ линейно зависимы от векторов - строк и векторов - столбцов матрицы A соответственно, а ранг возмущённой матрицы $A + \delta_{ij} e_i(m) e_j^T(n)$ не падает:

$$\text{rank}(A + \delta_{ij} e_i(m) e_j^T(n)) = \text{rank}(A). \quad (43)$$

Для этого случая

$$\Delta A^+(\delta_{ij}) = -\delta_{ij} \frac{p(i) p_{(j)}^T}{1 + p_{ji} \delta_{ij}}, \quad (44)$$

Замечание. Формулы возмущений получены также для проекционных и для взвешенных проекционных матриц..

Теорема 6. Невязка $\|Ax - b\|^2$ для СЛАУ $Ax = b$ определяется соотношением

$$\|Ax - b\|^2 = b^T Z(A^T)b. \quad (45)$$

Выводы

Совместное применение идеологий на основе МБМ и ПОМ при анализе ТПМ дает возможность:

1. контролировать или направлено изменять величину ранга матрицы ограничений системы;
2. проводить анализ свойств системы при изменении значений отдельных элементов и ее компонент;
3. использовать решения начальной системы при анализе возмущенной задачи;
4. анализировать корректность построения модели.

Благодарности

Статья частично финансирована из проекта ИТНЕА XXI Института Информационных теорий и Приложений FOI ИТНЕА и Консорциума FOIBulgaria (www.itea.org, www.foibg.com).

Литература

- [Волошин, 2004] Волошин О.Ф. Методи аналізу статичних балансових еколого-економічних моделей великої розмірності // Наукові записки, Т. VII, КПДВ "Педагогіка", Київ, 2004, - С. 43-55.
- [Кудин, 2008] Кудин В., Кудин Г., Волошин А. Анализ свойств модели Леонтьева при нечетко заданных параметрах с применением метода базисных матриц // Information Science & Computing, International Book Series, Number 7, Volume 7, ИТНЕА, SOFIA, p. 86-90, 2008.
- [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
- [Кириченко 2007] Кириченко Н.Ф., Донченко В.С. Псевдообращение в задачах кластеризации // Кибернетика и системный анализ.-2007.-№4.-С. 73-91.

Информация об авторах



Алексей Ф. Волошин – доктор технических наук, профессор, Киевский национальный университет имени Тараса Шевченко, Украина, 01017 Киев, ул. Владимирская, 64; e-mail: ovoloshin@unicyb.kiev.ua.



Григорий И. Кудин, Киев, Киевский национальный университет имени Тараса Шевченко, факультет кибернетики, Украина, к.ф.-м. н., зав. паб., E-mail: _Kudin@unicyb.kiev.ua

Владимир И. Кудин, Киев, Киевский национальный университет имени Тараса Шевченко, факультет кибернетики, Украина, д.т. н., с. н.с., E-mail: V.I.Kudin@mail.ru

МНОГОКРИТЕРИАЛЬНЫЙ МЕТОД РАСПРЕДЕЛЕНИЯ ОГРАНИЧЕННЫХ РЕСУРСОВ

Альберт Воронин

Аннотация. Рассмотрена проблема распределения заданного глобального ресурса системы при ограничениях, накладываемых на парциальные ресурсы. Показано, что проблема заключается в построении адекватной целевой функции для оптимизации процесса распределения ресурсов в условиях их ограниченности. Для решения рассматриваемой проблемы предпринимается подход многокритериальной оптимизации с применением нелинейной схемы компромиссов. Приведены модельные примеры.

Ключевые слова: многокритериальная оптимизация, распределение ресурсов, ограничения сверху и снизу, нелинейная схема компромиссов, принятие решений.

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

Содержание проблемы

В различных сферах управления и экономики актуальной является задача такого распределения ресурсов управляемой системы между отдельными элементами (объектами), при котором обеспечивается наиболее эффективное функционирование системы в заданных обстоятельствах. Часто эта задача решается субъективно, на основе опыта и профессиональной квалификации лица, принимающего решение (ЛПР). В простых случаях такой подход может оказаться оправданным. Однако при большом количестве объектов и в ответственных случаях резко возрастает цена ошибки управленческого решения. Становится необходимой разработка формализованных методов поддержки принятия решений для грамотного распределения ресурсов между объектами с учетом всех заданных обстоятельств.

Одним из таких обстоятельств обычно является ограниченность ресурсов. Наиболее распространен случай ограниченности сверху суммарного (глобального) ресурса системы, подлежащего распределению между отдельными объектами. В работе [1] рассматривается, в частности, задача перераспределения средств при уменьшении ранее запланированного объема финансирования проектов.

В практических случаях ограничения накладываются не только на глобальный ресурс, но и на парциальные ресурсы, выделяемые отдельным объектам. При этом ограничения могут накладываться как снизу, так и сверху. Такие ограничения или известны заранее, или определяются технико-экономическими расчетами или методами экспертных оценок. Следует различать ограничения условные (когда нарушение пределов нежелательно) и ограничения безусловные (когда их нарушение физически невозможно).

Пример 1. Для выполнения нескольких авиарейсов в разные города аэропорт располагает определенным ресурсом топлива, подлежащим распределению между самолетами. Для каждого рейса существует нижний предел, меньше которого выделять топливо бессмысленно, самолет просто не долетит до своего пункта назначения. В этом состоит суть ограничения снизу для каждого парциального ресурса. Если же данный рейс получает топливо сверх известного нижнего предела, то у него появляется возможность свободного маневрирования по эшелонам, обхода грозового фронта, ухода на запасной аэродром и т.п. С другой стороны, увеличивать парциальный ресурс неограниченно тоже нельзя, для него существует ограничение сверху. Это понятно хотя бы потому, что каждый самолет имеет определенную емкость

баков, больше которой принять топливо на борт он физически не может. Но обычно ограничение сверху вводится как условное и назначается полетным заданием. Учитывая заданный комплекс ограничений, требуется так распределить глобальный ресурс топлива между рейсами, чтобы обеспечивалась наиболее эффективная работа аэропорта в целом.

Пример 2. В проектно-конструкторскую организацию поступил заказ на разработку нескольких проектов. Для выполнения заказа обеспечен конкретный объем финансирования, подлежащий распределению между отдельными проектами. Для каждого проекта известен минимальный объем финансирования, меньше которого выполнение проекта невозможно. Обычно это защищенные статьи сметы – зарплата сотрудников, аренда помещений, коммунальные платежи, стоимость абсолютно необходимого оборудования и пр. Ясно, что при минимальном финансировании и качество выполнения проекта будет соответствующим. Увеличение средств делает разработку проекта более эффективной. Но повышать объем финансирования можно только до определенного предела, обусловленного полной сметной стоимостью проекта. Превышение этого предела называется нецелевым расходованием средств и грозит санкциями. Учитывая указанные ограничения снизу и сверху, требуется распределить глобальный объем финансирования между проектами так, чтобы работа проектно-конструкторской организации в целом была наиболее эффективной.

Нетрудно видеть, что сумма ограничений снизу для всех парциальных ресурсов представляет собой ограничение снизу для глобального ресурса, а сумма ограничений сверху ограничивает глобальный ресурс сверху.

Проблема заключается в построении адекватной целевой функции для оптимизации процесса распределения ресурсов в условиях их ограниченности. Простое равномерное распределение в данном случае не годится, так как может поставить некоторые объекты на грань невозможности их функционирования, в то время как другие объекты получают неоправданно большой ресурс.

В настоящей работе для решения рассматриваемой проблемы предпринимается подход многокритериальной оптимизации с применением нелинейной схемы компромиссов [2,3].

Постановка задачи

Поскольку рассматриваемая проблема актуальна для различных предметных областей, изложим постановку задачи в общем виде.

Задан подлежащий распределению глобальный ресурс R , а также $n \geq 2$ элементов системы (объектов), каждому из которых выделяется парциальный ресурс r_i , их совокупность составляет вектор

$r = \{r_i\}_{i=1}^n$. Формула для области определения этого вектора имеет вид

$$r \in X_r^\circ = \{r | 0 \leq r_i \leq R, i \in [1, n]\} \quad (1)$$

При этом выполняется условие

$$\sum_{i=1}^n r_i = R \quad (2)$$

Для каждого объекта известна (или определяется методом экспертных оценок) предельно допустимая величина выделяемого ресурса $B_{i \min}$, меньше которой данный объект функционировать не может.

Таким образом, задается система ограничений снизу

$$r_i \geq B_{i \min}, \sum_{i=1}^n B_{i \min} \leq R, i \in [1, n] \quad (3)$$

С другой стороны, для каждого объекта известна величина $B_{i \max}$, превышать которую ресурс объекта не может или не должен. Система ограничений сверху имеет вид

$$r_i \leq B_{i \max}, \sum_{i=1}^n B_{i \max} \geq R, i \in [1, n] \quad (4)$$

Из (3) и (4) следует, что

$$B_{i \max} \geq r_i \geq B_{i \min}, i \in [1, n] \quad (5)$$

и

$$\sum_{i=1}^n B_{i \max} \geq R \geq \sum_{i=1}^n B_{i \min} \quad (6)$$

С учетом (5) выражение (1) преобразуется к

$$r \in X_r = \{r | B_{i \max} \geq r_i \geq B_{i \min}, i \in [1, n]\} \quad (7)$$

Рассмотрим полярные (вырожденные) случаи неравенства (6). Если $R = \sum_{i=1}^n B_{i \min}$, то рассматриваемая задача сводится к такому распределению глобального ресурса, при котором каждый объект получает свой минимально допустимый парциальный ресурс: $r_i^* = B_{i \min}, i \in [1, n]$.

Если же глобальный ресурс позволяет полностью удовлетворять потребности объектов, т.е.

$$R = \sum_{i=1}^n B_{i \max}, \text{ то задача решается как } r_i^* = B_{i \max}, i \in [1, n].$$

Таким образом, в полярных случаях неравенства (6) рассматриваемая проблема имеет тривиальные решения. И только если выражение (6) становится **строгим** неравенством

$$\sum_{i=1}^n B_{i \max} > R > \sum_{i=1}^n B_{i \min} \quad (8)$$

задача оптимизации распределения ограниченных ресурсов приобретает смысл.

Ставится задача: в условиях (8) определить такие парциальные ресурсы $r^* \in X_r$, при которых выполняется требование (2) и приобретает экстремальное значение некоторая целевая функция $Y(r)$, вид которой следует выбрать и обосновать.

Метод решения

В задаче оптимизации распределения ограниченных ресурсов ограничение сверху $r_i \leq B_{i \max}, i \in [1, n]$ рассматривается как простое оптимизационное ограничение, приближение к которому обычно ничем особенным для системы не грозит. Совсем другой смысл имеет ограничение

снизу $r_i \geq B_{i \min}, i \in [1, n]$. Приближение ресурса к этому своему ограничению угрожает самой возможности функционирования соответствующего объекта. Можно сказать, что ограничение снизу является «критериеобразующим».

Поэтому выражение искомой целевой функции должно: 1) включать в себя ограничения снизу в явном виде, 2) штрафовать систему за приближение парциальных ресурсов к этим ограничениям и 3) быть дифференцируемым по своим аргументам. Простейшей целевой функцией, удовлетворяющей указанным требованиям, является

$$Y(r) = \sum_{i=1}^n B_{i \min} (r_i - B_{i \min})^{-1} \quad (9)$$

Анализ формулы (9) показывает, что это не что иное, как выражение скалярной свертки максимизируемых частных критериев $r_i, i \in [1, n]$ по нелинейной схеме компромиссов (НСК) в задаче многокритериальной оптимизации [2]. Действительно, в рассматриваемой задаче ресурсы $r_i, i \in [1, n]$ имеют двоякую природу. С одной стороны, их можно рассматривать как независимые переменные, *аргументы оптимизации* целевой функции $Y(r)$. С другой стороны, для каждого из объектов логично стремление максимизировать свой парциальный ресурс, уйти как можно дальше от опасного ограничения $B_{i \min}$ для повышения эффективности своего функционирования. С этой точки зрения ресурсы $r_i \geq B_{i \min}, i \in [1, n]$ могут рассматриваться как частные *критерии* качества функционирования соответствующих объектов. Эти критерии подлежат максимизации, они ограничены снизу, неотрицательны и противоречивы (увеличение одного ресурса возможно только за счет уменьшения других).

Концепция НСК основана на принципе «подальше от ограничений». Предполагается, что функция полезности ЛПР оценивает как предпочтительные те решения, которые дают большее удаление критериев от опасных ограничений. Скалярная свертка $Y(r)$ представляет собой модель функции полезности и включает в себя в явном виде разность $r_i - B_{i \min}$ как характеристику напряженности ситуации принятия решения. Это позволяет штрафовать критерии за приближение к своим ограничениям. На основании изложенного, задача векторной оптимизации распределения ограниченных ресурсов с учетом изопериметрического ограничения для аргументов приобретает вид

$$r^* = \arg \min_{r \in X_r} Y(r) = \arg \min_{r \in X_r} \sum_{i=1}^n B_{i \min} (r_i - B_{i \min})^{-1}, \sum_{i=1}^n r_i = R \quad (10)$$

Задачу (10) можно решать как аналитически, используя метод неопределенных множителей Лагранжа, так и численными методами, если аналитическое решение оказывается затруднительным.

Аналитическое решение предусматривает построение функции Лагранжа в виде

$$L(r, \lambda) = Y(r) + \lambda \left(\sum_{i=1}^n r_i - R \right),$$

где λ – неопределенный множитель Лагранжа, и решение системы уравнений

$$\frac{\partial L(r, \lambda)}{\partial r_i} = 0, i \in [1, n]$$

$$\frac{\partial L(r, \lambda)}{\partial \lambda} = \sum_{i=1}^n r_i - R = 0$$

Для решения многокритериальных задач численными методами с применением концепции НСК и с ограничениями на аргументы и критерии разработаны алгоритмы и составлена компьютерная программа [2]. Для ее использования в практических расчетах необходимо привести исходную многокритериальную задачу к каноническому виду.

Механизм НСК в данной программе разработан для минимизируемых критериев. Чтобы использовать его в нашем случае, необходимо применить монотонное (теорема Гермейера) преобразование, приводящее максимизируемые критерии $r_i, i \in [1, n]$ к виду минимизируемых. Таким преобразованием может быть

$$y_i(r) = 1/r_i, A_i = 1/B_i, i \in [1, n].$$

Понятно, что критерии $y_i(r)$ ограничены сверху величинами $A_i, i \in [1, n]$.

Тогда целевая функция приобретает канонический вид скалярной свертки по нелинейной схеме компромиссов в унифицированной форме для минимизируемых критериев [2]:

$$Z(r) = \sum_{i=1}^n A_i [A_i - y_i(r)]^{-1}.$$

Скалярная свертка выступает как инструмент композиции критериев для различных альтернатив $r \in X_r$ и задача векторной оптимизации распределения ограниченных ресурсов с учетом изопериметрического ограничения на аргументы приобретает вид

$$r^* = \arg \min_{r \in X_r} Z(r) = \arg \min_{r \in X_r} \sum_{i=1}^n A_i [A_i - y_i(r)]^{-1}, \sum_{i=1}^n r_i - R = 0.$$

Эта задача может быть решена посредством универсальной программы многокритериальной оптимизации. Рассмотрим вкратце некоторые ее возможности.

Программа векторной оптимизации

Для решения широкого спектра оптимизационных задач разработана и изложена в работе [2] программа векторной оптимизации TURBO-OPTIM. Программа выполнена на языке BorlandC++3.1 с использованием библиотеки TurboVision, что обеспечивает эффективное использование ресурсов ЭВМ, стандартизованную и удобную среду для пользователя, простоту модификации и отладки.

Для работы с программой необходимо выполнить следующие **этапы**:

- выделить набор частных критериев так, чтобы все они принимали неотрицательные значения и требовали *минимизации*;
- определить допустимое предельное значение для каждого критерия;
- выделить набор параметров (независимых переменных), от которых зависят частные критерии;

- определить диапазон изменения каждого параметра (минимальное, стартовое и максимальное значения);
- задать ограничения для параметров, имеющие вид неравенств $g_j(r) \leq 0, j \in [1, k]$, где k – количество ограничений;
- определить вид зависимости частных критериев от параметров.

Программа позволяет решать задачи оптимизации для следующих случаев связи частных критериев с аргументами оптимизации (параметрами):

- критерии выражаются через параметры явно, известны аналитические зависимости;
- критерии являются некоторыми функционалами и для расчета их значений требуется решение системы дифференциальных уравнений;
- зависимости критериев от параметров не известны и для определения значений параметров необходимо проведение экспериментов;
- значения критериев можно получить, выполнив написанную пользователем программу;
- есть таблица зависимости частных критериев от параметров.

В каждом из перечисленных случаев программа представляет пользователю средства нахождения минимума обобщенного критерия, построенного по нелинейной схеме компромиссов, одним из методов оптимизации: 1) метод симплекс-планирования в модификации Нелдера-Мида и 2) нелокальный метод нелинейного программирования (дуальный метод оптимизации).

Иллюстрационные примеры

1. Для выполнения двух рейсов ($n=2$) аэропорт располагает топливом общим объемом $R=12$ тонн (цифры условные). Минимальная потребность первого рейса составляет $r_1 \geq B_{1\min} = 2$ тонны, второго рейса $r_2 \geq B_{2\min} = 5$ тонн. Это ограничения снизу для парциальных ресурсов. Емкость баков первого самолета $B_{1\max} = 7$ тонн, второго самолета $B_{2\max} = 10$ тонн. Это ограничения сверху.

Условие (8) в виде строгого неравенства (размерности опущены)

$$B_{1\min} + B_{2\min} = 7 < R = 12 < B_{1\max} + B_{2\max} = 17$$

соблюдается. Значит, задача оптимизации распределения ограниченных ресурсов может быть поставлена и решение будет нетривиальным.

Ставится задача: получить аналитическое решение компромиссно-оптимального распределения топлива между рейсами.

Строим функцию Лагранжа

$$L(r, \lambda) = B_{1\min} (r_1 - B_{1\min})^{-1} + B_{2\min} (r_2 - B_{2\min})^{-1} + \lambda (r_1 + r_2 - R).$$

Получаем систему уравнений

$$\begin{aligned}\frac{\partial L(r, \lambda)}{\partial r_1} &= -B_{1\min} (r_1 - B_{1\min})^{-2} + \lambda = 0 \\ \frac{\partial L(r, \lambda)}{\partial r_2} &= -B_{2\min} (r_2 - B_{2\min})^{-2} + \lambda = 0 \\ r_1 + r_2 - R &= 0\end{aligned}$$

Подставляя числовые данные

$$\begin{aligned}-2(r_1 - 2)^{-2} + \lambda &= 0 \\ -5(r_2 - 5)^{-2} + \lambda &= 0 \\ r_1 + r_2 - 12 &= 0\end{aligned}$$

и решая эту систему методом Гаусса (последовательного исключения переменных), получаем

$$r_1^* = 3,94 \text{ тонн}, r_2^* = 8,06 \text{ тонн}.$$

Поставленная задача решена в предположении, что относительная важность обоих рейсов для ЛПП одинакова. Если же нет, то в целевую функцию вводятся весовые коэффициенты α_1 и α_2 , отражающие индивидуальные предпочтения ЛПП. Эти коэффициенты нормированы и определены на симплексе:

$$\alpha_1, \alpha_2 \in X_\alpha = \left\{ \alpha_i \mid \alpha_i \geq 0, \sum_{i=1}^{n=2} \alpha_i = 1, i \in [1;2] \right\}.$$

2. В конструкторское бюро поступил заказ на проектирование и изготовление полунатурных макетов самолетов трех видов ($n=3$): 1) пассажирский лайнер, 2) транспортный самолет, 3) спортивно-тренировочный самолет. Для выполнения заказа обеспечено финансирование общим объемом $R = 10$ миллионов гривен (здесь и далее цифры условные).

Рассчитана полная финансовая смета для каждого проекта (ограничения сверху):

$$r_1 \leq B_{1\max} = 7 \text{ млн грн}; r_2 \leq B_{2\max} = 5 \text{ млн грн}; r_3 \leq B_{3\max} = 4 \text{ млн грн}.$$

Экономическими расчетами определены минимальные объемы финансирования отдельных проектов, ниже которых проектирование невозможно (ограничения снизу):

$$r_1 \geq B_{1\min} = 2 \text{ млн грн}; r_2 \geq B_{2\min} = 1 \text{ млн грн}; r_3 \geq B_{3\min} = 0,5 \text{ млн грн}.$$

Условие (8) представляет собой строгое неравенство (размерности опущены)

$$\sum_{i=1}^n B_{i\max} = 16 > R = 10 > \sum_{i=1}^n B_{i\min} = 3,5,$$

поэтому изложенная методика может быть применена для оптимизации распределения ограниченных ресурсов.

Ставится задача: используя программу векторной оптимизации TURBO-OPTIM, найти компромиссно-оптимальные значения парциальных объемов финансирования r_1^* , r_2^* и r_3^* для проектирования и

изготовления полунатурных макетов пассажирского лайнера, транспортного самолета и спортивного самолета соответственно.

В соответствии с этапами работы с программой, устанавливаем: режим «аналитика», метод оптимизации «симплекс-планирование» (по умолчанию) и далее вводим числовые данные (размерности опущены):

$$r_{1\min} = B_{1\min} = 2; r_{1\text{start}} = 3; r_{1\max} = B_{1\max} = 7$$

$$r_{2\min} = B_{2\min} = 1; r_{2\text{start}} = 3; r_{2\max} = B_{2\max} = 5$$

$$r_{3\min} = B_{3\min} = 0.5; r_{3\text{start}} = 3; r_{3\max} = B_{3\max} = 4$$

$$r_1 + r_2 + r_3 - 10 = 0$$

$$y_1 = 1/r_1; y_2 = 1/r_2; y_3 = 1/r_3$$

$$y_{1\max} = A_1 = \frac{1}{B_{1\min}} = 0.5; y_{2\max} = A_2 = \frac{1}{B_{2\min}} = 1; y_{3\max} = A_3 = \frac{1}{B_{3\min}} = 2$$

После этого даем команду «выполнить» и программа определяет искомые значения парциальных объемов финансирования по проектам:

$$r_1^* = 4,945 \text{ млн грн}; r_2^* = 3,083 \text{ млн грн}; r_3^* = 1,972 \text{ млн грн}.$$

Полученный результат соответствует унифицированной версии свертки по нелинейной схеме компромиссов, которая применяется для широкого использования. Если желательно учесть индивидуальные предпочтения ЛПП, то программа содержит соответствующую опцию.

Благодарности

Статья частично финансирована из проекта **ITHEA XXI** Института Информационных теорий и приложений FOI ITHEA и Консорциума FOIBulgaria (www.ithea.org, www.foibg.com).

Библиография

1. Воронин А.Н. Векторная оптимизация иерархических структур // Проблемы управления и информатики. – 2004. – № 6. – С. 26-34.
2. Воронин А.Н., Зиятдинов Ю.К., Козлов А.И. Векторная оптимизация динамических систем. – Киев: Техніка, 1999. – 284 с.
3. Воронин А.Н. Нелинейная схема компромиссов в многокритериальных задачах оценивания и оптимизации // Кибернетика и системный анализ. – 2009. – № 4. – С. 106-114.

Сведения об авторе

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

ПРИМЕНЕНИЕ МЕТОДА «МАСКА» ДЛЯ ГРУППОВОЙ ЭКСПЕРТНОЙ КЛАССИФИКАЦИИ НАУЧНЫХ ПРОЕКТОВ ПО МНОГИМ КРИТЕРИЯМ

Петровский Алексей Борисович, Ройзензон Григорий Владимирович,
Тихонов Игорь Петрович, Балышев Александр Владимирович

Аннотация: В работе излагается новый метод «МАСКА» групповой классификации многопризнаковых объектов, описываемых качественными признаками. Метод основан на теории мультимножеств. Разработаны алгоритмы для построения согласованных правил отнесения объектов к классам решений и выделения противоречиво классифицированных объектов. Рассмотрен пример применения метода для конкурсного отбора научных проектов в Российском фонде фундаментальных исследований.

Ключевые слова: групповой вербальный анализ решений, многокритериальная порядковая классификация, агрегирование признаков, интегральные показатели оценки, результативность научных проектов

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

Введение

Многокритериальная классификация рассматриваемых объектов по их свойствам, выраженных признаками (атрибутами), в индивидуальном принятии решений основывается на субъективных предпочтениях единственного лица, принимающего решение (ЛПР). При коллективном принятии решений необходимо учитывать одновременно различные интересы многих участников, разнообразие и несовпадение их целей и способов выражения их предпочтений. Примером такой задачи служит конкурсный отбор научных проектов, оцененных несколькими экспертами по многим качественным критериям, в котором необходимо разделить проекты на группы, в той или иной степени отвечающие целям конкурса и им не соответствующие.

Известно достаточно много подходов к индивидуальной классификации объектов, характеризующихся многими количественными и/или качественными признаками [Ларичев, 2006], [Doumpos, Zorounidis, 2002], [Greco et al., 2002], [Köksalan et al., 2003], [Roy, Bouyssou, 1993]. Методы групповой многокритериальной классификации в настоящее время менее развиты [Петровский, 2003], [Petrovsky, 2001]. Главные трудности при коллективном принятии решений связаны с необходимостью учета большого числа вербальных и количественных данных и обработкой этих данных, не прибегая к дополнительным преобразованиям типа усреднения, смешивания, взвешивания, которые могут привести к необоснованным и необратимым искажениям исходной информации. Воспользовавшись аппаратом теории мультимножеств для представления многопризнаковых объектов, можно преодолеть указанные выше трудности.

В работе рассматривается модификация метода МАСКА (МногоАспектная Согласованная Классификация Альтернатив) для групповой классификации многопризнаковых объектов, которые характеризуются многими разнородными (количественными и качественными) признаками и могут существовать в нескольких экземплярах с отличающимися, в частности, противоречивыми значениями признаков [Петровский, 2003], [Petrovsky, 2001].

Метод базируется на теории метрических пространств мультимножеств и позволяет строить согласованное групповое правило для классификации объектов, которое аппроксимирует

индивидуальные правила экспертной сортировки объектов. Этот метод допускает использование различных, в том числе и противоречивых, данных для описания объектов и позволяет находить несогласованные индивидуальные правила классификации.

Предложенный подход к построению согласованного решающего правила для классификации многопризнаковых объектов был апробирован на базе данных конкурсов целевых фундаментальных исследований, выполняемым по грантам Российского фонда фундаментальных исследований (РФФИ) в интересах Федеральных агентств и ведомств России [Петровский, Тихонов, 2009]. Для различных научных направлений были построены групповые решающие правила классификации, основанные на многокритериальных оценках проектов многими экспертами, которые позволили выделить наиболее важные для отбора проектов критерии и выявить расхождения в индивидуальных правилах сортировки проектов, применявшихся экспертами.

Конкурсный отбор научных проектов

По установленному в РФФИ порядку экспертиза научных проектов, в том числе целевых фундаментальных исследований, производится при конкурсном отборе заявок. Оценку заявок выполняют эксперты, работающие в различных научно-исследовательских институтах, высших учебных заведениях, научно-производственных организациях. Каждый проект независимо оценивается несколькими экспертами, как правило, тремя без согласования их мнений. Для оценки содержания заявки используются специальные анкеты, которые содержат критерии, имеющие словесные шкалы оценок с развернутыми формулировками градаций качества. По каждому критерию эксперт выбирает только одну из оценок (такая экспертиза называется закрытой). Кроме того, эксперт дает свое заключение о целесообразности поддержки проекта. На основе заключений экспертов Экспертный совет РФФИ принимает решение о принятии проекта и объемах его финансирования.

Экспертная оценка содержания целевых фундаментальных исследований проводится по 12 качественным критериям, представленным в экспертной анкете и объединенным в группы: Научная характеристика проекта; Оценка возможностей практической реализации проекта.

В группу «Научная характеристика проекта» входят 9 критериев:

Q_{11} . Уровень фундаментальности проекта; Q_{12} . Ориентированность результата; Q_{13} . Цели исследования; Q_{14} . Методы достижения цели проекта; Q_{15} . Характер исследований; Q_{16} . Научная значимость проекта; Q_{17} . Степень новизны предлагаемых решений; Q_{18} . Потенциал исполнителей; Q_{19} . Техническая оснащенность.

Группа «Оценка возможностей практической реализации проекта» состоит из 2 критериев:

Q_{21} . Завершающая стадия фундаментальных исследований, предлагаемых в проекте;
 Q_{22} . Масштабы применимости результатов заявленных исследований.

Каждый критерий имеет порядковую или номинальную шкалу оценок с развернутыми словесными формулировками градаций качества. Например, шкала критерия Q_{17} . «Степень новизны предлагаемых решений» выглядит следующим образом:

q_{17}^1 - решения сформулированы впервые и существенно превосходят уровень имеющихся решений;
 q_{17}^2 - решение находится на уровне имеющихся решений;
 q_{17}^3 - решение уступает имеющимся решениям.

По каждому проекту каждый эксперт давал также свое заключение о поддержке проекта:

r_1 - безусловно поддержать (оценка «5»),

r_2 - целесообразно поддержать (оценка «4»),

r_3 - поддержка возможна (оценка «3»),

r_4 - не поддерживать (оценка «2»).

Совокупность экспертных заключений можно считать еще одним критерием R , характеризующим проекты. Таким образом, каждый проект представляет собой объект, описываемый многими качественными (вербальными) признаками.

Принятие того или иного заключения из множества R может рассматриваться как решение задачи классификации рассматриваемых объектов по их свойствам, выраженных признаками (атрибутами) Q_{11} – Q_{22} , которая состоит в разбиении множества объектов на несколько заранее заданных упорядоченных групп (классов решений) D_1 – D_4 . Рекомендации экспертов являются индивидуальными правилами предварительной классификации (сортировки) рассматриваемых проектов, которые могут совпадать или различаться.

Для отбора проектов нужно построить агрегированное групповое правило, которое обобщает индивидуальные решения всех экспертов и позволяет отнести проект в один из заданных классов. Это решающее правило должно формулироваться на естественном языке и учитывать экспертные оценки заявок по многим критериям. Построение такого решающего правила в при наличии многих несоответствующих индивидуальных заключений экспертов относится к числу сложных задач многокритериального принятия решений. Вместе с тем наличие такого решающего правила позволяет формировать разные стратегии конкурсного отбора проектов, проводить анализ качества работы экспертов, выявлять различные точки зрения.

Для построения решающих правил конкурсного отбора целевых фундаментальных исследований использована оригинальная методология группового вербального анализа решений, которая успешно применялась при решении различных практических задач [Ларичев, 2006], [Петровский, 2003], [Петровский, 2009]. В вербальном анализе решений рассматриваемые варианты и классы решений описываются с помощью качественных критериев, имеющих словесные формулировки градаций на шкалах оценок. Обычно вводится небольшое число таких градаций (3-5), чтобы обеспечить ясную различимость соседних оценок. Именно такой способ формализации экспертных оценок проектов принят в РФФИ.

В вербальном анализе решений численные коэффициенты важности критериев и ценности вариантов не рассчитываются и не применяются, а качественные оценки признаков не преобразуются в какие-либо числовые показатели. Даже при небольшом числе градаций оценок на шкалах критериев оказывается возможным описать достаточно сложные свойства объектов. Тем самым, используя только качественные измерения, на множестве кортежей многокритериальных оценок можно задать отношения превосходства и эквивалентности вариантов решения, с помощью которых осуществляется их порядковая классификация, частичное упорядочение или выделение лучшего варианта.

Вербальный анализ решений отличается активным участием ЛПР в анализе и решении стоящей проблемы, позволяя разносторонне и достаточно подробно выражать предпочтения ЛПР, уточнять и корректировать их в ходе решения задачи, генерировать и обосновывать новые варианты решения. Методы группового вербального анализа решений дают возможность при решении задач коллективного выбора одновременно учитывать различные интересы многих участников, разнообразие и несоответствие их целей и способов выражения их предпочтений. В целом вербальные методы более «прозрачны», мало чувствительны к ошибкам измерения и менее трудоемки для человека.

Представление многопризнаковых объектов

Многопризнаковые объекты $A_i, i=1, \dots, n$ обычно принято представлять как векторы или кортежи $q_i=(q_{i1}^{e_1}, \dots, q_{im}^{e_m})$ в пространстве $Q=Q_1 \times \dots \times Q_m$, где $Q_s=\{q_s^{e_s}\}$ – непрерывная или дискретная шкала s -го признака, $e_s=1, \dots, h_s, s=1, \dots, m$. Ситуация существенно усложняется, если одному и тому же объекту A_i может соответствовать не один, а несколько m -мерных векторов. В таких случаях объект A_i представляется в пространстве Q группой, состоящей из k векторов $\{q_i^{(1)}, \dots, q_i^{(k)}\}$ вида $q_i^{(j)}=(q_{i1}^{e_1(j)}, \dots, q_{im}^{e_m(j)})$, $j=1, \dots, k$, которая должна рассматриваться как единое целое. Измеренные разными способами значения параметров или индивидуальные оценки экспертов, могут быть похожими, различающимися и даже противоречивыми, что может приводить к несравнимости m -мерных векторов $q_i^{(j)}=(q_{i1}^{e_1(j)}, \dots, q_{im}^{e_m(j)})$, характеризующих один и тот же объект A_i . Совокупность таких многомерных объектов может иметь в пространстве Q сложную структуру, достаточно трудную для анализа.

Эти трудности можно преодолеть, если воспользоваться иным способом представления многопризнаковых объектов, основанным на формализме мультимножеств, который позволяет одновременно учесть все комбинации значений количественных и качественных признаков, а также различное число значений каждого из этих признаков. Введем вместо прямого произведения m шкал признаков $Q=Q_1 \times \dots \times Q_m$ множество $G=Q_1 \cup \dots \cup Q_m$ – обобщенную шкалу признаков, состоящую из m групп признаков, и представим объект в таком символическом виде:

$$A_i = \{k_{A_i}(q_1^{e_1}) \circ q_1^{e_1}, \dots, k_{A_i}(q_1^{h_1}) \circ q_1^{h_1}, \dots, k_{A_i}(q_m^1) \circ q_m^1, \dots, k_{A_i}(q_m^{h_m}) \circ q_m^{h_m}\}, \quad (1)$$

где число $k_{A_i}(q_s^{e_s})$ указывает, сколько раз признак $q_s^{e_s} \in Q_s$ встречается в описании объекта A_i , а знак \circ обозначает кратность признака $q_s^{e_s}$. Множество G характеризует свойства совокупности объектов $A=\{A_1, \dots, A_n\}$. Такая запись объекта A_i представляет его как мультимножество или множество с повторяющимися элементами [Петровский, 2003], [Petrovsky, 2001].

Определяются следующие операции над мультимножествами: объединение $A \cup B$, пересечение $A \cap B$, сложение $A + B$, вычитание $A - B$, симметрическая разность $A \Delta B$, дополнение $\bar{A} = Z - A$, умножение на целое число $c \cdot A, c > 0$, арифметическое умножение $A \cdot B$, арифметическая n -ая степень A^n , прямое произведение $A \times B$, прямая n -ая степень $(\times A)^n, Z$ – максимальное мультимножество с $k_Z(x) = \max_{A \in A} k_A(x)$.

Новые типы операций над мультимножествами открывают новые возможности для агрегирования многопризнаковых объектов. Например, группа X_a объектов может быть получена как сумма $X_a = \sum_i A_i$, объединение $X_a = \cup_i A_i$ или пересечение $X_a = \cap_i A_i$ мультимножеств A_i , описывающих объекты A_i , либо как линейная комбинация различных мультимножеств вида $X_a = \sum_i c_i \cdot A_i, X_a = \cup_i c_i \cdot A_i$ или $X_a = \cap_i c_i \cdot A_i$. Заметим кстати, что и само мультимножество A_i вида (2) формально можно представить как сумму $A_i = \sum \{q_i^{(j)}\}$ k различных множеств (m -мерных векторов или кортежей) $q_i^{(j)}$, характеризующих один и тот же объект A_i .

На семействе мультимножеств $A=\{A_1, \dots, A_n\}$ можно ввести новые классы метрических пространств мультимножеств (A, d) , которые задаются следующими видами метрик:

$$d_{1p}(A, B) = \left(\sum_{x_i \in G} w_i |k_A(x_i) - k_B(x_i)| \right)^{1/p}, \quad p > 0;$$

$$d_{2p}(A, B) = \left(\sum_{x_i \in G} w'_i |k_A(x_i) - k_B(x_i)| \right)^{1/p}, \quad w'_i = w_i / \sum_{j=1}^h w_j k_Z(x_j);$$

$$d_{3p}(\mathbf{A}, \mathbf{B}) = \left(\frac{\sum_{x_i \in G} w_i |k_A(x_i) - k_B(x_i)|}{\sum_{x_i \in G} w_i \max[k_A(x_i), k_B(x_i)]} \right)^{1/p}$$

Основная метрика $d_{1p}(\mathbf{A}, \mathbf{B})$ является метрикой типа Хемминга при $p=1$, используемой во многих приложениях. Полностью усредненная метрика $d_{2p}(\mathbf{A}, \mathbf{B})$ характеризует различие между двумя мультимножествами A и B , отнесенное к расстоянию, максимально возможному в исходном пространстве. Локально усредненная метрика $d_{3p}(\mathbf{A}, \mathbf{B})$ задает различие, отнесенное к максимально возможной «общей части» $A \cup B$ только этих двух мультимножеств в исходном пространстве.

Метод согласованной групповой классификации объектов

Процедура классификации объектов в рамках формальной логики может быть описана как совокупность решающих правил, которые представляются выражениями вида:

$$\text{ЕСЛИ } \langle \text{условия} \rangle, \text{ ТО } \langle \text{решение} \rangle \quad (2)$$

При прямой классификации терм $\langle \text{условия} \rangle$ включает названия объектов или перечень значений признаков, описывающих объекты класса. При непрямой классификации один или несколько термов $\langle \text{условия} \rangle$ конструируются как отношения между различными признаками и/или их значениями. Терм $\langle \text{решение} \rangle$ в обоих случаях означает, что объект принадлежит к определенному классу. Заметим, что подобным образом формируются и базы знаний экспертных систем продукционного типа.

Рассмотрим задачу групповой классификации объектов в следующей постановке. Задано множество объектов (вариантов, альтернатив) A_1, \dots, A_n , которые описываются m дискретными признаками Q_1, \dots, Q_m , имеющими количественные и/или качественные шкалы оценок. Каждая группа признаков $Q_s = \{q_s^1, \dots, q_s^{h_s}\}$, $e_s = 1, \dots, h_s$, $s = 1, \dots, m$ отражает содержательное качество объектов. Например, $q_s^{e_s}$ может быть значением показателя, определяющего какое-либо свойство объекта, или оценкой объекта по критерию, и тому подобное. Объекты A_1, \dots, A_n предварительно рассортированы по нескольким классам D_1, \dots, D_f путем прямой классификации. Принадлежность объекта A_i , $i = 1, \dots, n$ к некоторому классу D_t , $t = 1, \dots, f$ выражается атрибутом R , которое можно считать еще одним качественным признаком со шкалой значений $R = \{r_{it}\}$. Совокупность признаков $\{q_i^{e_1}, \dots, q_m^{e_m}; r_{it}\}$ представляет собой индивидуальное правило классификации вида (2), относящее объект A_i к классу D_t .

Любой объект A_i существует в k экземплярах, которые отличаются наборами признаков, его характеризующих. Например, когда объект независимо оценивается и сортируется по классам k экспертами или его параметры измеряются k различными способами. Тем самым имеется k различных и, возможно, противоречивых индивидуальных правил классификации объектов. Требуется построить одно или несколько агрегированных групповых правил вида (2), которые наилучшим (в некотором смысле) образом аппроксимируют совокупность всех индивидуальных правил классификации объектов и относят объекты в заданные классы с допустимой точностью.

Для решения задачи групповой классификации многопризнаковых объектов использован метод МАСКА (МногоАспектная Согласованная Классификация Альтернатив), основанный на теории мультимножеств [Петровский, 2003], [Petrovsky, 2001]. Метод предназначен для построения согласованных групповых правил классификации реально имеющих объектов, оцененных несколькими экспертами по многим критериям, которые агрегируют индивидуальные правила сортировки объектов, и выявления противоречивых индивидуальных правил классификации.

Сопоставим каждому многопризнаковому объекту мультимножество

$$\mathbf{A}_i = \{k_{A_i}(q_1^1) \circ q_1^1, \dots, k_{A_i}(q_1^{h_1}) \circ q_1^{h_1}, \dots, k_{A_i}(q_m^1) \circ q_m^1, \dots, k_{A_i}(q_m^{h_m}) \circ q_m^{h_m}, k_{A_i}(r_1) \circ r_1, \dots, k_{A_i}(r_f) \circ r_f\} \quad (3)$$

над доменом $G = Q_1 \cup \dots \cup Q_m \cup R$, где $k_{A_i}(q_s^{es})$ и $k_{A_i}(r_t)$ равны числу экспертов, давших объекту A_i оценку q_s^{es} и рекомендацию r_t . Принадлежность объекта A_i к некоторому классу D_t определяется, например, правилом большинства голосов, в соответствии с которым объект A_i считается принадлежащим к классу D_t , если $k_{A_i}(r_t) > k_{A_i}(r_p)$ для всех $p \neq t$. Представление каждого объекта как мультимножества (3) есть, по сути, групповое правило классификации вида (2), объединяющее k индивидуальных правил сортировки. Нужно найти одно или несколько агрегированных правил классификации всей совокупности объектов вида (2), которые наилучшим образом были бы согласованы с групповыми правилами классификации.

Для простоты положим, что результатом классификации должно быть разложение совокупности объектов только на два класса: D_a («хорошие») и D_b («плохие»). При необходимости рассортировать объекты на большее число классов, можно сначала разбить совокупность объектов на две группы, затем одну из них или обе группы – на подгруппы, и так далее. Сопоставим каждому классу решений D_a и D_b мультимножества X_a и X_b , которые представляют собой соответственно совокупности хороших и плохих объектов, сформированных путем сложения соответствующих им мультимножеств

$$\begin{aligned} \mathbf{X} &= \sum_{i \in I_t} \mathbf{A}_i = \{k_{X_t}(q_1^1) \circ q_1^1, \dots, k_{X_t}(q_1^{h_1}) \circ q_1^{h_1}, \dots, k_{X_t}(q_m^1) \circ q_m^1, \dots, k_{X_t}(q_m^{h_m}) \circ q_m^{h_m}, \\ k_{X_t}(r_a) \circ r_a, k_{X_t}(r_b) \circ r_b\}, \quad k_{X_t}(q_s^{es}) &= \sum_{i \in I_t} k_{A_i}(q_s^{es}), \quad k_{X_t}(r_t) = \sum_{i \in I_t} k_{A_i}(r_t), \quad t = a, b, \end{aligned} \quad (4)$$

где подмножества индексов $I_a \cup I_b = \{1, \dots, n\}$, $I_a \cap I_b = \emptyset$. При суммировании мультимножеств учитываются значения всех признаков, характеризующих все объекты, входящие в группу.

Процедура нахождения агрегированного правила, аппроксимирующего большое число противоречивых правил сортировки многопризнаковых объектов заключается в следующем. Образует новые мультимножества

$$\begin{aligned} \mathbf{R}_a &= \{k_{R_a}(X_a) \circ X_a, k_{R_a}(X_b) \circ X_b\}, \quad \mathbf{R}_b = \{k_{R_b}(X_a) \circ X_a, k_{R_b}(X_b) \circ X_b\}, \\ \mathbf{Q}_{as} &= \sum_{j \in J_{as}} \mathbf{Q}_j, \quad \mathbf{Q}_{bs} = \sum_{j \in J_{bs}} \mathbf{Q}_j, \quad \mathbf{Q}_j = \{k_j(X_a) \circ X_a, k_j(X_b) \circ X_b\}, \end{aligned}$$

которые назовем «категориальными» и «содержательными», где подмножества индексов $J_{as} \cup J_{bs} = \{1, \dots, h_s\}$, $J_{as} \cap J_{bs} = \emptyset$, $s = 1, \dots, m$. Здесь аргументы X_a и X_b соответствуют классам \mathbf{X}_a и \mathbf{X}_b .

«Категориальные» мультимножества \mathbf{R}_a и \mathbf{R}_b задают наилучшую из всех возможных декомпозиций рассматриваемой совокупности объектов на классы D_a и D_b для имеющегося набора первичных правил сортировки. При идеальных предварительных сортировках объектов, в которых отсутствуют противоречия между индивидуальными правилами экспертов, расстояние $d^* = d(\mathbf{R}_a, \mathbf{R}_b)$ между мультимножествами \mathbf{R}_a и \mathbf{R}_b будет максимально возможным расстоянием между объектами, входящими в разные классы. При идеальных предварительных сортировках объектов, в которых отсутствуют противоречия между индивидуальными правилами экспертов, расстояние d^* будет равно соответственно $d_{1p}^* = [kn]^{1/p}$, $d_{2p}^* = [1/h]^{1/p}$, $d_{3p}^* = 1$, где $h = h_1 + \dots + h_m + f$ – общее число значений всех признаков, описывающих объекты.

Аналогичным образом для каждой s -ой группы признаков Q_s нужно сформировать такие пары «содержательных» мультимножеств \mathbf{Q}_{sa}^* и \mathbf{Q}_{sb}^* , которые находились бы на максимально возможном расстоянии в метрическом пространстве мультимножеств и с достаточной точностью совпадали бы с разложением $\{\mathbf{R}_a, \mathbf{R}_b\}$. Задача нахождения классифицирующих признаков $q_s^j \in \mathbf{Q}_{as}^*$, $q_s^j \in \mathbf{Q}_{bs}^*$ сводится, таким образом, к решению m оптимизационных задач вида

$$d(\mathbf{Q}_{as}, \mathbf{Q}_{bs}) \rightarrow \max d(\mathbf{Q}_{as}, \mathbf{Q}_{bs}) = d(\mathbf{Q}_{as}^*, \mathbf{Q}_{bs}^*). \quad (5)$$

Решение каждой из указанных задач задает наилучшее разложение $\{\mathbf{Q}_{sa}^*, \mathbf{Q}_{sb}^*\}$ имеющейся совокупности объектов по s -ой группе признаков. Классифицирующие признаки для различных групп признаков можно упорядочить по величине $L_s = d(\mathbf{Q}_{as}^*, \mathbf{Q}_{bs}^*) / d(\mathbf{R}_a, \mathbf{R}_b)$, которая определяет точность аппроксимации и в некотором смысле характеризует относительную значимость s -ой группы признаков \mathbf{Q}_s в агрегированном групповом правиле классификации объектов.

Разные комбинации классифицирующих признаков, входящих в «содержательные» и «категориальные» мультимножества, дадут разные варианты агрегированных групповых правил для классификации совокупности многопризнаковых объектов:

$$\text{IF } \langle (q_{i1} \in Q_{au}^*) \text{ AND } (q_{i2} \in Q_{av}^*) \text{ AND } \dots \text{ AND } (q_{in} \in Q_{aw}^*) \rangle, \text{ THEN } \langle \text{Object } A_i \in D_a \rangle, \quad (6)$$

$$\text{IF } \langle (q_{i1} \in Q_{bu}^*) \text{ AND } (q_{i2} \in Q_{bv}^*) \text{ AND } \dots \text{ AND } (q_{in} \in Q_{bw}^*) \rangle, \text{ THEN } \langle \text{Object } A_i \in D_b \rangle. \quad (7)$$

Среди объектов, отобранных по агрегированному правилу в заданный класс D_a или D_b , будут как правильно, так и неправильно классифицированные объекты. Желательно найти такие согласованные групповые правила, которые обеспечат максимальную разность между числом правильно и числом неправильно классифицированных объектов. Комбинации соответствующих классифицирующих признаков включаются в согласованные групповые правила отнесения объектов к уточненным классам $D_a \setminus D_{ac}$ («безусловно хорошие») и $D_b \setminus D_{bc}$ («безусловно плохие»), которые записываются в виде:

$$\text{IF } \langle (\sum_{q \in Q_{au}^*} k_{A_i}(q) > \sum_{q \in Q_{bu}^*} k_{A_i}(q)) \text{ AND } (\sum_{q \in Q_{av}^*} k_{A_i}(q) > \sum_{q \in Q_{bv}^*} k_{A_i}(q)) \text{ AND } \dots \text{ AND } (k_{A_i}(r_a) > k_{A_i}(r_b)) \rangle, \text{ THEN } \langle \text{Object } A_i \in D_a \setminus D_{ac} \rangle, \quad (8)$$

$$\text{IF } \langle (\sum_{q \in Q_{au}^*} k_{A_i}(q) < \sum_{q \in Q_{bu}^*} k_{A_i}(q)) \text{ AND } (\sum_{q \in Q_{av}^*} k_{A_i}(q) < \sum_{q \in Q_{bv}^*} k_{A_i}(q)) \text{ AND } \dots \text{ AND } (k_{A_i}(r_a) < k_{A_i}(r_b)) \rangle, \text{ THEN } \langle \text{Object } A_i \in D_b \setminus D_{bc} \rangle. \quad (9)$$

Одновременно формируется класс противоречиво классифицированных объектов $D_c = D_{ac} \cup D_{bc}$, которые имеют противоречивые индивидуальные правила сортировки

$$\text{IF } \langle [(\sum_{q \in Q_{au}^*} k_{A_i}(q) > \sum_{q \in Q_{bu}^*} k_{A_i}(q)) \text{ AND } \dots \text{ AND } (k_{A_i}(r_a) < k_{A_i}(r_b))] \text{ OR } [(\sum_{q \in Q_{au}^*} k_{A_i}(q) < \sum_{q \in Q_{bu}^*} k_{A_i}(q)) \text{ AND } \dots \text{ AND } (k_{A_i}(r_a) > k_{A_i}(r_b))] \rangle, \text{ THEN } \langle \text{Object } A_i \in D_c \rangle. \quad (10)$$

Эти объекты нуждаются в дополнительном анализе.

Алгоритмы построения агрегированных групповых правил

Процедура построения согласованных групповых правил для классификации многопризнаковых объектов, которые агрегируют большое число индивидуальных правил сортировки, включает в себя следующие алгоритмы:

алгоритм 1 – поиск классифицирующих признаков и построение агрегированных групповых правил отнесения объектов к классам D_a и D_b ;

алгоритм 2 – построение согласованных групповых правил отнесения объектов к уточненным классам $D_a \setminus D_{ac}$, $D_b \setminus D_{bc}$ и к классу противоречиво классифицированных объектов $D_c = D_{ac} \cup D_{bc}$.

Алгоритм 1 состоит из следующих шагов.

- Построить таблицу решений $K = \|k_{A_i}(q)\|_{k \times n}$, строки которой представляют собой мультимножества A_i вида (3).

• Сформировать путем сложения мультимножеств A_i группы объектов A_i , относящихся к заданным классам D_a («хорошие») и D_b («плохие»), которым соответствуют мультимножествам X_a и X_b (4). Отнесение объекта в один из двух классов производится по правилу большинства голосов на основе индивидуальных рекомендаций экспертов по признаку R без учета оценок по критериям Q_1, \dots, Q_{22} . Заметим, что объекты, получившие одинаковые оценки по каждому из критериев, но отличающиеся оценками по признаку R , попадают в разные классы. Построить агрегированную таблицу решений $K = \|k_{xi}(q)\|_{2 \times n}$, строки которой представляют собой мультимножества X_a и X_b .

• Найти классифицирующие признаки $q_s^j \in Q_{as}^*$, $q_s^j \in Q_{bs}^*$ по каждой s -ой группе признаков Q_s , которые являются решением оптимизационной задачи (5) с метрикой типа Хемминга. Проранжировать классифицирующие признаки по убыванию уровня аппроксимации L_s .

• Построить агрегированные групповые правила классификации объектов (6) и (7):

Укажем основные шаги алгоритма 2 применительно к классу лучших объектов D_a .

• По каждому критерию Q_s^* , входящему в групповое правило отнесения объектов к классу D_a , найти правильно классифицированные объекты, которые удовлетворяют условию

$$(\sum_{q \in Q_{au}} k_{Ai}(q) > \sum_{q \in Q_{bu}} k_{Ai}(q)) \text{ AND } (k_{Ai}(r_a) > k_{Ai}(r_b)),$$

и неправильно классифицированные объекты, которые удовлетворяют условию

$$(\sum_{q \in Q_{au}} k_{Ai}(q) > \sum_{q \in Q_{bu}} k_{Ai}(q)) \text{ AND } (k_{Ai}(r_a) < k_{Ai}(r_b)).$$

• Выбрать критерий Q_u^* , обеспечивающий максимальную разность между числом правильно и числом неправильно классифицированных объектов, который включается в согласованное групповое правило отнесения объектов к уточненному классу $D_a \setminus D_{ac}$.

• Рассмотреть все комбинации критерия Q_u^* с остальными критериями, входящими в групповое правило отнесения объектов к классу D_a , и найти правильно и неправильно классифицированные объекты, которые удовлетворяют условиям

$$(\sum_{q \in Q_{au}} k_{Ai}(q) > \sum_{q \in Q_{bu}} k_{Ai}(q)) \text{ AND } (\sum_{q \in Q_{av}} k_{Ai}(q) > \sum_{q \in Q_{bv}} k_{Ai}(q)) \text{ AND } (k_{Ai}(r_a) > k_{Ai}(r_b)),$$

$$(\sum_{q \in Q_{au}} k_{Ai}(q) > \sum_{q \in Q_{bu}} k_{Ai}(q)) \text{ AND } (\sum_{q \in Q_{av}} k_{Ai}(q) > \sum_{q \in Q_{bv}} k_{Ai}(q)) \text{ AND } (k_{Ai}(r_a) < k_{Ai}(r_b)).$$

• Выбрать пару критериев Q_u^* , Q_v^* , обеспечивающих максимальную разность между числом правильно и числом неправильно классифицированных объектов, которые включаются в согласованное групповое правило отнесения объектов к уточненному классу $D_a \setminus D_{ac}$.

• Рассмотреть все комбинации пары критериев Q_u^* , Q_v^* с остальными критериями, входящими в групповое правило отнесения объектов к классу D_a , и найти правильно и неправильно классифицированные объекты.

• Выбрать тройку, затем четверку и т.д. критериев Q_u^* , Q_v^* , Q_w^* , ... обеспечивающих максимальную разность между числом правильно и числом неправильно классифицированных объектов, которые включаются в согласованное групповое правило отнесения объектов к уточненному классу $D_a \setminus D_{ac}$.

• Включить критерии Q_u^* , Q_v^* , Q_w^* , ..., обеспечивающие максимальную разность между числами правильно и неправильно классифицированных объектов, в согласованное групповое правило отнесения объектов к уточненному классу решений $D_a \setminus D_{ac}$ (8)

$$\begin{aligned} & \text{IF } \langle (\sum_{q \in Q_{a2}} k_{Ai}(q) > \sum_{q \in Q_{b2}} k_{Ai}(q)) \text{ AND } (\sum_{q \in Q_{a3}} k_{Ai}(q) > \sum_{q \in Q_{b3}} k_{Ai}(q)) \\ & \text{AND } (\sum_{q \in Q_{a1}} k_{Ai}(q) > \sum_{q \in Q_{b1}} k_{Ai}(q)) \text{ AND } (\sum_{q \in Q_{a5}} k_{Ai}(q) > \sum_{q \in Q_{b5}} k_{Ai}(q)) \\ & \text{AND } (k_{Ai}(r_a) > k_{Ai}(r_b)) \rangle, \text{ THEN } \langle \text{Object } A_i \in D_a \setminus D_{ac} \rangle. \end{aligned}$$

Аналогичным образом с помощью алгоритма 2 строится согласованное групповое правило отнесения объектов к уточненному классу решений $D_b \setminus D_{bc}$ (9)

$$\begin{aligned} & \text{IF } \langle (\sum_{q \in Q_{a2}} k_{A_i}(q) < \sum_{q \in Q_{b2}} k_{A_i}(q)) \text{ AND } (\sum_{q \in Q_{a3}} k_{A_i}(q) < \sum_{q \in Q_{b3}} k_{A_i}(q)) \\ & \text{AND} (\sum_{q \in Q_{a1}} k_{A_i}(q) < \sum_{q \in Q_{b1}} k_{A_i}(q)) \text{ AND } (\sum_{q \in Q_{a5}} k_{A_i}(q) < \sum_{q \in Q_{b5}} k_{A_i}(q)) \\ & \text{AND } (k_{A_i}(r_a) < k_{A_i}(r_b)) \rangle, \text{ THEN } \langle \text{Object } A_i \in D_b \setminus D_{bc} \rangle. \end{aligned}$$

После этого формируется класс $D_c = D_{ac} \cup D_{bc}$ объектов, которые имеют противоречивые индивидуальные правила сортировки (10)

$$\begin{aligned} & \text{IF } \langle [(\sum_{q \in Q_{a2}} k_{A_i}(q) > \sum_{q \in Q_{b2}} k_{A_i}(q)) \text{ AND } (\sum_{q \in Q_{a3}} k_{A_i}(q) > \sum_{q \in Q_{b3}} k_{A_i}(q)) \\ & \text{AND} (\sum_{q \in Q_{a1}} k_{A_i}(q) > \sum_{q \in Q_{b1}} k_{A_i}(q)) \text{ AND } (\sum_{q \in Q_{a5}} k_{A_i}(q) > \sum_{q \in Q_{b5}} k_{A_i}(q)) \\ & \text{AND } (k_{A_i}(r_a) < k_{A_i}(r_b))] \\ & \text{OR } [\sum_{q \in Q_{a2}} k_{A_i}(q) < \sum_{q \in Q_{b2}} k_{A_i}(q) \text{ AND } (\sum_{q \in Q_{a3}} k_{A_i}(q) < \sum_{q \in Q_{b3}} k_{A_i}(q)) \\ & \text{AND} (\sum_{q \in Q_{a1}} k_{A_i}(q) < \sum_{q \in Q_{b1}} k_{A_i}(q)) \text{ AND } (\sum_{q \in Q_{a5}} k_{A_i}(q) < \sum_{q \in Q_{b5}} k_{A_i}(q)) \\ & \text{AND } (k_{A_i}(r_a) < k_{A_i}(r_b))] \rangle, \text{ THEN } \langle \text{Object } A_i \in D_c \rangle. \end{aligned}$$

Провести дополнительный анализ таких объектов для устранения противоречий и принятия окончательного решения.

Апробация метода

Предложенный подход к построению согласованного решающего правила для классификации многопризнаковых объектов был апробирован для анализа итогов конкурса РФФИ целевых фундаментальных исследований 2006 года. Экспериментальные расчеты проводились на модельной базе данных, которая включала в себя экспертные оценки поддержанных и отклоненных проектов по областям следующим знаний:

02. Физика и астрономия – всего 127 проектов, в том числе 39 поддержанных и 88 отклоненных проектов,
04. Биология и медицинская наука – всего 252 проекта, в том числе 68 поддержанных и 184 отклоненных проектов.

Были приняты следующие правила агрегирования индивидуальных правил сортировки проектов. Проект безусловно включался в число поддержанных, когда все эксперты давали оценку $r_1 \ll 5$ (то есть $k_{A_i}(r_2) = k_{A_i}(r_3) = k_{A_i}(r_4) = 0$) или число оценок $r_1 \ll 5$ было большее или равно числу оценок $r_2 \ll 4$ (то есть $k_{A_i}(r_1) \geq k_{A_i}(r_2)$, $k_{A_i}(r_3) = k_{A_i}(r_4) = 0$). Проект безусловно включался в число отклоненных, когда ни один из экспертов не давал оценку $r_1 \ll 5$ (то есть $k_{A_i}(r_1) = 0$). Проекты, не удовлетворявшие этим правилам, были исключены из таблиц решений и отнесены к проектам, требующих дополнительного анализа. В эту же группу были включены проекты, которые имели одинаковые заключения всех экспертов и были отнесены к классам либо поддержанных D_a , либо отклоненных D_b .

Совокупности классифицирующих признаков $Q_a^* = \cup_s Q_{as}^*$ и $Q_b^* = \cup_s Q_{bs}^*$, построенных с помощью алгоритма 1, которые определяют принадлежность проекта к классам поддержанных D_a и отклоненных D_b проектов, оказались одинаковыми по областям 02 и 04 и выглядят следующим образом (признаки расположены в порядке убывания их относительной значимости):

$$Q_a^* = Q_{a16}^* \cup Q_{a22}^* \cup Q_{a21}^* \cup Q_{a15}^* = \{q_{16}^1, q_{16}^2; q_{22}^1, q_{22}^2; q_{21}^1, q_{21}^2; q_{15}^3\},$$

$$Q_b^* = Q_{b16}^* \cup Q_{b22}^* \cup Q_{b21}^* \cup Q_{b15}^* = \{q_{16}^3; q_{22}^3; q_{21}^3, q_{21}^4; q_{15}^1, q_{15}^2\}.$$

Множество классифицирующих признаков Q_b^* дополняет множество Q_a^* .

Последовательное рассмотрение классифицирующих признаков с помощью алгоритма 2 показало, что построенная комбинация признаков дает максимальное количество правильно классифицированных проектов и минимальное количество неправильно классифицированных. Согласованное агрегированное групповое правило для отнесения проектов к уточненному классу $D_a \setminus D_{ac}$ (безусловно поддержанные) имеет вид:

$$\text{для критериев } Q_{16}, Q_{22} - k_{A_i}(q_s^1) + k_{A_i}(q_s^2) > k_{A_i}(q_s^3),$$

$$\text{для критерия } Q_{15} - k_{A_i}(q_s^1) + k_{A_i}(q_s^2) < k_{A_i}(q_s^3),$$

$$\text{для критерия } Q_{21} - k_{A_i}(q_{21}^1) + k_{A_i}(q_{21}^2) > k_{A_i}(q_{21}^3) + k_{A_i}(q_{21}^4).$$

При применении этих правил следует также учитывать выполнение приведенных выше правил агрегирования индивидуальных правил сортировки проектов по заключениям экспертов. Согласованное агрегированное групповое правило для отнесения проектов к уточненному классу $D_b \setminus D_{bc}$ (безусловно отклоненные) получается заменой знаков на противоположные.

В рассмотренных областях знаний конечный результат классификации проектов достигается уже при комбинации только нескольких первых критериев из полученных перечней, а именно:

$$\text{по области } O2 - \text{критериев } Q_{16}^*, Q_{22}^*;$$

$$\text{по области } O4 - \text{критериев } Q_{16}^*, Q_{22}^*, Q_{21}^*.$$

Таким образом, наиболее важными для отбора проектов в рассмотренных областях знаний являются критерии Q_{16} . Научная значимость проекта, Q_{22} . Масштабы применимости результатов заявленных исследований, Q_{21} . Завершающая стадия фундаментальных исследований, предлагаемых в проекте. Фактически, экспертные оценки только по этим критериям и следует принимать во внимание при разбиении проектов по уточненным классам решений и выявлении проектов, требующих дополнительного рассмотрения из-за расхождения индивидуальных заключений экспертов. Учитывая эти факты, можно значительно сократить время, используемое для классификации проектов.

Каждое из правил классификации можно записать на естественном языке, что является отличительной чертой вербального анализа решений. Согласованное агрегированное групповое правило для отбора безусловно поддержанных проектов выглядит следующим образом:

«Проект имеет исключительно высокую или значительную научную значимость (оценки q_{16}^1, q_{16}^2); массовый или междисциплинарный масштабы применимости результатов заявленных исследований (оценки q_{22}^1, q_{22}^2); предлагаемые в проекте исследования завершаются лабораторным образцом или ключевыми элементами разработки (оценки q_{21}^1, q_{21}^2)».

Построенные правила классификации проектов позволяют провести анализ работы экспертов, определить критерии, которые оказывают решающее влияние на отбор проектов, оценить качество и согласованность экспертных оценок, выявить существенные расхождения в их мнениях.

Уровень согласованности индивидуальных экспертных оценок проектов по критериям и заключений экспертов по поддержке проектов оказался невысоким. Выявлено следующее число проектов, имеющих расхождения между оценкой их содержания и заключениями экспертов и требующих дополнительного анализа:

по области $O2$ – 33 проекта или 26% от общего числа конкурсных заявок, включенных в таблицу решений (в том числе 15 проектов на предварительной стадии и 18 проектов по результатам работы алгоритмов);

по области 04 – 84 проекта или 33% от общего числа конкурсных заявок, включенных в таблицу решений (в том числе 44 проекта на предварительной стадии и 40 проектов по результатам работы алгоритмов).

Эти факты свидетельствуют о недостаточном внимании и аккуратности некоторых экспертов при проведении экспертизы проектов.

Важное значение для работы алгоритмов построения правил классификации проектов имеет число экспертов, назначаемых для проведения экспертизы. Если заявка оценивается только двумя экспертами, то не исключена возможность возникновения ситуации, когда заключения экспертов будут противоположными: один эксперт поддерживает проект, а другой отклоняет. Для большей достоверности выводов при классификации проектов целесообразно, чтобы минимальное число экспертов, оценивающих заявку, было не менее трех, а желательно – чтобы заявка оценивалась четырьмя экспертами. Построение решающих правил упрощается, если каждый проект оценивается одним и тем же одинаковым числом экспертов.

Заключение

Проблемы классификации объектов, которые описываются многими количественными и качественными признаками, причем каждый из объектов может существовать в нескольких «экземплярах», являются достаточно трудными для решения и до настоящего момента плохо разработанными. Преодолеть трудности оказалось возможным, используя методы вербального анализа решений и аппарат теории мультимножеств, которые не содержат необоснованных преобразований исходной информации и не приводят к потере или искажению данных.

Изложен новый метод согласованной групповой классификации многопризнаковых объектов, основанный на их представлении с помощью мультимножеств. Метод обеспечивает построение согласованного группового правила для классификации многопризнаковых объектов, которое аппроксимирует большое число противоречивых индивидуальных правил сортировки.

Подход апробирован на модельной базе данных, содержащей результаты экспертной оценки и конкурсного отбора проектов целевых фундаментальных исследований в РФФИ. Сформированы решающие правила для отнесения объектов в заданные классы. Решающие правила позволили выделить наиболее важные для отбора проектов критерии по каждой области знания и выявить расхождения в индивидуальных правилах сортировки проектов, применявшихся экспертами.

Предлагаемый подход обеспечивает систематизацию имеющейся информации, анализ причин принятия окончательного решения, получение его обоснования. Полученные результаты могут быть использованы при проведении экспертного анализа итогов конкурсов целевых фундаментальных исследований, а также могут найти применение при анализе конкурсов инициативных и иных проектов (с учетом их специфики), выполняемых в РФФИ.

Благодарности

Работа поддержана программами фундаментальных исследований президиума РАН «Интеллектуальные информационные технологии, математическое моделирование, системный анализ и автоматизация» и ОНИТ РАН «Информационные технологии и методы анализа сложных систем», Российским фондом фундаментальных исследований (проекты 08-01-00247, 08-07-13532, 09-07-00009, 09-07-12111).

Список литературы

- [Ларичев, 2006] Ларичев О.И. Вербальный анализ решений / Под ред. А.Б.Петровского. — М.: Наука, 2006.
- [Петровский, 2003] Петровский А.Б. Пространства множеств и мультимножеств. — М: Едиториал УРСС, 2003.
- [Петровский, 2009] Петровский А. Б. Теория принятия решений. — М.: Издательский центр «Академия», 2009.
- [Петровский, Тихонов, 2009] Петровский А. Б., Тихонов И. П. Фундаментальные исследования, ориентированные на практический результат: подходы к оценке эффективности // Вестник РАН. — 2009.— Т. 79, № 11.— С. 1006–1011.
- [Doumpos, Zopounidis, 2002] Doumpos M., Zopounidis C. Multicriteria Decision Aid Classification Methods. — Dordrecht: Kluwer Academic Publishers, 2002.
- [Greco et al., 2002] Greco S., Matarazzo B., Slowinski R. Rough sets methodology for sorting problems in presence of multiple attributes and criteria // European Journal of Operational Research. — 2002. — Vol. 138, no. 2. — Pp. 247-259.
- [Köksalan et al., 2003] Köksalan M., Ulu C. An interactive approach for placing alternatives in preference classes // European Journal of Operational Research. — 2003. — Vol. 144, no. 2. — Pp. 429–439.
- [Petrovsky, 2001] Petrovsky A.B. Multi-Attribute Sorting of Qualitative Objects in Multiset Spaces. // Multiple Criteria Decision Making in the New Millenium. — Berlin: Springer-Verlag, 2001. — Pp.124-131.
- [Roy, Bouyssou, 1993] Roy B., Bouyssou D. Aide Multicritere a la decision: Methodes et cas — Paris: Economica, 1993.

Сведения об авторах

Петровский Алексей Борисович – д.т.н., заведующий лабораторией Института системного анализа РАН, Россия, Москва 117312, пр-т 60-летия Октября, 9, тел. (495)135-8503, e-mail: pab@isa.ru

Ройзензон Григорий Владимирович – к.т.н., старший научный сотрудник Института системного анализа РАН, Россия, Москва 117312, пр-т 60-летия Октября, 9, тел. (495)135-8503, e-mail: rgv@isa.ru

Тихонов Игорь Петрович – к.т.н., начальник отдела целевых фундаментальных исследований Российского фонда фундаментальных исследований, Россия, Москва 117334, Ленинский пр-кт, 32А, e-mail: tikhonov@rfbr.ru

Балышев Александр Владимирович – к.биол.н., заместитель начальника отдела целевых фундаментальных исследований Российского фонда фундаментальных исследований, Россия, Москва 117334, Ленинский пр-кт, 32А, e-mail: bav@rfbr.ru

СРАВНЕНИЕ АЛЬТЕРНАТИВ В ПОДХОДЕ ОБОБЩЕННЫХ ИНТЕРВАЛЬНЫХ ОЦЕНОК

Михаил Стернин, Геннадий Шепелёв, Алла Рябова

Abstract: A method of “direct” comparison of alternatives described by polyinterval expert estimations is proposed. The method is based on approach of generalized interval estimations and method of comparison of monointerval estimations that were elaborated by authors earlier. For given concrete configuration of comparable pair of polyinterval estimations expert, comparing squares of favorable (green), unfavorable (red) and neutral (yellow) zones and using own preferences, may select preferable alternative or make conclusion that alternatives are incomparable. So called assurance factor as a measure of comparing mentioned squares is used in the process of comparison. An example of uprising problem of comparison of alternatives described by polyinterval expert estimations is given. This example concerning construction of polyinterval estimations of two investment projects on base of two monointerval estimations of costs and revenues for each project received from two experts. It is showed that analogue of Hurwicz (point) estimation for monointerval estimations may be constructed for polyinterval ones. A system of methods for comparing alternatives described by polyinterval expert estimations was built as a result of these studies. Using the methods an expert may take into account different levels of uncertainty for problem of comparing alternatives. If there is not information concerning chances of scenario realization in the framework of polyinterval estimations method of “direct” comparison of alternatives may be recommended. If there is information concerning chances of realization for different bands of scenarios in polyinterval estimations method of comparing monointerval estimations combined with Monte-Carlo method may be recommended. If additionally there is information concerning chances of realization for different values of target indicator in the framework of scenarios in polyinterval estimations method of reducing generalized interval estimations to monointerval ones by means of averaging may be recommended.

Keywords: decision making, generalized interval estimations, expert analysis, comparison of alternatives, generalized probability distributions, aggregation of information.

ACM Classification Keywords: H.4.2 Types of Systems – Decision support; G.1.0 Numerical Analysis - Interval Arithmetic, G.3 Probability and Statistics – Distribution functions.

Введение

При анализе проблемных ситуаций, возникающих в задачах принятия решений многих предметных областей, часто используются результаты расчетов на математических моделях, параметры которых являются числовыми данными. В условиях неопределенности такие данные представляются интервальными оценками. Для получения значений результирующих показателей моделей в этом случае необходимо привлечение методов интервального анализа и (если подобная информация имеется) различных гипотез о шансах на реализацию тех или иных значений исходных параметров модели, лежащих в пределах их оценок.

Наиболее распространенным математическим аппаратом формализации таких гипотез являются теоретико-вероятностные методы. Результирующие показатели, получаемые в рамках этого формализма, например, методом статистических испытаний, также имеют тогда интервально-вероятностное представление. При наличии нескольких альтернатив, характеризующихся интервальными или интервально-вероятностными показателями, выбор «лучшей» альтернативы может быть получен как решение задачи принятия решений лишь с привлечением предпочтений человека – «владельца

проблемы» [Петровский, 2009] с использованием методов сравнения интервальных оценок [Стернин, Шепелев, 2009 а] или сравнения интервально-вероятностных оценок [Chugunov et al., 2008].

С усложнением проблемных ситуаций значительно возросла роль специалистов-экспертов соответствующих предметных областей, проявляющаяся, в частности, в том, что из-за отсутствия требуемой информации значения исходных данных задаются преимущественно в виде экспертных интервальных оценок. Однако в ряде случаев эксперту, как оказалось, затруднительно выразить свои знания об анализируемом параметре, указывая единственную интервальную оценку: интервал излишнего размаха снижает ценность знаний эксперта, а зауженный интервал довольно часто ведет к ошибкам предсказания [Slovic et al, 2001; Shepelyov, Sternin, 2003].

Целесообразно поэтому дополнительно к моноинтервальному подходу разработать такой подход к выявлению и представлению экспертных знаний о неопределенных количественных данных анализируемых задач, который дал бы эксперту возможность выразить свои знания о параметрах задачи, задавая совокупность интервальных оценок. Отказ от использования только одноинтервальных оценок позволяет эксперту формализовать свои представления о возможной неточности длины и положения интервалов-оценок значений каждого параметра.

Такой подход - подход обобщенных интервальных оценок (ОИО) [Стернин, Шепелев, 2010; Chugunov et al., 2008] – разработан нами как непосредственное развитие традиционного моноинтервального подхода. В подходе ОИО осуществляется замена первичного объекта моноподхода, точечной оценки, интервальной оценкой, интервала как средства описания неопределенности в моноподходе совокупностью интервалов, названной полиинтервальной оценкой (ПИО), а функции распределения вероятностей, заданной на моноинтервальной оценке, совместной функцией распределения вероятностей, заданной на вышеуказанной совокупности. Подход ОИО находится на стыке нескольких научных областей, таких как теория принятия решений, инженерия знаний, теория вероятностей, системы поддержки экспертных решений. Этот подход, расширяя возможности более полного выявления знаний экспертов об исходных данных задачи, приводит к более адекватному учету неопределенности и повышению качества принимаемых решений.

К настоящему времени в подходе ОИО развит ряд методов выявления и формализации экспертных знаний. Краткое изложение полученных в этом направлении результатов дано в следующем разделе работы, который можно рассматривать также как введение в проблематику ОИО, необходимое для представления последующего материала.

Для того чтобы сделать ОИО полноценным инструментом принятия решений, следует разработать методы сравнения альтернатив, показатели качества которых выражены как ОИО. Один из таких методов, основанный на сравнении одноинтервальных оценок, предложен ранее [Стернин, Шепелев, 2009 а]. Результаты разработки другого возможного метода сравнения альтернатив, основанного на сравнении ПИО, представлены в настоящей статье. Метод базируется на анализе лицом, принимающим решение, возможных конфигураций относительной локализации пар ПИО. Поскольку такой подход в методическом отношении следует подходу, использованному нами при сравнении моноинтервальных оценок, основные особенности последнего суммированы в разделе «Сравнение моноинтервальных оценок альтернатив».

Обобщенные интервальные оценки

Для представления в подходе ОИО своих знаний о параметре модели эксперт вначале формирует полиинтервальную оценку. Для этого эксперт может задать несколько характерных интервалов, входящих в совокупность интервалов-оценок, а затем восстановить по ним ПИО. Приведем простейший пример построения ПИО: эксперт задает в качестве начальной оценки параметра какой-либо интервал, а затем

«размывает» его (не обязательно симметрично) в обе стороны от границ. Естественно теперь считать, что и интервалы, промежуточные между начальным и получившимся в результате размыва, также входят в совокупность интервалов-оценок и характеризуют знания эксперта о параметре, являясь возможными реализациями (сценариями) анализируемой величины.

В этом случае ПИО параметра D наглядно представляется криволинейной трапецией на плоскости ($X = D$, $Y = h$), которая восстанавливается по экспертным оценкам оснований трапеции. Ось ординат $h \in [0, 1]$ служит осью меток упорядоченных левых границ интервалов совокупности. Наибольшему основанию трапеции («базовый» интервал) соответствует $h = 0$, а наименьшему («мини» интервал) отвечает $h = 1$. Таким образом, ПИО определяется положением и длинами минимального (верхнего $[D_{lu}, D_{ru}]$) и базового (нижнего $[D_{ld}, D_{rd}]$) интервалов совокупности, задаваемых своими левыми D_l и правыми D_r границами, и формой боковых границ трапеции. В приложениях, как правило, считают эти границы прямолинейными. Интерпретация оси h как оси меток интервалов, образующих ПИО, не единственна. Она обусловлена содержанием задачи. Так, эта ось имеет непосредственный «физический» смысл в задачах с зависимыми переменными. Например, в задаче оценки зависимости запасов природных ископаемых от цены на них каждой точечной оценке цены на оси h ПИО отвечает интервальная оценка объемов запасов на оси D .

По аналогии с моноинтервальным случаем суждения эксперта о шансах на реализацию значений анализируемой величины формализуются путем задания на ПИО плотности совместной функции распределения вероятностей $F(D, h) = F_1(h)F_2(D|h)$. Распределение $F_1(h)$ характеризует шансы реализации различных сценариев в их совокупности. Плотности $F_2(D|h)$ описывают шансы реализации тех или иных значений D на каждом из интервалов совокупности. В общем случае на разных (по h) полосах ПИО плотности $F_2(D|h)$ могут принадлежать к разным семействам распределений. Полученная конструкция – ПИО вместе с определенной на ней $F(D, h)$ – является обобщенной интервальной оценкой величины D . Подход ОИО может быть использован в задачах выявления и формализации экспертных знаний и в задачах принятия решений в нескольких направлениях.

Во-первых, для каждого исходного параметра можно получить усредненное распределение вероятности $F(D)$ на базовом интервале, интегрируя $F(D, h)$ по h при различных сечениях ПИО прямой $D = D_0$, и тем самым свести задачу расчета результирующих показателей модели к известному моноинтервальному случаю. Усредненное распределение вероятности $F(D)$ на базовом интервале, определяемое соотношением (1), является вероятностной смесью распределений на интервалах-сценариях ОИО со смешивающей функцией, задаваемой распределением на оси h ПИО. Для наиболее распространенных на практике прямолинейных границ ПИО в (1) $h_l(D) = (D - D_{ld}) / (D_{lu} - D_{ld})$, $h_r(D) = (D_{rd} - D) / (D_{rd} - D_{ru})$.

Отметим, что в результате такого усреднения возникают вероятностные распределения, обобщающие известные распределения. Например, при задании равномерных распределений на обеих осях ПИО в результате усреднения возникает обобщенное равномерное распределение, плотность которого имеет простое аналитическое выражение. Свойства обобщенного равномерного распределения вместе с тем существенно богаче свойств стандартного равномерного распределения, что позволяет использовать его в задачах принятия решений для аппроксимации знаний эксперта о шансах реализации анализируемого параметра модели в различных ситуациях.

$$F(D) = \begin{cases} \int_0^{h_l(D)} F_1(h)F_2(D|h)dh, & D \in [D_{ld}; D_{lu}) \\ \int_0^1 F_1(h)F_2(D|h)dh, & D \in [D_{lu}; D_{ru}) \\ \int_0^{h_r(D)} F_1(h)F_2(D|h)dh, & D \in [D_{ru}; D_{rd}) \end{cases} \quad (1)$$

Во-вторых, отказавшись от усреднения, для исходных параметров и результирующих показателей моделей могут быть рассчитаны «вероятностные трубки» (так называемые «probability bounds» [Ferson, Tuckey, 2006]), непосредственно использующие всю совокупность полученных от эксперта знаний, содержащуюся в ОИО. В вероятностных трубках отражена информация о вариабельности экспертных знаний на всех интервалах-сценариях. Трубки максимального размаха с достоверностью содержат в себе все возможные, с точностью до знаний эксперта, значения оцениваемых показателей моделей и соответствующих вероятностей. Распределения вероятностей на каждом из сечений такой трубки (по обеим осям ПИО) позволяют учесть неопределенность экспертных знаний, повышая тем самым надежность принимаемых решений.

Поскольку каждая из интервалов-оценок, входящих в ПИО, является возможным сценарием реализации, предложено использовать подход ОИО в сценарном анализе теории принятия решений, что потребовало известного развития аппарата обобщенных равномерных распределений [Стернин, Шепелев, 2008].

При построении ПИО требуются экспертные оценки нескольких (не менее двух) реперных интервалов, на базе которых восстанавливается ПИО. Эта особенность подхода ОИО позволяет использовать его в решении различных задач агрегирования экспертных данных, заданных интервально. Процедуры агрегирования нескольких моноинтервальных оценок, основанные на подходе ОИО, опробованы на примере оптимизационной модели пополнения фондов углеводородов различной степени изученности [Стернин, Шепелев, 2009; Shepelyov, Sternin, 2009]. Несколько моноинтервальных оценок для коэффициентов подтверждаемости запасов различных фондов естественно возникают в этой модели при пересчете названных коэффициентов в случае необходимости перехода от системы фондов с детализированной структурой к более агрегированным фондам. Для дальнейшего использования модели требуются одноинтервальные оценки коэффициентов, переход к которым осуществляется посредством построения вначале на базе исходных одноинтервальных оценок полиинтервальных оценок коэффициентов, а затем возврата к моноинтервальному подходу путем усреднения вероятностных распределений, специфицирующих ОИО, согласно (1).

Для дальнейшего развития метода ОИО как инструмента поддержки процессов принятия решений необходимо, как уже отмечалось, исследовать возможности сравнения ПИО для результирующих индикаторов, описывающих эффективность альтернативных вариантов. Одной из таких возможностей является сравнение получаемых методом Монте-Карло различных реализаций моноинтервальных оценок в парах сопоставляемых ПИО с накоплением статистики результатов сравнения и дальнейшим использованием предпочтений лица, принимающего решение (ЛПР). Важнейшая составляющая этого подхода – разработка и выбор метода сопоставления моноинтервальных оценок. Эта, последняя, задача имеет и самостоятельное значение, несвязанное с задачей сравнения ПИО. Вместе с тем использованный при ее решении подход, основанный на анализе возможных конфигураций относительного расположения интервалов в их сравниваемой паре, будет принят также и при решении задачи сравнения ПИО.

Сравнение моноинтервальных оценок альтернатив

В работе [Стернин, Шепелев, 2009 а] предложен метод непосредственного сопоставления интервальных оценок, процесс сравнения в котором трактуется как задача принятия решений. В соответствии с методом в паре сравниваемых интервалов выделяются три зоны: зона, благоприятствующая проверяемой гипотезе о предпочтительности одной из оценок; зона, неблагоприятная для истинности проверяемой гипотезы, и зона эквивалентности оценок. Вывод о предпочтительности или эквивалентности интервальных оценок делается на основе сопоставления длин зон и предпочтений ЛПР. Метод позволяет либо указать лучшую интервальную оценку из сравниваемых, либо рекомендовать ЛПР временно отказаться от принятия решения. В последнем случае, вероятно, целесообразно воздержаться от немедленного принятия решения и попытаться сузить диапазон неопределенности, получив дополнительную информацию. Мы полагаем, что метод непосредственного сравнения интервалов адекватно учитывает специфику задач принятия решений, для которых характерно наличие понятия несравнимости.

В рамках метода рассматривается задача сравнения двух числовых интервала $I_i = [L_i, R_i]$, заданных их левыми L_i и правыми R_i границами, $L_i < R_i$. С точностью до перестановки I_1 и I_2 возможны следующие четыре варианта их взаимного расположения (конфигурации):

а) $R_1 < L_2$. б) $L_1 < L_2 < R_1 < R_2$. в) $L_1 < L_2 < R_2 < R_1$. д) $L_1 = L_2, R_1 = R_2$.

Ищутся условия, при которых I_2 может быть предпочтительнее («больше»), чем I_1 .

Конфигурация а) не вызывает трудностей: при любых возможных будущих реализациях значений i_1 и i_2 в интервалах I_1 и I_2 соответственно, интервал I_1 не может превзойти I_2 , т.е. всегда $i_2 > i_1$. Конфигурация а) – единственная конфигурация, для которой интервальные оценки сравнимы в строгом смысле слова. В этом случае фактически нет задачи принятия решений, поскольку нет неопределенности: интервал I_2 строго больше интервала I_1 .

Конфигурация д) характеризуется наибольшей неопределенностью, интервалы эквивалентны друг другу по предпочтительности. В этой ситуации ЛПР не в состоянии принять рационального решения в пользу какого-либо из интервалов без дополнительной информации. Необходимо либо отказаться от выбора, либо, если сделать выбор все-таки нужно, а оперативно получить дополнительную информацию невозможно, воспользоваться случайным выбором. Из этого примера видно, что мерой неопределенности в рассматриваемой задаче сравнения может служить протяженность S пересечения $I_1 \cap I_2$.

Анализ конфигураций б) и в) позволяет выделить еще две меры, играющие важную роль при сравнении интервалов. Именно, из рассмотрения конфигурации б) следует, что если для I_1 текущая реализация $i_1 \in [L_1, L_2]$ или для I_2 текущая реализация $i_2 \in [R_1, R_2]$, то эти ситуации благоприятствуют тому, чтобы I_2 оказался предпочтительнее I_1 . Назовем зону благоприятствования зеленой зоной. Протяженность интервала $[L_2, R_1]$ характеризует размеры зоны неопределенности при сравнении I_1 и I_2 . Назовем зону неопределенности желтой зоной. Таким образом, в конфигурации б) шансы на то, что I_2 окажется предпочтительнее I_1 , зависят от соотношения длин желтой и зеленой зон. Можно видеть, что протяженность зеленой зоны $S_g = L_2 - L_1 + R_2 - R_1$, а желтой зоны $S_y = R_1 - L_2$.

В конфигурации в) появляется неблагоприятная («красная») зона протяженностью $S_r = R_1 - R_2$. Здесь протяженность зеленой зоны $S_g = L_2 - L_1$, а желтой зоны $S_y = R_2 - L_2$. Соотношения между длинами зон в сопоставляемых парах оценок разных конфигураций определяют возможные результаты сравнения интервалов. Указанные соотношения удобно представить в виде числовых коэффициентов.

Среди нескольких возможных числовых коэффициентов, характеризующих ситуацию принятия решений, один показатель K , названный нами коэффициентом уверенности, имеет ясный экономический смысл.

Этот коэффициент равен $K = [Sg - Sr]/S(I_1U_2)$, т.е. доле разности длин зеленой и красной зон в общей протяженности сравниваемых интервалов с учетом их возможного пересечения. Содержательно он характеризует относительный прирост возможного максимального выигрыша за счет верного принятия решений. Действительно, в случае конфигурации b), например, при выборе интервала I_2 в качестве предпочтительного максимум возможного выигрыша от такого выбора равен $R_2 - L_1$. При выборе интервала I_1 эта величина равна $R_1 - L_2$. Таким образом, относительный прирост возможного максимального выигрыша за счет верного принятия решений равен $[(R_2 - L_1) - (R_1 - L_2)]/(R_2 - L_1)$, т.е. K . Отметим, что если бы проверялась гипотеза о предпочтительности интервала I_1 , зеленая и красная зоны поменялись бы местами. Вычисленное в конкретной ситуации сравнения интервальных величин значение коэффициента уверенности K может быть использовано ЛПР при принятии решений. Именно, если при проверке гипотезы о предпочтительности интервала I_2 значение коэффициента K для сравниваемой пары интервальных оценок окажется «достаточно большим» и согласуется с представлениями ЛПР о приемлемой величине риска, который связан с принятием решения и который измеряется пороговым (индивидуальным для каждого ЛПР и ситуации принятия решений) значением K_{th} , то принимается, что $I_2 \succ I_1$, если $K \geq K_{th}$.

Принятый в методе непосредственного сравнения моноинтервальных оценок подход может быть перенесен на случай сравнения полиинтервальных оценок.

Сравнение полиинтервальных оценок альтернатив

Прежде всего, приведем пример ситуаций, в которых могут возникать задачи о сравнении полиинтервальных оценок. Пусть имеются два инвестиционных проекта, эффективность которых сравнивается по значению чистого дисконтированного дохода (ЧДД). Если, из-за неопределенности, затраты и результаты по каждому из проектов в каждом году сроков их реализации имеют интервальные оценки, то и для ЧДД будут получены интервальные оценки [Стернин, Шепелев, 2009 а]. В случае двух экспертов, занимающихся оценками затрат и результатов по проектам, по каждому из проектов будут получены по две интервальные оценки ЧДД. Между этими двумя реперными сценариями, своими для каждого проекта, лежат промежуточные сценарии. Формализация этой картины производится построением ПИО для ЧДД проектов на основе реперных сценариев. Для сравнения проектов необходимо сравнить их ПИО.

Итак, рассматривается задача сравнения двух полиинтервальных оценок P_1 и P_2 , заданных их угловыми точками (вершинами) $D_{idl}, D_{idr}, D_{iul}, D_{iur}, i = 1, 2$ и боковыми границами трапеций, представляющих ПИО. Будем для простоты анализировать случай пары ПИО с вложенными интервальными оценками ($D_{idl} < D_{iul}, D_{iur} < D_{idr}$) и считать боковые границы прямолинейными, что не влияет на окончательную формулу для коэффициента уверенности K , имеющему смысл, аналогичный смыслу коэффициента уверенности для пары сравниваемых монооценок. Как и ранее, ищутся условия, при которых P_2 может быть предпочтительнее («больше»), чем P_1 .

Рассмотрим теперь возможные варианты взаимного расположения ПИО P_1 и P_2 в их паре. Конфигурация с $D_{1dr} < D_{2dl}$ не вызывает трудностей: при любых возможных будущих реализациях значений d_1 и d_2 в ПИО P_1 и P_2 соответственно, оценка P_1 не может превзойти P_2 , т.е. всегда $P_2 \succ P_1$. Эта конфигурация – единственная конфигурация, для которой полиинтервальные оценки сравнимы в строгом смысле слова. В этом случае фактически нет задачи принятия решений, поскольку нет неопределенности: оценка P_2 строго больше оценки P_1 .

Вырожденная, в известном смысле, конфигурация двух ПИО с совпадающими вершинами характеризуется наибольшей неопределенностью, оценки эквивалентны друг другу по

предпочтительности. В этой ситуации ЛПР не в состоянии принять рационального решения в пользу какой-либо из оценок без дополнительной информации. Необходимо либо отказаться от выбора, либо, если сделать выбор все-таки нужно, а оперативно получить дополнительную информацию невозможно, воспользоваться случайным выбором. Из этого примера видно, что мерой неопределенности в рассматриваемой задаче сравнения может служить площадь S пересечения $\Pi_1 \cap \Pi_2$.

Таблица 1. Возможные конфигурации пар ПИО.

| № п.п | D_{2ul} | D_{2ur} | D_{2dl} | D_{2dr} |
|-------|-----------|-----------|-----------|-----------|
| 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 1 | 1 | 2 |
| 3 | 1 | 1 | 1 | 3 |
| 4 | 1 | 1 | 2 | 2 |
| 5 | 1 | 1 | 2 | 3 |
| 6 | 1 | 2 | 1 | 2 |
| 7 | 1 | 2 | 1 | 3 |
| 8 | 1 | 2 | 2 | 2 |
| 9 | 1 | 2 | 2 | 3 |
| 10 | 1 | 3 | 1 | 2 |
| 11 | 1 | 3 | 1 | 3 |
| 12 | 1 | 3 | 2 | 2 |
| 13 | 1 | 3 | 2 | 3 |

| № п.п | D_{2ul} | D_{2ur} | D_{2dl} | D_{2dr} |
|-------|-----------|-----------|-----------|-----------|
| 14 | 2 | 2 | 1 | 2 |
| 15 | 2 | 2 | 1 | 3 |
| 16 | 2 | 2 | 2 | 2 |
| 17 | 2 | 2 | 2 | 3 |
| 18 | 2 | 3 | 1 | 2 |
| 19 | 2 | 3 | 1 | 3 |
| 20 | 2 | 3 | 2 | 2 |
| 21 | 2 | 3 | 2 | 3 |
| 22 | 3 | 3 | 1 | 2 |
| 23 | 3 | 3 | 1 | 3 |
| 24 | 3 | 3 | 2 | 2 |
| 25 | 3 | 3 | 2 | 3 |
| 26 | 3 | 3 | 3 | 3 |

Если зафиксировать Π_1 , то, помимо рассмотренных, возможны другие конфигурации, определяемые взаимным положением вершин сравниваемых ПИО. Примеры возможных конфигураций приведены в таблице 1. Пусть цифрой 1 кодируется ситуация, когда в текущей конфигурации анализируемая вершина нижнего основания Π_2 лежит левее D_{1dl} или верхнего основания левее D_{1ul} , цифрой 2 ситуация, когда в текущей конфигурации анализируемая вершина нижнего основания Π_2 лежит в пределах нижнего основания Π_1 или анализируемая вершина верхнего основания Π_2 в пределах верхнего основания Π_1 . Наконец, цифрой 3 ситуация, когда в текущей конфигурации анализируемая вершина нижнего основания Π_2 лежит правее D_{1dr} или верхнего основания правее D_{2ur} .

В каждой из этих конфигураций присутствуют, вообще говоря, три зоны: зона, благоприятствующая реализации события, состоящего в том, что $\Pi_2 \succ \Pi_1$, или зеленая зона; зона, неблагоприятствующая реализации этого события, или красная зона; зона неопределенности, или желтая зона. За меру последней ранее была выбрана площадь S пересечения $\Pi_1 \cap \Pi_2$. Меры первых двух зон это соответствующие площади фигур, образованные боковыми границами и основаниями ПИО. Формула для вычисления коэффициента уверенности K аналогична подобной формуле для коэффициента уверенности при сравнении монооценок: $K = [Sg - Sr] / S(\Pi_1 \cup \Pi_2)$, т.е. K равен доле разности площадей зеленой и красной зон в общей площади сравниваемых полиоценок с учетом их возможного пересечения.

Обозначим коэффициент уверенности в ситуации проверки гипотезы $\Pi_2 \succ \Pi_1$ через $K(\Pi_2, \Pi_1)$, а для проверки противоположной гипотезы $K(\Pi_1, \Pi_2)$. Можно видеть, что $-1 \leq K(\Pi_2, \Pi_1) \leq 1$. Если $K(\Pi_2, \Pi_1) = 1$, то Π_2 строго больше Π_1 , если $K(\Pi_2, \Pi_1) = -1$, то Π_1 строго больше Π_2 . Равенство $K(\Pi_2, \Pi_1) = 0$ означает отсутствие предпочтения. Будем говорить, что Π_2 теоретически предпочтительнее Π_1 , если $K(\Pi_2, \Pi_1) > 0$.

Таким образом, $K(\Pi_2, \Pi_1)$ задает степень предпочтения альтернативы Π_2 альтернативе Π_1 . Поскольку в ситуациях проверки пары противоположных гипотез зеленая и красная зоны меняются местами, то коэффициент уверенности обладает свойством антисимметричности: $K(\Pi_2, \Pi_1) = -K(\Pi_1, \Pi_2)$. При фиксации положения $\Pi_1 K(\Pi_2, \Pi_1)$ возрастает при сдвиге Π_2 как целого вправо и убывает при сдвиге Π_2 влево. Отношение предпочтения со степенью предпочтения $K(\Pi_2, \Pi_1)$ транзитивно: если $K(\Pi_2, \Pi_1) > 0$ и $K(\Pi_3, \Pi_2) > 0$, то $K(\Pi_3, \Pi_1) > 0$.

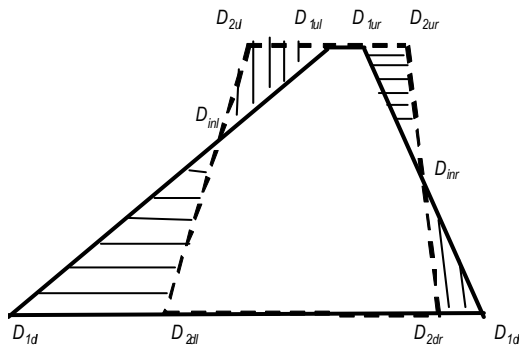


Рис.1. Конфигурация сравниваемых ПИО

треугольники с вершинами $D_{1dl}D_{1il}D_{2dl}$, $D_{1ul}D_{2ul}D_{1ur}$ (горизонтальная штриховка), желтая зона – фигура с вершинами $D_{2dl}D_{1il}D_{1ul}D_{2ul}D_{1ir}D_{2dr}$. Все остальное – это красная зона (вертикальная штриховка). В этом нетрудно убедиться, используя метод сравнения моноинтервалов. Действительно, если для Π_1 и Π_2 реализуются сценарии-моноинтервалы, принадлежащие полосе значений h от 0 до h_{inl} , отвечающего D_{inl} , то для этих сценариев фигура с вершинами $D_{1dl}D_{1il}D_{2dl}$ – это зеленая зона. Графически обсуждаемая конфигурация представлена на рисунке 1. Ситуация здесь сходна с ситуацией при исчислении геометрических вероятностей [Гнеденко, 2007], а именно, в обоих случаях устанавливается взаимно-однозначное соответствие между множеством элементарных исходов, благоприятствующих интересующему нас событию, и множеством точек плоской фигуры, являющейся частью другой (объемлющей) фигуры.

Подсчет площадей зон требует дополнительных усилий и использования элементарной планиметрии. Коэффициент уверенности в методе непосредственного сравнения полиоценок имеет тот же содержательный смысл, что и для метода сравнения монооценок: он характеризует относительный прирост возможного максимального выигрыша за счет верного принятия решений.

Теоретического предпочтения одной полиоценки другой может оказаться недостаточно при принятии решений с использованием коэффициента K : следует учесть восприятие ЛПР риска. Именно, если при проверке гипотезы о предпочтительности оценки Π_2 вычисленное значение коэффициента $K(\Pi_2, \Pi_1)$ для сравниваемой пары полиинтервальных оценок окажется «достаточно большим» и согласуется с представлениями ЛПР о приемлемой величине риска, который измеряется пороговым (индивидуальным для каждого ЛПР и ситуации принятия решений) значением K_{th} , то принимается, что $\Pi_2 \succ \Pi_1$, если $K(\Pi_2, \Pi_1) \geq K_{th}$.

Здесь следует отметить еще одно обстоятельство. При попарном сравнении трех полиинтервальных оценок Π_1, Π_2, Π_3 может случиться так, что для пары Π_1, Π_2 оценка Π_2 предпочтительнее Π_1 с, например, $K = 65\% > K_{th} = 60\%$, а для пары Π_1, Π_3 полиинтервальная оценка Π_3 предпочтительнее Π_1 с $K = 80\% > K_{th}$. Это

Для каждой конкретной конфигурации пары сравниваемых ПИО получаются свои зеленая, красная и желтая зоны. Процедура их нахождения требует пояснения, которое мы проведем на следующем примере. Пусть имеется 12-я, согласно таблице 1, конфигурация сравниваемых ПИО. Таким образом, Π_1 лежит внутри Π_2 , за исключением части Π_1 , лежащей слева от пересечения левых боковых границ ПИО, и части Π_1 , лежащей справа от пересечения правых боковых границ ПИО. Если обозначить точку пересечения левых боковых границ ПИО D_{inl} , а правых D_{inr} , то, при проверке гипотезы о том, что Π_2 предпочтительнее Π_1 , зеленые зоны это

не означает, что для ЛПР P_3 предпочтительнее P_2 , потому что при непосредственном сравнении P_3 и P_2 может оказаться, что P_3 «лучше» P_2 с $K < K_{th}$, т.е. для ЛПР, принявшего для выбора величину K_{th} , эти оценки несравнимы. Таким образом, при принятии решений всегда необходимо прибегать к непосредственному попарному сравнению полиинтервалов-альтернатив.

При сравнении моноинтервальных оценок часто используют полученные методом Гурвица [Hurwicz, 1951] точечные оценки, «эквивалентные» исходным интервальным. Подчеркнем, что такая замена не осуществляется автоматически, а базируется на предпочтениях ЛПР, ее/его склонности к риску. При сравнении полиинтервальных оценок подобные точечные оценки, «эквивалентные» исходным полиинтервальным, также могут быть введены. Однако вместо одного коэффициента Гурвица, определяющего компромисс оптимизма – пессимизма ЛПР, надо ввести как обобщение подхода Гурвица пару коэффициентов. Один, для оси ординат ПИО (оси сценариев), определяет мнение ЛПР о степени неопределенности задачи: чем ближе этот коэффициент к 1, тем уже выбираемый ЛПР интервал-сценарий и тем меньше неопределенность в задаче. Второй коэффициент аналогичен коэффициенту Гурвица на выбранном интервале-сценарии. Вместе эта пара задает искомую точечную оценку.

Хотя точечные оценки должны отражать предпочтения ЛПР, переход от интервальных оценок к точечным не содержит процедур, явным образом отражающих связь получаемых точечных оценок с предпочтениями ЛПР. Недостаточная наглядность такого перехода не позволяет ЛПР в полной мере получить представление об обоснованности выбора той или иной эквивалентной точечной оценки, особенно в случае полиинтервальной оценки. Развитие этого направления потребовало бы создания специальных процедур выявления предпочтений ЛПР в терминах пары обобщенных коэффициентов Гурвица.

Заключение

Предложенный в работе метод непосредственного сравнения полиинтервальных оценок позволяет на основе предпочтений ЛПР либо указать лучшую интервальную оценку из сравниваемых, либо рекомендовать ЛПР временно отказаться от принятия решения из-за несравнимости интервалов и опасности совершить ошибку второго рода. Мы полагаем, что этот метод способствует достаточно корректному выявлению предпочтений ЛПР при сравнении альтернатив, описываемых полиинтервальными оценками. Выделение зеленой, красной и желтой зон цветом в компьютерных системах поддержки экспертных решений позволяет, ввиду наглядности, организовать продуктивный диалог с ЛПР.

В результате исследований, проведенных в этой и в предыдущих работах авторов, разработана совокупность методов сравнения альтернатив, описываемых обобщенными интервальными оценками. Каждый из этих методов применим к задачам с разной степенью неопределенности. Если информация, касающаяся шансов на реализацию сценариев-моноинтервалов, содержащихся в ПИО, отсутствует, следует использовать метод непосредственного сравнения альтернатив, предложенный в настоящей статье. Если подобная информация имеется, следует использовать метод статистических испытаний с разыгрыванием интервалов-сценариев, содержащихся в ПИО, а затем метод сравнения моноинтервальных оценок [Стернин, Шепелев, 2009 а]. Если, наконец, у эксперта дополнительно имеется информация и о шансах на реализацию различных значений в пределах сценариев ПИО, следует использовать метод сведения задачи в полиинтервальной постановке к задаче в моноинтервальной постановке в соответствии с соотношением (1).

Благодарности

Работа поддержана программами фундаментальных исследований президиума РАН «Интеллектуальные информационные технологии, математическое моделирование, системный анализ и автоматизация» и ОНИТ РАН «Информационные технологии и методы анализа сложных систем», Российским фондом фундаментальных исследований (проекты 08-01-00247, 08-07-13532, 09-07-00009, 09-07-12111, 10-07-00242).

Bibliography

- [Гнеденко, 2007] Б. Гнеденко. Курс теории вероятностей. М.: УРСС. 2007.
- [Петровский, 2009] А. Петровский. Теория принятия решений. М: «Академия». 2009.
- [Стернин, Шепелев, 2008] М. Стернин, Г. Шепелев. Анализ сценариев в методе обобщенных интервальных оценок. // Таврический вестник информатики и математики. Т. 2, сс. 195-201. 2008.
- [Стернин, Шепелев, 2009 а] М. Стернин, Г. Шепелев. Сравнение полиинтервальных оценок в методе ОИО. // Intelligent Support of Decision Making. International Book Series N 10. Supplement to Information Technologies and Knowledge. V. 3, pp. 83 – 88. 2009.
- [Стернин, Шепелев, 2009 б] М. Стернин, Г. Шепелев. Агрегирование информации методом обобщенных интервальных оценок на примере задачи о пополнении фондов запасов нефти. // Труды третьей международной конференции «Системный анализ и информационные технологии» САИТ – 2009. М. сс. 349-356. 2009,.
- [Стернин, Шепелев, 2010] М. Стернин, Г. Шепелев. Обобщенные интервальные экспертные оценки в принятии решений. // Доклады академии наук (сер. математика, информатика). Т. 432, сс. 1 – 2. 2010.
- [Chugunov et al., 2008] N. Chugunov, G. Shepelyov, M. Sternin. The generalized interval estimations in decision making under uncertainty. // Int. J. Technology, Policy and Management. V. 8, pp. 298 – 321. 2008.
- [Ferson, Tucker, 2006] S. Ferson, W. Tucker. Sensitivity analysis using probability bounding // Reliability Engineering & System Safety. V. 91, pp. 1435 – 1442. 2006.
- [Hurwicz, 1951] L. Hurwicz. Optimality criteria for decision making under ignorance. 'Cowles Commission Discussion Paper', Statistics, 1951, # 370, New Haven.
- [Shepelyov, Sternin, 2003] G. Shepelyov, M. Sternin. Method of generalized interval estimations for intelligent DSS // DSS in the uncertainty of the Internet age. Proceedings of the conference. Katowice. University of economics. pp. 367-377. 2003
- [Shepelyov, Sternin, 2009] G. Shepelyov, M. Sternin. Aggregation of Information by GIE Method: Model of Supplementing Oil Reserves. // Advances in Decision Technology and Intelligent Information System. K. J. Engemann, G. E. Lasker (Eds.). // The International Institute for Advanced Studies in Systems Research and Cybernetics. V. X, pp. 11 – 15. 2009.
- [Slovic et al, 2001] P. Slovic, B. Fischhoff, S. Lichtenstein. Facts vs. fears: understanding perceived risk. // In Kahneman D., Slovic P., Tversky A (Eds.). Judgment Under Uncertainty: Heuristics and Biases // Cambridge: Cambridge University Press. 2001.

Authors' Information



Михаил Стернин – старший научный сотрудник Института системного анализа РАН. Россия, 117312, Москва, просп. 60-летия Октября, ИСА РАН; e-mail: mister@isa.ru



Геннадий Шепелёв – заведующий лабораторией Института системного анализа РАН. Россия, 117312, Москва, просп. 60-летия Октября, ИСА РАН; e-mail: gis@isa.ru



Алла Рябова – инженер - исследователь Института системного анализа РАН. Россия, 117312, Москва, просп. 60-летия Октября, ИСА РАН; e-mail: ryabova@isa.ru

ПРОЦЕДУРЫ НАХОЖДЕНИЯ СТРОГОГО РЕЗУЛЬТИРУЮЩЕГО ОТНОШЕНИЯ НА МНОЖЕСТВЕ АЛЬТЕРНАТИВ В ЗАДАЧЕ ЛИНЕЙНОГО УПОРЯДОЧЕНИЯ АЛЬТЕРНАТИВ

Павел П. Антосяк, Алексей Ф. Волошин

Аннотация: Разработаны новые процедуры последовательного анализа для задачи линейного упорядочения альтернатив. В результате применения предлагаемых процедур происходит сужение множества возможных вариантов, на котором гарантируется существование оптимального варианта. Работа процедур иллюстрируется на численном примере.

Ключевые слова: задача линейного упорядочения альтернатив, комбинаторная оптимизация, булево программирование, последовательный анализ вариантов.

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

Вступление

Задача линейного упорядочения альтернатив (ЗЛУА) имеет широкое практическое применение в принятии коллективных решений в экономике, теории расписаний, проектировании вычислительной техники и т.п. Она является NP-сложной комбинаторной проблемой, ей посвящены многочисленные публикации [Reinelt,1985, Laguna,1999, Campos,2001].. В [Antosiak,2009] для решения ЗЛУА были предложены процедуры последовательного анализа и отсеивания вариантов, общая идеология которого предложена в [Mikhalevich,1965] и развита в [Mikhalevich,1977, Volkovich, Voloshin,., 1978,1984] для решения разнообразных оптимизационных задач. В частности, для решения различных классов многовариантных задач [Mikhalevich,1977] был разработан широко известный алгоритм «Киевский веник».

В работе [Antosiak, 2009] процедуры одного из эффективных алгоритмов последовательного анализа для решения дискретных оптимизационных задач W [Volkovich, Voloshin,., 1978] были применены для ЗЛУА. Отмечались, в частности, сложности применения алгоритма W в случае формулирования ее как задачи булевого программирования.

В данной работе предлагаются процедуры предварительного сужения множества допустимых вариантов задачи путем фиксации элементов перестановки, которые не могут участвовать в конструировании транзитивных отношения и тем самым образовывать оптимальный вариант.

Постановка задачи

Рассматривается множество альтернатив $A = \{a_1, \dots, a_n\}$ и перестановка $p: A \rightarrow A$. Каждая перестановка $p = (p_1, \dots, p_n)$ однозначно определяет некоторое линейное упорядочение альтернатив. Обозначим через e_{ij} , $i, j \in N = \{1, \dots, n\}$ цену размещения альтернативы a_i перед альтернативой a_j в линейном порядке, а через E квадратную матрицу цен порядка n . Тогда ЗЛВА состоит в нахождении такой перестановки p , при которой достигается максимальная суммарная цена:

$$E(p) = \sum_{i=1}^{n-1} \sum_{j=i+1}^n e_{p_i, p_j} \quad (1)$$

Задачу (1) можно также представить как задачу булевого программирования [Reinelt, 1985]:

$$E(x) = \sum_{i=1}^{n-1} \sum_{j=i+1}^n d_{ij} x_{ij} \rightarrow \max \quad (2)$$

$$x_{ij} + x_{jk} - x_{ik} \leq 1, \quad 1 \leq i < j < k \leq n \quad (3)$$

$$-x_{ij} - x_{jk} + x_{ik} \leq 0, \quad 1 \leq i < j < k \leq n, \quad (4)$$

$$x_{ij} \in X_{ij}, \quad 1 \leq i < j \leq n, \quad (5)$$

где $d_{ij} = e_{ij} - e_{ji}$, $X_{ij} = \{0,1\}$ – множество возможных вариантов значений компоненты x_{ij} , $X = \prod_{i>j} X_{ij}$ – множество всех возможных вариантов.

Процедуры фиксации отношений между двумя альтернативами для ЗЛУА

Обозначим $a_i \succ a_j$, $i, j \in N$, $i \neq j$, если в оптимальном решении (p_1^*, \dots, p_n^*) задачи (1) для некоторых $t < h$ справедливо $p_t^* = i$, $p_h^* = j$ либо для оптимального варианта x^* задачи (2)–(5) выполнено $x_{ij}^* = 1$.

Для каждой альтернативы $a_i \in A$, $i \in N$ введем обозначения:

$$N_i^- = \{z \in N \mid d_{iz} < 0, z \neq i\}, \quad N_i^+ = \{z \in N \mid d_{iz} \geq 0, z \neq i\},$$

$$S_i^- = \sum_{z \in N_i^-} d_{iz}, \quad S_i^+ = \sum_{z \in N_i^+} d_{iz}.$$

Утверждение 1.

1) Если $d_{ij} > -S_i^-$, то $\forall v^* \in \text{Arg max}_p E(p) \Rightarrow a_i \succ a_j$.

2) Если $d_{ij} \geq -S_i^-$, то $\exists v^* \in \text{Arg max}_p E(p) \Rightarrow a_i \succ a_j$.

Доказательство. Пусть $d_{ij} > -S_i^-$. Предположим противоположное. Пусть существует оптимальный вариант v^* такой, что $v^* = (p_1^*, \dots, p_{t-1}^*, j, p_{t+1}^*, \dots, p_{h-1}^*, i, p_{h+1}^*, \dots, p_n^*)$, $1 \leq t < h \leq n$. Для этого варианта выполняются необходимые условия оптимальности задачи (1) [Антосяк, 2008] и, в частности, следующие:

$$\sum_{k=1}^{h-t} d_{ip_{h-k}^*} \leq 0. \quad (6)$$

Своего наименьшего значения сумма в левой части (6) примет, если $\sum_{k=1}^{h-t-1} d_{ip_{h-k}^*} = S_i^-$. Тогда

$$\sum_{k=1}^{h-t} d_{ip_{h-k}^*} = d_{ij} + \sum_{k=1}^{h-t-1} d_{ip_{h-k}^*} \geq d_{ij} + S_i^- > 0,$$

что противоречит (6).

Пусть теперь $d_{ij} \geq -S_i^-$. Выберем произвольный оптимальный вариант v^0 задачи (1) такой, что

$v^0 = (p_1^0, \dots, p_{t-1}^0, j, p_{t+1}^0, \dots, p_{h-1}^0, i, p_{h+1}^0, \dots, p_n^0)$, $1 \leq t < h \leq n$. Возьмем вариант

$v^* = (p_1^*, \dots, p_{t-1}^*, i, j, p_{t+1}^*, \dots, p_{h-1}^*, p_{h+1}^*, \dots, p_n^*)$ и рассмотрим разность:

$$E(v^*) - E(v^0) = \sum_{k=1}^{h-t} d_{ip_{h-k}^*} = d_{ij} + \sum_{k=1}^{h-t-1} d_{ip_{h-k}^*} \geq d_{ij} + S_i^- \geq 0. \quad (7)$$

Из (7) вытекает, что вариант v^* эквивалентен всякому оптимальному варианту задачи (1) по значению целевой функции. □

Следующее утверждение обеспечивает "противоположный" результат.

Утверждение 2.

- 1) Если $d_{ij} < -S_i^+$, то $\forall v^* \in \text{Arg max}_p E(p) \Rightarrow a_j \succ^* a_i$.
- 2) Если $d_{ij} \leq -S_i^+$, то $\exists v^* \in \text{Arg max}_p E(p) \Rightarrow a_j \succ^* a_i$.

Этот факт доказывается аналогично утверждению 1.

Строгое результирующее отношение между двумя альтернативами как следствие использования процедур локализации интервалов изменения оптимальных рангов

В работе [Антосяк, 2008] введено понятие интервала изменения оптимальных рангов (ИИОР) для ЗЛУА.

Пусть $[R_i^{(k_i^{loc})}, r_i^{(k_i^{loc})}]$ интервал, который локализует интервал $[R_i^*, r_i^*]$ изменения оптимальных рангов i -ой альтернативы. Из определений ИИОР и процедур локализации ИИОР очевидным становится следующий результат:

$$\text{если } r_i^{(k_i^{loc})} \leq R_j^{(k_j^{loc})}, \text{ то } \forall v^* \in \text{Arg max}_p E(p) \Rightarrow a_j \succ^* a_i.$$

Из необходимых условий оптимальности для варианта $v^* = (p_1^*, \dots, p_{t-1}^*, j, p_{t+1}^*, \dots, p_{h-1}^*, i, p_{h+1}^*, \dots, p_n^*)$, $1 \leq t < h \leq n$, и определения интервала изменения рангов альтернатив получаем неравенство $1 \leq R_j^{(k_j^{loc})} \leq t < h \leq r_i^{(k_i^{loc})} \leq n$. Возьмем произвольный индекс q , $t < q < h$. Тогда, если $z = i_q^*$, то для ИИОР альтернативы с индексом z выполняется система неравенств:

$$\begin{cases} R_j^{(k_j^{loc})} < R_z^{(k_z^{loc})} < r_i^{(k_i^{loc})}, \\ R_j^{(k_j^{loc})} < r_z^{(k_z^{loc})} < r_i^{(k_i^{loc})}. \end{cases} \quad (8)$$

(8) означает, что для альтернативы a_z еще не установлено $a_z \succ^* a_j$ и $a_i \succ^* a_z$, или в терминах эквивалентной постановки (2)–(5) $|X_{zj}^{(s)}| \neq \{1\}$ и $|X_{iz}^{(s)}| \neq \{1\}$, $s = 0, 1, \dots$. Поэтому в дальнейшем введем в рассмотрение множество индексов альтернатив

$$N_{ij}^- = \left\{ z \in N_i^- : |X_{zj}^{(s)}| \neq \{1\}, |X_{iz}^{(s)}| \neq \{1\}, s = 0, 1, \dots \right\}, \text{ и обозначим } S_{ij}^- = \sum_{z \in N_{ij}^-} d_{iz}.$$

Сказанное делает справедливым следующий результат, который обобщает утверждение 1.

Утверждение 3.

- 1) Если $d_{ij} > -S_{ij}^-$, то $\forall v^* \in \text{Arg max}_p E(p) \Rightarrow a_i \succ^* a_j$.
- 2) Если $d_{ij} \geq -S_{ij}^-$, то $\exists v^* \in \text{Arg max}_p E(p) \Rightarrow a_i \succ^* a_j$.

По аналогии, если для ИИОР альтернативы с индексом $Z = i_q^*$ выполняется система неравенств:

$$\begin{cases} R_i^{(k_i^{loc})} < R_z^{(k_z^{loc})} < r_j^{(k_j^{loc})}, \\ R_i^{(k_i^{loc})} < r_z^{(k_z^{loc})} < r_j^{(k_j^{loc})}, \end{cases} \quad (9)$$

Из (9) следует, что для альтернативы a_z еще не установлено $a_z \succ^* a_i$ и $a_j \succ^* a_z$, или в терминах эквивалентной постановки (2)–(5) $|X_{zi}^{(s)}| \neq \{1\}$ и $|X_{jz}^{(s)}| \neq \{1\}$, $s = 0, 1, \dots$. Далее аналогично, введя в рассмотрение обозначения

$$N_{ij}^+ = \{z \in N_i^+ : |X_{zi}^{(s)}| \neq \{1\}, |X_{jz}^{(s)}| \neq \{1\}, s = 0, 1, \dots\},$$

$$S_{ij}^+ = \sum_{z \in N_{ij}^+} d_{iz},$$

получаем обобщение утверждения 2.

Утверждение 4.

- 1) Если $d_{ij} < -S_{ij}^+$, то $\forall v^* \in \text{Arg max}_p E(p) \Rightarrow a_j \succ^* a_i$.
- 2) Если $d_{ij} \leq -S_{ij}^+$, то $\exists v^* \in \text{Arg max}_p E(p) \Rightarrow a_j \succ^* a_i$.

Далее для отсева значений, которые делают невозможным построение ни одного транзитивного отношения, можно применять процедуру последовательного анализа и отсеивание вариантов W_1 [Antosiak, 2009], которая, в силу выше сказанного, уже не может закончиться аварийно. Проведенные вычислительные эксперименты на данных [LOLIB] показали высокую эффективность предложенных процедур.

Пример

Проиллюстрируем работу алгоритма на примере, сгенерированного случайным образом [Mitchell, 2000], матрицу выигрышей которого в удобной для нас форме будет записано в следующем виде:

| | a_1 | a_2 | a_3 | a_4 | a_5 | a_6 |
|-------|-------|-------|-------|-------|-------|-------|
| a_1 | 0 | 3 | -5 | 7 | 6 | 4 |
| a_2 | -3 | 0 | 4 | 7 | -3 | 2 |
| a_3 | 5 | -4 | 0 | -10 | -4 | 0 |
| a_4 | -7 | -7 | 10 | 0 | -5 | -3 |
| a_5 | -6 | 3 | 4 | 5 | 0 | 6 |
| a_6 | -4 | -2 | 0 | 3 | -6 | 0 |

Результат работы первого шага процедуры фиксации представлено ниже.

| i | j | d_{ij} | N_{ij}^- | $-S_{ij}^-$ | N_{ij}^+ | $-S_{ij}^+$ | Зафиксированное значение \bar{x}_{ij} |
|-----|-----|----------|-----------------|-------------|-----------------|-------------|--|
| 1 | 2 | 3 | {3} | 5 | – | – | – |
| 1 | 3 | -5 | – | – | {2,4,5,6} | -20 | – |
| 1 | 4 | 7 | {3} | 5 | – | – | 1 |
| 1 | 5 | 6 | {3} | 5 | – | – | 1 |
| 1 | 6 | 4 | {3} | 5 | – | – | – |
| 2 | 1 | -3 | – | – | {3,6} | -6 | – |
| 2 | 3 | 4 | {1,5} | 6 | – | – | – |
| 2 | 4 | 7 | {5} | 3 | – | – | 1 |
| 2 | 5 | -3 | – | – | {3,4,6} | -13 | – |
| 2 | 6 | 2 | {1,5} | 6 | – | – | – |
| 3 | 1 | 5 | {2,4,5} | 18 | – | – | – |
| 3 | 2 | -4 | – | – | {1} | -5 | – |
| 3 | 4 | -10 | – | – | {1} | -5 | 0 |
| 3 | 5 | -4 | – | – | {1} | -5 | – |
| 3 | 6 | 0 | {2,4,5} | 18 | {1} | -5 | – |
| 4 | 5 | -5 | – | – | {3} | -10 | – |
| 4 | 6 | -3 | – | – | {3} | -10 | – |
| 5 | 2 | 3 | {1} | 6 | – | – | – |
| 5 | 3 | 4 | {1} | 6 | – | – | – |
| 5 | 4 | 5 | { \emptyset } | 0 | – | – | 1 |
| 5 | 6 | 6 | {1} | 6 | – | – | 1 |
| 6 | 1 | -4 | – | – | { \emptyset } | 0 | 0 |
| 6 | 2 | -2 | – | – | { \emptyset } | 0 | 0 |
| 6 | 3 | 0 | {1,2,5} | 12 | {4} | -3 | – |
| 6 | 4 | 3 | { \emptyset } | 0 | – | – | 1 |

Итак, в результате получим такое сокращенное множество допустимых вариантов:

$$X^{(1)} = \prod_{j>i} |X_{ij}^{(1)}| = \\ = \{0,1\} \times \{0,1\} \times \{1\} \times \{1\} \times \{1\} \times \{0,1\} \times \{1\} \times \{0,1\} \times \{1\} \times \{0\} \times \{0,1\} \times \{0,1\} \times \{0\} \times \{0\} \times \{1\}. \quad (10)$$

Применим к сокращенному множеству (10) процедуру последовательного анализа W_1 . Сначала по критерию

$$\bar{x}_{13} < -\max_{x_{34}^{(1)}} x_{34} + \min_{x_{14}^{(1)}} x_{14}$$

отсеивается значение $\bar{x}_{13} = 0$. А потом последовательно значения $\bar{x}_{23} = 0$, $\bar{x}_{35} = 1$, $\bar{x}_{36} = 1$ по соответствующими критериям отсева:

$$\bar{x}_{23} < -\max_{x_{34}^{(1)}} x_{34} + \min_{x_{24}^{(1)}} x_{24}, \bar{x}_{35} > \max_{x_{34}^{(s)}} x_{34} + \max_{x_{45}^{(s)}} x_{45}, \bar{x}_{36} > \max_{x_{34}^{(s)}} x_{34} + \max_{x_{46}^{(s)}} x_{46}.$$

Таким образом, получаем следующее сокращенное множество допустимых вариантов:

$$X^{(1)} = \{0,1\} \times \{1\} \times \{1\} \times \{1\} \times \{1\} \times \{1\} \times \{1\} \times \{0,1\} \times \{1\} \times \{0\} \times \{0\} \times \{0\} \times \{0\} \times \{0\} \times \{1\}. \quad (11)$$

Применение процедур фиксации на втором шаге к сокращенному множеству (11) дает:

| i | j | d_{ij} | N_{ij}^- | $-S_{ij}^-$ | N_{ij}^+ | $-S_{ij}^+$ | Зафиксированное значение \bar{x}_{ij} |
|---|---|----------|-----------------|-------------|-----------------|-------------|--|
| 1 | 2 | 3 | $\{\emptyset\}$ | 0 | - | - | 1 |
| 2 | 5 | -3 | - | - | $\{\emptyset\}$ | 0 | 0 |

Итак, получаем сокращение множества допустимых вариантов, с единственным допустимым вариантом, который и есть оптимальной перестановкой альтернатив $v^* = (1,5,2,6,4,3)$.

Заключение

Задача линейного упорядочения альтернатив имеет широкое практическое применение в принятии коллективных решений в экономике, теории расписаний, проектировании вычислительной техники и т.п. Она является NP-сложной комбинаторной проблемой, ей посвящены многочисленные публикации. Задача линейного упорядочения альтернатив может непосредственно рассматриваться как оптимизационная комбинаторная задача и как эквивалентная ей задача булевого программирования. Каждый из этих подходов не лишен недостатков, которые в данной работе предлагается преодолевать «комбинаторным» применением идеологии последовательного анализа и отсеивания вариантов – предварительным анализом задачи в комбинаторной постановке с последующим ее решением, как задачи булевого программирования, с помощью алгоритма последовательного анализа, конструирования и отсеивания W [Volkovich, Voloshin, 1978].

Благодарности

Статья частично финансирована из проекта ITNEA XXI Института Информационных теорий и Приложений FOI ITNEA и Консорциума FOIBulgaria (www.itea.org, www.foibg.com).

Библиография

- [Antosiak, 2009] Antosiak P., Voloshin O. Procedures of sequential analysis and sifting of variants for the linear ordering problem / International Book Series "Information science computing", – Volume 7, Number 3, – 2009. - P. 100–105.
- [Reinelt, 1985]. Reinelt G. The linear ordering problem: algorithms and applications. Research and Exposition in Mathematics. –Berlin, Germany: Heldermann Verlag, 1985.
- [Grotschel, 1984] Grotschel M., Jünger M., Reinelt G. A cutting plane algorithm for the linear ordering problem// Operations Research. –1984. –vol. 2.– №6.–P. 1195-1220.
- [Chanas, 1996] Chanas S., Kobylanski P. A new heuristic algorithm solving the linear ordering problem// Computational Optimization and Applications.– 1996.– vol. 6.– P. 191-205.
- [Laguna,1999] Laguna M., Marti R., Campos V. Intensification and diversification with elite tabu search solutions for the linear ordering problem// Computers & Operations Research.–1999. –vol. 26.- P. 1217-1230.
- [Campos, 2001] Campos V., Glover F., Laguna M., Marti R. An experimental evaluation of a scatter search for the linear ordering problem// Journal of Global Optimization. – 2001.– vol. 21.–№ 4.– P. 397-414.
- [Mikhalevich, 1965] Mikhalevich V.S. Consecutive optimization algorithms and their application. I. III// Cybernetics. – 1965. – №1. – P. 45-55; – №2. – P. 85-88 .
- [Mikhalevich, 1977] Михалевич В.С., Шор Н.З. и др. Вычислительные методы выбора оптимальных решений.-Киев: Наукова думка,1977.-178с.
- [Volkovich, Voloshin, 1978] Волкович В.Л., Волошин А.Ф. Об одной схеме метода последовательного анализа и отсеивания вариантов//Кибернетика,1978,№4.-С.99-105.
- [Volkovich, Voloshin,, 1984] Волкович В.Л., Волошин А.Ф. и др. Методы и алгоритмы автоматизированного проектирования сложных систем управления. – Киев: Наукова думка, 1984. – 216 с.
- [Volkovich, Voloshin, 1993] Волкович В.Л., Волошин А.Ф. и др. Модели и методы оптимизации надежности сложных систем.- Киев: Наукова думка,1993.-312с.
- [Антосяк, 2008] Антосяк, П. П. Локалізація інтервалів зміни оптимальних рангів об'єктів у задачі знаходження медіани Кемені – Снелла // Вісник Київського національного університету імені Тараса Шевченка. –2008. – № 8. – С. 4 - 7.
- [Mitchell, 2000] Mitchell John E., Brian Borchers. Solving linear ordering problems with a combined interior point/simplex cutting plane algorithm // High Performance Optimization / Ed. by H. Frenk et al. –Dordrecht, The Netherlands: KluwerAcademicPublishers, 2000. –P. 345–366.

Информация об авторах



Антосяк Павел Павлович –Ассистент, Ужгородский национальный университет, математический факультет. Ужгород, Украина. E-mail: antosp@ukr.net

Основные области научных исследований: коллективное принятие решений, методы дискретной оптимизации, «мягкие» вычисления.



Волошин Алексей Федорович – Профессор, Киевский национальный университет имени Тараса Шевченко, факультет кибернетики. Киев, Украина. E-mail: ovoloshin@unicyb.kiev.ua

Основные области научных исследований: теория принятия решений, методы оптимизации, математическая экономика, системы поддержки принятия решений, экспертные системы, е-образование

МЕТОДОЛОГИЧЕСКИЕ ПРИНЦИПЫ РАСПРЕДЕЛЕНИЕ КВОТ НА ВЫБРОСЫ ПАРНИКОВЫХ ГАЗОВ

Алексей Волошин, Ирина Горицына, Сергей Мащенко

Abstract: *Анализируются проблемы смягчения последствий изменения климата. Выделяются критерии и показатели для распределения квот на выбросы парниковых газов. Предлагаются методики расчета основных показателей и игровые модели выбора компромиссных решений.*

Keywords: *climate protection, emissions of greenhouse gases, global limit on emissions, distribution methods, vector criterium, Nash equilibrium, individual optimal equilibrium.*

ACM Classification Keywords: *I. Computing Methodologies – I.6. Simulation and modelling (Time series analysis) – I.6.8 Types of Simulation – Gaming.*

Введение

Изменение климата представляет собой одну из величайших угроз цивилизации. Сокращение выбросов парниковых газов (ПГ) называют смягчением последствий изменения климата, что было центральным пунктом международных переговоров. Поскольку следующий раунд переговоров фокусируется на вопросе о том, что могут сделать в целях смягчений последствий изменения климата развивающиеся страны, тематика остается в высшей степени актуальной.

Наука сегодня «недвусмысленно считает», что человеческая деятельность способствует изменению климата, и что результаты воздействия уже наблюдаются во всех секторах жизнедеятельности — продовольствие, вода, здравоохранение, сельское хозяйство, энергия и пр.

Современная цивилизация фактически попала в ловушку: если даже США, которые отказались подписать Киотский протокол (КП) и сократить в соответствии с ним выбросы CO₂, присоединятся к нему, согласятся ли сделать это Китай и Индия. Эти два новых могучих индустриальных государства (каждое — с населением свыше миллиарда) переживают период бурного экономического роста. Если эти государства решат затормозить свое развитие, за которым последуют огромные выбросы ПГ, но помогает миллиардам людей вырваться из бедности, то это может привести к социальным потрясениям. А если нынешние тенденции сохранятся, то рост выбросов ПГ приведет к постоянным неурожаем и вызовет голод, который опять-таки вызовет социальные проблемы. Результаты последней конференции ООН по вопросам изменения климата в Копенгагене оказались неутешительными.

Системы критериев и показателей

Политика по климату сама по себе не решит климатическую проблему. Необходима также политика по технологиям, промышленности, сельскому хозяйству, энергии, строительству, транспорту и т.д. Подход к смягчению последствий, ориентированный на развитие, представляет особый интерес для развивающихся стран, где бедность и развитие занимают более приоритетные позиции в повестке дня, чем политика по климату.

Основополагающими принципами в глобальной политике по смягчению последствий изменения климата должны быть: 1) особые обстоятельства развивающихся стран; 2) принятие предупредительных мер (т.е., недостаточная научная определенность не должна использоваться в качестве причины для бездействия); 3) право на содействие устойчивому развитию; 4) устойчивый экономический рост.

Если кто-либо желает выразить в количественном отношении ответственность и возможности, то имеет значение, какая система мер будет использоваться для выражения этих концепций. Числовые результаты у стран будут отличаться, в зависимости от того, рассматриваем ли мы: а) определенные газы (только CO₂ или все шесть газов, перечисленных Киотским протоколом); б) источники выбросов (только энергия или также землепользование, смена землепользования и лесное хозяйство); в) интервал времени (годовые или суммарные выбросы); г) масштаб (национальный или выбросы на душу населения).

Можно принимать во внимание много других критериев, например, богатство страны природными ресурсами или население на квадратный километр, или же использовать различные вариации вышеперечисленных критериев. На приемлемость критерия может влиять то, применяется ли он только к развивающимся странам или ко всем странам. Например, критерии могут включать исторически более длинный период в отношении суммарных выбросов, чем таковые, которые применимы только к развивающимся странам. Взвешивание критериев зависит от того, можно ли использовать рыночные механизмы для выполнения обязательств. Если их использовать можно, то тогда становится более уместной платежеспособность, а потенциал сокращения выбросов становится менее важен, поскольку нет необходимости сокращать внутренние выбросы.

Таблица 1.

Основные индикаторы оценки ответственности, возможности и потенциала сокращения выбросов

| Индикаторы | Единицы измерения |
|--|--|
| Годовые выбросы | Выброс шести ПГ, МтCO ₂ -eq |
| | % общемирового количества |
| Квоты на душу населения | Выбросы на душу населения шести ПГ, МтCO ₂ -eq |
| Историческая ответственность | Суммарные выбросы 1950 — 2000, МтCO ₂ -eq |
| | % общемирового количества |
| Платежеспособность | НВП на душу, Int'l \$, ППП 2000 \$ |
| Потенциал сокращения, интенсивность выбросов | CO ₂ / НВП, кг CO ₂ / int'l \$ НВП ППП |

Существует следующие четыре «школы мышления» или подхода (табл.2):

Таблица 2

| ШКОЛА МЫШЛЕНИЯ | ЗАДАЧА |
|-----------------------------|--|
| Атмосфера в первую очередь | Стабилизация концентрации парниковых газов (ПГ) |
| Равенство в первую очередь | Обеспечение справедливости распределения ответственности за смягчение последствий (исторический вклад) |
| Развитие в первую очередь | Придание развитию большей устойчивости |
| Технология в первую очередь | Развитие и передача технологий с низким выбросом ПГ |

Предлагается огромное разнообразие подходов в определении будущих обязательств. Подход, ставящий «атмосферу первой», может быть выражен как распространение Киотских целей на возможно большее число стран. Выдвижение «равенства на первый» план может означать несколько вещей, по крайней мере две из которых — равные права для каждого и историческая ответственность — отображаются в подходах с использованием показателей на душу населения, а также в Бразильском предложении. В других подходах утверждается, что право на (устойчивое) развитие — это также вопрос равенства. В

самом деле, равенство относится не только к мерам смягчения, но также к адаптации, финансированию и технологиям. Специальные подходы, которые на передний план выдвигают развитие, включают НВП как единицу измерения развития в целях интенсивности, начиная с политики по устойчивому развитию или аспектов чистого развития. В секторальных подходах на передний план выдвигается технология.

Поставленные в КП цели имеют форму согласованного сокращения выбросов в процентном отношении по сравнению с годовыми выбросами в базовом, 1990 году. Подсчитано абсолютное число тонн CO₂, которые необходимо сократить. Начиная с собственных выбросов страны, подход устанавливает квоты на выбросы с учетом существующей между странами разницы в выбросах. Задача для большинства Стран состоит скорее в возвращении к уровню базового года, чем к согласованному сокращению.

Развивающиеся страны отдают предпочтение исчислению ответственности, исходя из показателей на душу населения [UNDP report, 2008]. Права «на душу населения» в качестве отправной точки исходят из равного права каждого жителя планеты на использование атмосферы как глобального достояния. В чистом виде подход на душу населения не рассматривает нынешние уровни выбросов, а скорее глобальный бюджет, выделенный в равной мере для стран с учетом численности их населения. Таким образом, цели в отношении абсолютных выбросов в тоннах CO₂ радикально отличаются от целей в Киотском стиле. Торговля квотами на выбросы осуществляется в большинстве предложений, рассматривающих подходы на душу населения, результатом чего станет увеличение выгод для густонаселенных стран с низкими уровнями выбросов на душу населения. Стоит упомянуть, что Индия и Китай заявили о том, что они не будут рассматривать никакие другие подходы, кроме таковых, основанных на подходе с показателями на душу населения. Этот подход менее привлекателен для менее густонаселенных стран. Бразильское предложение основывается на распределении ответственности исходя из суммарных выбросов за предыдущие десятилетия.

Подход, основанный на учете интенсивности выбросов, требует сокращения количества выбросов, связанных с выходом экономической продукции (ПГ/НВП). Подход допускает рост выбросов в случае экономического роста. С целью учета разных национальных условий обязательства можно сформулировать как уменьшение в процентном отношении интенсивности производимых страной выбросов. Цели в отношении интенсивности выбросов выполнить сложнее в случае, когда экономический рост ниже ожидаемого. В случае успеха, сниженная интенсивность способствует вычитанию выбросов из экономического роста. Такой подход зачастую рассматривается как «более мягкий» в сравнении с абсолютными целями, поскольку он подсчитывает выбросы за относительные сроки.

В международных соглашениях представлено достаточно много предложений по методам распределения квот на выбросы ПГ. Вот основные из них:

- Равные отчисления на душу населения.
- Долгосрочная стабилизация и сближение показателей по странам.
- Дифференцированное сближение.
- Скорректированные отчисления на душу населения.
- Выдача разрешений на выбросы.
- Историческая ответственность (Бразильское предложение).
- Платежеспособность.
- Равные затраты на меры по смягчению последствий.
- Мультисекторальная конвергенция.
- Мультикритериальность

Сотрудники Потсдамского института исследований воздействия на климат разработали "Климатическую стратегию максимума двух градусов" ("2C maxClimateStrategy"): документ посвящен новому подходу к ограничению выбросов, который позволит добиться поставленной в Копенгагене цели. Предлагается ввести глобальное ограничение по объемам выбросов парниковых газов: таким образом, по их словам, можно будет учесть интересы стран, не желающих вводить национальные схемы ограничения выбросов и торговли ими. Этот общий лимит, по замыслу авторов, затем будет распределяться между странами исходя из численности населения.

В качестве первого глобального лимита немецкие эксперты предлагают ввести норматив в 35 миллиардов тонн CO₂-эквивалента в 2015 году. Исходя из данных о численности населения в 2010 году, это соответствует примерно пяти тоннам на одного человека.

Обратим внимание на последний метод распределения квот на выбросы ПГ: мультикритериальность.

Авторы данной работы предлагают положить в основу распределения квот концепцию мультикритериальности и предлагают для обсуждения разработанную ими иерархическую процедуру определения доли в планетарном лимите на выбросы, используя все доступные показатели экономического, экологического, демографического и социального развития стран.

Модели и методы

До сих пор все сложности подписания глобального протокола упирались в определение квот на выбросы ПГ. Алгоритм распределения должен удовлетворить все страны: богатые страны, учитывая их платежеспособность и историческую ответственность, должны взять на себя основное бремя расходов. Наикрупнейшие развивающиеся страны (Индия, Китай) готовы участвовать в мероприятиях, если их доля будет определяться исходя из количества выбросов на душу населения. Бедные страны должны получить такой лимит, чтобы выйти на углеродный рынок с предложением свободных квот. Бедность страны определяется соотношением их ВВП на душу населения со средним значением этой величины по планете. Как только показатель ВВП на душу населения в развивающейся стране превысит на несколько процентов среднемировое значение, страна должна участвовать в мероприятиях по снижению выбросов ПГ. Также ограничение выбросов не должно препятствовать развитию страны, поскольку рост ВВП всегда сопровождается ростом выбросов. В этом процессе главным должен стать показатель динамики выбросов на единицу ВВП: он должен уменьшаться.

Уровень богатства региона/страны можно определить соотношением их величины ВВП на душу населения со среднемировым значением. На сегодня он составляет приблизительно 10 тыс. дол. США (с учетом паритета покупательной способности - ППС).

С учетом вышеизложенного выделим два основополагающих критерия: численность населения и выбросы ПГ в абсолютном исчислении (рис.1, таб.3).

Введем следующие обозначения:

EQ_i - квота на выбросы ПГ в i -ом регионе (стране);

GEL - глобальный лимит выбросов ПГ, млрд.т CO₂ eq.;

SHP_i - удельный вес населения i -го региона(страны),%;

SHE_i - удельный вес выбросов ПГ i -го региона(страны), %;

NPC_i - ВВП на душу населения i -го региона(страны), тыс. \$US (ППС);

ENP_i - выбросы на единицу ВВП i -го региона(страны), т.CO₂ eq. на тыс. \$US (ППС);

EC_i - выбросы на душу населения i -го региона(страны), т;

AVE - выбросы на единицу ВВП в среднем по планете т.CO2 eq. на тыс. \$US (ППС);

$AVNP$ - ВВП на душу населения в среднем по планете тыс. \$US (ППС);

$AVEC$ - средний глобальный выброс на душу населения, т.

Таблица 3

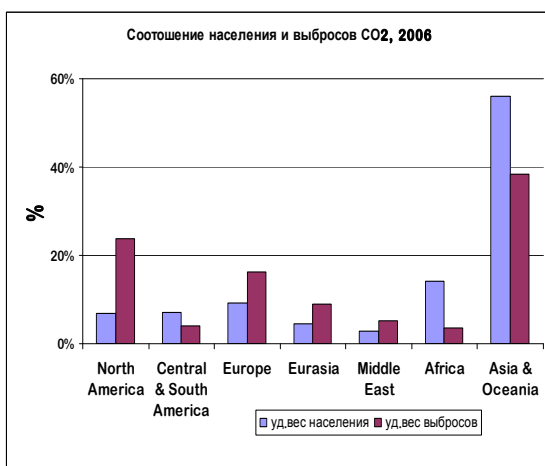


Рисунок 1

| Регионы | Уд.вес населения | Уд.вес выбросов CO2 |
|------------------------|------------------|---------------------|
| Сев. Америка | 6,73% | 23,82% |
| Центр. и Южная Америка | 6,96% | 3,90% |
| Европа | 9,06% | 16,17% |
| Евразия | 4,38% | 8,91% |
| Средний Восток | 2,87% | 5,16% |
| Африка | 14,02% | 3,62% |
| Азия и Океания | 55,98% | 38,43% |

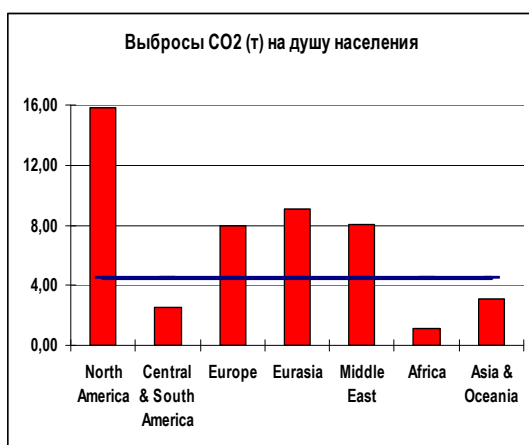


Рисунок 2

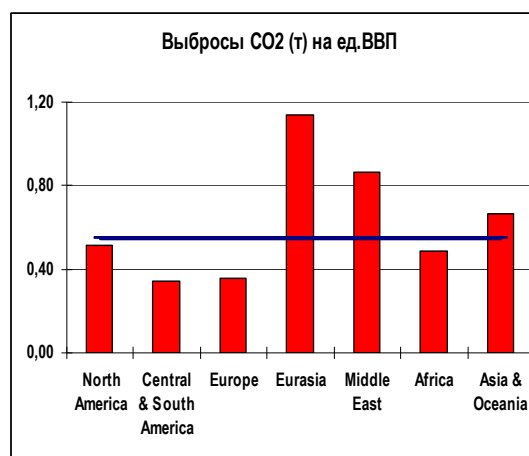


Рисунок 3

С точки зрения исторической ответственности и справедливости Северная Америка должна получить ограничение на выбросы в размере не менее 6,7% от величины мирового лимита на выбросы ПГ. С точки зрения региональной (американской) справедливости квота на выбросы не должна превышать 23,8% мирового лимит. Для Африки, например, историческая и региональная справедливость требует назначить долю в мировом лимите между 3,6% и 14,0%. Таким образом, два глобальных критерия (удельный вес

населения и удельный вес выбросов ПГ) могут обозначить диапазон разрешенных выбросов – $[SHP_i; SHE_i]$, если $SHP_i < SHE_i$, и $[SHE_i; SHP_i]$ в противном случае. Назовем эти критерии критериями первого уровня. Далее введем критерии второго уровня. Ими могут быть: 1) ВВП на душу населения (как индикатор благосостояния общества); 2) выбросы ПГ на единицу ВВП (как индикатор эффективности использования энергетических ресурсов) (рис. 2); 3) выбросы на душу населения (как косвенный индикатор здоровья общества) (рис.3); 4) потребление энергии (как индикатор уровня развития экономики); 5) соотношение темпов прироста выбросов ПГ и темпов прироста ВВП (динамика эластичности выбросов по величине ВВП).

Как видно из рис. 2 и 3, отношение региональных показателей к глобальному среднему (горизонтальная прямая) различно в различных регионах.

С учетом этих показателей и их соотношения с глобальными средними, предлагаем развитым (богатым) странам (ВВП на душу населения в которых превышает средний глобальный показатель в разы) такой алгоритм определения их квоты в выбросах ПГ в рамках обозначенного критериями первого уровня

диапазона: пусть $R_i = \frac{NPC_i}{AVNP}$ -- своего рода коэффициент богатства региона (страны),

$T_i = \frac{ENP_i}{AVE} \cdot \frac{EC_i}{AVEC}$ -- своего рода показатель эффективности экономики (индустриализации страны и

ответственности государства в сфере противодействия внешним эффектам производства). Тогда можно сузить диапазон разрешенных выбросов $[EQ_{\min}; EQ_{\max}]$ до такого:

$$[EQ_{\min}; EQ_{\max}] = [SHP_i \cdot \alpha(R_i); SHE_i \cdot \beta(T_i)], \alpha(R_i) = \ln(R_i + 1), \quad (1)$$

Коэффициент «индустриализации» должен соединить два относительных показателя: выбросы ПГ на душу населения и выбросы ПГ на ед. ВВП и сдвинуть правую границу на некоторую величину

$$\beta(T_i) = \frac{\ln(T_i) + \min \ln(T_i)}{\max \ln(T_i) + \min \ln(T_i)}. \quad (2)$$

Подобное распределение может стать первым приближением в процедуре дальнейших согласований региональных квот. Дальнейшую процедуру согласования квот можно представить в понятиях теории игр.

Основная задача при игровом подходе – определить понятие консенсуса. Как известно [Волошин, 2006], при решении задачи принятия решений (в частности, задачи выбора, рассматриваемой нами) консенсус ищется при выборе метода (алгоритма) либо при определении свойств получаемого решения (в виде аксиом, ограничений, условий и т.п.). В предыдущей работе этих же авторов [Волошин, 2009] рассматривались оба подхода – как выбор способа распределения квот на выбросы, так и требование к получаемому распределению. В данной работе предлагается, как указывалось выше, «иерархическая» процедура принятия решений – на первом уровне определяется метод распределения, на втором – требования к свойствам решения.

Выше отмечалось, что Китай и Индия (суммарно более три миллиарда жителей планеты) готовы участвовать в мероприятиях по реализации КП, если их доля в квотах будет определяться исходя из количества выбросов на душу населения.

Поэтому нами предлагается искать консенсус в выборе соотношения весовых коэффициентов α и β в (1). Например, в соответствии с пропорциональным соотношением количества населения (табл.) при расчете квот Китаю и Индии (при сравнении с Европой – соотношение населения 4:1, доходов на душу населения 8:1) соотношение коэффициентов установить равным $\lambda^* = \frac{\alpha}{\beta} = \ln(4 * 8) \approx 3,5$. На втором уровне

иерархии полученное распределение $EQ = (EQ_i(\lambda^*))_{i=1,n}$ анализируется на предмет удовлетворения его «интегральному» качеству, выражаемым функцией полезности (в частности, векторной [Мащенко, 2004]), учитывающей максимальное количество параметров (см. выше «Системы критериев и показателей»).

Последняя встреча в Копенгагене показала, что наиболее адекватно, по нашему мнению, проблемы КП описывает модель принятия решения в условия полной информации при некооперативном поведении игроков [Волошин, 2006]. В таких моделях наиболее обоснованным является концепция «равновесия по Нэшу» – ситуации, от которой невыгодно отклоняться любому (изолированному) игроку. И самое интересное, что (по нашему мнению) мировое сообщество именно и находится в подобной ситуации! Сказанное можно проиллюстрировать простым примером Самуэльсона [Самуэльсон, 2008], в котором для удобства изменим числовые значения «прибылей» игроков (без потери содержания). Пусть игра описывается табл. 4, где H_1, H_2 – соответственно стратегии игроков 1 и 2, ориентированных на низкий уровень загрязнения; B_1, B_2 – высокий. В ячейках (ситуациях) A–D таблицы указаны прибыли соответственно 1 и 2 игроков. Единственное некооперативным равновесием по Нэшу (ячейка D) оказывается неэффективным – ни одному из агентов не удастся увеличить прибыль, снизив уровень загрязнения. Осуществить переход в ячейку A (с той же прибылью игроков, но низким уровнем загрязнения) возможно (по Самуэльсону) путем вмешательства «правительства» (координирующего органа) установлением штрафов или квот на выбросы. Но никакое правительство не поможет, если в ячейке D значение прибыли поменять, например, на (100, 100)!

| | | 2 | |
|---|-------|-------------|-------------|
| | | H_2 | B_2 |
| 1 | H_1 | A 20; 20 | B -5; 15 |
| | B_1 | C 15; -5 | D 10; 10 |

| | | 2 | |
|---|-------|------------|-----------|
| | | H_2 | B_2 |
| 1 | H_1 | A 10; 7 | B 0; 9 |
| | B_1 | C 8; 0 | D 1; 1 |

| | | 2 | |
|---|-------|---------------|-----------|
| | | H_2 | B_2 |
| 1 | H_1 | A 8.9; 8.1 | B 0; 8 |
| | B_1 | C 8; 0 | D 1; 1 |

Какой же выход из сложившейся ситуации? Во-первых, получение «объективных» значений «выигрышей» (рассчитываемых с учетом параметров, о которых говорилось выше); во-вторых, необходимо отойти от концепции «некооперативного равновесия по Нэшу». Обоснование последнего утверждения вытекает из следующего простого обобщения модели Самуэльсона (табл. 5). В этой игре два равновесия по Нэшу (ячейки A и D), причем в равновесии A выигрыш каждого игрока в 2 раза больший (с учетом низкого загрязнения) и казалось бы, что агенты могли бы договориться о выборе равновесия A. Однако, если от ситуации A риск отклонения каждым из игроков очень высок (теряя всего при этом 25% выигрыша, одновременно принося убытки конкуренту), то отклоняться от ситуации D в принципе не выгодно каждому игроку. Концепция равновесия по Нэшу с учетом эффективности по выигрышу и эффективности по риску принадлежит нобелевским лауреатам за 1994 г. Дж. Харшани и Р. Зельтену [Харшани, 2001]. Но и этот подход не позволяет существенно изменить ситуацию принятия решений при реализации КП, что показывает следующий пример (табл. 5).

Здесь равновесие по Нэшу (ячейка D) единственное, но «очень неэффективное». Ячейка A дает каждому игроку выигрыши существенно большие, однако игроку 2 выгодно отклониться от ситуации A. Как «заставить» игроков выбрать ситуацию A? «Более богатый» игрок 1 может добровольно передать

второму игроку 1,1 единиц своего выигрыша. В результате ситуация А станет равновесной по Нэшу (табл.6)

Пожертвовав 11% выигрыша (или, другими словами, на 11% учтя интересы противоположной стороны) игрок 1 получает 7,9 дополнительной прибыли (на 790% большей по сравнению с ситуацией D!).

Формально концепция «компромисс ради решения конфликта» (каждый игрок выбирает стратегии индивидуально (некооперативно), но учитывает при этом интересы остальных агентов) разработана в [Мащенко, 2007, 2009], где предложен принцип «индивидуальной оптимальности». Он обоснован, в первую очередь, для одноцелевых игр, в которых у всех агентов цель одна, но она характеризуется для каждого своей функцией выигрыша (полезности).

Рассматривается задача коллективного принятия решений в следующей постановке:

$$\begin{aligned} u_i(x) \rightarrow \max, i \in N = \{1, n\}, u \geq 2, \\ x \in X = \prod_{i=1}^n X_i \end{aligned} \quad (3)$$

где u_i – скалярная функция полезности i -го агента; X_i – множество его стратегий; $x = (x_1, \dots, x_n)$ – ситуация принятия решений.

«Объективную» функция полезности каждого агента необходимо конструировать с учетом всех возможных аспектов. Стратегии поведения, и соответственно, ситуации игры определяются выбором принципов распределения квот на выбросы и анализа полученных распределений.

Принцип индивидуальной оптимальности базируется на следующем специальном отношении доминирования по Нэшу.

Определение 1. Ситуация y находится в отношении сильного NE-доминирования агента $i \in N$, если

$$u_j(y_i, x_{N \setminus i}) > u_j(x), \forall j \in N. \quad (4)$$

Обратим внимание, что в (4) анализируется влияние изменения стратегии i -м игроком (с x_i на y_i) на значения функций полезности всех игроков (в частности, i -го, что определяет равновесие по Нэшу).

Определение 2. Ситуация x^* называется слабым индивидуально-оптимальным равновесием, если не существует такого игрока $i \in N$ и ситуации $x \in X$, которая бы сильно доминировала x^* по Нэшу.

Применение слабых индивидуально-оптимальных равновесий мотивируется таким сценарием. Игроки заключают необязательный договор (игроки прислушиваются к научно-обоснованным рекомендациям экспертов) о том, что они будут придерживаться ситуации x^* . В том и только в том случае, когда основанием соглашения выступает слабая индивидуально-оптимальная ситуация, изменения любым игроком $i \in N$ согласованной с другими игроками стратегии x_i^* на другую, всегда будет приводить к ситуации, которая не будет лучшей x^* хотя бы для одного игрока. Поскольку эта ситуация может не быть наилучшей для всех игроков одновременно, то она и является компромиссом.

В [Мащенко, 2009] приводятся необходимые и достаточные условия индивидуальной оптимальности, критерий оценки ее «стабильности», принципе выбора и индивидуально-оптимальных равновесий.

Заключение

В работе проанализировано принципы неэффективности КП, предлагаются концепции, для которых должны базироваться механизмы распределений квот на выбросы. Утверждается, что наиболее адекватным принципом выбора конкретной ситуации принятия коллективных решений при распределении квот на выбросы является принцип индивидуальной оптимальности, соответствии с которым каждый

участник стремится к «компромиссу ради достижения конфликта», выбирая свои стратегии индивидуально (некооперативно), но учитывая при этом интересы остальных участников. Авторы надеются, что приемлемое для всех стран концепция распределения квот на выбросы будет базироваться на изложенном принципе индивидуальной оптимальности и призывают использовать ее для построения адекватных математических моделей.

Благодарности

Статья частично финансирована из проекта ITNEA XXI Института Информационных теорий и Приложений FOI ITNEA и Консорциума FOIBulgaria (www.itea.org, www.foibg.com).

Литература

- [Самуэльсон, 2008] Самуэльсон П., Нордхаус В. Экономика.–М.: Вильямс, 2008.–1360с.
- [Волошин, 2006] Волошин О.Ф., Мащенко С.О. Теорія прийняття рішень.– К.: ВПЦ «Київський ун-т», 2006.– 304 с.
- [UNDPreport, 2008] Переговоры по смягчению последствий изменения климата, с уделением особого внимания опциям для развивающихся стран <http://www.undp.org>
- [Волошин, 2009] Волошина А., Горицына И. Механизмы распределения квот на выбросы по Киотскому протоколу"; In: International Book Series "Information Science and Computing", 2009. -V. 3/2009, N10. - P. 175-181
- [Мащенко, 2004] Мащенко С.О., Павлюченко О.Г. Общие условия векторного равновесия по Нэшу (укр.) // Вестник Киевского университета. Серия: физ.-мат. науки, 2004, №4.– С.212-217.
- [Мащенко, 2007] Мащенко С.О. Исследование стабильности равновесий на основе принципа индивидуальной оптимальности // Кибернетика и системный анализ, 2007, №.– С.162-169.
- [Мащенко, 2009] Мащенко С.О. Индивидуально-оптимальные равновесия некооперативных игр в отношениях предпочтения // Кибернетика и системный анализ, 2009, №1.– С.171-179.
- [Харшаньи, 2009] Харшаньи Дж., Зельтен Р. Общая теория выбора равновесия в играх.–С.-Пб.: Экономическая школа, 2001.–206 с.

Информация об авторах



Волошин Алексей – доктор технических наук, профессор, Киевский национальный университет имени Тараса Шевченко, Украина, 01017 Киев, ул. Владимирская, 64; e-mail: ovoloshin@unicyb.kiev.ua

Сфера научных интересов: принятие решений, системы поддержки принятия решений, математическая экономика, экспертные системы, е-образование



Горицына Ирина– кандидат экономических наук, старший научный сотрудник; Киевский национальный университет имени Тараса Шевченко, Украина, 01017 Киев, ул. Владимирская, 64; e-mail: goritsyna@mail.ru

Сфера научных интересов: эколого-экономическое моделирование, принятие решений, информационные технологии



Мащенко Сергей– кандидат физ.-мат. наук, доцент; Киевский национальный университет имени Тараса Шевченко, Украина, 01017 Киев, ул. Владимирская, 64; e-mail: msomail@yandex.ru

Сфера научных интересов: методы оптимизации, принятие решений, системы поддержки принятия решений, теория игр

ИНСТРУМЕНТАЛЬНЫЙ КОМПЛЕКС АЛГОРИТМОВ И ПРОГРАММ ОПТИМАЛЬНОГО ПРОЕКТИРОВАНИЯ СЕТЕЙ С ТЕХНОЛОГИЕЙ MPLS

Юрий П. ЗАЙЧЕНКО, Елена ЗАЙЧЕНКО

Abstract Software tools for MPLS networks design are elaborated and presented in this work. The software tools are based on original methods and algorithms suggested by authors and includes algorithms for solution numerous tasks of channels capacities assignment, flows distribution, survivability analysis and structural synthesis. The elaborated models and algorithms take into account the specificity of MPLS technology. The results of experimental investigations and practical implementation of the suggested software tools are presented/

Keywords: MPLS technology, capacities assignment/, flows distribution, topological optimization, survivability analysis).

ACM Classification Keywords: A.0 General Literature - Conference C.1.0 General COMPUTER-COMMUNICATION NETWORKS C.2.1 Network Architecture and Design. Network communications:

Введение

Бурный рост сети Интернет, острые потребности пользователей в передаче разнотипных видов информации- аудио, видео и данных, в частности мультимедийной информации, потребовали разработки новой коммуникационной технологии, способной передавать эти виды информации на высоких и сверхвысоких скоростях и обеспечивать при этом требуемое качество обслуживания, различное для разных видов информации. Такой технологией явилась технология многопротокольной коммутации меток (MPLS), которая возникла на рубеже двух веков и в настоящее время интенсивно развивается, охватывая все новые и новые области приложений [1,2]. Технология MPLS стала унифицированным транспортным механизмом для большинства сетевых технологий, в частности, технологий IP, ATM, быстрой передачи кадров FrameRelay, а также технологии локальных сетей-Ethernet.

Специфическими особенностями технологии MPLS являются следующие:

- 1) Наличие различных классов сервиса (ClassofService –CoS) до 8 классов, которые обслуживаются в маршрутизаторах (LSR) с относительными приоритетами.
- 2) Введение показателей качества обслуживания (QoS) , а именно средняя задержка в доставке пакетов $T_{cp,k}$ для разных классов и доля потерянных пакетов – PLR (PacketsLossratio).

Появление новой технологии поставило на повестку дня разработку моделей и методов для анализа и оптимального проектирования сетей MPLS . Такие модели и методы должны учитывать особенности технологии MPLS. В работах авторов был разработан комплекс моделей и методов и инструментальных средств на их основе, для анализа и оптимизации характеристик и структур сетей с технологией MPLS [1,2].

Целью настоящего доклада является краткое изложение основных моделей, методов, а также функций инструментального комплекса для анализа и синтеза сетей “MPLSNetBuilder”.

Комплекс моделей и методов анализа и синтеза сетей с технологией MPLS

- 1) В работах [2,3] разработаны **аналитические модели оценки показателей качества сервиса–средней задержки $T_{cp,k}$** .

$$T_{cp,k} = \frac{1}{H_{\Sigma}^{(k)}} \sum_{(r,s) \in E} \frac{f_{rs}^{(k)} \sum_{i=1}^k f_{rs}^{(i)}}{\left(\mu_{rs} - \sum_{i=1}^{k-1} f_{rs}^{(i)} \right) \cdot \left(\mu_{rs} - \sum_{i=1}^k f_{rs}^{(i)} \right)} \quad (1)$$

и доли потерянных пакетов- PLR (PacketsLossRatio) для различных классов сервиса в сетях MPLS. Эти модели учитывают специфику технологии MPLS и используются для решения многих задач анализа и синтеза сетей MPLS.

В [3] сформулирована **задача распределения потоков** (РП) различных классов в сети MPLS, разработан и исследован алгоритм РП. Данная задача формулируется следующим образом:

Задана сеть с технологией MPLS в виде графа $G = (X, E)$, где $X = \{x_j\}_{j=1, \overline{n}}$ - узлы сети (коммутаторы MPLS) и $E = \{(r, s)\}$ - множество дуг (каналов связи). Имеем возможный набор пропускных способностей (скоростей передачи) каналов связи – $\{\mu_{rs}\}, (r, s) \in E$.

Пусть также введены классы обслуживания (CoS – Classofservice), задана матрица требований для k -го класса $H(k) = \|h_{i,j}(k)\|, k = \overline{1, K}$, где $h_{i,j}(k)$ - интенсивность потока, который необходимо передать от узла i в узел j . Кроме того, введены показатели качества QoS для каждого класса в виде ограничения на среднюю задержку - $T_{зад}(k)$.

Необходимо выбрать маршруты передачи $\{\pi_{ij}\}$ и найти такое распределение потоков всех классов $F(k) = [f_{rs}(k)]$, при которых обеспечивается передача заданных величин потоков со средней задержкой, которая не превышает $T_{зад}(k)$.

В работе исследованы свойства оптимального потока и предложен алгоритм ее решения.

2) В работе [2,3] рассмотрена комбинированная задача **выбора пропускных способностей и распределения потоков (ВПСРП)** в сетях MPLS. Она формулируется так.

Задана сеть MPLS в виде орграфа $G = (X, E)$, где $X = \{x_j\}_{j=1, \overline{n}}$ множество узлов сети, $E = \{(r, s)\}$ - множество каналов связи (КС), набор пропускных способностей (ПС) каналов $D = \{d_1, d_2, \dots, d_k\}$ и их удельных стоимостей $C = \{c_1, c_2, \dots, c_k\}$. Заданы также матрицы требований входящих потоков соответствующих классов $H = \|h_{ij}^{(k)}\|, i, j = \overline{1, n}, k = \overline{1, K}$ и ограничения на среднюю задержку для классов потоков $T_{cp,k}$,

Требуется выбрать такие ПС каналов связи $\{\mu_{rs}^{(0)}\}$ и найти распределение потоков всех классов $F(k) = [f_{rs}(k)]$, при которых стоимость сети будет минимальной, а установленные ограничения на задержки по классам будут выполняться полностью. Математическая модель данной задачи будет иметь следующий вид:

$$\text{Найти, } \min C_{\Sigma} = \sum_{(r,s) \in E} C_{rs}(\mu_{rs}) \quad (2)$$

при ограничениях

$$T_{cp,k}(F(k), \mu_{rs}) \leq T_{зад,k}, k = \overline{1, K} \quad (3)$$

В работе [2,3] предложен и исследован метод решения задачи ВПСПП для сетей MPLS.

3) Важным показателем эффективности функционирования сетей является живучесть. В работах авторов были впервые предложены **показатели живучести** (ПЖ) сетей с технологией MPLS. и разработан метод анализа и оптимизации сетей по этим показателям.

Поскольку основное назначение сети с технологией MPLS является передача заданных величин входящих потоков различных классов, то *живучесть сети MPLS предлагается оценивать величиной максимального потока, который возможно передать в сети при отказах ее элементов- каналов и узлов при сохранении заданных показателей качества.*

В работах [1] для анализа показателей живучести сетей MPLS был введен следующий комплексный показатель;

$$P\{H_{\Sigma}^{\Phi}(1) \geq r\% H_{\Sigma}^0(1)\}, P\{H_{\Sigma}^{\Phi}(2) \geq r\% H_{\Sigma}^0(2)\} \dots P\{H_{\Sigma}^{\Phi}(k) \geq r\% H_{\Sigma}^0(k)\} \quad (3.2)$$

где $H_{\Sigma}^0(k)$ - величина потока k -го класса в безотказном состоянии сети (номинального потока);

$H_{\Sigma}^{\Phi}(k)$ - фактическая величина потока класса k в случае действия отказов, $r=(50 \div 100)$, $k = \overline{1, K}$.

Таким образом, показатели живучести определяют распределение вероятностей передачи заданной доли номинального потока при отказах ее элементов. Предложен алгоритм оценки комплексного показателя живучести для сетей MPLS и проведены его исследования для различных примеров сетей.

4) Важной задачей завершающего этапа процесса проектирования сетей является **задача структурного синтеза сетей по критерию стоимости** при ограничениях на показатели QoS и живучести.

В работах [4] были впервые сформулирована постановка задача структурного синтеза сетей MPLS. предложен и исследован генетический алгоритм синтеза структуры .

Задано множество узлов сети $X = \{x_j\} j = \overline{1, n}$ - маршрутизаторов MPLS (так называемых LSR-LabelSwitchingRouters), их размещение по территории региона, набор пропускных способностей каналов связи $D = \{d_1, d_2, \dots, d_k\}$, из которых ведется синтез, их удельных стоимостей на ед. длины $C = \{c_1, c_2, \dots, c_k\}$, определены классы обслуживания CoS (ClassofService), известны матрицы входящих требований для k -го класса $H(k) = \|h_{ij}(k)\| i, j = \overline{1, n}; k = 1, 2, \dots, K$, где $h_{ij}(k)$ – интенсивность k -го класса, который необходимо передавать из узла i в узел j за единицу времени, Кроме того, введены ограничения на показатели качества (QoS) для каждого класса k в виде ограничения на среднюю задержку $T_{зад,k}$, $k = \overline{1, K}$.

Требуется найти структуру сети в виде набора каналов связи (КС) $E = \{(r, s)\}$, выбрать пропускные способности (ПС) каналов связи $\{\mu_{rs}\}$ и найти распределение потоков всех классов $F(k) = [f_{rs}(k)]$, таким образом, чтобы обеспечить передачу требований всех классов $H(k)$ в полном объеме и с

задержками T_{cp} , не превышающими заданные $T_{зад,k}$ и при этом бы выполнялись ограничения на долю потерянных пакетов CLP_k , а стоимость сети была бы при этом минимальной [1,4].

В работах [1,4] предложен и исследован генетический алгоритм синтеза структуры сетей MPLS.

Структура и функции инструментального программного комплекса “MPLSNetBuilder”

Программный комплекс “MPLS NetBuilder” состоит из следующих функциональных модулей [1]:

- Модуль оценки показателей качества (QoS), а именно, средней задержки $T_{cp,k}$ и доли потерянных пакетов PLR_k .
- Модуль анализа и оптимизации функциональных характеристик сетей MPLS, включающий программы РП и ВПС РП.
- Модуль анализа живучести сетей с технологией MPLS, включающий программы нахождения максимального потока (НМП) в сети при отказах её элементов.
- Модуль структурного синтеза сетей MPLS.

Информационные взаимосвязи между модулями приведены на рис 1.

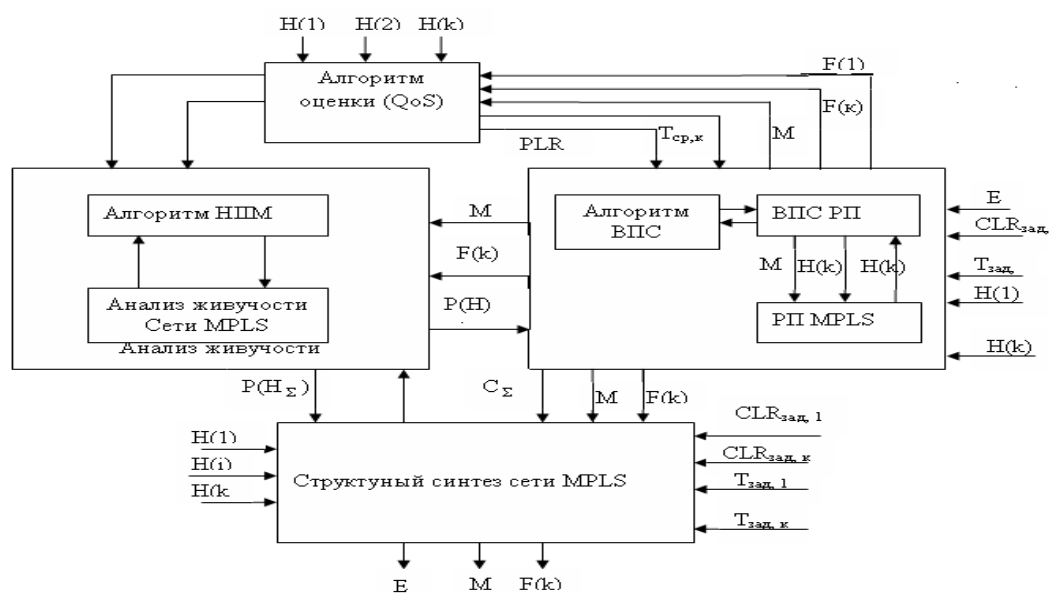


Рис.1. Структурная схема инструментального комплекса анализа и синтеза сетей “MPLS NetBuilder”

Кроме основных функциональных модулей, в состав комплекса входят служебные программные модули, включающие программный модуль интерфейса с пользователем – проектировщиком.

Этот программный модуль позволяет выводить синтерованную структуру в виде карты, осуществлять процедуры проектирования структуры в интерактивном режиме, выводить по указанию пользователя на экране монитора распределения потоков и маршруты передачи, характеристики каналов связи, распределения потоков в КС и их загрузки и характеристики узлов.

Для работы программного комплекса “MPLS NetBuilder” разработан и реализован банк данных, состоящий из отдельных баз данных, таких, как:

- База данных требований входящих потоков.

- База данных каналов связи (их стоимостных и функциональных характеристик, пропускных способностей).
- База данных узлов сети с их характеристиками.
- База данных распределения потоков различных классов и др.

Все функциональные программные модули взаимодействуют друг с другом через банк данных.

В качестве СУБД используется MySQLServer.

Программный комплекс “MPLS NetBuilder” позволяет решать широкий комплекс задач анализа, оптимизации и синтеза структуры сетей с технологией MPLS при обеспечении заданных показателей качества средней задержки T_{CP} и доли потерянных ячеек PLR для различных классов сервиса, а также анализировать показатели живучести и оптимизировать структуру сетей MPLS при ограничениях на установленные значения показателей живучести.

По своим функциональным возможностям и характеристикам данный комплекс не имеет аналогов.

Результаты экспериментальных исследований

Проведены широкие экспериментальные исследования методов и алгоритмов инструментального комплекса. Некоторые из них представлены ниже. Алгоритмы РП и ВПСП исследовались на сети Украины, структура которой приведена на рис.2.



Рис. 2

Рассматривалось три класса сервиса. В ходе экспериментов в алгоритме РП увеличивалась матрица требований по передаче соответствующего трафика с помощью умножения на коэффициент k . На этот коэффициент умножаются все требования матрицы $H(k)$ по передаче соответствующего трафика. При увеличении входящих требований по передаче трафика увеличиваются потоки в каналах связи, а, следовательно, увеличивается и задержка по типам сервисов. Зависимость задержки для первого класса (K_1) приводится на рис. 3, а для второго класса (K_2) - на рис. 4.

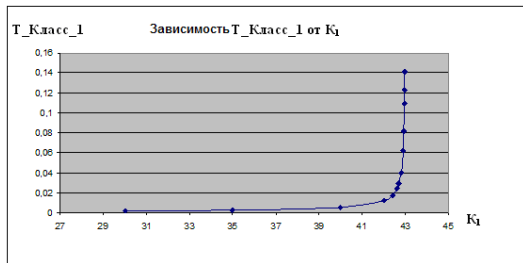


Рис.3

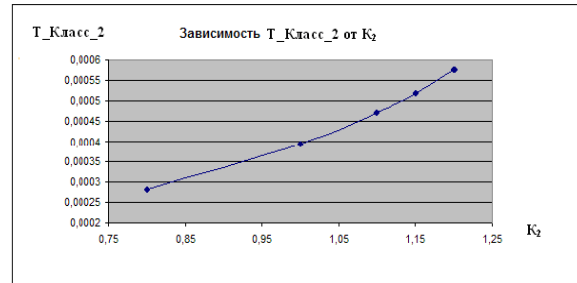


Рис.4

Из представленного графика на рис.3 видно, что время задержки начинает стремительно расти при достижении полной пропускной способности канала связи.

В следующей серии экспериментов исследовался алгоритм ВПСРП. В этих экспериментах исследовались зависимости стоимости сети от вариации ограничений на среднюю задержку для различных классов сервиса. Соответствующая зависимость для класса 2 приводится на рис. 5

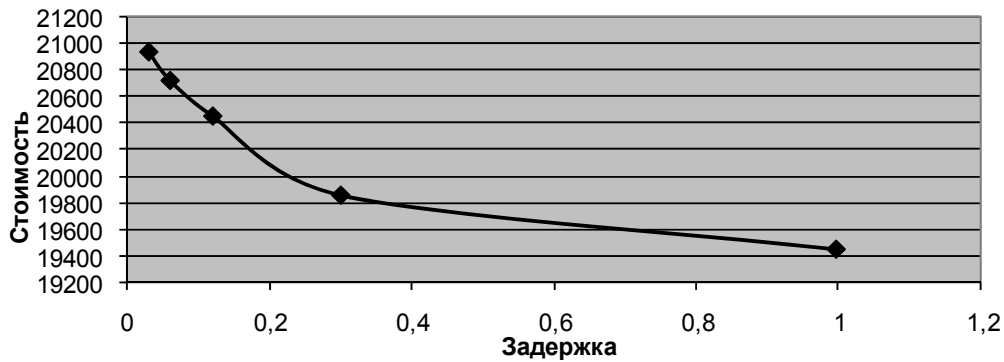


Рис. 5

Как и следовало ожидать, ужесточение требований по средней задержке приводит к увеличению стоимости сети. Это объясняется тем, что для обеспечения заданных значений $T_{ср}$ приходится увеличивать ПС каналов.

В следующем эксперименте исследовалась зависимость стоимости сети от изменения объемов входящих требований. При этом исходная матрица H умножалась на коэффициент пропорциональности K : Соответствующие результаты приведены на рис.6.. Полученный вид зависимости хорошо согласуется с теорией

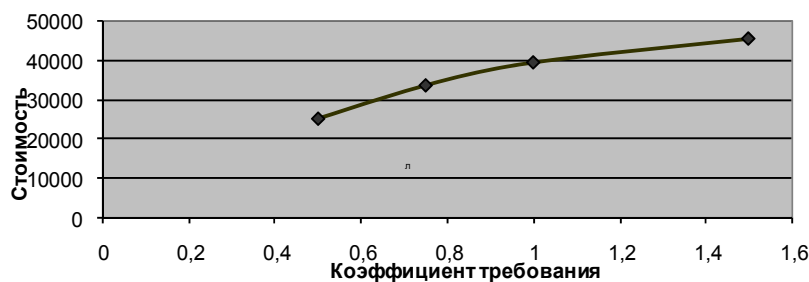


Рис. 6

В последующих экспериментах исследовался генетический алгоритм структурного синтеза сетей, При этом варьировались матрицы входящих требований и ограничения на показатели качества. Анализируя полученные структуры сети при варьировании коэффициента пропорциональности k в диапазоне 0.2-2.0 сделан вывод об *устойчивости базовой структуры (бекбона) сети* в широком диапазоне вариации входящей нагрузки (при десятикратном её увеличении). Это свидетельствует о высокой эффективности разработанного метода структурного синтеза сетей.

Заключение

В докладе описан инструментальный комплекс алгоритмов и программ анализа и синтеза сетей с технологией MPLS. Созданный комплекс предназначен для решения многих задач оптимального проектирования сетей MPLS: распределения потоков, выбора пропускных способностей каналов связи, оценки и анализа показателей живучести и структурного синтеза сетей MPLS.

Проведены широкие экспериментальные исследования алгоритмов анализа и синтеза сетей.

Разработанный комплекс не имеет аналогов и позволяет сократить время проектирования, повысить обоснованность принимаемых проектных решений, снизить капитальные затраты на создание сетей.

Благодарности

Статья частично финансирована из проекта **ITHEA XXI** Института Информационных теорий и Приложений FOI ITHEA и консорциума FOIBulgaria (www.ithea.org, www.foibg.com)

Литература

- 1) Зайченко Е.Ю. Зайченко Ю.П. Сети с технологией MPLS: моделирование, анализ и оптимизация. К.: Изд. «Политехника». 2008.-
- 2) Зайченко Е.Ю. Комплекс моделей и алгоритмов оптимизации характеристик сетей с технологией MPLS// Системні дослідження та інформаційні технології.- 2007.-№ 4. –с. 58-71
- 3) Зайченко Ю.П., Ахмед А.М. Шарадка. Оптимизация распределения потоков в сетях с технологией MPLS с учетом задержки в маршрутизаторах // Электроника и связь. -2007. -№1. -С.84 -87.
- 4) Зайченко Ю.П., Е. Ю. Зайченко, Мохаммадреа Моссавари, Ашраф Абдель-Карим Хилал Абу-Аин. Структурный синтез компьютерных сетей с технологией MPLS// Системні дослідження та інформаційні технології.- 2006.-№ 4.-с.65-70.

Информация об авторах

Зайченко Юрий Петрович, профессор, д.т.н., «Институт прикладного системного анализа». Киев, НТУУ «КПИ», ул. Политехническая 14. тел: +38(044)406-83-93, e-mail: baskervil@voliacable.com

Зайченко Елена Юрьевна, профессор, д.т.н., «Институт прикладного системного анализа». Киев, НТУУ «КПИ», ул. Политехническая 14. тел: +38(044)406-83-93, e-mail: syncmaster@bigmir.net, vojtko@lotus.uz.gov.ua

ПОСТРОЕНИЕ ОПТИМИЗАЦИОННОЙ БАЛАНСОВОЙ МОДЕЛИ С ПРИРОСТНЫМИ КОЭФФИЦИЕНТАМИ

Елена Единак

Abstract: Expansion of classical model of interbranch balance and its transformation in optimization model.

Keywords: Inter-branch balance model, Marginal variables, economic-mathematical modeling.

ACM Classification Keywords: ACM Classification Keywords: I. Computing Methodologies – I.6. Simulation and modeling (Time series analysis) – I.6.8 Types of Simulation – Gaming.

Введение

Классическая модель В. Леонтьева [1] «затраты–выпуск» (межотраслевой баланс) не дает возможности расчета наиболее перспективного развития отраслевой структуры. Расчеты данной модели ведутся по итерационному принципу. Они не определяют наиболее эффективные направления развития как экономики в целом, так и ее структуры. С целью расширения границ использования в классическую модель «затраты–выпуск» вводятся возможные приросты i -го вида экономической деятельности. Ввод приростов ведет к модификации модели В. Леонтьева [1] «затраты–выпуск» и превращает ее в оптимизационную.

Прирост объемов выпуска продукции по промышленным отраслям описывают изменения объемов производства за определенное время. Исходя из условий баланса – прирост хотя бы одного вида экономической деятельности обусловит и прирост всех макроэкономических показателей: валового внутреннего продукта, валовой прибыли, смешанного дохода и т.д.

Построение оптимизационной балансовой модели

Ввод в традиционную балансовую модель В. Леонтьева приростов (уменьшений) объемов выпуска по видам экономической деятельности описывается приростными (маржинальными) переменными (\tilde{x}_i) и соответствующими им коэффициентами (\tilde{a}_{ij}). Эти приросты дают возможность прогнозировать как рост макроэкономических показателей в целом, так и по каждому виду экономической деятельности и отдельным производственным отраслям.

Таким образом, применяя приростные переменные и коэффициенты можно записать оптимизационную балансовую модель с использованием таблиц «затраты–выпуск»:

$$\sum_{j=1}^n c_j x_j + \sum_{j=1}^n \tilde{c}_j \tilde{x}_j \rightarrow \max, \quad (1)$$

при сохранении условий баланса и не отрицательности основных переменных ($i = \overline{1, n}$):

$$x_i + \tilde{x}_i \geq \sum_{j=1}^n a_{ij} x_j + \sum_{j=1}^n \tilde{a}_{ij} \tilde{x}_j + y_i + \tilde{y}_i, \quad (2)$$

$$x_i \geq q_i, \quad (3)$$

$$\tilde{x}_i \leq \lambda_i^{(t)} x_i, \quad (4)$$

$$x_i \geq 0, \quad (5)$$

$$\tilde{x}_i \geq 0, \quad (6)$$

где:

a_{ij} – коэффициенты прямых затрат продукции i на единицу выпуска продукции j ;

\tilde{a}_{ij} – приростные коэффициенты прямых затрат продукции i на дополнительную единицу выпуска продукции;

c_j – коэффициент результативных показателей производства (ВВП, валовая прибыль, смешанные доход) на единицу измерения вида экономической деятельности;

\tilde{c}_j – приростные коэффициенты прироста результативных показателей производства (ВВП, валовой прибыли, смешанного дохода);

x_i – абсолютный размер объема выпуска i -го вида, или интенсивность развития i -го вида экономической деятельности;

\tilde{x}_i – прирост общего объема выпуска i -го вида, или возможный прирост интенсивности развития i -го вида экономической деятельности;

x_j – объемы выпуска продукции данного вида экономической деятельности;

\tilde{x}_j – прирост объема выпуска продукции данным видом экономической деятельности;

u_i – абсолютный размер объема конечного потребления i -го вида продукции;

\tilde{u}_i – прирост объемов конечного потребления i -го вида продукции;

q_i – объемы производства, достигнутые i -м видом экономической деятельности в базовом году;

$\lambda_{i(t)}$ – норма возможного прироста объемов производства i -й отрасли за t -ый период.

Эта модель линейна и проста в вычислениях – сводится к задаче линейного программирования, и, во-вторых, для нее имеется статическая информация.

Так же отметим, что существует принципиальная возможность включения в модель, как дополнительных целевых функций, так и ограничений. Данная модель легко превращается в эколого-экономическую при внесении соответствующих целевых функций и ограничений.

Методика расчетов приростных коэффициентов

Целью данной работы является описание методики расчетов приростных коэффициентов. Рассмотрим подробнее формирование данных коэффициентов и переменных, которым они соответствуют.

Как известно изменение коэффициентов прямых затрат a_{ij} связано с технологическими изменениями [1].

Однако, как показывает практика, в определенных производственных отраслях прирост объемов производства обеспечивается только лишь за счет простого увеличения объема выпуска продукции или.

При этом коэффициент a_{ij} становится нулевым. Величины приростных коэффициентов \tilde{a}_{ij} могут быть, как положительными, так и отрицательными. Отрицательные коэффициенты означают, что дополнительный прирост выпуска продукции сопровождается сокращением величины удельных затрат в расчете на дополнительную единицу выпуска продукции. И наоборот, положительный коэффициент означает, что данные затраты возрастают.

Что касается, результативных макроэкономических показателей, то положительные коэффициенты свидетельствуют о прирост эффективности функционирования данного вида экономической деятельности. Отрицательные – о снижении эффективности.

Величина приростных коэффициентов прямых затрат, а также результативных показателей выпуска продукции, предложено рассчитывать как средневзвешенные величины абсолютных приростов прямых затрат (результативных показателей) за весь период наблюдений к объемам выпуска продукции за этот период:

$$\tilde{a}_{ij} = \frac{\bar{x}_{ij}^{(t)} - \bar{x}_{ij}^{(0)}}{\bar{x}_j^{(t)} - \bar{x}_j^{(0)}} \quad (7)$$

\tilde{a}_{ij} – коэффициент прироста прямых затрат (результативных показателей) в расчете на дополнительную единицу выпуска продукции;

$\bar{x}_{ij}^{(0)}, \bar{x}_{ij}^{(t)}$ – абсолютные величины затрат i -го вида продукции на производство продукции j -м видом экономической деятельности соответственно в базовом и t -м периодах;

$\bar{x}_j^{(0)}, \bar{x}_j^{(t)}$ – объемы производства продукции j -м видом экономической деятельности в соответствующих периодах.

При расчетах приростных коэффициентов \tilde{a}_{ij} за базовый период $\bar{x}_j^{(0)}$ брались объемы выпуска продукции за 2000 г. За t -й период $\bar{x}_j^{(t)}$ – объемы выпуска продукции за 2005 г., соответственно $\bar{x}_{ij}^{(0)}, \bar{x}_{ij}^{(t)}$ – промежуточное потребление за эти соответствующие годы.

При этом сумма промежуточных затрат продукции i -го вида на весь объем производства j -го вида экономической деятельности в t -ом году с учетом приростов объемов производства будет иметь вид:

$$a_{ij}x_j + \tilde{a}_{ij}\tilde{x}_j \quad (8)$$

А затраты на единицу продукции на перспективный период рассчитывается следующим образом:

$$a_{ij}^{(t)} = \frac{a_{ij}x_j + \tilde{a}_{ij}\tilde{x}_j}{x_j + \tilde{x}_j} \quad (9)$$

В зависимости от количества годов, учитываемых в перспективном периоде, изменяются значения $\bar{x}_i^{(t)}$, и как следствие \tilde{y}_i . Аналогично рассчитываются и коэффициенты результативных показателей \tilde{c}_{ij} .

Приведем примеры расчетов приростных коэффициентов прямых затрат. Для упрощения вычислений некоторые виды экономической деятельности были агрегированы. В результате мы имеем 23 вида экономической деятельности, в соответствии им и были рассчитаны приростные коэффициенты прямых затрат и другие необходимые для построения модели коэффициенты.

Анализ расчетов показал, что средние темпы прироста выпуска продукции равны 31,5% (оптимальный вариант функционирования экономики Украины). Наиболее высокие темпы развития у:

- пищевая промышленность – прирост 43,0 %;
- отели и рестораны – 42,0 %;
- операции с недвижимостью – 41,4 %;
- торговля – 41,1 %

Такая ситуация привела к увеличению удельного веса данных видов экономической деятельности в структуре выпуска продукции Украины. Объемы же выпуска продукции большинства важных с точки зрения экономики отраслей увеличиваются только при увеличении потребления их продукции. И как следствие, это приводит к уменьшению удельного веса этих видов экономической деятельности в структуре выпуска продукции в стране.

Величины приростных коэффициентов

| Виды экономической деятельности | Валовая прибыль смешанный доход | ВВП. | Промежуточное потребление |
|---|---------------------------------|-------------|---------------------------|
| 1. Сельское хозяйство, охота и связанные с ними услуги | 0,02929 | 0,00273 | 0,83289 |
| 2. Лесное хозяйство и связанные с ним услуги | 0,8026608 | -0,042128 | -0,110864 |
| 3. Рыбное хозяйство | 0,25 | 0,1011905 | 0 |
| 4. Добыча энергетических материалов | 0,5850521 | 0,1205068 | 5,234728 |
| 5. Добыча неэнергетических материалов | 0,1431 | 0 | 1,23572 |
| 6. Пищевая промышленность, переработка сельскохозяйственных продуктов | 0,08066 | 0 | -0,29082 |
| 7. Текстильная и кожевенная промышленность | 0,09919 | 0 | 0,14296 |
| 8. Деревообрабатывающая и целлюлозно-бумажная промышленность, издательская деятельность | 0,12845 | 0 | -0,04563 |
| 9. Производство кокса, продуктов нефтепереработки и ядерного топлива | 0,04842 | 0 | 0,50522 |
| 10. Химическое производство, резиновые и пластмассовые изделия | 0,0895684 | 0,069299 | 1,267237 |
| 11. Производство неметаллических минеральных продуктов | 0,1105644 | 0,0522246 | 1,8626555 |
| 12. Metallургия и обработка металла | 0,1209293 | 0,0074161 | 3,3345192 |
| 13. Производство машин и оборудования | 0,0994508 | 0,0711469 | -0,100330 |
| 14. Производство электроэнергии, газа и воды | 0,4902276 | 0,1923912 | 3,7802622 |
| 15. Строительство | 0,3030313 | -0,056816 | 0,344542 |
| 16. Торговля | 0,1780667 | 0,0124041 | 0,1828724 |
| 17. Отели и рестораны | 0,1174945 | 0,1374426 | 0,0997407 |
| 18. Транспорт | 0,3097111 | 0,0328594 | 1,7391932 |
| 19. Операции с недвижимостью, сдача в аренду | 0,149284 | 0,0362064 | 0,6754931 |
| 20. Государственное управление | 0,6372434 | 0,0084993 | 0,5017664 |
| 21. Образование | 0,6777306 | 0,0127419 | 0,1048713 |
| 22. Охрана здоровья и социальная помощь | 0,5416412 | 0,018764 | 0,4419063 |
| 23. Другие виды деятельности | 0,276963378 | 0,043977598 | 1,4473*10 ⁻⁰⁶ |

Выводы

Осуществлено расширение классической модели межотраслевого баланса, и построение на ее основе оптимизационной модели. Модель дополнена маргинальными переменными и соответствующими им коэффициентами, что приводит к значительному расширению возможностей выбора рациональной отраслевой структуры экономики при одновременном сохранении ее сбалансированности.

Предложен метод расчетов приростных коэффициентов. В отличие от описанных в литературе методов, он предполагает расчет данных коэффициентов за каждый конкретный период наблюдений. Модель, дополненная приростными переменными и соответствующими им коэффициентами, значительно расширению возможностей выбора рациональной отраслевой структуры экономики.

Bibliography

[Леонтьев В. В., 1997] Леонтьев В. В. Межотраслевая экономика / пер. с англ.; науч. ред. А. Г. Грандберг. – М.: ОАО «Экономика». 1997. – 479 с.

Authors' Information



Елена Едина – кандидат экономических наук, инженер I категории; Киевский национальный университет имени Тараса Шевченко, Украина, 1017 Киев, ул Владимирская, 64; e-mail: elyedin@gmail.com

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ ОЦЕНКИ РИСКА БАНКРОТСТВА КОРПОРАЦИИ

Ольга Прокопенко

Аннотация: рассматривается проблема возможного риска банкротства предприятия, приведен общий обзор существующих методик анализа: метода Альтмана, метода Давыдовой-Беликова и нечеткого матричного метода, проведен сравнительный анализ описанных методов на основании полученных результатов.

Ключевые слова: нечёткая логика, прогнозирование банкротства

ACM Classification Keywords: G.1.0 Mathematics of Computing – NUMERICAL ANALYSIS – General – Error analysis I.2.3 Computing Methodologies - ARTIFICIAL INTELLIGENCE - Deduction and Theorem Proving - Uncertainty, “fuzzy”, and probabilistic reasoning; I.2.6 Computing Methodologies - ARTIFICIAL INTELLIGENCE – Learning - Connectionism and neural nets; I.5.1 Computing Methodologies - PATTERN RECOGNITION - Models - Fuzzy set

Вступление

В данной работе рассматривается проблема возможного риска банкротства предприятия и предлагается разработанная автором методика анализа риска банкротства фирмы с использованием аппарата нечетких множеств и классических методов – как наиболее пригодного математического аппарата для исследования.

Актуальность темы исследования состоит в том, что, несмотря на значительный объем накопленных в данной сфере знаний, процесс банкротства фирмы сегодня досконально не изучен. Не хватает единого системного подхода, с помощью которого можно было бы своевременно идентифицировать угрозу разорения и внедрить эффективные методы предотвращения банкротства предприятия.

Модель Альтмана

Модель Альтмана была разработана в 1968 году на основе статистического анализа финансовых показателей деятельности субъектов предпринимательской деятельности. Эту модель называют Z-моделью. Модель активно используется для экономики США.

Имея результаты хозяйственной деятельности 66 компаний (поровну - 33 успешных и столько же банкротов), рассчитав 22 финансовых коэффициента и воспользовавшись лишь пятью из них, Альтман вывел формулу, которая на сегодняшний день известна всем:

$$Z = 1.2K_1 + 1.4 K_2 + 3.3 K_3 + 0.6 K_4 + 1.0 K_5 \quad (1)$$

где

K1 - рабочий капитал / общая стоимость активов;

K2 - чистая прибыль / общая стоимость активов;

K3 - рыночная капитализация компании (рыночная стоимость акций) / сумма задолженности;

K4 - финансовая устойчивость;

K5 - объем продаж / общая стоимость активов.

Если значение $Z < 1,8$, то вероятность банкротства очень высока. Если значение $1,81 < Z < 2,7$, то вероятность банкротства высока. Если значение $2,71 < Z < 2,99$, то банкротство возможно. Если значение $Z > 3,0$ - вероятность банкротства очень низкая.

Метод Давыдовой-Беликовой

Российские ученые Давыдова и Беликов разработали и предложили свою модель, основанную на модели Альтмана, которая корректирует коэффициенты и параметры согласно условиям переходной экономики. Модель Давыдовой-Беликова имеет вид:

$$R = 8.38 \cdot K_1 + K_2 + 0.054 \cdot K_3 + 0.63 \cdot K_4 \quad (2)$$

где:

$$K_1 = \frac{\text{оборотный капитал}}{\text{сумма активов}}$$

$$K_2 = \frac{\text{чистая прибыль}}{\text{сумма собств. капитала}}$$

$$K_3 = \frac{\text{выручка от реализации}}{\text{сумма активов}}$$

$$K_4 = \frac{\text{чистая прибыль}}{\text{сумма активов}}$$

При $R < 0$ - вероятность банкротства максимальная (90-100%). При $0 < R < 0,18$ - вероятность банкротства высокая (60-80%). При $0,18 < R < 0,32$ - вероятность банкротства средняя (35-50%). При $0,32 < R < 0,42$ - вероятность банкротства низкая (15-20%). При $R < 0,42$ - вероятность банкротства минимальна (до 10%).

Нечетко-множественная модель

В предлагаемой нечетко-множественной модели финансовое состояние фирмы описывается набором количественных и качественных факторов финансового анализа общим числом N . При этом все факторы являются измеримыми, то есть, имеют носитель со своей областью определения на вещественной оси.

Нечеткие описания в структуре модели корпорации появляются в связи с неуверенностью эксперта, которая возникает в ходе классификации уровня факторов. Например, эксперт не может четко разграничить понятия, к примеру, «высокой» и «максимальной» вероятности. Или когда надо провести границу между средним и низким уровнем значения параметра. Тогда применение нечетких описаний означает следующее:

Эксперт фиксирует показатель (фактор) и его количественный носитель.

На выбранном носителе эксперт строит лингвистическую переменную со своим терм-множеством значений. Например, переменная «Уровень показателя X » может обладать терм-множеством значений «Очень низкий, Низкий, Средний, Высокий, Очень высокий».

Далее эксперт каждому значению лингвистической переменной (которое, по своему построению, является нечетким подмножеством значений интервала $(0,1)$ – области значений показателя уровня менеджмента) сопоставляет функцию принадлежности уровня менеджмента тому или иному нечеткому подмножеству.

Общеупотребительными функциями в этом случае являются трапецевидные функции принадлежности. Кроме того, нами в работе были применены треугольные функции.

Для целей компактного описания трапецевидные функции принадлежности $\mu(X)$ удобно описывать трапецевидными числами вида $B(a_1, a_2, a_3, a_4)$, где a_1 и a_4 – абсциссы нижнего основания, a_2 и a_3 – абсциссы верхнего основания трапеции, задающей $\mu(X)$ в области с ненулевой принадлежностью носителя x соответствующему нечеткому подмножеству.

Также формальный ввод классификатора предполагает введение набора узловых точек, которые являются абсциссами максимумов соответствующих функций принадлежности классификатора. В стандартном пятиуровневом классификаторе 5 симметрично расположенных на 01-носителе узловых точек: {0.1, 0.3, 0.5, 0.7, 0.9}.

Продемонстрируем, как применяется полученный классификатор при поэтапном конструировании нечетко-множественной модели корпорации.

На 1-м этапе зададимся лингвистическими переменными и соответствующими им нечеткими подмножествами.

Лингвистическая переменная E «Состояние корпорации» имеет пять значений:

- E_1 – нечеткое подмножество состояний "предельного неблагоприятия";
- E_2 – нечеткое подмножество состояний "неблагополучия";
- E_3 – нечеткое подмножество состояний "среднего качества";
- E_4 – нечеткое подмножество состояний "относительного благополучия";
- E_5 – нечеткое подмножество состояний "предельного благополучия";

Соответствующая переменной E лингвистическая переменная G «Риск банкротства» также имеет 5 значений:

- G_1 – нечеткое подмножество "предельный риск банкротства",
- G_2 – нечеткое подмножество "степень риска банкротства высокая",
- G_3 – нечеткое подмножество "степень риска банкротства средняя",
- G_4 – нечеткое подмножество "низкая степень риска банкротства",
- G_5 – нечеткое подмножество "риск банкротства незначителен".

Носитель множества G – показатель степени риска банкротства g – принимает значения от нуля до единицы по определению.

Для произвольного отдельного финансового или управленческого показателя X_i задаем лингвистическую переменную B_i «Уровень показателя X_i » на нижеследующем терм-множестве значений:

- B_{i1} – подмножество "очень низкий уровень показателя X_i ",
- B_{i2} – подмножество "низкий уровень показателя X_i ",
- B_{i3} – подмножество "средний уровень показателя X_i ",
- B_{i4} – подмножество "высокий уровень показателя X_i ",
- B_{i5} – подмножество "очень высокий уровень показателя X_i ",

На 2-м этапе введем набор отдельных показателей $X = \{X_i\}$ общим числом N , которые, по мнению эксперта-аналитика, с одной стороны, влияют на оценку риска банкротства предприятия, а, с другой стороны, оценивают различные по природе стороны деловой и финансовой жизни предприятия (во избежание

дублирования показателей с точки зрения их значимости для анализа). Примеры выбора системы показателей:

X1 – коэффициент автономии

X2 – коэффициент оборачиваемости всех активов и т.д.

На 3-м этапе сопоставляем каждому показателю X_i уровень его значимости для анализа r_i . Чтобы оценить этот уровень, нужно расположить все показатели по порядку убывания значимости так, чтобы выполнялось правило

$$r_1 \geq r_2 \geq \dots r_N \quad (3)$$

Если система показателей проранжирована в порядке убывания их значимости, то значимость i -го показателя r_i следует определять по правилу Фишберна:

$$r_i = \frac{2(N - i + 1)}{(N + 1)N} \quad (4)$$

Правило Фишберна отражает тот факт, что об уровне значимости показателей неизвестно ничего кроме (4). Тогда оценка (5) отвечает максимуму энтропии наличной информационной неопределенности об объекте исследования, т.е. позволяет лицу, принимающему решение, принимать наилучшие оценочные решения в наихудшей информационной обстановке.

Если же все показатели обладают равной значимостью (равнопредпочтительны или системы предпочтений нет), тогда

$$r_i = \frac{1}{N} \quad (5)$$

На 4-м этапе классифицируем степень риска банкротства. Строим классификатор текущего значения g показателя степени риска как критерий разбиения этого множества на нечеткие подмножества. Этот классификатор является стандартным пятиуровневым классификатором на 01-носителе.

В данной работе для анализа нами были предложены треугольные и трапецевидные функции принадлежности следующего вида:

Трапецевидные ФП

$$\text{ОН:} \quad \mu_1(x) = \begin{cases} 1, & 0 \leq x < 0.1 \\ (0.2 - x)/0.1, & 0.1 \leq x < 0.2 \\ 0, & 0.2 \leq x < 1 \end{cases} \quad (6)$$

$$\text{Н:} \quad \mu_2(x) = \begin{cases} 0, & 0 \leq x < 0.1 \\ (x - 0.1)/0.1, & 0.1 \leq x < 0.2 \\ 1, & 0.2 \leq x < 0.25 \\ (0.3 - x)/0.05, & 0.25 \leq x < 0.3 \\ 0, & 0.3 \leq x < 1 \end{cases} \quad (7)$$

$$\text{С:} \quad \mu_3(x) = \begin{cases} 0, & 0 \leq x < 0.25 \\ (x - 0.25)/0.05, & 0.25 \leq x < 0.3 \\ 1, & 0.3 \leq x < 0.45 \\ (0.5 - x)/0.05, & 0.45 \leq x < 0.5 \\ 0, & 0.5 \leq x < 1 \end{cases} \quad (8)$$

$$B: \mu_4(x) = \begin{cases} 0, & 0 \leq x < 0.45 \\ (x - 0.45) / 0.05, & 0.45 \leq x < 0.5 \\ 1, & 0.5 \leq x < 0.6 \\ (0.7 - x) / 0.1, & 0.6 \leq x < 0.7 \\ 0, & 0.7 \leq x < 1 \end{cases} \quad (9)$$

$$OB: \mu_5(x) = \begin{cases} 0, & 0 \leq x < 0.6 \\ (x - 0.6) / 0.1, & 0.6 \leq x < 0.7 \\ 1, & 0.7 \leq x < 1 \end{cases} \quad (10)$$

Треугольные ФП

$$OH: \mu_1(x) = \begin{cases} 1, & x = 0 \\ (0.2 - x) / 0.2, & 0 \leq x < 0.2 \\ 0, & 0.2 \leq x < 1 \end{cases} \quad (11)$$

$$H: \mu_2(x) = \begin{cases} x / 0.2, & 0 \leq x < 0.2 \\ 1, & x = 0.2 \\ (0.4 - x) / 0.2, & 0.2 \leq x < 0.4 \\ 0, & 0.4 \leq x < 1 \end{cases} \quad (12)$$

$$C: \mu_3(x) = \begin{cases} 0, & 0 \leq x < 0.2 \\ (x - 0.2) / 0.2, & 0.2 \leq x < 0.4 \\ 1, & x = 0.4 \\ (0.6 - x) / 0.2, & 0.4 \leq x < 0.6 \\ 0, & 0.6 \leq x < 1 \end{cases} \quad (13)$$

$$B: \mu_4(x) = \begin{cases} 0, & 0 \leq x < 0.4 \\ (x - 0.4) / 0.2, & 0.4 \leq x < 0.6 \\ 1, & x = 0.6 \\ (1 - x) / 0.4, & 0.6 \leq x < 1 \end{cases} \quad (14)$$

$$OB: \mu_5(x) = \begin{cases} 0, & 0 \leq x < 0.6 \\ (x - 0.6) / 0.4, & 0.6 \leq x < 1 \\ 1, & x = 1 \end{cases} \quad (15)$$

Сопоставив лингвистические переменные «Уровень фактора» (6)-(15) и «Степень риска банкротства предприятия», установим взаимно однозначное соответствие вида табл. 1

Таблица 1.

Соответствие лингвистических переменных

| № | Уровень фактора | Степень риска банкротства предприятия |
|---|-----------------|---------------------------------------|
| 1 | ОН | Запретельная (очень высокая) |
| 2 | Н | Опасная (высокая) |
| 3 | С | Пограничная (средняя) |
| 4 | В | Приемлемая (низкая) |
| 5 | СВ | Незначительная (очень низкая) |

На 5-м этапе классифицируем значения показателей. Сформируем набор классификаторов текущих значений x показателей X , как критерий разбиения полного множества их значений на нечеткие подмножества вида B . При этом в клетках таблицы стоят трапецевидные числа, характеризующие соответствующие функции принадлежности.

Таблица 2.

Классификация степени риска банкротства

| Интервал значений g | Классификация уровня параметров | Степень оценочной уверенности (функция принадлежности) |
|-------------------------|---------------------------------|--|
| $0 \leq g \leq 0,15$ | G5 | 1 |
| $0,15 < g < 0,5$ | G5 | $\mu_5 = 10 * (0,25 - g)$ |
| | G4 | $1 - \mu_5 = \mu_4$ |
| $0,25 \leq g \leq 0,35$ | G4 | 1 |
| $0,35 < g < 0,45$ | G4 | $\mu_4 = 10 * (0,45 - g)$ |
| | G3 | $1 - \mu_4 = \mu_3$ |
| $0,45 \leq g \leq 0,55$ | G3 | 1 |
| $0,55 < g < 0,65$ | G3 | $\mu_3 = 10 * (0,65 - g)$ |
| | G2 | $1 - \mu_3 = \mu_2$ |
| $0,65 \leq g \leq 0,75$ | G2 | 1 |
| $0,75 < g < 0,85$ | G2 | $\mu_2 = 10 * (0,85 - g)$ |
| | G1 | $1 - \mu_2 = \mu_1$ |
| $0,85 \leq g \leq 1$ | G1 | 1 |

На 6-м этапе производим оценку уровней показателей. Для этого оценки текущих уровней показателей сводим в таблицу 3.

Таблица 3.

Текущий уровень показателей

| Наименование показателя | Текущее значение |
|-------------------------|------------------|
| X1 | x1 |
| ... | ... |
| ... | ... |
| Xi | xi |
| ... | ... |
| XN | xN |

7 этап – распознавание уровня показателей на основе набора классификаторов. Результатом проведенной классификации является таблица 4, где λ_{ij} - уровень принадлежности носителя x_i нечеткому подмножеству B_j .

Таблица 4.

Уровни принадлежности носителей нечетким подмножествам

| Наименование показателя | Результат классификации по подмножествам | | | | |
|-------------------------|--|-----|-----|-----|----------------|
| | Bi1 | Bi2 | Bi3 | Bi4 | Bi5 |
| X1 | λ_{11} | ... | ... | ... | λ_{15} |
| ... | ... | ... | ... | ... | ... |
| X_i | λ_{i1} | ... | ... | ... | λ_{i5} |
| ... | ... | ... | ... | ... | ... |
| X_N | λ_{N1} | ... | ... | ... | λ_{N5} |

Тогда степень риска банкротства g :

$$g = \sum_{j=1}^5 g_j \sum_{i=1}^N r_i \lambda_{ij} \quad (16)$$

где

$$g_j = 0.9 - 0.2(j-1) \quad (17)$$

λ_{ij} определяется по таблице 4, а r_i - по формуле (4) либо (4).

Полученное значение сверяют с данными таблицы 2 и делают выводы о степени риска банкротства предприятия.

Экспериментальные исследования

Был проведен сравнительный анализ различных методов оценки риска банкротства предприятия, а именно: оценка риска с помощью методики Альтмана, Давыдовой-Беликова, матричного метода Недосекина. Экспериментальные исследования проводились на основе финансовых данных предприятий Украины 2004-2008 годы.

Для выявления наиболее эффективного метода именно для украинской экономики был проведен анализ полученных результатов. Для этого были рассчитаны средневзвешенные ошибки прогнозирования.

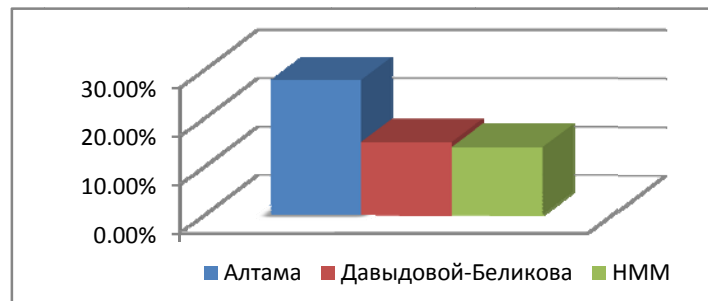


Рис 1. Средневзвешенная ошибка прогнозирования различными методами для 58 компаний (на период 2005 года - за 2 года до банкротства)

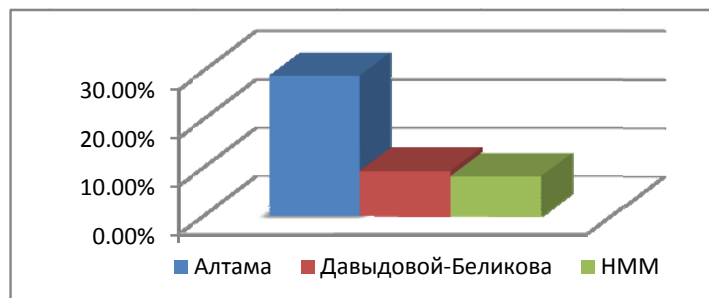


Рис 2. Средневзвешенная ошибка прогнозирования различными методами для 58 компаний (на период 2006 года - за год до банкротства)

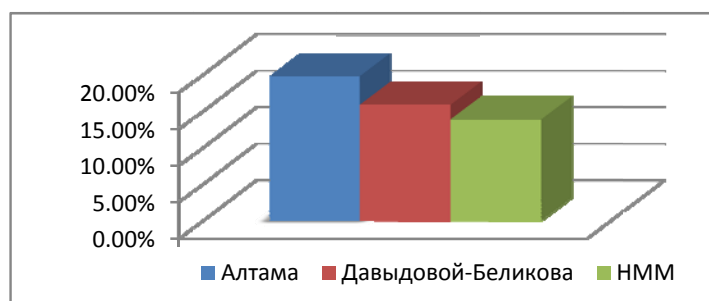


Рис 3. Средневзвешенная ошибка прогнозирования различными методами для 47 компаний (на период 2005 года - за 3 года до банкротства)

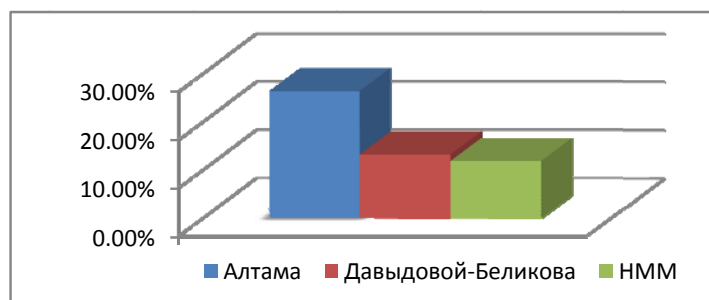


Рис 4. Средневзвешенная ошибка прогнозирования различными методами для 47 компаний (на период 2006 года - за 2 года до банкротства)

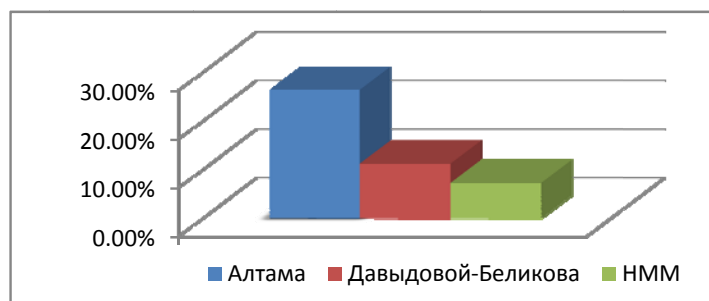


Рис 5. Средневзвешенная ошибка прогнозирования различными методами для 44 компаний (на период 2007 года - за год до банкротства)

Выводы

Лучший результат показал нечеткий матричный метод. Средняя ошибка классификации для рассмотренных групп составила (примерно) 11,2%.

На втором месте стоит метод Давыдовой-Беликова. Средневзвешенная ошибка в целом составляет 12,2%.

Средневзвешенная ошибка метода Альтмана составила 23,7%.

Если ситуация с методикой Альтмана была ожидаемой (ведь известно, что методика разрабатывалась именно для американских компаний), то относительно метода Давыдовой-Беликова, адаптированного к условиям переходной экономики, полученные результаты показывают нецелесообразность применения этой методики для украинских предприятий.

Благодарности

Статья частично финансирована из проекта **ITHEA XXI** Института Информационных теорий и Приложений FOI ITHEA и консорциума FOIBulgaria (www.ithea.org, www.foibg.com)

Литература

1. Зайченко Ю.П. Основы проектування інтелектуальних систем. Навчальний посібник. – К.: Видавничий Дім «Слово», 2004. – С. 352
 2. Зайченко Ю.П. Нечёткие модели и методы в интеллектуальных системах. Учебное пособие для студентов высших учебных заведений. – К.: «Издательский Дом «Слово», 2008. – С. 344
 3. Зайченко Ю.П. Дослідження операцій. Підручник. Сьоме видання, перероблене та доповнене. – К.: Видавничий Дім «Слово», 2006. – С. 816
-

Информация об авторах

Ольга Юрьевна Прокопенко – аспирант Национального технического университета Украины «КПИ»,
адрес электронной почты: prokopenko.o@gmail.com

Основные сферы научных исследований автора: анализ финансового состояния корпорации в условиях неопределенности

ИССЛЕДОВАНИЕ ДВОЙСТВЕННОЙ ЗАДАЧИ ОПТИМИЗАЦИИ ИНВЕСТИЦИОННОГО ПОРТФЕЛЯ В НЕЧЕТКИХ УСЛОВИЯХ

Юрий Зайченко, Ови Нафас Агаи Аг Гамиш

Abstract: *The dual problem of fuzzy portfolio optimization is considered and investigated. A mathematical model of this problem was constructed, explored and the sufficient conditions for its convexity were obtained. The sufficient optimality conditions of its solution are presented. The results of experimental investigations of solutions are presented and discussed*

Keywords: *fuzzy portfolio optimization, dual optimization problem/ convexity sufficient conditions,*

ACM Classification Keywords: *G.1.0 Mathematics of Computing– General – Error analysis; G.1.6 Mathematics of Computing – NUMERICAL ANALYSIS – Optimization - Gradient methods, Least squares methods; I.2.3 Computing Methodologies - ARTIFICIAL INTELLIGENCE - Uncertainty, “fuzzy”, and probabilistic reasoning; I.2.6 Computing Methodologies - ARTIFICIAL INTELLIGENCE – Learning - Connectionism and neural nets;*

Введение

В последние годы проблема оптимизации инвестиционных портфелей представляет значительный интерес в связи с развитием финансовых рынков в Украине и мире. Нахождение оптимального портфеля позволяет инвесторам и финансовым фондам распределять финансовые средства в портфели ценных бумаг с целью получения максимально возможной прибыли и сократить риск ошибочных решений. Особенностью данной проблемы является существенная неопределенность исходной информации относительно доходности ценных бумаг (ЦБ) в будущий момент времени. Новый подход к задаче оптимизации портфеля, который позволяет учесть неопределенность исходных данных и является альтернативой классической модели Марковица, базируется на применении аппарата нечетких множеств. Проблема нечеткой портфельной оптимизации была рассмотрена и исследована в работах [1,2,3]. В этих работах рассматривалась следующая постановка задачи: необходимо оптимизировать ожидаемую доходность портфеля при ограничениях на возможный риск. Алгоритм для решения этой задачи был предложен и исследован в [2]. В работе [3] было предложено использовать прогнозирование доходностей акций, что позволило повысить эффективность получаемых решений.

Целью настоящей работы является рассмотрение двойственной задачи нечеткой портфельной оптимизации, ее исследование и определение достаточных условий выпуклости этой задачи.

Двойственная задача нечеткой портфельной оптимизации

Исходная задача оптимизации нечеткого портфеля, которую естественно называть прямой, имеет следующий вид: [1,2]

Найти ожидаемую доходность нечеткого портфеля

$$\tilde{r} = \sum_{i=1}^N \tilde{r}_i x_i \rightarrow \max \quad (1)$$

при ограничениях на риск

$$\beta(x) \leq \beta_{\text{задан}} \quad 0 < \beta < 1 \quad (2)$$

$$\sum_{i=1}^N x_i = 1 \quad (3)$$

$$x_i \geq 0 \quad (4)$$

Рассмотрим случай, когда критериальное значение доходности r^* удовлетворяет условиям

$$\sum_{i=1}^N x_i r_{i1} \leq r^* \leq \sum_{i=1}^N x_i \tilde{r}_i = \tilde{r} \quad (5)$$

Тогда величина риска равна

$$\beta(x) = \frac{1}{\sum_{i=1}^N x_i r_{i2} - \sum_{i=1}^N x_i r_{i1}} \left[\left(r^* - \sum_{i=1}^N x_i r_{i1} \right) + \left(\sum_{i=1}^N x_i \tilde{r}_i - r^* \right) \ln \left(\frac{\sum_{i=1}^N x_i \tilde{r}_i - r^*}{\sum_{i=1}^N x_i \tilde{r}_i - \sum_{i=1}^N x_i r_{i1}} \right) \right] \quad (6)$$

Рассмотрим двойственную задачу оптимизации нечеткого портфеля относительно задачи (1)-(4)

Минимизировать

$$\beta(x) \quad (7)$$

при условиях

$$\tilde{r} = \sum_{i=1}^N x_i \tilde{r}_i \geq r_{\text{зад}} = r^* \quad (8)$$

$$\sum_{i=1}^N x_i = 1, x_i \geq 0 \quad (9)$$

Требуется доказать, что функция риска $\beta(x)$ является выпуклой, где

$$\beta(x) = \left(A(x) + B(x) \ln \frac{B(x)}{C(x)} \right) * D(x)$$

Для этого необходимо доказать, что функция

$$D(x) = \frac{1}{\sum_{i=1}^N x_i r_{i2} - \sum_{i=1}^N x_i r_{i1}}$$

– выпукла, и функция

$$A(x) + B(x) \ln \frac{B(x)}{C(x)}$$

выпукла. А кроме того обе функции являются убывающими по x_i и неотрицательными, где

$$A(x) = r^* - \sum_{i=1}^N x_i r_{i1}; B(x) = \sum_{i=1}^N x_i \tilde{r}_i - r^*; C(x) = \sum_{i=1}^N x_i \tilde{r}_i - \sum_{i=1}^N x_i r_{i1}$$

Действительно $A(x)$ - линейна и поэтому не строго выпукла, а функции $B(x)$ и $C(x)$ также линейны.

Кроме того $r_i \geq r_{i1}$, $i = \overline{1, N}$, $\sum_{i=1}^N x_i \tilde{r}_i - r^* > 0$; по предположению (условие (8)).

Рассмотрим функцию $D(x)$.

$$\frac{\partial D(x)}{\partial x_i} = - \frac{1}{(r_{i2} - r_{i1})x_i^2} \frac{\partial^2 D(x)}{\partial x_i^2} = \frac{2}{(r_{i2} - r_{i1})x_i^3}$$

И т.к. $r_{i2} \geq r_{i1}$, то $\frac{\partial^2 D(x)}{\partial x_i^2} > 0$, для всех $i = \overline{1, N}$. Следовательно, функция $D(x)$ выпуклая.

Вычислим

$$\begin{aligned} \frac{\partial}{\partial x_i} \left[B(x) \ln \frac{B(x)}{C(x)} \right] &= B'(x) \ln \frac{B(x)}{C(x)} + B(x) \frac{B'(x)C(x) - C'(x)B(x)}{C^2(x)} = \\ &= B'(x) \ln \frac{B(x)}{C(x)} + B'(x) - \frac{C'(x)B(x)}{C(x)} \end{aligned} \quad (10)$$

Поскольку

$$C_j'(x) = \frac{\partial}{\partial x_j} C(x) = \tilde{r}_j - r_{j1} B'(x) = \frac{\partial B(x)}{\partial x_i} = \tilde{r}_i$$

то подставляя в (10), получим:

$$\frac{\partial}{\partial x_i} \left[B(x) \ln \frac{B(x)}{C(x)} \right] = \tilde{r}_i \ln \frac{B(x)}{C(x)} + \tilde{r}_i - (\tilde{r}_i - r_{i1}) \frac{B(x)}{C(x)} \quad (11)$$

Найдем вторую частную производную

$$\begin{aligned} \frac{\partial^2}{\partial x_i^2} \left[B(x) \ln \frac{B(x)}{C(x)} \right] &= \tilde{r}_i \frac{C(x) B'(x) C(x) - C'(x) B(x)}{C^2(x)} \\ &= (\tilde{r}_i - r_{i1}) \frac{B'(x) C(x) - C'(x) B(x)}{C^2(x)} = \\ &= \tilde{r}_i \left(\frac{B'(x)}{B(x)} - \frac{C'(x)}{C(x)} \right) - (\tilde{r}_i - r_{i1}) \left[\frac{B'(x)}{C(x)} - \frac{C'(x) B(x)}{C^2(x)} \right] = \end{aligned} \quad (12)$$

$$= \tilde{r}_i \left[\frac{\tilde{r}_i}{\sum_{i=1}^N x_i \tilde{r}_i - r^*} - \frac{\tilde{r}_i - r_{i1}}{\sum_{i=1}^N x_i \tilde{r}_i - \sum_{i=1}^N x_i r_{i1}} \right] -$$

$$- (\tilde{r}_i - r_{i1}) \left[\frac{\tilde{r}_i}{\sum_{i=1}^N x_i (\tilde{r}_i - r_{i1})} - \frac{(\sum_{i=1}^N x_i \tilde{r}_i - r^*) (\tilde{r}_i - r_{i1})}{(\sum_{i=1}^N x_i (\tilde{r}_i - r_{i1}))^2} \right] =$$

$$- (\tilde{r}_i - r_{i1}) \left[\frac{\tilde{r}_i}{\sum_{i=1}^N x_i \tilde{r}_i - \sum_{i=1}^N x_i r_{i1}} - \frac{(\sum_{i=1}^N x_i \tilde{r}_i - r^*) (\tilde{r}_i - r_{i1})}{(\sum_{i=1}^N x_i (\tilde{r}_i - r_{i1}))^2} \right] = \quad (13)$$

$$\frac{\tilde{r}_i^2}{\sum_{i=1}^N x_i \tilde{r}_i - r^*} - \frac{2\tilde{r}_i (\tilde{r}_i - r_{i1})}{\sum_{i=1}^N x_i (\tilde{r}_i - r_{i1})} + \frac{(\tilde{r}_i - r_{i1})^2 (\sum_{i=1}^N x_i \tilde{r}_i - r^*)}{(\sum_{i=1}^N x_i (\tilde{r}_i - r_{i1}))^2} = \quad (14)$$

После приведения к общему знаменателю получим:

$$\begin{aligned} &= \frac{\tilde{r}_i^2 \left(\sum_{i=1}^N x_i (\tilde{r}_i - r_{i1}) \right)^2 - 2\tilde{r}_i (\tilde{r}_i - r_{i1}) \left(\sum_{i=1}^N x_i \tilde{r}_i - r^* \right) \sum_{i=1}^N x_i (\tilde{r}_i - r_{i1})}{\left(\sum_{i=1}^N x_i \tilde{r}_i - r^* \right) \left(\sum_{i=1}^N x_i (\tilde{r}_i - r_{i1}) \right)^2} + \\ &+ \frac{(\tilde{r}_i - r_{i1})^2 \left(\sum_{i=1}^N x_i \tilde{r}_i - r^* \right)^2}{\left(\sum_{i=1}^N x_i (\tilde{r}_i - r_{i1}) \right)^2 \left(\sum_{i=1}^N x_i \tilde{r}_i - r^* \right)} = \\ &= \frac{\left[\tilde{r}_i \left(\sum_{i=1}^N x_i (\tilde{r}_i - r_{i1}) \right) - (\tilde{r}_i - r_{i1}) \left(\sum_{i=1}^N x_i \tilde{r}_i - r^* \right) \right]^2}{\left(\sum_{i=1}^N x_i \tilde{r}_i - r^* \right) \left(\sum_{i=1}^N x_i (\tilde{r}_i - r_{i1}) \right)^2} \geq 0 \end{aligned} \quad (15)$$

А т.к. $\tilde{r}_i > (\tilde{r}_i - r_{i1})$ и $\sum_{i=1}^N x_i (\tilde{r}_i - r_{i1}) > \sum_{i=1}^N x_i \tilde{r}_i - r^*$, то выражение (15) строго больше 0. Таким образом все частные производные второго порядка

$$\Delta_{ii} = \frac{\partial^2}{\partial x_i^2} \left[B(x) \ln \frac{B(x)}{C(x)} \right] > 0$$

и соответственно

$$\frac{\partial^2}{\partial x_i^2} \left[A(x) + B(x) \ln \frac{B(x)}{C(x)} \right] > 0$$

Теперь необходимо показать, что все диагональные миноры вида

$$\begin{bmatrix} \Delta_{ii} & \Delta_{ij} \\ \Delta_{ji} & \Delta_{jj} \end{bmatrix} = \Delta_{ii}\Delta_{jj} - \Delta_{ij}\Delta_{ji} = \Delta_{ii}\Delta_{jj} - \Delta_{ji}^2 > 0 \quad (16)$$

Эти условия будут достаточные условия выпуклости функции $B(x) \ln \frac{B(x)}{C(x)}$, а следовательно и исходной функции $A(x) + B(x) \ln \frac{B(x)}{C(x)}$.

Вычислим смешанные частные производные $\frac{\partial^2}{\partial x_i \partial x_i} \left[B(x) \ln \frac{B(x)}{C(x)} \right]$

$$\begin{aligned} \frac{\partial^2}{\partial x_i \partial x_i} \left[B(x) \ln \frac{B(x)}{C(x)} \right] &= \frac{\partial}{\partial x_i} \left(B'(x) \ln \frac{B(x)}{C(x)} + B'(x) - \frac{C'(x)B(x)}{C(x)} \right) = \\ &= \tilde{r}_i \frac{C(x)B_j'(x)C(x) - C_j'(x)B(x)}{C^2(x)} - (\tilde{r}_i - r_{i1}) \frac{B_j'(x)C(x) - C_j'(x)B(x)}{C^2(x)} = \\ &= \tilde{r}_i \left(\frac{B_j'(x)}{B(x)} - \frac{C_j'(x)}{C(x)} \right) - (\tilde{r}_i - r_{i1}) \left[\frac{B_j'(x)}{C(x)} - \frac{C_j'(x)B(x)}{C^2(x)} \right] \end{aligned} \quad (17)$$

Где $B_j'(x) = \frac{\partial}{\partial x_j} B(x) = \tilde{r}_j$; $C_j'(x) = \frac{\partial}{\partial x_j} C(x) = \tilde{r}_j - r_{j1}$.

Подставляя эти значения в (17), получим

$$\begin{aligned} &\tilde{r}_i \left(\frac{\tilde{r}_j}{\sum_{i=1}^N x_i \tilde{r}_i - r^*} - \frac{\tilde{r}_j - r_{j1}}{\sum_{i=1}^N x_i (\tilde{r}_i - r_{i1})} \right) - \\ & - (\tilde{r}_i - r_{i1}) \left[\frac{\tilde{r}_j}{\sum_{i=1}^N x_i (\tilde{r}_i - r_{i1})} - \frac{(\tilde{r}_j - r_{j1}) \sum_{i=1}^N x_i \tilde{r}_i - r^*}{\sum_{i=1}^N x_i (\tilde{r}_i - r_{i1})^2} \right] = \\ &= \frac{\tilde{r}_i \tilde{r}_j (\sum_{i=1}^N x_i (\tilde{r}_i - r_{i1}))^2 - \tilde{r}_i (\tilde{r}_j - r_{j1}) (\sum_{i=1}^N x_i \tilde{r}_i - r^*) (\sum_{i=1}^N x_i (\tilde{r}_i - r_{i1}) - \tilde{r}_j (\tilde{r}_i - r_{i1})) -}{(\sum_{i=1}^N x_i \tilde{r}_i - r^*) (\sum_{i=1}^N x_i (\tilde{r}_i - r_{i1})^2)} \\ &= \frac{\sum_{i=1}^N x_i (\tilde{r}_i - r_{i1}) (\sum_{i=1}^N x_i \tilde{r}_i - r^*) + (\tilde{r}_i - r_{i1}) (\tilde{r}_j - r_{j1}) \sum_{i=1}^N x_i (\tilde{r}_i - r_{i1})^2}{(\sum_{i=1}^N x_i \tilde{r}_i - r^*) (\sum_{i=1}^N x_i (\tilde{r}_i - r_{i1})^2)} = \\ &= \frac{\tilde{r}_i \tilde{r}_j (\sum_{i=1}^N x_i (\tilde{r}_i - r_{i1}))^2 - (2\tilde{r}_i \tilde{r}_j - \tilde{r}_i r_{j1} - \tilde{r}_j r_{i1}) (\sum_{i=1}^N x_i (\tilde{r}_i - r_{i1}) (\sum_{i=1}^N x_i \tilde{r}_i - r^*)) +}{(\sum_{i=1}^N x_i \tilde{r}_i - r^*) (\sum_{i=1}^N x_i (\tilde{r}_i - r_{i1})^2)} \\ &+ \frac{(\tilde{r}_i - r_{i1}) (\tilde{r}_j - r_{j1}) (\sum_{i=1}^N x_i \tilde{r}_i - r^*)^2}{(\sum_{i=1}^N x_i \tilde{r}_i - r^*) (\sum_{i=1}^N x_i (\tilde{r}_i - r_{i1})^2)} \end{aligned} \quad (18)$$

Для удобства и сокращения выкладок обозначим знаменатель

$$\sum_{i=1}^N x_i \tilde{r}_i - r^* \sum_{i=1}^N x_i (\tilde{r}_i - r_{i1})^2 = E(x) \quad (19)$$

Подставим выражение для Δ_{ii} и Δ_{ij} из (15) и (18) в (16) и получим

$$\Delta_{ii}\Delta_{jj} - \Delta_{ji}^2 =$$

$$\begin{aligned}
& \frac{[\tilde{r}_i \sum_{i=1}^N x_i (\tilde{r}_i - r_{i1}) - (\tilde{r}_i - r_{i1}) (\sum_{i=1}^N x_i \tilde{r}_i - r^*)]^2}{E^2(x)} * \\
& * \left[\frac{\tilde{r}_j (\sum_{i=1}^N x_i (\tilde{r}_i - r_{i1}) - (\tilde{r}_j - r_{j1}) (\sum_{i=1}^N x_i \tilde{r}_i - r^*))}{E(x)} \right] - \\
& - \frac{\{\tilde{r}_i \tilde{r}_j (\sum_{i=1}^N x_i (\tilde{r}_i - r_{i1})^2 - (2\tilde{r}_i \tilde{r}_j - \tilde{r}_i r_{j1} - \tilde{r}_j r_{i1}) (\sum_{i=1}^N x_i (\tilde{r}_i - r_{i1}) (\sum_{i=1}^N x_i \tilde{r}_i - r^*))\}}{E^2(x)} + \\
& + \frac{(\tilde{r}_i - r_{i1}) (\tilde{r}_j - r_{j1}) (\sum_{i=1}^N x_i \tilde{r}_i - r^*)}{E^2(x)}
\end{aligned} \tag{20}$$

Для дальнейшего упрощения введем обозначения

$$\sum_{i=1}^N x_i (\tilde{r}_i - r_{i1}) = \tilde{r} - r_{min}; \quad \sum_{i=1}^N x_i \tilde{r}_i - r^* = \tilde{r} - r^*. \tag{21}$$

Подставляя их в (20) получим

$$\begin{aligned}
\Delta_{ii} \Delta_{jj} - \Delta_{ji}^2 &= \frac{[\tilde{r}_i (\tilde{r} - r_{min}) - (\tilde{r}_i - r_{i1}) (\tilde{r} - r^*)]^2}{E^2(x)} * \\
& * \left[\frac{\tilde{r}_j (\tilde{r} - r_{min}) - (\tilde{r}_j - r_{j1}) (\tilde{r} - r^*)}{E^2(x)} \right] - \\
& - \frac{\{\tilde{r}_i \tilde{r}_j (\tilde{r} - r_{min})^2 - (2\tilde{r}_i \tilde{r}_j - \tilde{r}_i r_{j1} - \tilde{r}_j r_{i1}) (\tilde{r} - r_{min}) (\tilde{r} - r^*)\}}{E^2(x)} + \\
& + \frac{(\tilde{r}_i - r_{i1}) (\tilde{r}_j - r_{j1}) (\tilde{r} - r^*)}{E^2(x)}
\end{aligned} \tag{22}$$

Далее обозначим

$$\begin{aligned}
\tilde{r}_i (\tilde{r} - r_{min}) - (\tilde{r}_i - r_{i1}) (\tilde{r} - r^*) &= F; \quad \Delta_{ji} = \frac{H}{E} \\
\tilde{r}_j (\tilde{r} - r_{min}) - (\tilde{r}_j - r_{j1}) (\tilde{r} - r^*) &= G
\end{aligned} \tag{23}$$

и подставляем в (23), получим

$$\Delta_{ii} \Delta_{jj} - \Delta_{ji}^2 = \frac{F^2 G^2 - H^2}{E^2} = \frac{(FG - H)(FG + H)}{E^2} > 0 \tag{24}$$

Условие неотрицательности (24) таково:

$$FG - H > 0$$

откуда

$$\begin{aligned}
FG - H &= [\tilde{r}_i (\tilde{r} - r_{min}) - (\tilde{r}_i - r_{i1}) (\tilde{r} - r^*)] [\tilde{r}_j (\tilde{r} - r_{min}) - (\tilde{r}_j - r_{j1}) (\tilde{r} - r^*)] - \\
& - \tilde{r}_i \tilde{r}_j (\tilde{r} - r_{min})^2 - (2\tilde{r}_i \tilde{r}_j - \tilde{r}_i r_{j1} - \tilde{r}_j r_{i1}) (\tilde{r} - r_{min}) (\tilde{r} - r^*) - (\tilde{r}_i - r_{i1}) (\tilde{r}_j - r_{j1}) (\tilde{r} - r^*)^2 \\
& = \tilde{r}_i \tilde{r}_j (\tilde{r} - r_{min})^2 - \tilde{r}_j (\tilde{r}_i - r_{i1}) (\tilde{r} - r^*) (\tilde{r} - r_{min}) - \tilde{r}_i (\tilde{r} - r_{min}) (\tilde{r}_j - r_{j1}) (\tilde{r} - r^*) + \\
& + (\tilde{r}_i - r_{i1}) (\tilde{r}_j - r_{j1}) (\tilde{r} - r^*) - \tilde{r}_i \tilde{r}_j (\tilde{r} - r_{min})^2 + (2\tilde{r}_i \tilde{r}_j - \tilde{r}_i r_{j1} - \tilde{r}_j r_{i1}) (\tilde{r} - r_{min}) (\tilde{r} - r^*) - \\
& - (\tilde{r}_i - r_{i1}) (\tilde{r}_j - r_{j1}) (\tilde{r} - r^*)^2 = \\
& = (\tilde{r} - r_{min}) (\tilde{r} - r^*) [-\tilde{r}_j (\tilde{r}_i - r_{i1}) - \tilde{r}_i (\tilde{r}_j - r_{j1}) - 2\tilde{r}_i \tilde{r}_j - \tilde{r}_i r_{j1} - \tilde{r}_j r_{i1}] = 0
\end{aligned} \tag{25}$$

Итак, мы получили следующие условия

$$\Delta_{ii} = \frac{\partial^2}{\partial x_i^2} \left[B(x) \ln \frac{B(x)}{C(x)} \right] > 0 \text{ для всех } i = \overline{1, N},$$

и кроме того, диагональные миноры $\mu_{i1} = \begin{bmatrix} \Delta_{ii} & \Delta_{ij} \\ \Delta_{ji} & \Delta_{jj} \end{bmatrix} \geq 0$

Это является достаточными условиями выпуклости функции $B(x) \ln \frac{B(x)}{C(x)}$ следовательно и функции

$$A(x) + B(x) \ln \frac{B(x)}{C(x)}.$$

Теперь остается показать, что произведение выпуклых функций $A(x) + B(x) \ln \frac{B(x)}{C(x)}$ и $D(x)$ будет так же выпуклым на интервале $x_i \in [0, 1], i = \overline{1, N}$ с учетом того, что

$$D(x) = \frac{1}{\sum_{i=1}^N x_i (r_{i2} - r_{i1})}$$

где $r_{i2} \geq r_{i1}, x_i \in [0, 1], \sum_{i=1}^N x_i = 1$.

Заметим, что $A(x) + B(x) \ln \frac{B(x)}{C(x)}$ и $D(x)$, как показано выше, положительны и $D(x)$ монотонно убывающая функция, поскольку

$$\frac{\partial D(x)}{\partial x_i} = -\frac{1}{(r_{i2} - r_{i1}) x_i^2} < 0$$

Для удобства обозначим $A(x) + B(x) \ln \frac{B(x)}{C(x)} = \varphi(x)$.

Докажем, что

$$\frac{\partial \varphi}{\partial x_i} = \varphi'(x) < 0$$

$$\begin{aligned} \frac{\partial \varphi}{\partial x_i} &= \frac{\partial}{\partial x_i} \left(A(x) + B(x) \ln \frac{B(x)}{C(x)} \right) = A'(x) + B'(x) + B(x) \frac{C(x) B'(x) C(x) - C'(x) B(x)}{C^2(x)} = \\ &= A'(x) + B'(x) \ln \frac{B(x)}{C(x)} + B'(x) - \frac{B'(x) C(x)}{C(x)} \end{aligned} \quad (26)$$

Подставив значения $A'(x)$ и $B'(x)$ в (26), получим

$$\frac{\partial \varphi}{\partial x_i} = -r_{i1} + \tilde{r}_i \ln \frac{B(x)}{C(x)} + \tilde{r}_i - (\tilde{r}_i - r_{i1}) \frac{B(x)}{C(x)} = \tilde{r}_i \left(1 + \ln \frac{B(x)}{C(x)} - r_{i1} - (\tilde{r}_i - r_{i1}) \frac{B(x)}{C(x)} \right) \quad (27)$$

Поскольку $\frac{B(x)}{C(x)} < 1$, то $-r_{i1} + \tilde{r}_i \frac{B(x)}{C(x)} < 0$

Отсюда после упрощения (27), мы получим

$$\frac{\partial \varphi}{\partial x_i} = \tilde{r}_i \left(1 + \ln \frac{B(x)}{C(x)} - \frac{B(x)}{C(x)} \right) \quad (28)$$

$$1 + \ln \frac{B(x)}{C(x)} - \frac{B(x)}{C(x)} = 1 + \ln \frac{\tilde{r} - r^*}{\tilde{r} - r_{min}} - \frac{\tilde{r} - r^*}{\tilde{r} - r_{min}} \quad (29)$$

Заметим, что $r^* > r_{min} = \sum_{i=1}^N x_i r_{i1}$ и $\tilde{r} > r^*$. Покажем, что выражение (29) меньше 0.

Обозначим $\tilde{r} - r^* = a$,

Тогда $\tilde{r} - r_{min} = \tilde{r} - r^* + (r^* - r_{min}) = a + y$, где $y = r^* - r_{min} > 0$.

Подставляя в (29), получим

$$1 + \ln \frac{\tilde{r} - r^*}{\tilde{r} - r_{min}} - \frac{\tilde{r} - r^*}{\tilde{r} - r_{min}} = 1 + \ln \frac{a}{a+y} - \frac{a}{a+y} \quad (30)$$

Покажем, что

$$\Delta = \ln \frac{a}{a+y} - \frac{a}{a+y} < 0 \text{ для всех } y > 0.$$

Очевидно, $\Delta = \ln \frac{a}{a+y} - \frac{a}{a+y} = 0$ при $y=0$ и кроме того функция $\Delta(y)$ монотонно убывающая, т.к.

$$\Delta'(y) = -\frac{1}{a+y} - \frac{a}{(a+y)^2} = -\frac{y}{(a+y)^2} < 0, \quad \text{для всех } y > 0.$$

Таким образом $\Delta(y) < 0$, для всех $y > 0$, и окончательно $\frac{\partial \beta(x)}{\partial x_i} < 0$.

Вычислим первые производные

$$\frac{\partial}{\partial x_i} (\varphi(x)D(x)) = \varphi'(x)D(x) + D'(x)\varphi(x) < 0 \quad (31)$$

$$\begin{aligned} \frac{\partial^2}{\partial x_i^2} (\varphi(x)D(x)) &= \varphi''(x)D(x) + D'(x)\varphi'(x) + D''(x)\varphi(x) + D'(x)\varphi'(x) = \\ &= \varphi''(x)D(x) + \varphi(x)D''(x) + 2D'(x)\varphi'(x) \end{aligned} \quad (32)$$

Но поскольку $D''(x) > 0, \varphi''(x) > 0, D'(x) < 0, \varphi'(x) < 0$, then (32) > 0

$$\frac{\partial^2}{\partial x_i^2} (\varphi(x)D(x)) > 0 \quad (33)$$

А условия (33) являются достаточными условиями того, что $\beta(x) = \varphi(x)D(x)$ - выпукла..

Итак, мы доказали что в случае, когда $\sum_{i=1}^N x_i r_{i1} \leq r^* \leq \sum_{i=1}^N x_i \tilde{r}_i = \tilde{r}$

то функция риска $\beta(x)$ является выпуклой.

Кроме, того если интервалы для нечетких доходностей (r_{i2}, r_{i1}) таковы, что удовлетворяют значению порядка, а именно:

если $\tilde{r}_1 < \tilde{r}_2 < \dots < \tilde{r}_i < \tilde{r}_{i+1} < \dots$, то $\tilde{r}_{11} \leq \tilde{r}_{21} \leq \dots \leq \tilde{r}_{i1} \leq \tilde{r}_{i+1,1} \leq$

то как было доказано раньше функция риска $\beta(x)$ монотонно убывающая.

Таким образом, для данного случая задача нечеткой портфельной оптимизации (7)-(8) является задачей выпуклого программирования.

Учитывая, что ограничение (8) линейны, составим функцию Лагранжа

$$L(x, \lambda, \mu) = \beta(x) + \lambda \left(r^* - \sum_{i=1}^N x_i \tilde{r}_i \right) + \mu \left(\sum_{i=1}^N x_i - 1 \right)$$

Условия оптимальности по Куну-Таккеру будут таковы

$$\frac{\partial L}{\partial x_i} = \frac{\partial \beta(x)}{\partial x_i} - \lambda r_i + \mu \geq 0; \quad i = \overline{1, N}$$

$$\frac{\partial L}{\partial \lambda} = -\sum_{i=1}^N x_i \tilde{r}_i + r^* \leq 0 \quad \frac{\partial L}{\partial \mu} = \sum_{i=1}^N x_i - 1 = 0$$

И условия дополняющей нежесткости

$$\frac{\partial L}{\partial x_i} x_i = 0 ; i = \overline{1, N} \quad \frac{\partial L}{\partial \lambda} \lambda = \lambda \left(- \sum_{i=1}^N x_i \tilde{r}_i + r^* \right) = 0 \quad x_i \geq 0 , \quad x \geq 0$$

где $\lambda \geq 0$ и μ - неопределенные множители Лагранжа. Эту задачу можно решать стационарными методами выпуклого программирования, например, методом Зойтендейка, или штрафных функций.

Экспериментальные исследования

Входными данными для экспериментов являются рыночные цены акций ведущих российских компании – ОАО РАО ЕЭС России, ОАО ГМК Норильский никель, ОАО ЛУКОЙЛ и ОАО Газпром за февраль 2008 года. Наиболее доходными за этот период являются акции ОАО ГМК Норильский никель, наименее доходными – акции ОАО «Газпром».

Исследуем зависимость уровня риска risk от заданного порогового значения доходности r^* для различных комбинации бумаг в портфеле.

Таблица 1. Результаты по портфелю из акций ОАО Газпром и ОАО РАО ЕЭС

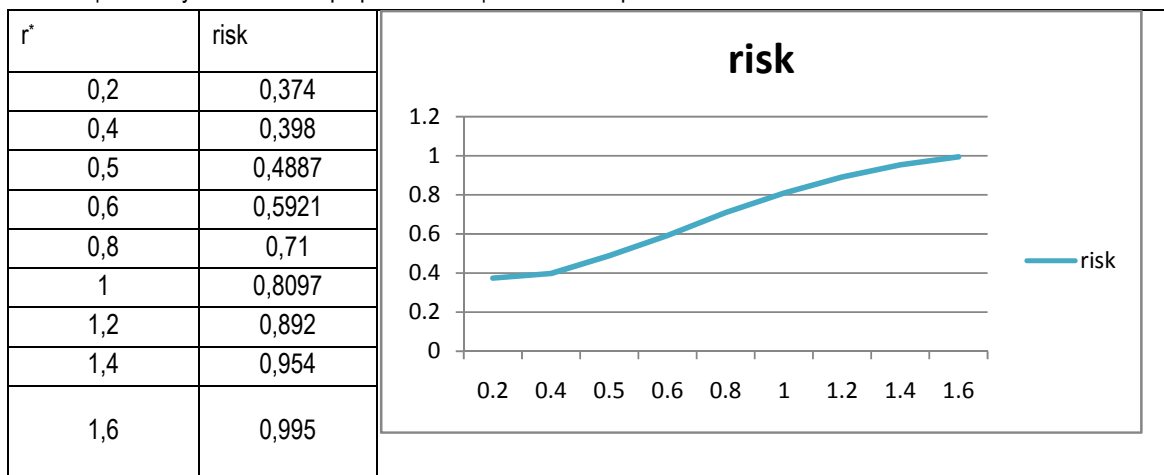


Таблица 2. Результаты по портфелю из акций ОАО ЛУКОЙЛ и ОАО РАО ЕЭС России

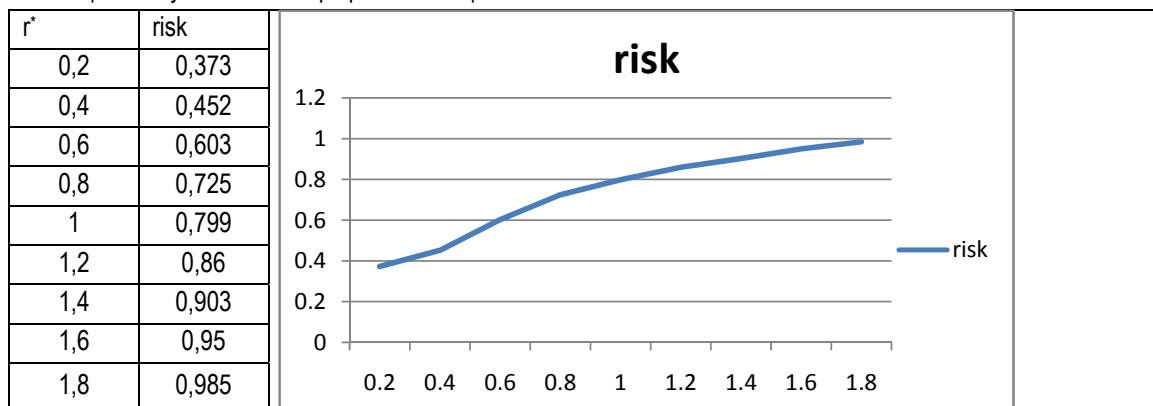


Таблица 3. Результаты по портфелю из акций ОАО Газпром и ОАО ЛУКОЙЛ

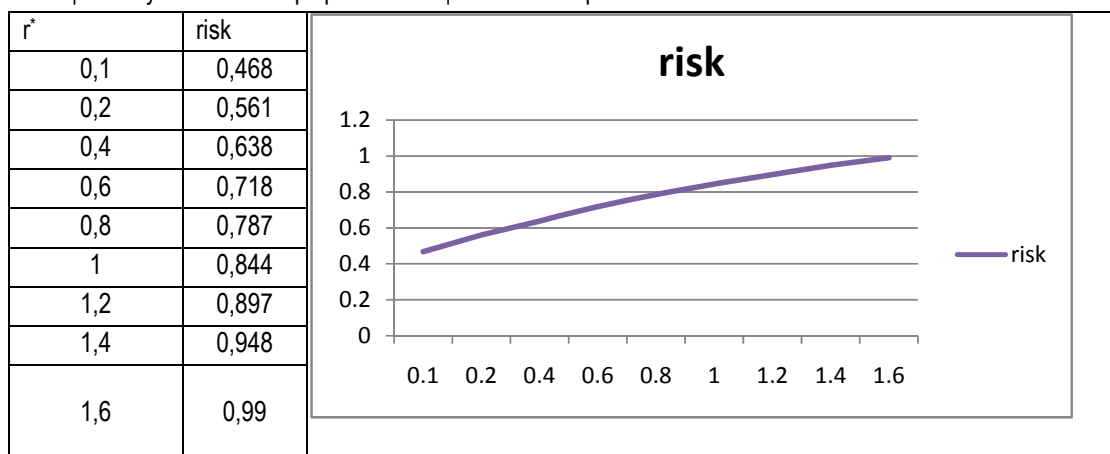
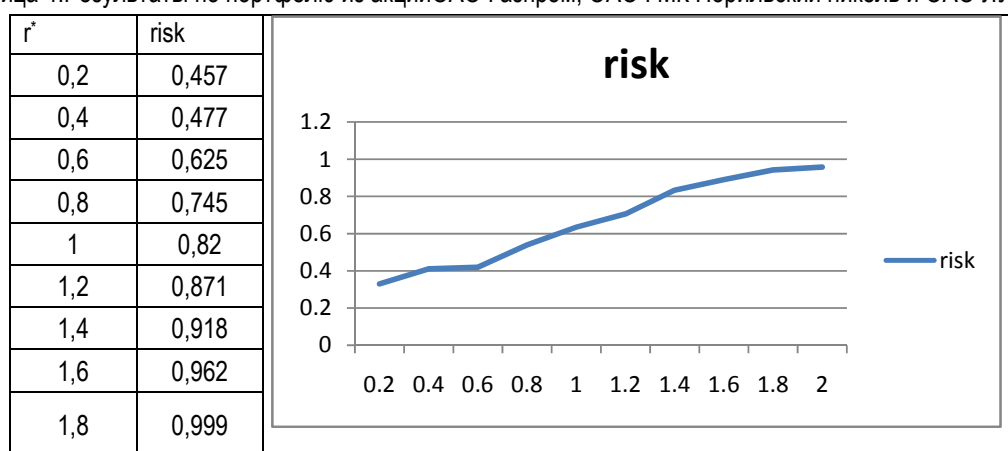


Таблица 4. Результаты по портфелю из акций ОАО Газпром, ОАО ГМК Норильский никель и ОАО ЛУКОЙЛ



Таким образом, судя по полученным графикам можно сделать вывод что при увеличении заданного порогового значения доходности, уровень риска увеличивается.

Исследуем зависимость уровня доходности портфеля $R=(R_1;R_2)$ от уровня риска risk.

На рис.1 и в таблице 5 приведены результаты по портфелю из акций ОАО Газпром и ОАО ПАО ЕЭС России

Таблица 5

| R1 | R ₂ | R2 | risk |
|----------|----------------|----------|--------|
| -3,27524 | 0,32628 | 4,88492 | 0,374 |
| -3,27693 | 0,325257 | 4,878173 | 0,398 |
| -3,6526 | 0,228595 | 4,396182 | 0,4887 |
| -3,77124 | 0,197885 | 4,242538 | 0,5921 |
| -3,88174 | 0,16928 | 4,09942 | 0,71 |
| -3,89545 | 0,165732 | 4,081668 | 0,8097 |
| -3,88902 | 0,167396 | 4,089994 | 0,892 |
| -3,90088 | 0,163151 | 4,065503 | 0,954 |
| -3,88659 | 0,168024 | 4,093136 | 0,995 |

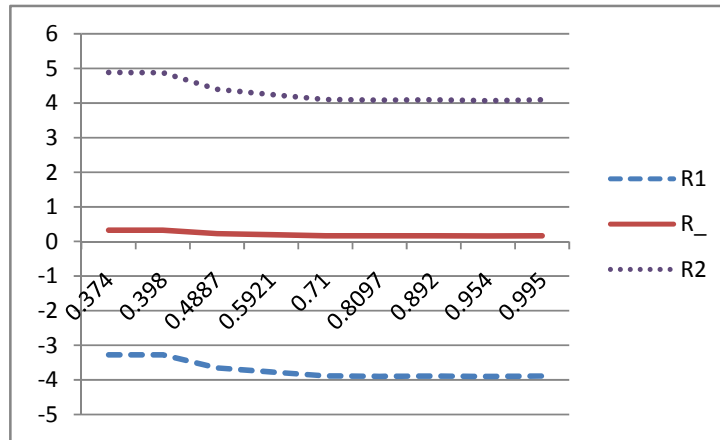


Рис. 1

На рис.2, 3 и в таблице бприведены результаты по портфелю из акцийОАО ГМК Норильский никель и ОАО Газпром

Таблица 6

| R1 | R_ | R2 | risk |
|----------|----------|----------|---------|
| -4,12801 | 0,42681 | 5,738035 | 0,391 |
| -3,72911 | 0,4227 | 5,39345 | 0,4391 |
| -3,73777 | 0,40998 | 5,32853 | 0,462 |
| -3,77676 | 0,35274 | 5,03639 | 0,55 |
| -3,84824 | 0,2478 | 4,5008 | 0,6837 |
| -3,89062 | 0,185578 | 4,183233 | 0,79737 |
| -3,90398 | 0,165963 | 4,083121 | 0,8836 |
| -3,906 | 0,163 | 4,068 | 0,954 |
| -3,906 | 0,163 | 4,068 | 0,999 |

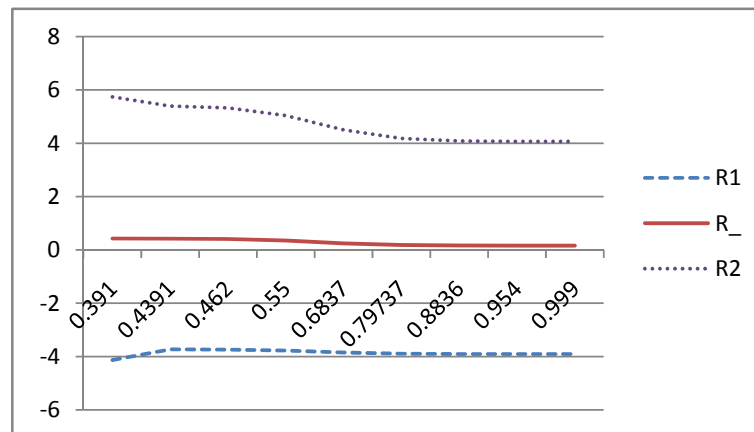


Рис. 2

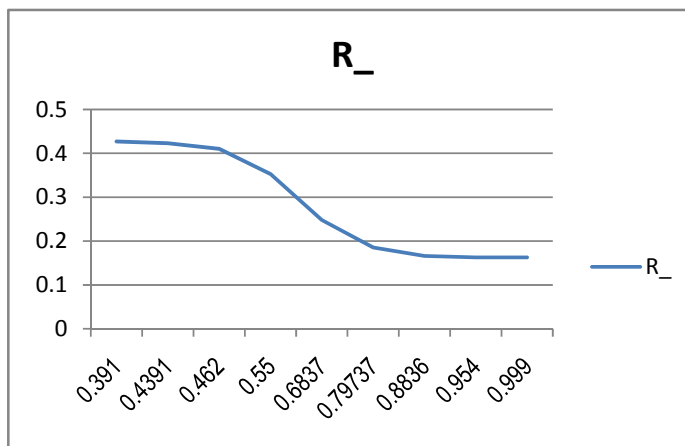


Рис. 3

На рис 4,5 и в таблице 7 приведены результаты по портфелю из акций ОАО ЛУКОЙЛ и ОАО ГМК Норильский никель

Таблица 7

| R1 | R_ | R2 | risk |
|----------|----------|----------|--------|
| -4,19568 | 0,463374 | 5,638832 | 0,342 |
| -4,26125 | 0,440213 | 5,524536 | 0,415 |
| -4,38483 | 0,396564 | 5,309132 | 0,462 |
| -4,67612 | 0,293676 | 4,801394 | 0,5924 |
| -4,78078 | 0,256708 | 4,61896 | 0,695 |
| -4,79717 | 0,250918 | 4,590386 | 0,7858 |
| -4,77927 | 0,257242 | 4,621598 | 0,8705 |
| -4,78709 | 0,254481 | 4,60797 | 0,937 |
| -4,80222 | 0,249136 | 4,581594 | 0,989 |

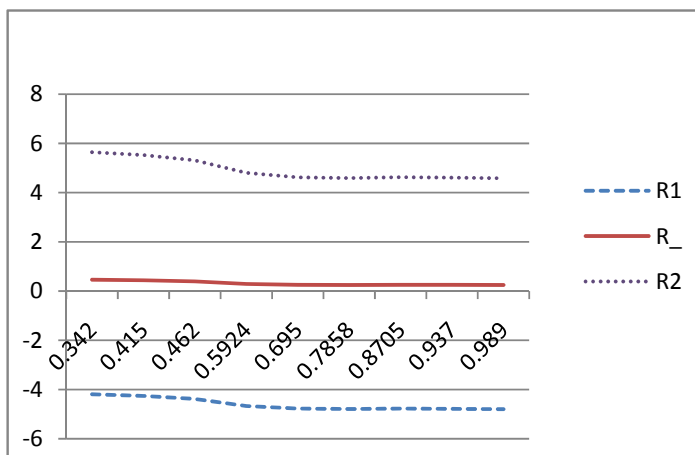


Рис. 4

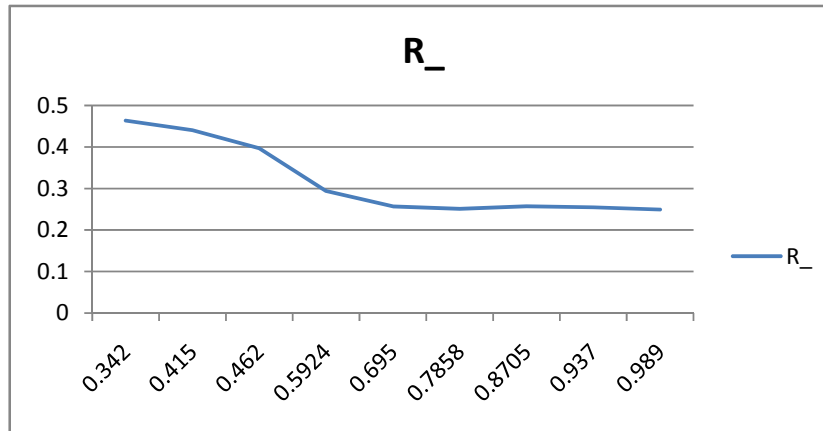


Рис. 5

На рис.6, и в таблице 8 приведены результаты по портфелю из акций ОАО Газпром, ОАО ГМК Норильский никель и ОАО ЛУКОЙЛ

Таблица 8

| R1 | R_ | R2 | risk |
|----------|----------|----------|-------|
| -3,76731 | 0,403705 | 5,295202 | 0,457 |
| -4,029 | 0,347101 | 5,023902 | 0,477 |
| -4,15759 | 0,315776 | 4,872436 | 0,625 |
| -4,64647 | 0,262933 | 4,63917 | 0,745 |
| -4,7979 | 0,250057 | 4,584974 | 0,82 |
| -4,7952 | 0,249805 | 4,58345 | 0,871 |
| -4,79988 | 0,248554 | 4,577364 | 0,918 |
| -4,79988 | 0,248454 | 4,577364 | 0,962 |
| -4,78547 | 0,252836 | 4,598328 | 0,999 |

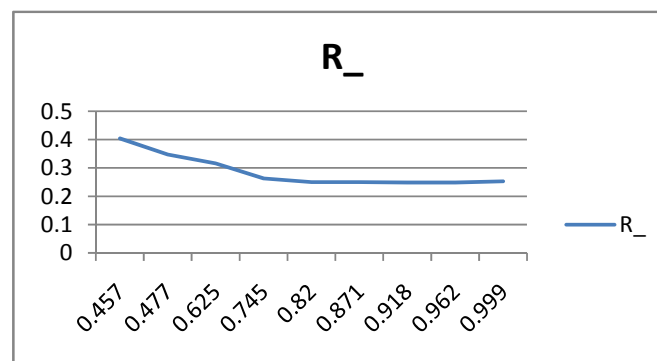


Рис. 6

На рис.7 и в таблице 9 приведены результаты по портфелю из акций ОАО РАО ЕЭС России, ОАО ГМК Норильский никель и ОАО ЛУКОЙЛ

Таблица 9

| R1 | R_ | R2 | risk |
|----------|----------|----------|-------|
| -3,61028 | 0,445731 | 5,522534 | 0,341 |
| -3,5609 | 0,434741 | 5,46456 | 0,42 |
| -4,11287 | 0,377739 | 5,19912 | 0,528 |
| -4,41087 | 0,290823 | 4,773968 | 0,701 |
| -4,41639 | 0,291458 | 4,77744 | 0,786 |
| -4,57568 | 0,272961 | 4,690976 | 0,853 |
| -4,66402 | 0,264004 | 4,64962 | 0,907 |
| -4,71599 | 0,258063 | 4,621886 | 0,955 |
| -4,72066 | 0,258482 | 4,624224 | 0,995 |

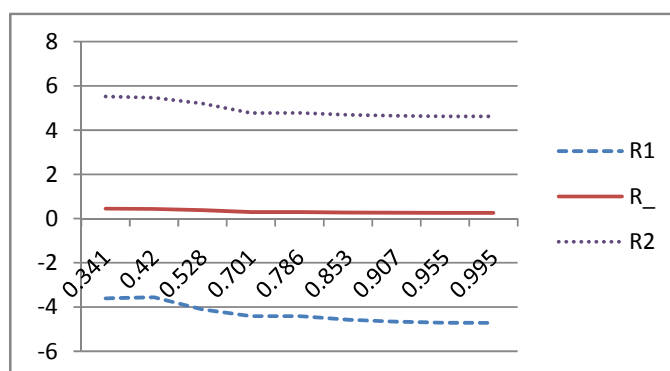


Рис. 7

На рис.8 и в таблице 10 приведены результаты по портфелю из акций ОАО РАО ЕЭС России, ОАО Газпром, ОАО ГМК Норильский никель и ОАО ЛУКОЙЛ

Таблица 10

| R1 | R_ | R2 | risk |
|----------|----------|----------|--------|
| -4,09352 | 0,452824 | 5,577784 | 0,34 |
| -3,89222 | 0,461952 | 5,610476 | 0,404 |
| -3,88378 | 0,439452 | 5,49371 | 0,5235 |
| -3,985 | 0,294998 | 4,751116 | 0,688 |
| -4,32256 | 0,292872 | 4,768776 | 0,799 |
| -4,42812 | 0,286567 | 4,744776 | 0,857 |
| -4,38233 | 0,214641 | 4,368052 | 0,954 |

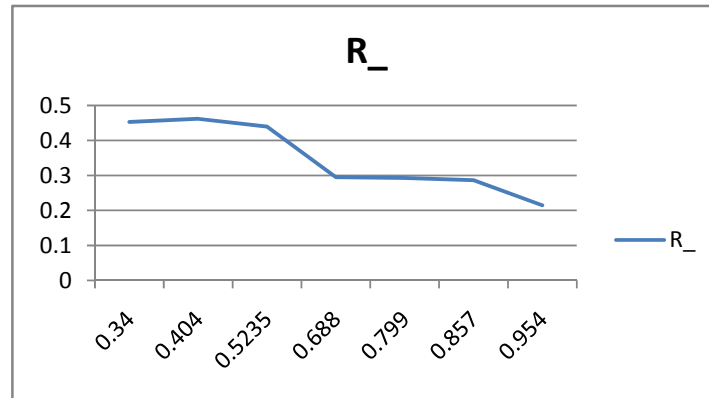


Рис. 8

Таким образом, судя по полученным графикам можно сделать вывод, что при увеличении уровня риска ожидаемая доходность инвестиционного портфеля падает.

Заключение

В работе рассмотрена и исследована двойственная задача нечеткой портфельной оптимизации. Определены достаточные условия, при которых данная задача является задачей выпуклого программирования. В этом случае данную задачу можно решать стандартными методами выпуклого программирования. Проведены экспериментальные исследования, в ходе которых построены зависимости риска портфеля от критериального значения доходности, а также ожидаемой доходности портфеля от величины риска.

Благодарности

Статья частично финансирована из проекта **ITHEA XXI** Института Информационных теорий и Приложений FOI ITHEA и консорциума FOIBulgaria (www.ithea.org, www.foibg.com)

Литература

- 1) Зайченко Юрий, Малихех Есфандиярфард. Анализ и сравнение результатов оптимизации инвестиционного портфеля при применении модели Марковитца и нечетко-множественного метода. //Proceedings of X111-th International Conference KDS-2007 "Knowledge, Dialogue Solution", Vol.1 , pp.278-286.
- 2) Зайченко Ю.П., Малихех Есфандиярфард. Оптимизация инвестиционного портфеля в условиях неопределенности// Системні дослідження та інформаційні технології.-№2-2008.-с. 59-76.
- 3) Зайченко Ю.П., д.т.н., проф., Малихех Есфандиярфард, Заика А.И. Анализ инвестиционного портфеля на основе прогнозирования курсов акций // Вісник національного технічного університету України «КПІ». Інформатика, управління та обчислювальна техніка.» Київ ТОО «БЕК+», №47 – 2007, - С. 168-179.

Информация об авторах

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

e-mail: baskervil@voliacable.com,

Ови Нафас Агаи Аг Гамиш (Иран) - аспирант НТУУ «КПИ»; 03056, Киев-56, Украина

e-mail: ovinafas@yahoo.com

МУЛЬТИ-АГЕНТНЫЕ ТЕХНОЛОГИИ ИНТЕЛЛЕКТУАЛЬНОГО УПРАВЛЕНИЯ В РОБОТОТЕХНИЧЕСКИХ, ТЕЛЕКОММУНИКАЦИОННЫХ И GRID-СЕТЯХ

Адилъ ВасильевичТимофеев

Аннотация: Рассматривается эволюция и перспективы развития пяти поколений систем управления в мехатронике и робототехнике. Описываются проблемы и методы интеллектуального сетевого управления информационными потоками в глобальных телекоммуникационных сетях и GRID-системах. Значительное внимание уделяется теории сетевых и нейросетевых агентов и мульти-агентных систем обработки информации и управления.

Ключевые слова: системы интеллектуального управления; сетевые и нейросетевые агенты; мульти-агентные технологии; GRID-системы.

1. Развитие теории управления и информационных технологий в мехатронике и робототехнике

Бурное развитие теории управления и информационных технологий привело в XX-м веке к созданию принципиально нового класса автоматических машин, названных роботами и мехатронными системами (МС). В XXI-м веке планируется разработать их микро- и наноподобия. Важное значение приобретает развитие принципов управления нанороботами и микро-МС, интеграции систем управления движением, навигации и функциональной диагностики, а также создание элементов теории мульти-агентных робототехнических систем (РС) и микро-МС, базирующиеся на групповой навигации и интеллектуальном управлении движением.

Архитектура мехатронных и робототехнических систем

С точки зрения структуры и основных функций любой робот или МС можно рассматривать как автоматическую машину, состоящую из следующих базовых компонент [1–3]:

- 1) Двигательная система (ДС) как объект управления;
- 2) Информационная система (ИС) как источник сенсорных и навигационных сигналов;
- 3) Управляющая система (УС) как средство автоматического управления;
- 4) Коммуникационная система (КС) как человеко-машинный интерфейс и средство связи..

Описанные разнородные компоненты конструктивно совмещены и являются обязательными подсистемами каждого робота или МС. При этом центральной интегрирующей компонентой является УС, играющая роль “мозга” и “нервной системы” робота или МС. КС играет важную роль для организации группового (коллективного) поведения роботов и МС (в том числе микро- и наносистем), а также для их связи с человеком.

В связи с последними достижениями микроэлектроники и нанотехнологий сегодня возникает возможность создания высокоточных и сверхминиатюрных МС и нанороботов. Такие микросистемы основываются на новых физических принципах и явлениях: акустические, оптические и пьезоэлектрические эффекты, сверхпроводимость при низких температурах, электрохимические явления в жидкой среде, электрические процессы в нейронах мозга и т.п. Поэтому разработка теории управления для микро-МС и нанороботов является одной из важнейших задач мехатроники и робототехники, а также связанных с ними нано- и информационных технологий.

1.2. Пять поколений систем управления движением

В [1–3] была предложена классификация УС роботов и МС на пять поколений:

1. Программные УС;
2. Адаптивные УС;
3. Интеллектуальные (когнитивные) УС;
4. Нейронные (нейросетевые) УС;
5. Творческие (креативные) УС.

Целью управления роботами и МС обычно является перевод ДС из начального состояния в заданное конечное состояние или осуществление (отслеживание) программного движения (ПД). Синтезируемые для этого программные УС должны обеспечить требуемые показатели качества (точность, быстродействие и т.п.) по всем управляемым координатам с учётом заданных ограничений на управления и состояния ДС. Искомые законы высококачественного программного управления для УС 1-го поколения могут быть синтезированы в аналитической форме (см., например, [1–8]).

Однако в ряде случаев этого недостаточно и требуется синтезировать адаптивные УС 2-го поколения для управления движением в условиях неопределённости. В [2–4, 9–12] были предложены различные методы дополнения законов управления алгоритмами адаптации, обеспечивающими идентификацию неизвестных параметров ДС, самонастройку параметров УС или коррекцию ПД по сигналам обратной связи от ИС.

Принцип интеллектуального (когнитивного) управления роботами и МС основывается на использовании в их УС элементов искусственного интеллекта (ИИ). В качестве интеллектуальных компонент УС можно использовать следующие элементы ИИ [1–3, 13–15]:

- моделирование среды (препятствий) на основе локальной сенсорной информации от ИС или средств виртуальной реальности;
- планирование кратчайших маршрутов и траекторий движения ДС в среде с известными или неизвестными препятствиями;
- распознавание образов сигналов от ИС и КС;
- диагностика состояний ДС по сигналам от ИС.

Эффективные методы проектирования интеллектуальных компонент для УС 3-го поколения описаны в работах [1–3, 13–15]. Реализация этих компонент осуществляется в базах данных (БД) и базах знаний (БЗ) интеллектуальной УС. При этом в УС формируются своеобразные машинно-ориентированные модели виртуальной реальности.

Сложность программных, адаптивных и интеллектуальных УС затрудняет компьютерное или микропроцессорное управление роботами и МС в реальном времени. Поэтому возникает необходимость в распараллеливании процессов обработки информации и управления.

Одним из наиболее эффективных средств такого распараллеливания являются нейронные сети (НС) [16–18]. Для проектирования нейронных УС 4-го поколения необходимо нейросетевое представление алгоритмов программного, адаптивного и интеллектуального управления и обучение НС по экспериментальным БД. Вычислительная сложность нейронных УС значительно меньше, чем программных и адаптивных УС. В то же время их быстродействие значительно выше, так как вычисление закона управления НС осуществляется за 2–3 такта за счёт параллельных вычислений в каждом слое НС независимо от сложности (числа степеней свободы и т.п.) и нелинейности ДС.

Особое значение в последние годы приобретают творческие (креативные) УС, которые могут основываться на нейронных сетях и нанотехнологиях.

Предложенные в [1–18] методы программного, адаптивного, интеллектуального и нейросетевого управления нашли применение в различных областях: были созданы транспортные роботы с гусеничным, одноосным двухколёсным и многоосным шестиколёсным шасси с адаптивным и интеллектуальным управлением, спроектированы системы управления бортовых манипуляторов орбитального комплекса “Буран”, внедрены адаптивные УС в РС для сварки, порошковой металлургии и лазерной обработки материалов, спроектированы интеллектуальные УС для нейрохирургических роботов, разработаны нейросетевые средства анализа воздушной обстановки и отслеживания самолётов на базе цифровых сигнальных процессоров в аэропорту “Пулково” и ведутся исследования мульти-агентных технологий навигации и управления движением летающих беспилотных роботов.

1.3. Мульти-агентные робототехнические системы

На практике роботы обычно интегрируются в РС для совместного достижения общих целей и решения сложных задач. При этом возникают новые проблемы группового управления и коммуникации, связанные с организацией “коллективного” поведения роботов. Традиционно эти проблемы решались на основе проектирования централизованных или децентрализованных УС для РС.

Компромиссный подход к управлению РС заключается в мульти-агентном управлении и навигации. В этом случае роботы рассматриваются как интеллектуальные мехатронные агенты РС, локальные УС которых имеют собственные (локальные) БД и БЗ и могут оперативно обмениваться информацией через КС.

Принцип действия мульти-агентных УС основывается на декомпозиции общей задачи на ряд локальных задач, возлагаемых на агентов РС, распределении этих задач между УС агентов, планировании коллективного поведения агентов, координации взаимодействия агентов на основе кооперации, реконфигурации, коммуникации и разрешения конфликтных ситуаций.

Задачи стратегического уровня обычно возлагаются на специального агента-координатора, а задачи тактического уровня параллельно решаются УС роботов как мехатронных агентов. В результате мульти-агентного управления и групповой навигации значительно увеличивается надёжность, адаптивность и быстродействие РС в изменяющейся среде с препятствиями. В последние годы были разработаны основы теории управления мульти-агентными РС (МАРС) и информационная технология мульти-агентной бесконфликтной навигации коллектива роботов, функционирующего в изменяющейся среде с препятствиями [3,14,18].

1.4. Управление коллективным движением роботов-агентов и разрешение конфликтов

Рассмотрим МАРС, в которой каждый мехатронный агент-робот может двигаться по своему маршруту, удовлетворяющему граничным условиям, причём скорость и ускорение его движения задаются агентом-координатором на супервизорном уровне управления [14-18]. Целью агента-координатора и локальных УС агентов является скорейшее прохождение всех маршрутов без столкновений (конфликтов) для выполнения локальных задач и общего (глобального) задания. Чтобы предотвратить столкновения (конфликты) агент-координатор может передавать по каналам связи КС команды о временной остановке агентов-роботов или об изменении скорости и ускорения их движения в зависимости от сложившейся ситуации. Эта динамическая ситуация оценивается агентом-координатором на основе запросов о текущем состоянии каждого агента-робота. Для прогнозирования возможных конфликтов агент-координатор должен знать маршруты и скорости движения агентов-роботов на заданное время вперёд или на всём

интервале движения. При этом он использует модель виртуальной реальности (ВР) и может заранее рассчитать время и место возможных столкновений роботов.

Значительный интерес для мульти-агентного управления МАРС представляют способы разрешения конфликтов, основанные на использовании мульти-агентных экспертных правил “дорожного движения” и нейросетевых алгоритмов распознавания дорожных ситуаций в виртуальном пространстве [18]. Эти правила должны быть обязательными для автономных УС роботов как мехатронных агентов МАРС.

Эволюция роботов и МС в последние годы осуществляется в направлении интеллектуализации и миниатюризации УС и интеграции в них процессов навигации, управления движением и функциональной диагностики. Эффективным средством миниатюризации и увеличения быстродействия УС является их реализация на базе НС и нанотехнологий [15]. Важное значение для организации группового целенаправленного поведения роботов и МС имеют методы мульти-агентной навигации и интеллектуального управления движением в динамической среде с использованием моделей ВР [14–18], а также алгоритмы адаптивной и нейросетевой маршрутизации потоков данных в распределённой мобильной КС с изменяющейся топологией связей [21].

2. Мульти-агентное управление информационными потоками в телекоммуникационных сетях и GRID-системах

Развитие информационных и телекоммуникационных систем (ИТКС) на современном этапе требует разработки теоретических основ проектирования интегрированных ИТКС нового поколения, включающих в себя телекоммуникационные системы (ТКС) и компьютерные сети (КС) или GRID-системы, в которых аккумулируются распределенные информационные и вычислительные ресурсы. Такие ИТКС предоставляют своим пользователям как внешним агентам высококачественные услуги для их массового удаленного доступа и эффективного использования распределенных информационных и вычислительных ресурсов с помощью IP-протоколов и других средств управляемой связи и передачи информации.

Совершенствование ИТКС связано прежде всего с развитием методологии автоматизации, адаптации и интеллектуализации систем сетевого управления информационными потоками на базе динамических моделей ТКС как сложных объектов управления с переменной структурой, методов оптимизации процессов маршрутизации информационных потоков и принципов адаптивного и интеллектуального управления трафиком с использованием мульти-агентных технологий и протоколов нового поколения (IPv6 и др.). На этом новом пути возможен как учет реальной динамики ТКС, т.е. фактического состояния или изменения структуры (топологии) и параметров (весов каналов связи) ТКС в реальном времени, так и адаптация к различным факторам неопределенности на основе мониторинга и функциональной диагностики ИТКС [15,21].

2.1 Архитектура мульти-агентных телекоммуникационных и GRID-систем

Архитектура глобальной ТКС или GRID-системы состоит из четырех основных (базисных) подсистем [14,15,21]:

- 1) Распределенная система связи (РСС);
- 2) Сетевая система управления (ССУ);
- 3) Распределенная информационная система (РИС);
- 4) Распределенная транспортная система (РТС).

Эти подсистемы взаимосвязаны и предназначены для управляемой передачи внешним агентам-пользователям ТКС по их запросам распределенных в КС или GRID-средах информационных и вычислительных ресурсов. Центральную роль в эффективной организации и управляемой передаче информационных потоков играют ССУ.

ССУ нового поколения должны быть адаптивными и интеллектуальными [15,21], т.е. обладать способностями к адаптации (автоматической самонастройке) по отношению к изменяющемуся количеству пользователей с учетом их запросов «по интересам» и персональных требований к качеству предоставляемых услуг, к изменяющимся структуре (топологии) ТКС и параметрам (весам) узлов и каналов связи. Интеллектуальная ССУ основана на обучении и самообучении новым функциям и правилам функционирования ИТКС, а также на самоорганизации структуры и функций ССУ в зависимости от изменений в ИТКС и предотвращении отказов и сетевых конфликтов.

2.2. Оптимизация и адаптация при сетевом управлении трафиком

Традиционная статическая постановка задачи планирования и оптимизации маршрутов передачи данных основывается на предположении, что структура (число узлов, топология связей) и параметры (стоимость каналов связи) ОТС известны и неизменны. При этом в роли внешнего агента-пользователя ТКС обычно выступает один клиент, формирующий запрос к одному из узловых компьютеров сети через РСС.

Динамическая постановка задачи исходит из того, что структура или параметры ТКС или GRID-системы могут изменяться с течением времени, но при этом остаются известными. В этом случае сетевая информация о РТС обновляется, что приводит к автоматическому изменению (пересчету) оптимальных маршрутов передачи потоков данных в ССУ.

При адаптивной постановке задачи маршрутизация осуществляется в условиях неопределенности, когда топология и параметры каналов связи РТС, а также трафик и число пользователей могут непредсказуемо изменяться. При этом доступная информация о ТКС обычно имеет локальный характер. Мониторинг и обновление сетевой информации по каналам обратной связи РИС позволяют адаптивно скорректировать маршруты и алгоритмы управления потоками данных [15,21].

2.3. Сетевые и нейросетевые агенты в глобальных ТКС и GRID-системах

Основные функции обработки информации, самоорганизации и управления информационными потоками по запросам внешних агентов-пользователей глобальной ТКС распределяются между внутренними агентами, роль которых выполняют сетевые или нейросетевые агенты. Архитектура этих внутренних сетевых агентов аналогична архитектуре ТКС. В этом проявляется фрактальность сетевых и нейросетевых агентов по отношению к ТКС и ее подсетям как автономным системам [21,22].

Каждый внутренний сетевой или нейросетевой агент имеет собственную локальную БД и локальную БЗ, а также средства связи с другими агентами через РСС для обмена информацией в процессе совместного принятия решений, самоорганизации по «интересам» и автоматического формирования сетевого управления РТС, обеспечивающего адресную доставку информационных и вычислительных ресурсов по запросам внешних агентов-пользователей глобальной ТКС или GRID-системы.

Нейросетевые агенты предназначены прежде всего для параллельной передачи и обработки сложных мультимедийных сигналов и образов (2D- или 3D-изображения и т.п.). В результате обучения по множеству прецедентов из обучающей БД осуществляется настройка архитектуры (топологии сетевых нейронов) и параметров (синаптических весов) нейронных агентов на решаемую задачу [16,17,21,22]. В последнее время разработаны модели нейросетевых агентов для адаптивной маршрутизации (агенты-маршрутизаторы) и автоматической классификации WEB-сайтов на естественном (русском) языке

(агенты-классификаторы). Программная реализация и имитационное моделирование этих агентов свидетельствует об их эффективности и преимуществах по отношению к традиционным информационным технологиям [21–23].

Предлагаемые математические модели и оптимизационные методы динамической, адаптивной, нейросетевой и мульти-агентной (много-адресной и многопоточковой) маршрутизации информационных потоков для глобальных ИТКС нового поколения представляются важным шагом в направлении создания теории адаптивного мульти-агентного (массового) обслуживания глобальных информационных и телекоммуникационных сетей [21]. Эта новая теория должна прийти на смену традиционной статистической теории массового обслуживания. Полученные результаты могут быть полезны для создания нового поколения IP-сетей или для организации адаптивного мульти-агентного (массового) обслуживания GRID-систем различного масштаба и назначения.

2.4. Логические и нейросетевые методы распознавания сложных изображений и сцен

Использование логических описаний образов позволяет свести задачи распознавания сложных изображений и интеллектуального анализа сцен к поиску логического вывода в исчислении предикатов. При этом решаются три типа задач распознавания образов [2,3,28,29]:

- идентификация объекта из заданного класса на сложном изображении или 3D-сцене;
- классификация всех объектов на сложном изображении или 3D-сцене;
- анализ и локализация всех объектов на сложном изображении или 3D-сцене.

Для увеличения эффективности поиска предлагается иерархический способ формирования таких логических описаний. Нейросетевое представление иерархических логических описаний образов и решающих правил обеспечивает массовый параллелизм и высокое быстродействие при распознавании сложных изображений и анализа сцен [28,29].

Предложенные логические методы и нейросетевые технологии успешно использовались при решении следующих прикладных задач [2,3,20,25]:

- оценка потенциальной террористической опасности людей и предметов на вокзалах;
- распознавание и оценка террористической опасности транспортных средств, движущихся вблизи газопроводов.

Работа выполнена при поддержке грантов РФФИ № 09–08–00767-а и РФФИ–ГФЕН № 10-08-91159-а и Проекта № 1.6 Программы № 1 Президиума РАН.

Список литературы

- Тимофеев А.В. Роботы и искусственный интеллект. – М.: Наука, 1978, 192 с.
- Тимофеев А.В. Управление роботами. – Л.: Издательство ЛГУ, 1985, 217с.
- Тимофеев А.В. Адаптивные робототехнические комплексы. — Л.: Машиностроение, 1988, 332 с.
- Тимофеев А.В., Экало Ю.В. Системы цифрового и адаптивного управления роботом. СПб.: Изд-во СПбГУ, 1999, 248с.
- Попов Е.П., Тимофеев А.В. Принцип скоростного управления в задаче аналитического синтеза автоматов стабилизации. – Доклады АН, 1981, т.256, № 5, с.1073–1076
- Попов Е.П., Тимофеев А.В. Параметрическая самонастройка и оптимизация автоматов управления программным движением. – Доклады АН, 1984, т. 274, № 2, с.78–81

Тимофеев А.В. Спектральный синтез диофантовых и нелинейных регуляторов обратимых динамических систем. – Доклады АН, 1997, т.353, № 2, с.173–176.

Тимофеев А.В. Управляемость, робастность и инвариантность обратимых динамических систем с нелинейной динамикой. – Доклады АН, 1998, т.359, № 2, с.171–174.

Тимофеев А.В. Адаптивная стабилизация программных движений и оценка времени адаптации. – Доклады Академии наук СССР, 1979, т.248, № 3, с.545–549.

Тимофеев А.В. Параметрическая оптимизация программных движений и адаптивное терминальное управление. – Доклады АН, 1981, т.256, № 2, с.310–313.

Тимофеев А.В. Управляемость на подпространстве и адаптивные модальные регуляторы. – Доклады АН, 1983, т. 273, № 5, с. 1070–1073.

Тимофеев А.В. Синтез адаптивных регуляторов с помощью функций Ляпунова. – Доклады АН, 1984, т. 274, № 2, с. 276–279.

Timofeev A.V. Intelligent Control Applied to Non-Linear Systems and Neural Networks with Adaptive Architecture. – Journal of Intelligent Control, Neurocomputing and Fuzzy Logic, 1996, v.1, № 1, pp.1–18.

А.В.Тимофеев. Мульти-агентное и интеллектуальное управление сложными робототехническими системами. – Юбилейный сборник “Теоретические основы и прикладные задачи интеллектуальных информационных технологий”, посвященный 275-летию РАН и 20-летию СПИИ РАН – СПб, СПИИРАН, 1999, с.71–81.

Тимофеев А.В. Методы высококачественного управления, интеллектуализации и функциональной диагностики автоматических систем. Часть I, Часть II. – Мехатроника, автоматизация, управление, 2003, № 5, 2004, № 2.

Каляев А.В., Тимофеев А.В. Методы обучения и минимизации сложности когнитивных нейромодулей супер-макро-нейрокомпьютера с программируемой архитектурой. – Доклады АН, 1984, т. 337, № 2, с. 180–183

Тимофеев А.В. Методы синтеза диофантовых нейронных сетей минимальной сложности. – Доклады АН, 1995, т. 345, № 1, с.32–35.

Timofeev A.V. Neural Control, Multi-Agent Navigation and Virtual Reality Models of Robots. – CD-ROM Proceedings of NOLCOS' 01 5-th IFAC Symposium “Non-Linear Control Systems”, Saint-Petersburg, July 4-6, 2001.

Тимофеев А.В. Эволюция нейроинформатики: от перцептронов к квантовым нейрокомпьютерам. — Нейрокомпьютеры: разработка и применение, 2002, № 5-6, с. 107-115.

Тимофеев А.В. Оптимальный синтез и минимизация сложности генно-нейронных сетей по генетическим базам данных. — Нейрокомпьютеры: разработка и применение, 2002, № 5-6, с. 34-39.

Тимофеев А.В., Сырцев А.В. Модели и методы маршрутизации потоков данных в телекоммуникационных системах с изменяющейся динамикой. – Москва: Издательство «Новые технологии», 2005 г., С. 85.

Тимофеев А.В. Проблемы и методы адаптивного управления потоками данных в телекоммуникационных системах. – «Информатизация и связь», 2003, №№1-2, с. 68-74.

Тимофеев А. В., Борисова П. В., Мышков П. С. Разработка Web-классификаторов и нейронных агентов для инфотелекоммуникационных сетей. – Материалы научно-технической конференции «Модели устойчивого регионального развития». 2005. С. 212–215.

Timofeev A.V., Gulenko I.E., Litvinov M.V. Analysis, Processing and Transfer of Dynamic Images in Virtual Reality Models. – Pattern Recognition and Image Analysis, vol.16, no.1, pp.97–99, 2006.

Timofeev A. V. Parallel Structures and Self-Organization of Heterogeneous Polynomial Neural Networks for Pattern Recognition and Diagnostics of States. – Pattern Recognition and Image Analysis, 2007, Vol. 17, No. 1, pp. 163–169.

Тимофеев А.В., Шеожев А.М., Шибзухов З.М. Мульти-агентные диофантовые нейронные сети в задачах распознавания и диагностики.– Нейрокомпьютеры: разработка и применение, 2005, № 10–11, с. 69–74.

Timofeev A. V , Azaletskiy P. S. , Myshkov P. S, Kesheng Wang. Neural Network System for Knowledge Discovery in Distributed Heterogeneous Data. – Knowledge Enterprise: Intelligent Strategies in Product Design, Manufacturing, and Management, vol. 207, 2006, pp. 144–151.

Косовская Т.М., Тимофеев А.В. Иерархическое описание классов и нейросетевое распознавание сложных образов – Нейрокомпьютеры: разработка и применение. 2007. № 6. С. 30–33.

Timofeev A.V., Kosovskaya T.M. Conditions of Effectiveness of Pattern Recognition Problem Solution Using Logical Level Descriptions of Classes. – International Scientific Journal "Information Theories and Applications" (IJ ITA) Vol.15 / 2007. С. 572–576.

Амбарян Т.Р., Тимофеев А.В. Модели квантовых и нейронных вычислений в задачах обработки информации – Известия вузов. Приборостроение. 2005 г., Т. 48, № 7, С. 35-40.

Информация об авторе



Адил Васильевич Тимофеев, заведующий лабораторией информационных технологий в управлении и робототехнике Учреждения Российской Академии наук Санкт-Петербургский институт информатики и автоматизации РАН, доктор технических наук, профессор, Заслуженный деятель науки РФ. Область исследования: теория оптимального, робастного, адаптивного, интеллектуального и нейронного управления роботами, мехатронными и аэрокосмическими системами; модели виртуальной реальности и оптимизация баз знаний; теория полиномиальных и гетерогенных нейронных сетей с самоорганизующейся архитектурой; методы синтеза многозначных решающих правил минимальной сложности для распознавания образов и диагностики состояний; модели генетического кода и генно-нейронных сетей; мульти-агентные системы и технологии навигации, управления и интеллектуального анализа потоков информации в робототехнических, компьютерных, телекоммуникационных и GRID-сетях.

МОДЕЛИРОВАНИЕ РЕАЛИСТИЧЕСКИХ ДВИЖЕНИЙ И МИМИКИ ДЛЯ ЗАДАЧ ВИЗУАЛИЗАЦИИ ЖЕСТОВОЙ ИНФОРМАЦИИ

Юрий Крак, Юрий Кривonos, Александр Бармак,
Антон Тернов, Богдан Троценко

Abstract: *Complex information technology for visualization communications gesture and mimicry is created.*

Keywords: *modelling, sign language, computer system.*

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

Введение

Задача исследования закономерностей и анимации движений является актуальной. Решение этой проблемы в виде создания инновационных технологий может существенно влиять на такие сферы человеческой деятельности как медицина, робототехника, компьютерная графика и системы виртуальной реальности.

Развитие современной науки, компьютеризация общества, использование мультимедийных и Интернет технологий создали достаточные условия для разработки компьютерных систем коммуникации для людей с дефектами слуха в формах и образах близких и понятных для них и для окружающего социума.

Люди с дефектами слуха для общения между собой используют жестовый язык. Основными визуальными средствами для воспроизведения этого языка есть фотографические и видео изображения жестов. Создавать с помощью этих средств современные учебные и коммуникационные компьютерные системы очень проблематично. Фотографическое изображение не дает нужной динамики жестов, а видео изображение достаточно громоздкое и в нем отсутствует нужная интерактивность (нельзя увидеть воспроизведение жеста с разных ракурсов, кроме того, в котором произведена видеозапись).

Эти существенные ограничения существующих средств воспроизведения жестового языка побуждают к разработке более гибких технологий, с помощью которых можно было бы создавать новые компьютерные системы обучения и коммуникации для людей с дефектами слуха. В развитие этого предложена концепция [Кривonos, 2008, 2009] информационной технологии невербального общения людей с дефектами слуха. Комплексная информационная технология включает в себя кроме другой и функциональность по синтезу движений жестового языка и дактильной азбуки на трехмерной модели человека.

Реализация возможности генерации анимации процесса общения с помощью жестового языка с использованием виртуальных моделей людей требует разработки соответствующих информационных и математических моделей. Исходя из этого, сформулирована следующая постановка задачи:

- необходимо разработать информационную и математическую модель для фиксации морфем (минимально значимых единиц) жестового языка;
- в рамках этой модели необходимо разработать технологию и соответствующее программное обеспечение для получения, сохранения и воспроизведения жестов;
- необходимо предложить алгоритмическое решение для расчета человекоподобных траекторий движения рук и корпуса модели при переходах от жеста к жесту.

Модель для фиксации единиц жестового языка

Процесс воспроизведения жеста трехмерной моделью человека можно считать анимацией с соответствующей частотой разных состояний упрощенной скелетной модели человека.

Скелетная модель человека упрощенно воссоздает скелет живого человека. Ее можно формализовать как иерархическую структуру, состоящую из соединенных кинематических пар, которые воссоздают основные кости человеческого скелета. Современные пакеты трехмерного моделирования (Poser, 3D Studio Max) умеют генерировать анимацию с помощью виртуальной статической модели и информации о изменении соответствующих углов скелета. Для формального описания процесса фиксации жеста можно использовать множество, отображающее упрощенный скелет человека (рис. 1) и изменения значений углов Эйлера и порядка их применения для соответствующих костей этого скелета с течением времени (дискретно, с соответствующей частотой (1/30 сек, например):

$$H = \{H_i : H_i = \{k, d_i, M_i \in M\}\}, \quad (1)$$

где H_i – i -я кость скелета ($i = 0, \dots, N - 1$, N – количество костей в скелете); k – индекс кости-предка; $d_i = [x_i, y_i, z_i]^T$ – координаты точки – конца кости в системе координат, связанной с началом этой кости;

$$M = \{M_i : M_i = (order_i, \theta_i)\}, \quad (2)$$

где M_i – для i -й кости значения углов Эйлера и порядок применения поворотов для кости с течением времени; $order_i \in \{1, \dots, 6\}$ – порядок применения поворотов вокруг соответствующих координатных осей для i -й кости (1-XYZ, 2-XZY, 3-YXZ, 4-YZX, 5-ZXY, 6-ZYX); $\theta = (\theta^j)$, $\theta^j = \{\theta_i^j : \theta_i^j = (\varphi_{iX}^j, \varphi_{iY}^j, \varphi_{iZ}^j)\}$ – множество изменений углов Эйлера для i -й кости с течением времени ($j = 0, \dots, K - 1$, K – количество кадров для воспроизведения движения с заданной частотой).

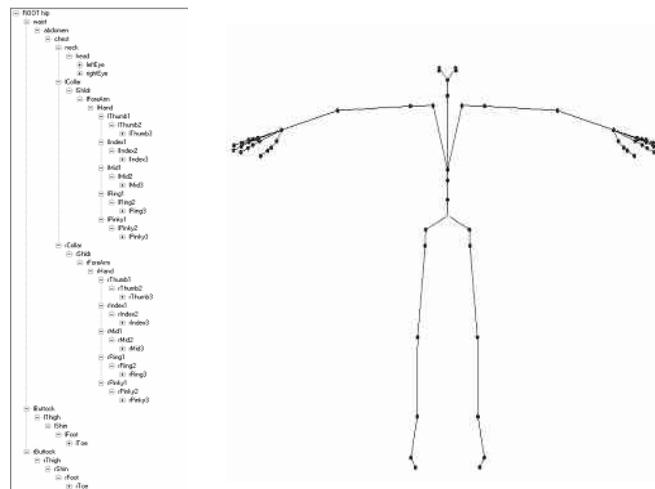


Рис. 1. Упрощенный скелет человека

Технология получения и сохранения жестов

Для получения множества углов, характеризующих изменения положения костей от начального состояния скелета предложено использовать технологию захвата движения (motion capture).



Рис.2. Приложения для видеозахвата движения.

Создано соответствующее программное обеспечение (рис. 2), с помощью которого возможно оцифровывать жесты.

Синтез процесса анимации жестового языка

Для синтеза процесса анимации жестов и мимики трехмерной моделью человека предложено следующее формальное описание, использующее соответствующие множества параметров и алгоритмы работы с ними. Трехмерная модель человека, на которой будет реализоваться процесс анимации жестов и мимики, имеет следующие атрибуты: $V = \{v_i : v_i = (x, y, z)\}$ - множество вершин треугольников для триангуляции поверхности трехмерной модели человека; $N = \{n_i : n_i = (x, y, z)\}$ - множество нормалей к вершинам; $T = \{t_i : t_i = (t, s)\}$ - множество текстурных координат к вершинам; $V^{ind} = \{V_i^{ind} : V_i^{ind} = (k_1, k_2, k_3)\}$ - множество индексов, указывающих порядок построения треугольников из множества вершин; $I = \{I_i : I_i = \{img\}\}$ - множество фотографических изображений элементов модели – текстуры.

Для моделирования скелетной анимации необходимо уметь рассчитывать новые значения вершин треугольников (V). Для этого предлагается использовать механизм скининга. Скининг можно определить как алгоритм привязки множества вершин треугольников, определяющих поверхность модели к значениям углов скелета. Тогда модель скелетной анимации можно формализовать следующим образом:

$MH = \{MH_i : HM_i = \{k, \{l_1, \dots, l_m\}, d_i, Glb_i, Order_i\}\}$ - описание упрощенного скелета человека (иерархия костей) для реализации скелетной анимации, где MH_i - i -я кость скелета ($i = 0, \dots, N - 1$, N - количество костей в скелете); k - индекс кости-предка; $\{l_1 \dots l_m\}$ - множество индексов дочерних костей, $d_i = [x_i, y_i, z_i]^T$ - координаты точки - конца кости в системе координат, связанной с началом этой кости; Glb - вектор для определения координат кости в глобальной системе координат, $Order_i$ - порядок применения поворотов.

$Skin = \{Skin_i : Skin_i = \{(IndexVertex_1, Weight_1), \dots\}\}$ - множество вершин, влияющих на текущую вершину при изменении углов.

Скининг рассчитывается для каждой вершины V следующим образом:

$$v'_j = \sum_{i=0}^N \left\{ (v_j * IBM_{H_i} * JM_{H_i}) * JW_{H_i} \right\}, \quad (3)$$

где: n - количество костей, связанных с вершиной v ; IBM_{H_i} - инверсная bind-pose матрица для кости H_i ; JM_{H_i} - матрица перемещения для кости H_i ; JW_{H_i} - весовой коэффициент для применения влияния точек кости H_i на вершину v .

Для моделирования анимации проговаривания и эмоционального окрашивания предложено использовать механизм морфинга. Морфинг можно определить как алгоритм плавного перехода от одного состояния объекта к другому. При морфинге используются только опорные состояния, с помощью которых рассчитываются промежуточные состояния и моделируется процесс анимации. Модель морфинга мимики проговаривания можно формализовать следующим образом:

Воспроизведение или построение мимики на лице трехмерной модели достигается при применении относительного (сегментного) морфинга к модели головы и мимического процесса. Формула относительного морфинга для M морфов в формализме модели имеет следующий вид:

$$V' = V + \sum_{m=1}^M \omega_m \cdot TM_m, \quad (4)$$

где ω_m - весовые коэффициенты, V - меш базовой модели без морфинга, TM_m - входной меш (морф) для блендинга (blending). Результатом операции есть линейная комбинация мешей модели и мимики.

Метод сегментного морфинга дает возможность: формировать несколько выражений лица на основе небольшого количества морфов и изменять состояние лица при анимации проговаривания. Дополнительным преимуществом сегментного морфинга есть то, что можно анимировать челюсть независимо от губ и глаз (моргание), независимо от эмоциональных проявлений на лице.



Рис. 3. Морфинг-мишени визем и эмоций

В работе предлагается осуществить синтез посредством морфинга визем украинской речи (рис. 3) в процессе проговаривания губами передаваемой словесной информации. Морфинг с использованием морфинг-мишеней визем осуществляется с учетом весовых коэффициентов:

- определяется количество кадров для анимации;
- определяется набор эмоций присутствующих в слове;
- по фонетической транскрипции рассчитывается набор визем для визуализации процесса артикуляции;

- для визем v_i рассчитываются длительности этапов анимации и точки появления виземы – минимальный номер кадра в котором $w_{v_i} > 0$.
- Для синхронизации анимации жеста и мимики полагалось, что начало и конец анимации мимики артикуляции должны совпадать с началом и концом анимации жеста.

Реализация воспроизведения процесса анимации

Для воспроизведения процесса анимации жестов и мимики трехмерной моделью человека создано приложение, реализующее скелетную (для воспроизведения жеста) и морфемную (для воспроизведения мимики проговаривания и эмоций) анимацию. Реализованная соответствующая программная функциональность, которая, используя трехмерное API OpenGL, воспроизводит по заданным атрибутам модель человека (рис. 4). Модель, с использованием алгоритмов скининга и морфинга, воспроизводит анимацию (рис. 5).



Рис. 4. Трехмерная модель человека



Рис. 5. Кадры анимации жеста «Ты»

Для реализации украинского жестового языка создано приложение (рис. 6), которое воссоздает методику преподавания жестового языка в специализированных общеобразовательных школах для глухих детей.

Функционально приложение состоит из трех информационных блоков (темы, слова и предложения) и блока воспроизведения жеста виртуальной моделью.

Блок воспроизведения жеста виртуальной моделью несет особенную функцию. С помощью него стало возможным демонстрировать в учебном процессе динамику жеста. Беря во внимание то, что жесты оцифровываются носителями жестового языка, они, по сути, делаются стандартами воспроизведения жеста. Реализованная в блоке функциональность для покадрового показа жеста служит средством, с помощью которого станет возможным изучение жеста без особенностей показа конкретным преподавателем. Это станет основой того, что изученные детьми жесты будут одинаковыми для разных школ, регионов. Фактически, жестовый язык (в основном своем множестве) станет стандартным на территории Украины.



Рис. 6. Приложение «Український жестовий язык».

Выводы

Используя модель для фиксации движений, которые воспроизводят украинский жестовый язык, было оцифровано множество из 50 жестов. Воспроизведение жестов из этого множества (с помощью трехмерной модели) показало способность предложенной технологии весьма реалистично воспроизводить на трехмерной модели движения, полученные с видео-изображения конкретного человека – носителя жестового языка.

Реализовано приложение, воспроизводящее уроки украинского жестового языка.

Созданная технология и программное обеспечение может стать основой создания стандарта жестового языка. Станет возможным решить проблему отличий для одних и тех самых жестов, которая возникает от того, что дети изучают жест, который содержит особенности конкретного преподавателя.

Дальнейшие исследования направлены на усовершенствование предложенной технологии:

- наполнения базы данных жестов основным множеством жестов украинского жестового языка – создание стандарта жестового языка;
- создание средства для семантического связывания предложений украинского языка с предложениями на жестовом языке.

Благодарности

Статья частично финансирована из проекта ITHEA XXI Института Информационных теорий и Приложений FOI ITHEA и Консорциума FOIBulgaria (www.itea.org, www.foibg.com).

Библиография

- [Кривонос, 2008] Кривонос Ю.Г., Крак Ю.В., Бармак О.В., Тернов А.С. Информационная технология невербального общения людей с проблемами слуха // Искусственный интеллект. – 2008. №3. – ст. 325-331
- [Кривонос, 2009] Кривонос Ю.Г., Крак Ю.В., Бармак О.В., Тернов А.С., Троценко Б.А. Информационная технология для моделирования украинского жестового языка // Искусственный интеллект. – 2009. №3. – ст. 186-198

Authors' Information



Iurii Krak – *V.M.GlushkovCybernetics Institute of NASU, senior scientist*
krak@unicyb.kiev.ua



Iurii Kryvonos - *V.M.GlushkovCybernetics Institute of NASU, Director Deputy,*



Alexander Barmak – *V.M.GlushkovCybernetics Institute of NASU, senior scientist,*
barmak@svitonline.com



Anton Ternov – *V.M.GlushkovCybernetics Institute of NASU, junior scientist,*
anton.ternov@gmail.com



Bohdan Trotsenko– *V.M.GlushkovCybernetics Institute of NASU, junior scientist,*
bohdan@trotsenko.com.ua

address: 40 Glushkov ave., Kiev, Ukraine, 03680

ПЕРЕМЕННАЯ РАЗРЕШАЮЩАЯ СПОСОБНОСТЬ КАК ФАКТОР РАЗВИТИЯ СРЕДСТВ ОБРАБОТКИ ВИЗУАЛЬНОЙ ИНФОРМАЦИИ

Владимир Калмыков

Abstract *The vision physiology investigations and computer vision information technologies creation objectively are the interconnected areas of human activity. One of principle differences between the vision and the computer vision is the variable sizes of the receptive fields excitation zones. Using the systemology analysis of the neurophysiology sources the new model of the neuron operation is offered. The variable resolving ability to be used the new tasks of visual information processing can be solved.*

Keywords: *переменная разрешающая способность, нейрон, рецептивные поля, визуальная информация, распознавание*

ACM Classification Keywords: *I.5.1 Models, I.3.5 Computational Geometry and Object Modeling.*

Введение

Исследования в области физиологии зрительного восприятия и создания информационных технологий автоматической обработки визуальной информации (иначе говоря, технического – компьютерного зрения) объективно являются взаимосвязанными областями человеческой деятельности. Действительно, предметом обеих научных дисциплин является исследование механизмов зрительного восприятия. Если предметом физиологии зрения является исследование процессов зрительного восприятия человека и животных, то одним из предметов компьютерных наук является создание и исследование средств технического зрения. Прогресс в одной из этих областей должен был бы инициировать прогресс в другой области.

К научным направлениям из области нейрофизиологии, непосредственно определяющим развитие понимания процессов зрительного восприятия, по-видимому, относятся:

- клеточная и молекулярная биология нейрона;
- синаптическая передача сигналов;
- преобразование зрительной информации в сетчатке глаза;
- преобразование информации нейронными структурами (в частности, в латеральном колленчатом теле, зрительной коре).

В области компьютерных наук при создании систем технического зрения и средств обработки визуальной информации основные решения определяются следующими научными направлениями:

- вычислительная геометрия;
- дискретная топология;
- распознавание образов и обработка изображений;
- нейросетевые технологии.

К сожалению взаимного междисциплинарного обмена результатами не происходит, по-видимому, из-за очень слабого взаимодействия между специалистами, представляющими разные научные направления, которое, в свою очередь, определяется существенно различными методами, характерными для перечисленных научных направлений.

Многие известные исследователи в области распознавания образов и обработки изображений вообще не упоминают о системе зрительного восприятия как о естественном прототипе технического зрения [1], [2].

Другие авторы используют в качестве прототипов упрощенные модели [3], в которых не отображены важные, уже известные элементы механизмов зрительного восприятия.

Если "от машины требуется чтение рукописного текста или расшифровка биомедицинских фотографий, ... приходится иметь дело с гораздо более сложными задачами их интерпретации.

Та видимая легкость, с которой животные и даже насекомые справляются с задачами восприятия, одновременно и ободряет, и обескураживает. Психологические и физиологические исследования дали ряд интересных результатов, касающихся процессов восприятия животными. Тем не менее, этого пока недостаточно для воспроизведения процессов восприятия с помощью ЭВМ. Особую привлекательность этому вопросу придает и то обстоятельство, что восприятие есть нечто известное по опыту каждому, а вместе с тем на деле никем не понятное. Беспольными оказываются и попытки исследования сущности восприятия посредством самоанализа из-за того, что, по-видимому, большинство обычных процессов восприятия протекает подсознательно. Парадоксально, что все мы хорошо владеем восприятием, но никто из нас не знает о нем достаточно." [5]

Одно из отличий зрительного восприятия от технического зрения

Обычно предполагается [2], что исходное изображение, подлежащее обработке, задано в аналоговой форме. Как правило, изображение ограничено прямоугольником приемлемых для обработки размеров. Рассматриваемое исходное изображение обязательно должно полностью помещаться в этом прямоугольнике и, по возможности, заполнять его.

Первой из операций, которую выполняют над обрабатываемым изображением, является дискретизация и квантование, причем следующие параметры системы технического зрения выбираются обычно из практических соображений:

- разрешающая способность – количество дискретных, обычно, квадратных элементов изображения – пикселей – помещающихся в одной единице длины (дюйме, сантиметре) и, соответственно, некоторый размер изображения в пикселях;
- множество значений яркости, которое может принимать каждый пиксел для каждого из принятых базовых цветов (градации серого, RGB и т.п.).

Предполагается, что разрешающая способность системы технического зрения наилучшим образом соответствует классу обрабатываемых изображений. При этом не возникают ненужные подробности при слишком большой разрешающей способности (например, не искажаются контурные линии объектов) и не пропадают существенные детали объектов при слишком малой разрешающей способности. Точно также и вид функции квантования и количество градаций яркости должны соответствовать обрабатываемым изображениям в смысле возможности отображения существенных деталей объектов.

Таким образом изображение можно представить в виде двумерного массива $V(N,M)$ шириной N и высотой M , причем каждому элементу этого массива $v(n,m)$ соответствует либо значение яркости элемента изображения – пиксела с координатами n,m для однотонного изображения, либо значения яркостей базовых цветов, например, красного, синего, зеленого для цветного изображения. Этот массив можно рассматривать в виде вектора или матрицы, в зависимости от выбранного для обработки математического аппарата.



а) фон нейтральный;

б) фон – прямоугольная решетка

Рис.1 Примеры произвольного текста



а)

б)

Рис.2 Тексты с произвольно наложенной решеткой, цвет линий которой совпадает с цветом фона.

Тексты а) и б) отличаются толщиной линий решетки.

Разработано и успешно применяются большое количество методов и алгоритмов для обработки изображений, представленных в виде вектора или матрицы. Большинство этих методов успешно применяются для обработки полутоновых изображений, если удачно выбраны параметры дискретизации и квантования для данного класса изображений. В то же время необходимо отметить следующее. При обработке изображений методами статистического распознавания, например, при определении меры сходства двух изображений как $f(V_1, V_2)$, используются значения яркостей всех пикселей, образующих изображение. Однако пиксели в поле изображения обычно относятся либо к объекту, либо к фону. В таком случае результат обработки зависит не только от значений яркости пикселей интересующего объекта, но и от значений яркости пикселей фона, что во многих случаях неприемлемо. В качестве примера можно привести изображение произвольного текста на однотонном фоне и на фоне некоторой произвольной решетки (рис.1). Текст на рис. 1а может быть распознан как статистическими, так и структурными методами распознавания. Такой же текст, но наложенный на произвольную решетку (рис.1б) представляет, с точки зрения распознавания, гораздо более сложную задачу. В случае попытки использования статистических методов результат вычисления сходства с эталонными изображениями будет искажен за счет наличия в поле изображения пикселей, относящихся к фону, отображающих произвольно наложенную решетку, причем пространственное расположение линий решетки меняется от изображения к изображению. В случае же попытки использования структурных методов на изображениях вида 1б невозможно получить контуры объектов. Вместо контуров объектов будут выделены контуры клеток решетки. В настоящее время внимание специалистов привлекают задачи распознавания текстур,

частным случаем которых является и наложенная на текст решетка. Однако методы распознавания текстур, как правило, значительно превосходят по сложности задачи распознавания текста, лиц и т.п., что затрудняет или исключает возможность их применения при решении подобных задач. Похожая, но несколько другая ситуация возникает при произвольном наложении на текст решетки цвета, совпадающего с цветом фона – рис.2. В данном случае при использовании статистических методов результат вычисления сходства с эталонными изображениями будет искажен за счет наличия в поле изображения пикселей, относящихся к объектам, отображающих произвольно наложенную решетку, причем пространственное расположение линий решетки, как и ранее, меняется от изображения к изображению. Так же как и для рис. 1б попытки использования структурных методов на изображениях вида рис. 2 невозможно получить контуры объектов – вместо контуров объектов будут выделены контуры клеток решетки.

В то же время зрение человека справляется с подобными задачами совершенно незаметно для нас, по-видимому, на подсознательном уровне.

Зрение– изменение размеров зон возбуждения рецептивных полей

В данном и следующем разделах приводятся известные сведения из области нейрофизиологии, необходимые для дальнейшего рассмотрения.

Как известно [6], визуальная информация в зрительной системе проецируется на сетчатку (рис.3). В рецепторах – колбочках и палочках происходит преобразование визуальной информации в электрические сигналы, которые, после некоторой предварительной обработки в слоях горизонтальных, биполярных и амакриновых клеток, поступают в ганглиозные клетки. Электрические сигналы ганглиозных клеток поступают в наружное коленчатое тело (НКТ). Наружное коленчатое тело также является парным слоистым органом – входящие в него нейроны образуют 6 слоев, где происходит сравнение (сопоставление) сигналов левого и правого глаз, выделение некоторых первичных признаков изображения, которые используются, в частности, при управлении глазодвигательными мышцами. Сигналы нейронов НКТ поступают в первичную зрительную кору головного мозга.

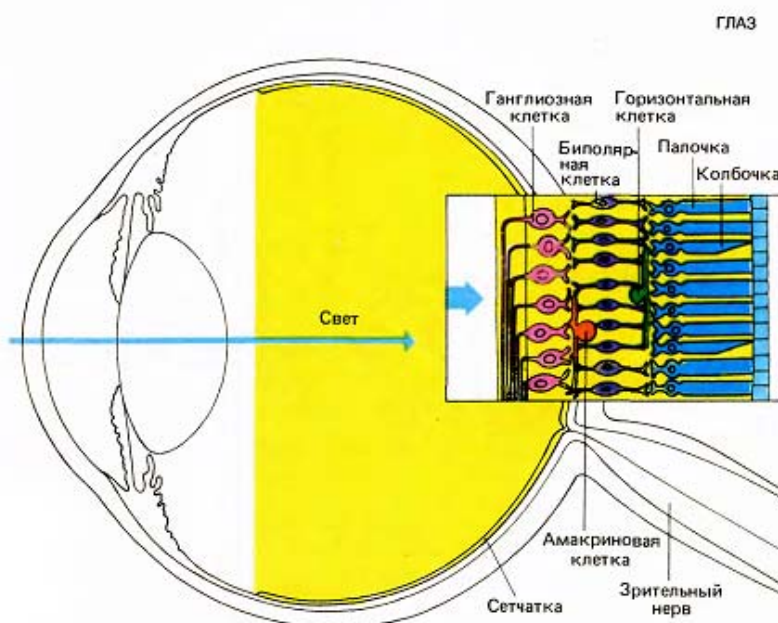


Рис.3

Относительное расположение трех слоев сетчатки внутри глаза.

Прежде чем свет достигнет палочек и колбочек, он проходит через слои ганглиозных, биполярных клеток.

На каждую ганглиозную клетку сетчатки поступают сигналы некоторого количества рецепторов. Эти рецепторы называют рецептивным полем данной ганглиозной клетки [6]. Каждый рецептор может быть связан с несколькими ганглиозными клетками. Рецептивные поля ганглиозных клеток имеют форму круга. Рецептивные поля соседних ганглиозных клеток перекрываются (рис.4 Хьюбел).

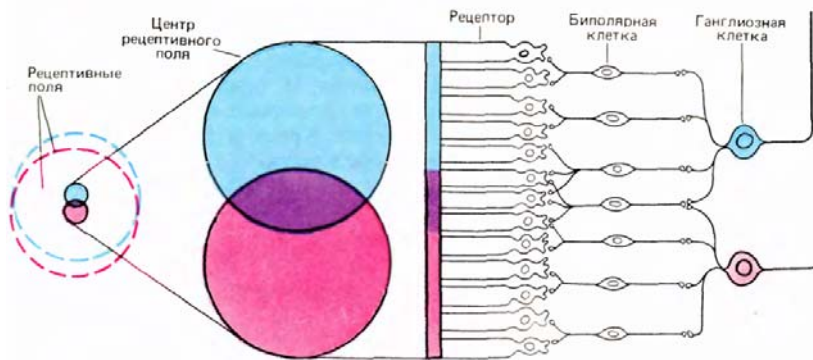
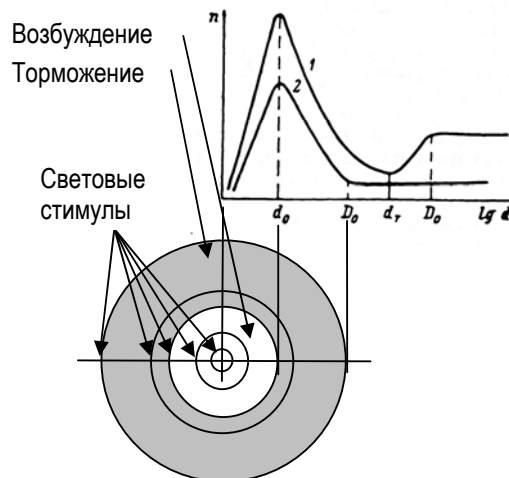


Рис.4.
Рецептивные поля соседних нейронов перекрываются

По аналогии с ганглиозными клетками, которые (опосредованно) связаны с рецепторами, совокупность связей, сигналов, нервных волокон нейронов, приходящих на данный нейрон, называют рецептивным полем данного нейрона. Учитывая, что геометрические соотношения по отношению к сетчатке в последующих слоях НКТ, сохраняются, рецептивные поля нейронов НКТ также имеют форму круга. Центральная часть рецептивных полей называется зоной возбуждения, периферийная часть (кольцо вокруг центральной части) – зоной торможения (рис.5). При подаче светового стимула на зону возбуждения рецептивного поля нейрона, генерируется определенное количество импульсов. Чем больше стимул, тем большее количество импульсов будет сгенерировано. Если размер светового стимула превысит диаметр зоны возбуждения, занимая часть зоны торможения, количество импульсов будет уменьшаться.



n – количество импульсов в ответе на стимул данного диаметра;
 d_0 – диаметр зоны возбуждения
 D_0 – диаметр зоны торможения
 1,2 – графики реакций нейронов различных типов

Рис.5 Реакция нейрона на стимулы различных размеров

Зрительная система человека работает в режиме следующих друг за другом зрительных актов. Зрительный акт длится примерно 150 мс. После чего происходит саккада – подергивание глазодвигательных мышц, в результате чего изображение на сетчатке смещается и начинается следующий зрительный акт.

В результате исследований размеров зон возбуждения рецептивных полей было установлено [7], что размер их не остается постоянным в течение зрительного акта. Если в начале зрительного акта этот размер имеет максимальное значение, то к концу зрительного акта уменьшается до минимально возможного – в случае ганглиозной клетки – это 1-2 рецептора. Таким образом, можно считать, что для зрительной системы существует **максимальная разрешающая способность**, определяемая количеством рецепторов в центральной части поля зрения и изменяемая в течение зрительного акта, **переменная разрешающая способность**, определяемая размером возбуждающей зоны рецептивного поля нейрона.

Этот механизм зрительного восприятия, имеет отношение к рассматриваемой проблеме распознавания изображений на неоднородном фоне.

Открытое [7] и исследованное явление изменения зон возбуждения нейронов зрительной системы не дает, однако, ответа на вопрос о механизме изменения диаметров зон возбуждения, что является очень существенным для возможной технической или программной реализации. Где и как реализуется этот механизм: в самом нейроне или в особых структурных свойствах нейронов, образующих его рецептивные поля.

Ответ на этот вопрос дает рассмотрение функционирования нейрона – основного структурного элемента зрительной системы. Вначале приведем традиционное представление о функционировании нейрона [6],[8].

Функционирование нейрона

Информационными входами нейрона являются сигналы, получаемые от других нейронов через межклеточные соединения *синапсы* – химически опосредованная передача с электрической развязкой или через каналы в соприкасающихся мембранах, которые соединяют внутриклеточное содержимое двух клеток – электрическая передача. Для специализированных нейронов – фоторецепторов информационными входами являются элементы, содержащие родопсин – вещество, чувствительное к свету.

Информационным выходом нейрона является серия импульсов – спайков в случае выполнения условий его возбуждения импульсами или потенциалами, поступающими по информационным входам.

Тело нейрона имеет характерную клеткам форму с внешней оболочкой – клеточной *мембраной* (рис. 6). От тела клетки отходит главный отросток – *аксон* – нервное волокно, передающее сигналы другим нейронам. Вблизи своего окончания аксон обычно разделяется на ветви, концы которых очень близко подходят к телам или дендритам других нервных клеток, но не соприкасаются с ними вплотную. В этих областях – синапсах, информация передается от одной нервной клетки, *пресинаптической*, к следующей – *постсинаптической*. Информацию от других нейронов каждый данный нейрон получает через *дендриты* и мембрану. Нервная клетка омывается солевым раствором и содержит его внутри. В число солей входят хлористый натрий, хлористый калий, хлористый кальций и ряд других солей. Большинство молекул соли диссоциировано, так что жидкости внутри и снаружи клетки содержат ионы хлора, калия, натрия и кальция (Cl^- , K^+ , Na^+ и Ca^{2+}).

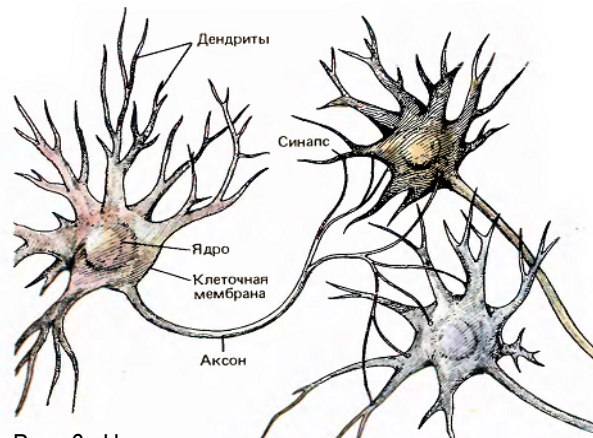


Рис. 6. Нервная клетка – это её тело, ограниченное клеточной мембраной, содержащее ядро и другие органеллы, аксон, передающий импульсы от клетки, и дендриты, к которым приходят импульсы от других

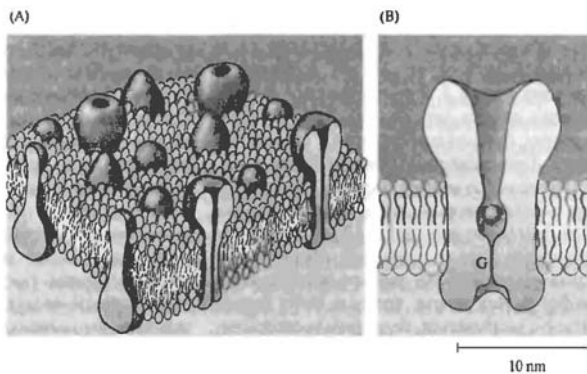


Рис. 7. Клеточная мембрана и ионный канал. (А) Клеточные мембраны состоят из жидкой фазы липидов и встроенных в липиды белковых молекул. Пронизывающие мембрану (трансмембранные) белки образуют ионные каналы. (В) Схематичное представление ионного канала с центральной водной порой и

Мембрана нервной клетки состоит из жирового слоя. Она содержит "отверстия", через которые ионы веществ могут переходить с одной её стороны на другую. Можно выделить три типа таких отверстий. Первые – это белковые молекулы в форме трубок, насквозь пронизывающих жировое вещество мембраны. В других случаях это не просто трубки, а миниатюрные белковые механизмы – *насосы*, которые способны избирательно перемещать ионы одного из типов из или внутрь клетки за счет энергии, которую клетка получает в процессе окисления глюкозы. Третьи – *ионные каналы*, – это «клапаны», которые могут открываться и закрываться под влиянием определенных видов воздействий. Некоторые из них изменяют свое состояние под влиянием мембранного потенциала, другие открываются или закрываются при наличии определенных веществ во внутренней и/или наружной жидкости.

В состоянии покоя электрические потенциалы внутри и снаружи клетки различаются на величину 30 – 100 мВ причем плюс находится снаружи.

В состоянии возбуждения импульсы – быстрые изменения мембранного потенциала вызваны движением ионов через клеточную мембрану. Направленное внутрь клетки движение положительно заряженных ионов натрия снижает общий отрицательный заряд мембраны или, другими словами, вызывает деполяризацию. Наоборот, результатом движения положительно заряженных ионов калия из клетки

является рост общего отрицательного заряда, то есть гиперполяризация. Гиперполяризация может быть обусловлена также движением внутрь клетки отрицательно заряженных ионов хлора.

Импульсы в нейроне возникают в точке аксона, близкой к месту его соединения с телом клетки; они передаются вдоль аксона, удаляясь от тела клетки, и доходят до области концевых разветвлений. Из окончаний аксона информация передается через синапсы следующей клетке или клеткам.

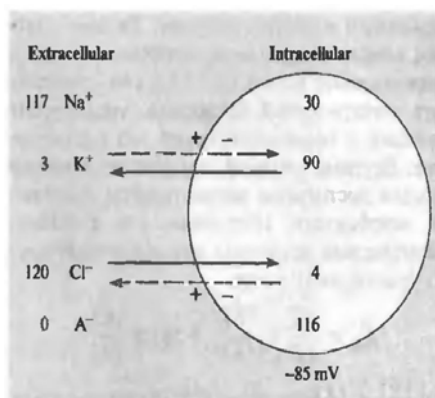


Рис. 8. Распределение ионов в идеальной клетке. Мембрана проницаема для K⁺ и Cl⁻, но непроницаема для Na⁺ и внутриклеточных анионов (A⁻). Концентрационный градиент ионов калия способствует выходу этих ионов из клетки (черная стрелка); градиент потенциала стремится перенести ионы калия внутрь клетки (серая стрелка). В состоянии покоя эти две силы уравновешивают друг друга. Концентрационный и электрический градиенты для хлора имеют противоположные направления. Ионные концентрации выражены в миллимолях (ммоль).

Информация передается в виде импульсов. Частота импульсов варьирует от одного в каждые несколько секунд или еще ниже до максимума около тысячи в секунду.

Поддержание мембранного потенциала покоя показано на модели идеальной клетки, показанной на рис. 8. В этой клетке содержатся ионы калия, натрия и хлора, а также крупные анионы. Клетка помещена в раствор, содержащий ионы хлора, натрия и калия. Концентрации ионов внутри и снаружи клетки и мембранный потенциал характерны для клеток лягушки. Клеточная мембрана в состоянии покоя проницаема для ионов калия и хлора, но не для натрия и внутриклеточных анионов. Стабильное состояние клетки обеспечивается выполнением условий:

1. Как внутриклеточная, так и внеклеточная среды должны быть электрически нейтральными.
2. Клетка должна находиться в осмотическом равновесии.
3. Суммарный заряд каждого отдельно взятого иона, переносимый через мембраны клетки, должен равняться нулю.

На рис. 8 показано, что ионы находятся в положении обратной пропорциональности: ионы калия более концентрированы внутри клетки, а ионы хлора снаружи. Хотя мембрана клетки проницаема для ионов калия, эти ионы не диффундируют из клетки наружу. Если бы ионы калия покидали клетку, то снаружи накапливался положительный

заряд, а внутри клетки образовался бы избыток отрицательного заряда. Возникший таким образом электрический потенциал при достижении определенного уровня привел бы к полному прекращению переноса. При этом равновесном потенциале для калия (E_K) электрический градиент полностью уравновешивает градиент химический, в результате чего движение ионов прекращается. Для данной модели соотношение концентраций хлора, которое, как и для калия, равно 1 : 30, и хлорный равновесный потенциал равен также -85 мВ. Как и в случае с калием, мембранный потенциал в точности уравновешивает стремление ионов хлора двигаться в направлении их концентрационного градиента, т. е. *внутри* клетки. То есть сколь угодно долго сохраняется динамическое равновесие.

Под возбуждением нейрона понимают генерацию нейроном импульса – потенциала действия. Основная роль в возбуждении принадлежит другому типу ионных каналов, при открытии которых ионы натрия устремляются в клетку. Благодаря постоянной работе насосных каналов концентрация натриевых ионов вне клетки примерно в 50 раз больше, чем в клетке, поэтому при открытии натриевых каналов ионы

натрия устремляются в клетку, а ионы калия через открытые калиевые каналы начинают выходить из клетки. Для каждого типа ионов – натрия и калия – имеется свой собственный тип ионного канала. Движение ионов по этим каналам происходит по концентрационным градиентам, т.е. из места высокой концентрации в место с более низкой концентрацией.

В покоем нейроне натриевые каналы мембраны закрыты и на мембране, как это уже описывалось выше, регистрируется потенциал покоя порядка -70 мВ (минус — внутри нейрона). Если потенциал мембраны деполяризовать (уменьшить поляризацию мембраны) примерно на 10 мВ, натриевый ионный канал открывается (рис. 9). Действительно, в канале имеется своеобразная заслонка, которая реагирует на потенциал мембраны, открывая этот канал при достижении потенциала определенной величины.

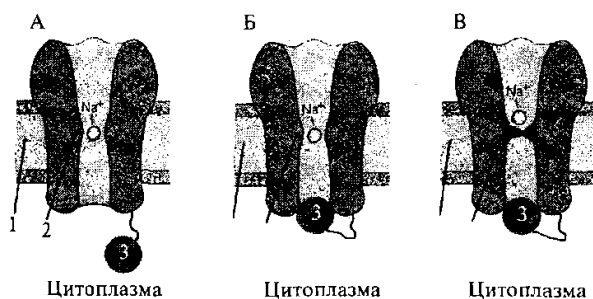


Рис. 9. Работа натриевого потенциалзависимого канала.
 А — канал открыт; Б — канал инактивирован; В — канал инактивирован и закрыт; 1 — мембрана; 2 — потенциалзависимый канал; 3 — инактивирующая частица.

Такой канал называется *потенциалзависимым*. Как только канал открывается, в цитоплазму нейрона по концентрационному градиенту устремляются из межклеточной среды ионы натрия. Другими словами, через мембрану будет протекать входящий ток ионов натрия, который будет смещать потенциал мембраны в сторону *деполяризации*, т. е. уменьшать поляризацию мембраны. Чем больше ионов натрия войдет в цитоплазму нейрона, тем больше его мембрана деполяризуется. Потенциал на мембране будет увеличиваться, открывая все большее количество натриевых каналов. Потенциал будет расти до тех пор, пока не станет равным примерно $+55$ мВ. Этот потенциал называют *натриевым равновесным потенциалом*. Учитывая, что в покое мембрана имела потенциал -70 мВ, то абсолютная амплитуда потенциала составит величину около 125 мВ. После того как потенциал на мембране достигнет своего максимального значения $+55$ мВ, натриевый ионный канал со стороны, обращенной в цитоплазму, закупоривается специальной белковой молекулой. Это так называемая «натриевая инактивация» (см. рис. 9); она наступает примерно через $0,5-1$ мс и не зависит от потенциала на мембране. Мембрана становится непроницаемой для натриевых ионов. Возвращение потенциала мембраны к исходному состоянию — состоянию покоя, осуществляется выходом ионов калия через открытые калиевые каналы. В клетке в состоянии покоя соотношение концентраций ионов калия внутри и вне клетки составляет $30:1$, поэтому при открывании калиевых каналов эти ионы покидают нейрон. В результате этих процессов мембрана нейрона возвращается к состоянию покоя (-70 мВ) и нейрон готовится к следующему акту возбуждения.

Таким образом, выражением возбуждения нейрона является генерация на мембране нейрона импульса — *потенциала действия*. Его длительность в нервных клетках составляет величину около $1/1000$ (1 мс). Описанная последовательность событий [9] приведена на рис. 10.

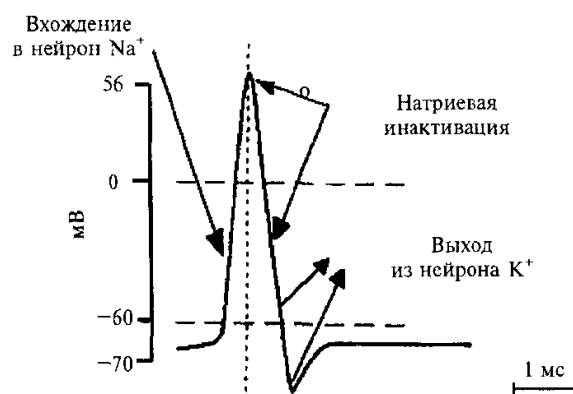


Рис. 10 Схема событий при возбуждении нейрона.

Первичная деполяризация мембраны — до 10 мВ происходит следующим образом. Возбуждение передается от одной клетки к другой через синапс. При возникновении в *пресинаптической* клетке электрический сигнала, или *импульса* последний перемещается по аксону к его синаптическим окончаниям. В каждом окончании в результате этого импульса открываются кальциевые каналы. Из каждого окончания под воздействием ионов кальция в узкий (0,02 мкм) заполненный жидкостью промежуток, отделяющий одну клетку от другой, — *синаптическую цель* — высвобождается химическое вещество *медиатор*, которое диффундирует ко второй — *постсинаптической* — клетке и воспринимается специальной белковой молекулой — *рецептором*. Под влиянием рецепторов открываются связанные с ними ионные, в данном случае, натриевые каналы и участок мембраны этой второй клетки деполяризуется. Если количество возбуждающих сигналов достаточно для открытия потенциалзависимых натриевых каналов, происходит процесс возбуждения нейрона, описанный выше.

Механизм изменения размеров зон возбуждения рецептивных полей

Если рассмотреть описанные процессы с позиций системологии, то такое описание не является полным. Действительно, почему то совсем не описано поведение и влияние ионов хлора в процессе генерации потенциала действия, которые активно участвуют в сохранении состояния покоя. Кроме того, не рассматривается функционирование возбужденного нейрона при генерации серии спайков, в процессе зрительного акта.

Под воздействием постсинаптического потенциала возбуждающих рецепторов, (а также при отсутствии воздействия тормозных рецепторов) начинают открываться потенциалзависимые натриевые каналы, нарастает натриевый ток и увеличивается деполяризация, в результате чего открывается еще больше натриевых каналов и внутриклеточный потенциал мембраны возрастает вплоть до величины натриевого равновесного потенциала.

Во время нарастания натриевого тока и увеличивающейся деполяризации ионы хлора поступают через поры по концентрационному градиенту и при уменьшении противодействующего электрического потенциала. Кроме того, наступает момент (возможно, это момент максимального значения потенциала), когда не могут не открыться потенциалзависимые хлорные каналы, которые также пропускают внутрь клетки ионы хлора. После достижения максимального значения уровня деполяризации начинается процесс выхода ионов калия и натрия, которые возвращают потенциал мембраны к исходному состоянию.

Однако за счет поступления ионов хлора потенциал покоя при генерации каждого следующего импульса смещается на некоторую величину в сторону гиперполяризации (рис.11).

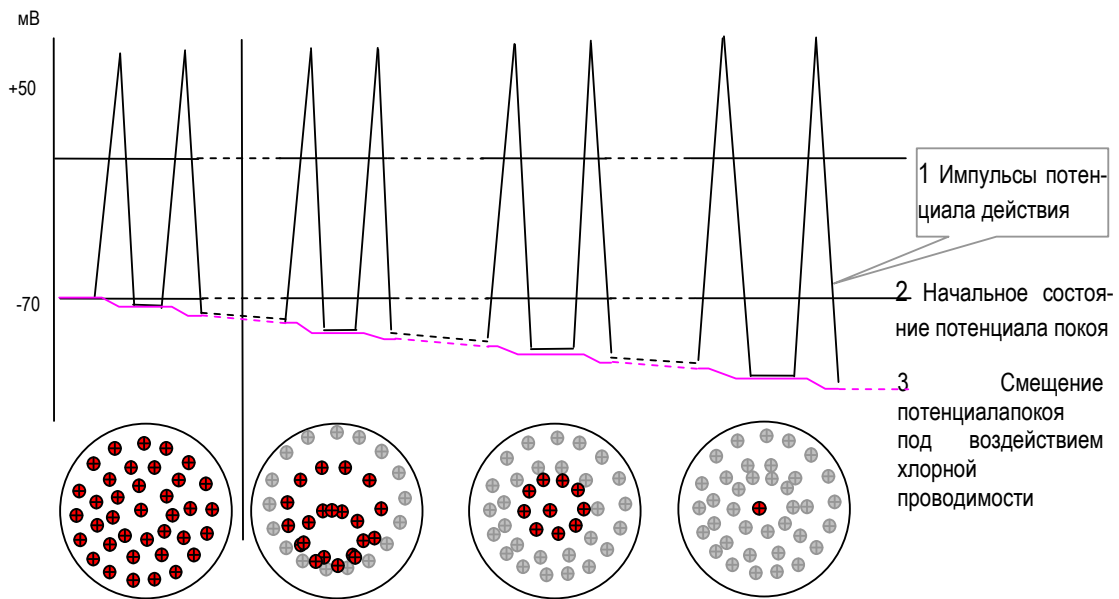


Рис.11 Изменение зоны возбуждения рецептивного поля

”В крупных ... нейронах возбудимость различных участков мембраны неодинакова. В области начального сегмента нервной клетки (аксонного холмика и начальной ... части аксона) имеется низкопороговая зона, мембрана которой обладает в несколько раз более высокой возбудимостью, чем на других участках клетки (порог возбудимости мембраны начального сегмента равен 10 мв, а порог возбудимости соматодендритической мембраны — 20—30 мв).“ [9]. Под «возбудимостью различных участков мембраны» имеется в виду порог открывания потенциалзависимых натриевых каналов, который увеличивается по мере удаления от аксонного холмика.

Итак, после каждого очередного (n-го) импульса потенциала действия происходит смещение потенциала покоя в сторону гиперполяризации. В результате в начале следующего (n+1-го) импульса потенциалзависимые натриевые каналы, которые имеют наибольший порог открывания (наиболее удалены от аксонного холмика), перестают открываться под воздействием постсинаптического потенциала возбуждающих рецепторов, находящихся в зоне расположения этих натриевых каналов. Это значит, что потенциал порога открывания упомянутых натриевых каналов больше, чем исходный потенциал покоя плюс смещение. То есть некоторая часть потенциалзависимых натриевых каналов уже не могут участвовать в накоплении заряда для генерации n+1-го импульса.

Можно также предположить, что расстояние возбуждающего рецептора от аксонного холмика соответствует расстоянию в поле зрения от соответствующей данному рецептору точки до центра рецептивного поля. Так что «неучастие» некоторого количества возбуждающих рецепторов в процессе генерации соответствует уменьшению зоны возбуждения рецептивного поля. С другой стороны, чем меньше потенциалзависимых натриевых каналов открывается при генерации очередного импульса, тем большее время необходимо для накопления заряда, достаточного чтобы этот импульс сформировался. Видимо, поэтому частота генерации импульсов со временем уменьшается.

Наличие такого, хотя и непроверенного экспериментально механизма уменьшения зон возбуждения нейронов зрительной системы во время зрительного акта полностью соответствует и подтверждает результаты, полученные в работе [7] с использованием метода временных срезов.

Таким образом, в процессе зрительного восприятия, точнее, одного зрительного акта, в зрительной системе существует последовательно информация о наблюдаемом изображении с различной разрешающей способностью, вначале низкой, затем все возрастающей и достигающей максимального значения к концу зрительного акта.

Демонстрационный эксперимент с использованием программы FineReader

Возвратимся к вопросу о том, почему изображения знаков на рис. 1а, 1б, 2а, 2б воспринимаются зрительной системой человека практически одинаково уверенно, одновременно с распознаванием фона, в то время как для современных средств обработки визуальной информации задача распознавания изображений, таких как рис. 1б, 2а, 2б, неразрешима, при этом рис. 1а распознается уверенно. Можно предположить, что именно обработка наблюдаемых изображений с низкой разрешающей способностью в начале зрительного акта дает возможность уверенного зрительного восприятия изображений знаков на различном фоне. Если это так, тогда можно предположить, что если некоторая программа распознавания текстов успешно обрабатывает изображения знаков (рис. 1а), отсканированные с низкой разрешающей способностью, то она также успешно обрабатывает и изображения (рис. 1б, 2а, 2б), отсканированные с такой же – низкой разрешающей способностью.

Проведем простой демонстрационный эксперимент с широко известной программой обработки и распознавания текстов FineReader. В качестве исходных используем изображения на рис. 1, 2 с разрешением соответственно 1а – 896x264, 1б – 895x275, 2а – 899x276, 2б – 892x262. Выбрана не имеющая смысла последовательность знаков А65Н7, чтобы исключить влияние использования словарей на результат распознавания. Результат распознавания исходных изображений: рис. 1а – успешно, остальные – отказ от распознавания по причине невозможности найти (определить) объект в поле изображения.

На рис. 12 представлены те же самые изображения, причем дискретизация выполнена с существенно меньшим разрешением. Результат распознавания всех этих изображений – успешно.



Рис. 12 Изображения рис. 1, 2: дискретизация выполнена при низком разрешении.

Заклучение

1. Проведенный системологический анализ работы нейронов зрительной системы позволил уточнить функционирование нейрона во время генерации потенциала действия и, в дополнение к ранее открытому явлению [7] уменьшения зон возбуждения рецептивных полей зрительной системы, обнаружить механизм, реализующий это явление.

2. Можно считать, что для зрительной системы существует **максимальная разрешающая способность**, определяемая количеством рецепторов в центральной части поля зрения и изменяемая в течение зрительного акта, **переменная разрешающая способность**, определяемая размером возбуждающей зоны рецептивного поля нейрона.

3. Результат проведенного простого демонстрационного эксперимента позволяет сделать обоснованное предположение, что восприятие – распознавание изображений в зрительной системе происходит на всех этапах зрительного акта, в том числе и на начальном этапе, когда изображение "расфокусировано", но зато и мешающие факторы, как например, в рассмотренном случае, в виде текстурных помех не оказывают влияния. Использование этого результата в средствах обработки визуальной информации позволяет совершенно естественным образом достигнуть инвариантности относительно масштаба объектов, выполнять распознавание объекта на фоне, искаженном высокочастотными помехами, также выполнять распознавание объектов, например, письменных знаков, элементы которых представляют мозаику или образованы элементами различных текстур.

Библиография

- [1] М.Шлезингер, В.Главач Десять лекций по статистическому и структурному распознаванию. –Київ: Наукова думка, 2004. – 535с.
- [2] Т.Павлидис Алгоритмы машинной графики и обработки изображений. М Радио и связь 1986г. 400с
- [3] У.Прэтт Цифровая обработка изображений: Пер. с англ.—М.: Мир, 1982.— Кн.1—312 с, ил.
- [4] Р. Гонсалес, Р.Вудс Цифровая обработка изображений М.: Техносфера, 2005. – 1072 с.
- [5] Richard O. Duda, Peter E. Hart Pattern classifications and scene analysis New York, London, Sydney, Toronto, A Wiley-Interface Publication, 1973. – 507 с.
- [6] David H. Hubel Eye, brain, and vision New York : Scientific American Library : Distributed by W.H. Freeman, 1988.—240p.
- [7] Н.Ф.Подвигин Динамические свойства нейронных структур зрительной системы Ленинград: Наука, 1979. –158 с. :
- [8] John G. Nichols, A. Robert Martin, Bruce G. Wallace, Paul A. Fuchs From Neuron to Brain Sunderland, Massachusetts USA: Sinauer Associates, Inc., 2001. – 672 p. : ill.
- [9] Физиология человека: Учебник для институтов физической культуры / Под ред. Н.В.Зимкина.- М.: Физкультура и спорт, 1975.- 496 с.

Authors' Information



Vladimir Kalmykov - senior researcher, candidate of engineering sciences, Institute of problems of mathematical machines and systems, prosp. akad. Glushkova 42, 03680, Kiev 187, Ukraine; e-mail: vl.kalmykov@gmail.com , kvq@immisp.kiev.ua

ОБОБЩЕННАЯ ТАБЛИЧНАЯ АЛГЕБРА, ОБОБЩЕННОЕ ИСЧИСЛЕНИЕ СТРОК И ИХ ЭКВИВАЛЕНТНОСТЬ

Дмитрий Буй, Ирина Глушко

Аннотация: В данной статье проведено обобщение табличной алгебры, построенной на основе реляционных алгебр Кодда; обобщение состоит в том, что вместо таблиц рассматриваются пары, состоящие из таблиц и схем таблиц. Представлено обобщение классического результата об эквивалентности реляционной алгебры Кодда и исчисления строк (кортежей). Исчисление строк пополнено произвольными предикатными и функциональными сигнатурами на универсальном домене. Доказано, что при этом обобщении исчисление строк остается не менее выразительным, чем табличная алгебра.

Ключевые слова: реляционные (табличные) базы данных, исчисление строк, табличная алгебра.

АСМ классификация ключевых слов: H.2 Database Management (E.5) H.2.4 Systems – Relational databases

Введение

Уточнение реляции в терминах именных множеств осуществлено Редьком В.Н., Броной Ю.И., Бум Д.Б., Поляковым С.А. в монографии [Редько, Брона, Буй, Поляков, 2001]. В данной работе обобщены полученные результаты на таблицы, которым приписывается определенная схема. Часть статьи посвящена обобщению классического результата об эквивалентности реляционной алгебры Кодда и исчисления строк (кортежей), пополненного произвольными предикатными и функциональными сигнатурами на универсальном домене. Доказано, что при этом исчисление строк остается не менее выразительным, чем табличная алгебра.

Обобщение табличной алгебры

Все неопределенные здесь понятия и обозначение понимаем в смысле [Редько, Брона, Буй, Поляков, 2001]. Рассматриваем два множества: \mathbf{A} – множество атрибутов и \mathbf{D} – универсальный домен. Под табличной алгеброй понимаем алгебру $\langle \mathbf{T}, \Omega_{P, \Xi} \rangle$, где \mathbf{T} – множество всех таблиц, $\Omega_{P, \Xi} =$

$\{\cup_R, \cap_R, \setminus_R, \sigma_{p, R}, \pi_{X, R}, \otimes_{R_1, R_2}, \div_{R_1, R_2}^{R_1}, Rt_{\xi, R}, \sim_R\}_{\substack{p \in P, \xi \in \Xi \\ X, R, R_1, R_2 \subseteq \mathbf{A}}}$ – сигнатура, P, Ξ – множества параметров.

Произвольное конечное множество атрибутов $R \subseteq \mathbf{A}$ назовём схемой. Строкой схемы R называется именованное множество на паре R, \mathbf{D} , проекция которого по первой компоненте равна R (т.е. рассматривается функция вида $s: R \rightarrow \mathbf{D}$). Под (обобщенной) таблицей понимаем пару $\langle t, R \rangle$, где $t \in \mathbf{T}(R)$ – таблица фиксированной схемы R (в смысле [Редько, Брона, Буй, Поляков, 2001]). Тогда $\mathbf{T}(R) = \{\langle t, R \rangle \mid t \in \mathbf{T}(R)\}$ – множество всех таблиц схемы R , а $\mathbf{T} = \cup \mathbf{T}(R)$ – множество всех таблиц.

Отличие от определения таблицы в смысле [Редько, Брона, Буй, Поляков, 2001] заключается в том, что каждой таблице приписывается ее схема. По сути это влияет только на случай пустой таблицы t_\emptyset , поскольку по непустой таблице схема восстанавливается однозначно. Запись $\langle t_\emptyset, R \rangle$ обозначает пустую таблицу схемы R .

Выражением табличной алгебры называется любое выражение, построенное из таблиц множества \mathbf{T} при использовании операций множества $\Omega_{p,\Xi}$.

Зададим операции. Определим сначала теоретико-множественные операции. Под объединением \cup_R (пересечением \cap_R , разностью \setminus_R) таблиц схемы R понимается бинарная частичная операция, полученная ограничением теоретико-множественного объединения (соответственно пересечения, разности) на множество всех таблиц схемы R .

$$\cup_R : \mathbf{T} \times \mathbf{T} \xrightarrow{\sim} \mathbf{T}, \quad \text{dom } \cup_R = \mathbf{T}(R) \times \mathbf{T}(R), \quad \langle t_1, R \rangle \cup_R \langle t_2, R \rangle = \langle t_1 \cup t_2, R \rangle;$$

$$\cap_R : \mathbf{T} \times \mathbf{T} \xrightarrow{\sim} \mathbf{T}, \quad \text{dom } \cap_R = \mathbf{T}(R) \times \mathbf{T}(R), \quad \langle t_1, R \rangle \cap_R \langle t_2, R \rangle = \langle t_1 \cap t_2, R \rangle;$$

$$\setminus_R : \mathbf{T} \times \mathbf{T} \xrightarrow{\sim} \mathbf{T}, \quad \text{dom } \setminus_R = \mathbf{T}(R) \times \mathbf{T}(R), \quad \langle t_1, R \rangle \setminus_R \langle t_2, R \rangle = \langle t_1 \setminus t_2, R \rangle; \text{ где } t_1, t_2 \in \mathbf{T}(R).$$

Пусть $p : S \xrightarrow{\sim} \{true, false\}$ – частичный предикат на множестве строк. Под селекцией по предикату p таблиц схемы R понимается унарная частичная параметрическая операция $\sigma_{p,R}$, которая таблице сопоставляет её подтаблицу, содержащую строки, на которых предикат p истинный. Следовательно, $\sigma_{p,R} : \mathbf{T} \xrightarrow{\sim} \mathbf{T}$, $\text{dom } \sigma_{p,R} = \{\langle t, R \rangle \mid t \subseteq \text{dom } p\}$, $\sigma_{p,R}(\langle t, R \rangle) = \langle \{s \mid s \in t \wedge p(s) \simeq true\}, R \rangle$, где $t \in \mathbf{T}(R)$. Здесь и далее \simeq – обобщенное (по другой терминологии сильное) равенство, то есть, обе части или одновременно не определены или одновременно определены и равны.

Пусть $X \subseteq \mathbf{A}$ – конечное множество атрибутов. Под проекцией по множеству атрибутов X таблиц схемы R понимается унарная параметрическая операция $\pi_{X,R}$, значениями которой есть таблицы схемы $R \cap X$, состоящие из ограничений по множеству атрибутов X строк исходных таблиц. Следовательно, $\pi_{X,R} : \mathbf{T}(R) \rightarrow \mathbf{T}(R \cap X)$, $\pi_{X,R}(\langle t, R \rangle) = \langle \{s \mid X \mid s \in t\}, R \cap X \rangle$, где $t \in \mathbf{T}(R)$.

Под соединением таблиц схем R_1, R_2 понимается бинарная операция \otimes_{R_1, R_2} , значениями которой являются таблицы схемы $R_1 \cup R_2$, состоящие из всех объединений совместных строк исходных таблиц.

$$\text{Следовательно, } \otimes_{R_1, R_2} : \mathbf{T}(R_1) \times \mathbf{T}(R_2) \rightarrow \mathbf{T}(R_1 \cup R_2),$$

$$\langle t_1, R_1 \rangle \otimes_{R_1, R_2} \langle t_2, R_2 \rangle = \langle \{s_1 \cup s_2 \mid s_1 \in t_1 \wedge s_2 \in t_2 \wedge s_1 \approx s_2\}, R_1 \cup R_2 \rangle, \text{ где } t_1 \in \mathbf{T}(R_1), t_2 \in \mathbf{T}(R_2).$$

Под делением таблиц схемы R_1 на таблицы схемы R_2 , где $R_2 \subseteq R_1$, понимается бинарная параметрическая частичная операция $\div_{R_2}^{R_1}$, значениями которой являются таблицы схемы $R_1 \setminus R_2$, состоящие из определенных строк, полученных в результате проекции по множеству атрибутов $R_1 \setminus R_2$ таблиц схемы R_1 (см. ниже). Следовательно, $\div_{R_2}^{R_1} : \mathbf{T} \times \mathbf{T} \xrightarrow{\sim} \mathbf{T}$, $\text{dom } \div_{R_2}^{R_1} = \mathbf{T}(R_1) \times \mathbf{T}(R_2)$,

$$\langle t_1, R_1 \rangle \div_{R_2}^{R_1} \langle t_2, R_2 \rangle = \langle \{s \in \pi_{R_1 \setminus R_2}(t_1) \mid \{s\} \otimes_{R_1, R_2} t_2 \subseteq t_1\}, R_1 \setminus R_2 \rangle, \text{ где } t_1 \in \mathbf{T}(R_1), t_2 \in \mathbf{T}(R_2).$$

Под активным дополнением таблиц схемы R понимается унарная операция \sim_R , которая таблице сопоставляет дополнение в её насыщении. Следовательно, $\sim_R : \mathbf{T}(R) \rightarrow \mathbf{T}(R)$, $\sim \langle t, R \rangle = \langle C(t) \setminus t, R \rangle$, где $t \in \mathbf{T}(R)$. Для замкнутости текста отметим, что $C(t) = \otimes_{A \in R} \pi_{\{A\}}(t)$.

Под переименованием таблиц схемы R , соответствующим инъективной частичной функции переименования атрибутов $\xi: \mathbf{A} \rightarrow \mathbf{A}$, понимается унарная параметрическая операция $Rt_{\xi,R}$, $\text{dom } Rt_{\xi,R} = \{\langle t, R \rangle \mid t \in T_{\xi}(R)\}$, где $T_{\xi}(R)$ – множество таблиц схемы R , а схема R ξ -допустима, значения которой задаются равенством $Rt_{\xi,R}(\langle t, R \rangle) = \langle Rs_{\eta,R}[t], \eta[R] \rangle$, $t \in T_{\xi}(R)$, где $\eta = \xi \cup \text{id}_{\text{Aidom}_{\xi}}$, $Rs_{\eta,R}(s) = \{\langle \eta(A), s(A) \rangle \mid A \in R\}$, $s \in S(R)$.

Обобщенное исчисление строк

В основе большинства реляционных языков запросов лежит реляционное исчисление, поскольку в отличие от реляционной алгебры, исчисление выражает лишь то, каким должен быть результат и не предусматривает определение того, как его получить. Реляционное исчисление основывается на исчислении предикатов первого порядка. Есть две формы реляционного исчисления: исчисление с переменными строками (по другой терминологии кортежами) и исчисление с переменными на доменах. Эти формы предложены Е. Коддом (E. Codd) [Codd, 1972] и М. Лакруа (M. Lacroix) с А. Пиротте (A. Pirotte) [Lacroix, Pirotte 1977] соответственно. Данный вопрос также рассматривали в своих работах Д. Мейер (D. Maier) [Мейер, 1987], Дж. Д. Ульман (J. D. Ullman) [Ульман, 1983], К. Дж. Дейт (C. J. Date) [Дейт, 2005], Т.М. Коннолли (T. M. Connolly) и К. Бегг (C. E. Begg) [Коннолли, Бегг, 2003], В. В. Пасичник и В. А. Резниченко [Пасичник, Резниченко, 1977].

Выражения исчисления строк имеют вид $\{x(R) \mid P(x)\}$, где P – некоторый предикат над переменной строкой x , а R – схема. Это выражение обозначает таблицу $\langle t, R \rangle$, $t \in T(R)$, состоящую из всех строк, на которых предикат P истинен.

Введем множество так называемых разрешенных формул исчисления строк при использовании:

- множества атрибутов \mathbf{A} и универсального домена \mathbf{D} ;
- множества предметных переменных (переменных строк) x_1, x_2, \dots ;
- множества предметных констант d_1, d_2, \dots ;
- множества функциональных символов $f_1^{n_1}, f_2^{n_2}, \dots$;
- множества предикатных символов $p_1^{m_1}, p_2^{m_2}, \dots$.

Областью интерпретации предметных констант есть универсальный домен \mathbf{D} , предметных переменных – множество всех строк. Применяем \mathbf{x} как синтаксическую переменную, областью изменения которой являются переменные; $\mathbf{f}(\mathbf{p})$ как синтаксическую переменную, областью изменения которой являются функциональные (соответственно предикатные) символы; \mathbf{d} как синтаксическую переменную, областью изменения которой являются константы; наконец, \mathbf{A} как синтаксическую переменную, областью изменения которой являются атрибуты.

Следующие выражения являются термами (индукция по длине термов):

- a) всякая предметная константа есть терм;
- b) $\mathbf{x}(\mathbf{A})$ – терм;
- c) если $\mathbf{u}_1, \dots, \mathbf{u}_n$ – термы, \mathbf{f} – n -арный функциональный символ, то $\mathbf{f}(\mathbf{u}_1, \dots, \mathbf{u}_n)$ – терм;
- d) выражение является термом только в том случае, если это следует из правил a), b), c).

Будем применять u как синтаксическую переменную, областью изменения которой являются термы. Сформулируем правила построения формул. Атомарные формулы (атомы) бывают двух типов.

- a1. Пусть $\langle t, R \rangle$ – таблица, а x – переменная строка. Тогда $t(x)$ – атом, означающий, что $x \in t$.
- a2. Пусть u_1, \dots, u_m – термы, а p – m -арный предикат на универсальном домене D . Тогда $p(u_1, \dots, u_m)$ – атом.

Используем логические связки \neg, \wedge, \vee , кванторы \exists, \forall и скобки $(,)$ для построения формул из атомов.

- f1. Каждый атом – формула
- f2. Если P – формула, то $\neg P$ – формула.
- f3. Если P и Q – формулы, то $P \wedge Q, P \vee Q$ – формулы.
- f4. Пусть x – переменная строка, P – формула, $R \subseteq A$ – схема, тогда $\exists x(R)P$ – формула.
- f5. Пусть x – переменная строка, P – формула, $R \subseteq A$ – схема, тогда $\forall x(R)P$ – формула.
- f6. Если P – формула, то (P) – формула.
- f7. Других формул нет.

В общем случае переменные строки в формулах могут быть свободные или связанные. Смысл этих понятий такой же как и в исчислении предикатов: вхождение переменной x в данную формулу называется связанным, если x является переменной входящего в эту формулу квантора (\exists или \forall) или находится в области действия входящего в эту формулу квантора; в противном случае вхождение переменной x в данную формулу называется свободным. Переменная называется свободной (связанной) переменной в данной формуле, если существуют свободные (соответственно связанные) ее вхождения в эту формулу [Мендельсон, 1971].

Для каждой переменной строки x определим схему (конечное множество атрибутов) $scheme(x, P)$ и множество атрибутов $attr(x, P)$, с которыми строка x встречается в формулах. Выражения $scheme(x, P)$ и $attr(x, P)$ определены, если строка x имеет свободное вхождение в формулу P , причем имеет место включение $attr(x, P) \subseteq scheme(x, P)$ (что следует из последующих определений при условии определения выражений).

Выделим класс разрешенных формул, используя понятия свободных и связанных переменных строк, схемы и множества атрибутов, с которыми переменная строка встречается в формулах. Определим выражение $attr$ сначала для термов:

1. если $u = d$, то $attr(x, u) = \emptyset$;
2. если $u = x(A)$, то $attr(x, u) = \{A\}$, а $attr(x, y(A)) = \emptyset$, где $x \neq y$;
3. если $u = f(u_1, \dots, u_n)$, где u_i – термы, то $attr(x, u) = \bigcup_{i=1}^n attr(x, u_i)$.

Пусть формула P – атом, тогда

- a1. если $P = t(x)$, то (единственное) вхождение переменной строки x свободно в формуле P и $scheme(x, P) = attr(x, P) = R$, где R – схема таблицы t ;

a2. если $P = p(u_1, \dots, u_m)$, где u_i – термы, причем x_1, \dots, x_k – все переменные этих термов, то вхождения этих переменных строк свободны в формуле P , схема $scheme(x_i, P)$ не определена, а $attr(x_i, P) = \bigcup_{j=1}^m attr(x_i, u_j)$, $i = 1, \dots, k$.

Атомарные формулы всегда разрешены. Для построения всех разрешенных формул проведем индукцию по длине формул. Предположим, что G и Q разрешенные формулы.

f2. Если $P = \neg G$, то формула P – разрешена, а вхождения переменных в P свободны или связаны в зависимости от того, свободны или связаны вхождения этих же переменных в формулу G . Если x входит в формулу G свободно, то $scheme(x, P) \simeq scheme(x, G)$ и $attr(x, P) = attr(x, G)$.

f3. Если $P = G \wedge Q$ или $P = G \vee Q$, то вхождения переменных в формулу P свободны или связаны в зависимости от того, свободны или связаны вхождения этих переменных в G или Q . Пусть переменная строка x входит в подформулы G и/или Q свободно. Определим схему и множество атрибутов, с которыми строка x встречается в формулах, для формулы P . Имеют место следующие случаи.

a. Схемы формул $scheme(x, G)$ и $scheme(x, Q)$ определены. Для того чтобы формула P была разрешена, должно выполняться равенство $scheme(x, G) = scheme(x, Q)$. Полагаем по определению $scheme(x, P) = scheme(x, G)$.

b. Схема определена только для одной из подформул. Пусть схема $scheme(x, G)$ определена, а схема $scheme(x, Q)$ – не определена. Для того чтобы формула P была разрешена, должно быть выполнено включение $attr(x, Q) \subseteq scheme(x, G)$. Полагаем по определению $scheme(x, P) = scheme(x, G)$.

c. Схема не определена для обоих подформул. В этом случае и схема $scheme(x, P)$ не определена.

В любом случае $attr(x, P) = attr(x, G) \cup attr(x, Q)$.

f4. Если $P = \exists x(R)G$ и переменная x входит в формулу G свободно, то формула P – разрешена. Кроме того, если $scheme(x, G)$ определена, то должно выполняться равенство $scheme(x, G) = R$ при условии выполнения включения $attr(x, G) \subseteq R$. Так как переменная x не входит свободно в формулу P , то $scheme(x, P)$ и $attr(x, P)$ не определены. Если $y \neq x$, то любое вхождение переменной y в P свободно или связано, в соответствии с тем, свободно или связано вхождение y в G . Если y входит в P свободно, то $scheme(y, P) \simeq scheme(y, G)$ и $attr(y, P) = attr(y, G)$.

f5. Если $P = \forall x(R)G$, то все определения и ограничения такие же, как и в случае f4 для квантора существования.

f6. Если $P = (G)$, то P – разрешена, а свободные и связанные вхождение переменных, схема и множество атрибутов, с которыми переменная строка встречается в формулах, остаются теми же, что и у формулы G .

После введения множества разрешенных формул можем дать окончательное определение выражения исчисления строк. А именно, выражение исчисления строк имеет вид $\{x(R) | P(x)\}$, где

1. формула P – разрешена;
2. переменная x – единственная свободно входящая в формулу P переменная;
3. если $scheme(x, P)$ определена, то $scheme(x, P) = R$, иначе $attr(x, P) \subseteq R$.

Пусть $P(x)$ – разрешенная формула, $R \subseteq A$. В результате подстановки конкретной строки s схемы R вместо переменной x в формулу P получим формулу обозначенную $P(s/x)$. Определим сначала истинностные значения атомов:

- a1. пусть строка x в подформуле $t(x)$ свободна в P . Атом $t(x)$ приобретает значение истина при подстановке конкретной строки s вместо x , если $s \in t$, иначе атом $t(x)$ приобретает значение ложь;
- a2. пусть строка x в подформуле $p(u_1, \dots, u_m)$ свободна в P , тогда при подстановке конкретной строки s_i вместо x , заменим $x(A_i)$ на $d_i \in D$, где $\langle A_i, d_i \rangle \in s_i$ (т.е. d_i – значение атрибута A_i в строке s_i). Атом $p(u_1, \dots, u_m)$, где u_i – термы (предметные константы), приобретает значение истина, при подстановке конкретной строки, если предикат p истинный на соответственных значениях, иначе атом приобретает значение ложь.

Набор значений истинности всех атомов формулы называют ее интерпретацией. Пусть формула P разрешенная формула без свободных переменных. Интерпретация формулы P определяется следующим образом.

- f2. Если $P = \neg G$, то в G нет свободных переменных. Формула P истинная, когда G ложная, и ложная, когда формула G истинная.
- f3. Если $P = G \wedge Q$ или $P = G \vee Q$, то в G или Q нет свободных переменных. Если $P = G \wedge Q$, то формула P истинная, тогда когда формулы G и Q обе истинные, в противном случае формула P ложная. Если $P = G \vee Q$, то формула P ложная, тогда когда формулы G и Q обе ложные, в противном случае формула P истинная.
- f4. Если $P = \exists x(R)G$, то x – единственная свободно входящая в формулу G переменная. Формула P истинная, если существует строка схемы R , $s \in S(R)$, такая, что формула $G(s/x)$ – истинная, иначе формула P ложная.
- f5. Если $P = \forall x(R)G$, то x – единственная свободно входящая в формулу G переменная. Формула истинная, если для каждой строки схемы R , $s \in S(R)$ формула $G(s/x)$ истинная, иначе формула P ложная.
- f6. Если $P = (G)$, то формула P истинная, если формула G истинная, и ложная, если формула G ложная.

Пусть $E = \{x(R) | P(x)\}$ – выражение исчисления строк. Значением выражения E назовем таблицу схемы R , содержащую все строки $s \in S(R)$, такие, что формула $P(s/x)$ истинная.

Теорема. Если E – выражение табличной алгебры, то можно эффективно (и равномерно) построить эквивалентное ему выражение F исчисления строк.

Доказательство. При доказательстве теоремы рассматриваем выражения табличной алгебры, которые содержат только операции объединения, пересечения, разницы, селекции, проекции, соединения и переименования потому, что (оставшиеся) операции деления и активного дополнения можно выразить через данные операции: $E_1 \dot{\div}_{R_2}^{R_1} E_2 = \pi_{R'}(E_1) \setminus_{R'} \pi_{R'}(\pi_{R'}(E_1) \otimes E_2 \setminus_{R_1} E_1)$, где $R_2 \subseteq R_1$, $R' = R_1 \setminus R_2$;

$\tilde{E}_1 = C(E_1) \setminus E_1$, где $C(E_1) = \pi_{A_1}(E_1) \otimes \dots \otimes \pi_{A_n}(E_1)$, $R = \{A_1, \dots, A_n\}$ – схема таблицы, которая является значением выражения E_1 (см., например, [1, 4]).

Доказательство будем проводить индукцией по числу операций в E .

Базис индукции (нет операций). Имеют место два случая. $E = t$, где t – таблица схемы R , положим $F = \{\mathbf{x}(R) | t(\mathbf{x})\}$. E – постоянная таблица $t = \{s_1, \dots, s_n\}$ схемы $\{A_1, \dots, A_m\}$, причем $s_i = \langle \langle A_1, d_{i_1} \rangle, \dots, \langle A_m, d_{i_m} \rangle \rangle = \bigcup_{j=1}^m \langle \langle A_m, d_{i_j} \rangle \rangle$. Положим $F = \{\mathbf{x}(\{A_1, \dots, A_m\}) | \bigwedge_{i=1}^n \bigwedge_{j=1}^m x(A_j) = d_{i_j}\}$. \square

Шаг индукции. Предположим, что утверждение теоремы справедливо для любого выражения табличной алгебры, содержащего менее k операций. Пусть выражение E содержит k операций.

Случай 1(объединение). $E = E_1 \cup_R E_2$. Тогда E_1 и E_2 имеют менее k операций, поэтому существуют выражения исчисления строк $\{\mathbf{x}(R) | P(\mathbf{x})\}$ и $\{\mathbf{x}(R) | Q(\mathbf{x})\}$, эквивалентные E_1 и E_2 соответственно. Положим F равным $\{\mathbf{x}(R) | P(\mathbf{x}) \vee Q(\mathbf{x})\}$. \square

Случай 2(разность). $E = E_1 \setminus_R E_2$. Тогда как и в случае 1 существуют выражения исчисления строк $\{\mathbf{x}(R) | P(\mathbf{x})\}$ и $\{\mathbf{x}(R) | Q(\mathbf{x})\}$, эквивалентные E_1 и E_2 соответственно. Положим F равным $\{\mathbf{x}(R) | P(\mathbf{x}) \wedge \neg Q(\mathbf{x})\}$. \square

Случай 3(пересечение). $E = E_1 \cap_R E_2$. Тогда существуют выражения исчисления строк $\{\mathbf{x}(R) | P(\mathbf{x})\}$ и $\{\mathbf{x}(R) | Q(\mathbf{x})\}$, эквивалентные E_1 и E_2 соответственно. Положим F равным $\{\mathbf{x}(R) | P(\mathbf{x}) \wedge Q(\mathbf{x})\}$. \square

Случай 4(селекция). $E = \sigma_{\tilde{p}, R}(E_1)$. Пусть $\{\mathbf{x}(R) | P(\mathbf{x})\}$ – выражение исчисления строк, эквивалентное E_1 . Положим F равным $\{\mathbf{x}(R) | P(\mathbf{x}) \wedge p(\mathbf{x}(A_1), \dots, \mathbf{x}(A_m))\}$, где $R = \{A_1, \dots, A_m\}$ схема таблицы, которая есть значением выражения E_1 . Здесь предполагается, что предикат-параметр селекции задан как $\tilde{p}(s) = T \Leftrightarrow p(s(A_1), \dots, s(A_m)) = T$, $s \in S(R)$, где p – m -арный сигнатурный предикатный символ. \square

Случай 5(проекция). $E = \pi_{X, R}(E_1)$. Пусть $\{\mathbf{x}(R) | P(\mathbf{x})\}$ – выражение исчисления строк, эквивалентное E_1 . Положим F равным $\{\mathbf{y}(X \cap R) | \exists \mathbf{x}(R)(P(\mathbf{x}) \wedge \bigwedge_{A \in X \cap R} \mathbf{y}(A) = \mathbf{x}(A))\}$. \square

Случай 6(соединение). $E = E_1 \otimes_{R_1, R_2} E_2$. Пусть $\{\mathbf{x}(R_1) | P(\mathbf{x})\}$ и $\{\mathbf{y}(R_2) | Q(\mathbf{y})\}$ – выражения исчисления строк, эквивалентные E_1 и E_2 соответственно. Положим F равным $\{\mathbf{z}(R_1 \cup R_2) | \exists \mathbf{x}(R_1) \exists \mathbf{y}(R_2)(P(\mathbf{x}) \wedge Q(\mathbf{y}) \wedge \bigwedge_{A \in R_1} \mathbf{z}(A) = \mathbf{x}(A) \wedge \bigwedge_{A \in R_2} \mathbf{z}(A) = \mathbf{y}(A))\}$. \square

Случай 7(переименование). $E = Rt_{\xi, R}(E_1)$, где $\xi: \mathbf{A} \xrightarrow{\sim} \mathbf{A}$ – инъективная функция переименования атрибутов. Тогда существует выражение исчисления строк $\{\mathbf{x}(R) | P(\mathbf{x})\}$, эквивалентное E_1 . Положим F равным $\{\mathbf{y}(R_2) | \exists \mathbf{x}(R)(P(\mathbf{x}) \wedge \bigwedge_{C \in R \setminus \text{dom} \xi} \mathbf{y}(C) = \mathbf{x}(C) \wedge \bigwedge_{A \in R \cap \text{dom} \xi} \mathbf{x}(A) = \mathbf{y}(\xi(A)))\}$, где $R_2 = R \setminus \text{dom} \xi \cup \xi[R]$. $\square \square$

Выводы

В работе проведено обобщение табличной алгебры на таблицы, которым приписывается определенная схема. Заданы основные операции над обобщенными таблицами. Классическое исчисления строк

пополнено произвольными предикатными и функциональными сигнатурами на универсальном домене (в то время как обычно рассматривают лишь бинарные предикаты, а функциональная сигнатура вообще пуста). Определено синтаксис термов, атомов и формул исчисления строк; выделен класс разрешенных формул, используя понятие свободных и связанных переменных строк, введено понятие схемы $scheme(x, P)$ и множества атрибутов $attr(x, P)$, с которыми переменная строка встречается в формулах. Доказано, что исчисление строк остается не менее выразительным, чем табличная алгебра. Задача для последующей работы установить дуальный результат.

Литература

- [Редько, Брона, Буй, Поляков, 2001] Реляционные базы данных: табличные алгебры и SQL-подобные языки / В. Н. Редько, Ю. Й. Брона, Д. Б. Буй, С. А. Поляков. Киев: Издательский дом «Академперіодика», 2001.
- [Codd, 1972] E. F. Codd. Relational Completeness of Data Base Sublanguages. In: Data Base Systems. New York: Prentice-Hall, 1972. – P. 65-93.
- [Lacroix, Pirotte, 1977] M. Lacroix, A. Pirotte. Domain-oriented relational languages / M. Lacroix, A. Pirotte. In: Proc. 3rd Int. Conf on Very Large Data Bases. Tokyo, October, 1977. – P. 370-378.
- [Мейер, 1977] Д. Мейер. Теория реляционных баз данных. Москва: Мир, 1987.
- [Ульман, 1977] Дж. Ульман. Основы систем баз данных. Москва: Финансы и статистика, 1983.
- [Дейт, 2005] К. Дж. Дейт. Введение в системы баз данных: [8-е изд.: пер. с англ.]. Москва: Издательский дом «Вильямс», 2005.
- [Коннолли, Бегг, 2003] Т. Коннолли, К. Бегг. Базы данных. Проектирование, реализация и сопровождение. Теория и практика: [3-изд.: пер. с англ.]. Москва: Издательский дом «Вильямс», 2003.
- [Пасичник, Резниченко, 1977] В. В. Пасичник, В. А. Резниченко. Организация баз данных и знаний. Киев: Издательская группа ВНУ, 2006.
- [Мендельсон, 1977] Э. Мендельсон. Введение в математическую логику. Москва: Наука, 1971.
- [Цаленко, 1977] М. Ш. Цаленко. Моделирование семантики в базах данных. Москва: Наука, 1989.

Информация об авторах

Дмитрий Буй – доктор физико-математических наук, старший научный сотрудник, Киевский национальный университет имени Тараса Шевченко, факультет кибернетики: Украина, Киев, 03680, пр. Глушкова 2, корп. 6; e-mail: buy@unicyb.kiev.ua.

Ирина Глушко – аспирантка, Киевский национальный университет имени Тараса Шевченко, факультет кибернетики: Украина, Киев, 03680, пр. Глушкова 2, корп. 6; e-mail: glushkoim@gmail.com.

ТИРАЖИРОВАНИЕ ДАННЫХ В ДИНАМИЧЕСКИ НАСТАИВАЕМЫХ РАСПРЕДЕЛЕННЫХ ИНФОРМАЦИОННЫХ СИСТЕМАХ

Людмила Лядова, Михаил Стрелков

Аннотация: Рассматривается подход к тиражированию данных в распределенных информационных системах, допускающих динамическую реструктуризацию данных, расширение функциональности системы в ходе ее эксплуатации. Описывается модель информационной системы, управляемой метаданными, приводятся описания схемы и алгоритмов тиражирования, обсуждаются некоторые особенности их реализации. Предлагаемые средства позволяют реплицировать данные, размещенные в локальной базе данных системы-источника, в соответствии со схемой, определяемой пользователем, во внешние подсистемы-приёмники в формате, пригодном для тиражирования данных в гетерогенных системах. Восстановление данных в системе-приёмнике происходит в автоматическом и «ручном» режиме.

Keywords: реплицирование данных, схема тиражирования, распределенные информационные системы, гетерогенные системы, динамическая адаптация, метаданные, метамоделирование.

ACM Classification Keywords: C. Computer Systems Organization – C.2 Computer-Communication Networks: C.2.4 Distributed Systems – Distributed applications, Distributed databases; H. Information Systems – H.2 Database Management: H.2.4 Systems – Distributed databases; H.2.5 Heterogeneous Databases – Data translation.

Введение

Распределенная информационная система (ИС) в общем случае предполагает как *распределенное хранение*, так и *распределенную обработку* данных. Данные, используемые при выполнении типовых операций, автоматизируемых с помощью программных средств ИС, должны быть приближены к местам их обработки. Таким образом, при создании крупных распределенных информационных систем, включающих территориально удаленные подсистемы, работающие с собственными копиями данных, размещенными в локальных базах данных (БД) этих подсистем, на локальных серверах, необходимо обеспечить возможность *взаимодействия* ИС, их *интеграции*. При этом приходится решать проблемы, связанные с *фрагментацией* данных, необходимостью их *тиражирования*, передачи из одной подсистемы в другую и *синхронизации* локальных копий. Асинхронный подход к тиражированию данных приемлем для многих ИС, так как зачастую в них не требуется обеспечивать идентичность локальных БД различных подсистем в любое время. Однако при передаче данных из одной подсистемы в другую должна решаться задача их согласования.

Вопросам интеграции распределенных источников данных уделяется большее внимание. Большинство современных промышленных СУБД имеют встроенные средства реплицирования данных несколькими способами [1-4], но при интеграции *разнородных информационных систем* использовать возможности, обеспечиваемые различными СУБД, зачастую не представляется возможным, поэтому задача тиражирования решается и при создании прикладных систем [5]. Кроме того, во многих случаях уже после внедрения ИС возникает необходимость *изменения модели предметной области, реструктуризации данных, расширения функциональности ИС*. Информационные системы, допускающие возможность динамической настройки на меняющиеся условия и потребности пользователей в ходе эксплуатации, обеспечивают максимальную гибкость при установке и оперативную адаптацию к происходящим изменениям. Однако эти возможности ИС должны быть обеспечены соответствующими средствами.

В частности, необходимо предусмотреть возможность динамического изменения модели системы, структуры данных, настройки пользовательского интерфейса на внесенные в модель изменения и т.п. Кроме того, могут модифицироваться шаблоны документов (отчетов), генерируемых системой, перечень операций, выполняемых над данными, размещенными в БД и пр. Динамически адаптируемая ИС должна обеспечивать возможность *оперативного тиражирования всех изменений, выполненных в одной подсистеме, в другие подсистемы*. Наличие настраиваемого на изменения компонента тиражирования не только данных, но и моделей ИС является обязательным для современных информационных систем, отвечающих требованию адаптируемости [6]. Если требуется организовать информационный обмен между системами, реализованными на разных платформах, использование средств, предоставляемых в распоряжение разработчиков системами управления БД, становится невозможным. Таким образом, разрабатываемые средства тиражирования должны быть *платформенно-независимы*.

В данной статье описывается подход к решению задачи тиражирования данных в динамически настраиваемых ИС, основанных на метамоделировании. Работа выполнена в рамках создания CASE-системы METAS, предназначенной для разработки динамически настраиваемых гетерогенных ИС.

Задача тиражирования данных в CASE-системе METAS

Предметная область системы, построенной на основе технологии METAS, описывается на *логическом уровне* и представляется в виде объектов различных типов и связей между ними [6]. Модель, созданная разработчиком, представляется с помощью многоуровневых *метаданных*. Метаданные содержат также информацию о пользовательском интерфейсе системы (*презентационный уровень*), включают описания запросов к БД и шаблоны документов для формирования отчетов и пр. Возможно расширение функциональности системы за счет создания новых типов данных, разработки и подключения новых операций над объектами ИС. Кроме того, возможности пользовательского интерфейса могут быть расширены подключением нестандартных элементов управления. Операции и элементы управления также описывается метаданными ИС. Возможность динамической адаптации обеспечивается функционированием системы в режиме интерпретации построенных моделей.

В данной статье рассматриваются средства тиражирования данных в гетерогенных ИС, созданных на основе технологии METAS, их передачи из одной подсистемы в другую с обеспечением синхронизации, согласования данных с учетом изменений, которые могли произойти в модели системы [7].

Отличительной особенностью предлагаемого подхода к тиражированию является отсутствие требования постоянного оперативного (on-line) соединения между узлами. Это может оказаться необходимым в условиях ненадежных каналов связи и позволяет передавать пакет тиражирования любым удобным способом (на носителе, по электронной почте и т.д.). Возможность тиражирования данных ИС в *асинхронном режиме* – одна из основных задач при разработке программных средств.

Разрабатываемые средства должны обеспечить возможность *тиражирования взаимосвязанных данных в соответствии со схемами, создаваемыми пользователями* в зависимости от их потребностей в обмене информацией. Реализация компонентов тиражирования данных об объектах ИС в рамках CASE-технологии позволяет оперировать с метаданными системы напрямую, что обеспечивает возможность *выбора данных для разработкой схемы реплицирования в терминах предметной области ИС* и делает возможным *автоматическое включение в пакет тиражирования данных о связанных между собой объектах*.

При тиражировании данных возможна реализация различных *моделей: тиражирование данных об объектах системы; тиражирование транзакций* (передача и повторение операций, выполненных в одной подсистеме с момента последней репликации, в другую подсистему). Каждая из моделей обладает

определенными преимуществами. Тиражирование транзакций основывается на ведении *журнала произведенных транзакций* в системе. Пакет тиражирования формируется на основе этого журнала и включает в себя все произведенные в системе изменения с момента последней репликации в конкретную подсистему. Тиражирование транзакций уменьшает сетевой трафик. Минусом технологии является неустойчивость к потерям пакетов репликации. Тиражирование данных об объектах – *тиражирование «снимка» БД*, например, таблиц или горизонтальных и/или вертикальных «вырезов» из них. Метод тиражирования снимка хорош для организации процесса начального наполнения баз данных. К недостаткам технологии можно отнести то, что пакет тиражирования имеет больший размер, чем при тиражировании транзакций. Кроме того, невозможно удаление объектов на другом узле.

При более детальном анализе задачи тиражирования выявляется ряд проблем, которые препятствуют созданию универсального механизма реплицирования в динамически адаптируемых системах:

Объем передаваемых данных. При тиражировании имеет смысл передавать не всю базу данных, а только те данные, которые были «заказаны» и изменились со времени передачи последней их реплики. Кроме того, при тиражировании «в обратном направлении» не следует переносить в базу-приёмник записи, полученные из нее же.

Проблема первичных ключей. В разных подсистемах распределенных ИС в БД вносится информация об объектах (записи, каждой из которых назначается некоторый первичный ключ), которая впоследствии тиражируется в другие БД, где, возможно, уже есть записи с такими же значениями ключей, но соответствующие другим объектам предметной области.

Однозначная идентификация объектов. Чтобы алгоритм тиражирования действовал эффективно, необходимо обеспечить однозначную идентификацию объектов, данные о которых получают из разных источников. В общем случае решается задача однозначной идентификации объектов по совокупности значений различных атрибутов.

Неполнота и/или ошибочность данных. При работе с БД важную роль играет человеческий фактор. При занесении информации в БД люди вносят в нее массу ошибок, причем как случайно, так и преднамеренно, поэтому возникают проблемы в процессе идентификации объектов ИС, определения, какие из значений атрибутов ошибочны в случае их несовпадения.

Поддержание целостности данных. Возможна ситуация, когда при восстановлении реплики в системе-приёмнике выполняется попытка включить данные в таблицы, связанные отношениями с другими таблицами, в которые еще не внесены соответствующие записи.

Различная структура данных в базах, участвующих в процессе тиражирования. В этом случае сложно автоматически установить соответствие между атрибутами одного и того же объекта. Проблему можно решать двумя способами: первый метод состоит в приведении структур данных к одному виду, второй предполагает «ручную» (при участии человека) установку соответствия.

Различные форматы хранения данных. Различные СУБД в гетерогенных распределенных информационных системах могут по-разному хранить данные одного и того же типа.

Удаленные записи. При тиражировании данных необходимо передавать сведения об удаленных записях. Это связано с тем, что в базе-приёмнике нужно удалять не все записи, которые отсутствуют в базе-источнике, а только те, что действительно были удалены именно из нее и их копии не могли быть получены из других источников.

Реализация подсистемы тиражирования данных в рамках технологии METAS требует решения следующих задач: построение модели ИС и компонентов тиражирования; разработка метода тиражирования данных об объектах, алгоритмов создания копий данных и их восстановления,

основанных на построенной модели; разработка метода тиражирования транзакций, обеспечивающего независимость от используемой в системе СУБД; реализация программных компонентов тиражирования в рамках CASE-технологии METAS. Эти задачи решаются с учетом перечисленных выше проблем, различных подходов к их решению.

Базовая модель системы, построенной на основе CASE-технологии METAS

Системы, созданные с помощью CASE-инструментария METAS, функционируют в режиме интерпретации многоуровневых моделей, описывающих систему с различных точек зрения, на разных уровнях детализации. В основе системы – графовые модели, представляющие модель (схему) данных ИС, её пользовательский интерфейс и пр.

Математическая модель данных логического уровня – граф $G_l(V_l, E_l)$, где $V_l = (e_1, \dots, e_p)$, $p \in N$ – множество вершин, представляющих объекты предметной области (сущности); $E_l = (r_1, \dots, r_q)$, $q \in N$ – множество дуг, представляющих связи между ними ($r_i = (e_j, e_k)$, $i = 1..q$; $e_j, e_k \in V_l$), направление дуги определяется типом связи между сущностями. С каждой вершиной графа $e \in V_l$ свяжем множество, $Attr(e) = \{a_{e,0}, a_{e,1}, \dots, a_{e,m}\}$, представляющее атрибуты сущности e , каждый атрибут представляется парой $\langle name, ref \rangle$, где $name$ – имя атрибута, а ref указывает таблицу-справочник, если значение атрибута выбирается из справочника: $a_{ref} \in V_{ph}$, в противном случае – пустое значение $a_{ref} = \emptyset$.

Физический уровень модели системы (уровень представления данных в реляционной БД) описывается графом $G_{ph}(V_{ph}, E_{ph})$, в котором $V_{ph} = \{t_1, t_2, \dots, t_{pn}\}$, $pn \in N$ – множество таблиц в БД; $E_{ph} = \{r_1, r_2, \dots, r_{pm}\}$, $pm \in N$ – множество связей между ними ($r_i = (t_j, t_k)$, $i = 1..pm$; $t_j, t_k \in V_{ph}$). Направление дуги от t_j (начальной вершины) к t_k (конечной вершине) определяется типом связи между таблицами. Дуги графа могут быть только однонаправленными. С каждой вершиной графа $t \in V_{ph}$ связывается множество $Fields(t)$, обозначающее поля таблицы t . $Fields(t) = \{f_{t,0}, f_{t,1}, \dots, f_{t,n_t}\}$, где каждый элемент представляет собой четверку $\langle name, type, req, pe \rangle$ ($name$ – имя поля таблицы; $type$ – тип поля; req – признак, является ли значение обязательным; pe указывает родительскую таблицу ($f_{pe} \in V_{ph}$), если поле содержит внешний ключ (первичной ключ родительской таблицы), иначе pe имеет пустое значение ($f_{pe} = \emptyset$). Будем считать, что элемент $f_{t,0}$ представляет ключевое поле таблицы t . Таким образом, каждая вершина $t \in V_{ph}$ представляет структуру таблицы в БД.

Экземпляр данных (строка) o таблицы t включает множество значений полей. Обозначим значение ключевого поля (идентификатора) строки o $Id(o) = Val(f_{t,0}, o) \in N$.

Зададим местоположение данных (строк таблицы) $O(t, place)$, $t \in V_{ph}$, $place \in Place$ ($place$ задает местоположение объектов – значение из множества $Place = \{db, pk\}$ – база данных ИС или пакет тиражирования соответственно). Будем обозначать $O(t, db) = O_{db}(t)$ и $O(t, pk) = O_{pk}(t)$. Обозначим через $CAV(r, o)$, $r = (t_1, t_2) \in E_{ph}$, $o \in O(t_2, place)$ значение дочернего поля связи указанного объекта.

Пусть $MT(e) = t$, $e \in V_l$, $t \in V_{ph}$ – главная таблица сущности e ; $S(e)$, $e \in V_l$ – множество справочников сущности e .

Отношение «М:М» между сущностями реализуется в системе при помощи вспомогательной таблицы. Ей будет соответствовать функция:

$$Mid(r = (e_1, e_2)) = \begin{cases} t_m, (e_2, e_1) \in E_l \\ \emptyset, (e_2, e_1) \notin E_l \end{cases}$$

Опишем соотношения вершин и дуг графа логической модели и элементов графа физической модели.

Любой вершине (сущности) графа логической модели соответствует множество таблиц на физическом уровне, состоящее из главной таблицы и таблиц-справочников:

$$(\forall e \in V_l)(\exists T_e \subset V_{ph}) : T_e = MT(e) \cup S(e)$$

Любой связи «1:М» или «М:1» на логическом уровне соответствует одна связь на физическом.

Связь «М:М» реализуется через промежуточную таблицу, при этом главные таблицы сущностей, состоящих в связи, имеют отношение «1:М» с промежуточной таблицей:

$$\begin{aligned} & (\forall r = (e_1, e_2) \in E_l, e_1, e_2 \in V_l)(\exists E_r \subset E_{ph}) : \\ & E_r = \{(t_1, t_2) \in E_{ph} \mid t_1, t_2 \in V_{ph}, t_1 = MT(e_1), t_2 = MT(e_2), Mid(r) = \emptyset\} \vee \\ & \vee \{(t_1, t_m) \in E_{ph}, (t_2, t_m) \in E_{ph} \mid t_1, t_2, t_m \in V_{ph}, t_1 = MT(e_1), t_2 = MT(e_2), t_m = Mid(r) \neq \emptyset\} \end{aligned}$$

Все множество вершин на физическом уровне можно представить следующим образом:

$$V_{ph} = \left\{ \bigcup_{e \in V_l} MT(e) \cup S(e) \cup \left(\bigcup_{\substack{d \in V_l, \\ (e,d) \in E_l}} Mid(e, d) \right) \right\}$$

Это объединение главных таблиц и справочников всех сущностей, а также всевозможных вспомогательных таблиц, реализующих связь «М:М».

Между атрибутами сущности и полями главной таблицы можно установить взаимно однозначное соответствие.

Понятие схемы тиражирования

При тиражировании создаётся копия данных, представляющих информацию о некоторых объектах предметной области ИС, размещенных в локальной БД ее подсистемы. Средства реплицирования позволяют извлекать (восстанавливать) данные из этой копии в базы данных других подсистем ИС.

В динамически настраиваемой системе, допускающей реструктуризацию данных в БД, меняется и структура данных, передаваемых из одной подсистемы в другую. Поэтому необходимо обеспечить пользователя средствами, позволяющими не только определить перечень передаваемой информации, но и структуру пакета тиражирования. При создании копии пользователь должен выбрать экземпляр сущности (*начальный*, или *корневой*, объект), информацию о котором требуется передать на другие узлы, а также сформировать список *дочерних сущностей*, связанных с *корневой*, информация о которых также должна быть передана в другую БД. Для каждой сущности из этого списка должны быть переданы в пакет только те экземпляры, которые имеют отношение к выбранному объекту. При этом начальная сущность может иметь обязательные связи (ссылки) на *родительские сущности*. Для создания корректной копии при включении сущности в список вместе с ней должны быть добавлены все ее родительские сущности.

Вводится понятие *схемы тиражирования*, которая включает:

список сущностей, выбранных пользователем $V_{sc} \subset V_l$;

список атрибутов каждой выбранной сущности, необходимых для сопоставления экземпляров (записей) при восстановлении данных $\forall e \in V_{sc} : \exists IdentAt(e) \subset Attr(e)$, причем список включает все обязательные атрибуты сущности плюс, возможно, некоторые другие:

$$IdentAt(e) = req(e) \cup nodefa(e), nodefa(e) \subset Attr(e);$$

- для каждого атрибута задается точность сопоставления $\forall a \in IdentAt(e) : \exists EqAc(a) \subset Accuracy$ (здесь множество *Accuracy* задает всевозможные наборы параметров сравнения значений (полное, с учетом регистра, с учетом пробела и др.); факт того, что значения атрибутов (или полей) совпадают с точностью *ea*, будем записывать следующим образом: $Val(f, o) \stackrel{ea}{\cong} Val(f, p)$, $o \in O(t, pl1)$, $p \in O(t, pl2)$, $f \in Fields(t)$, $t \in V_{sc}$;
- правила сопоставления объектов: требуется, чтобы выполнялось равенство значений всех атрибутов или хотя бы одного; правило задается так: $EqType(e) \in EqTypes = \{all, one\}$, $e \in V_{sc}$;

корневую сущность $e_{start} \in V_{sc}$ и корневой объект $o_{start} \in O(e_{start})$.

Модель данных, образованная списком сущностей V_{sc} , представляет собой некоторую подсхему логической модели. Графом схемы тиражирования назовём подграф графа логической модели G_l , полученный путём редукции вершин (сущностей), не включенных в схему тиражирования (из множества вершин V удаляются все «лишние» вершины, а из множества дуг – все дуги, инцидентные этим вершинам): $G_{sc}(V_{sc}, E_{sc})$, где $V_{sc} = (e_1, \dots, e_n) \subset V_l$, $n \in N$, $n \leq p$ – множество сущностей, заданных схемой; $E_{sc} = (r_1, \dots, r_m) \subset E_l$, $m \in N$, $m \leq q$ – множество связей между ними; $r_i = (e_j, e_k)$, $i = 1..m$; $e_j, e_k \in V_{sc}$, $E_{sc} = E_l \setminus \{r = (e_j, e_k) \in E_l \mid e_j \in V_l \setminus V_{sc} \vee e_k \in V_l \setminus V_{sc}\}$, т.е. из множества дуг графа логической модели G_l удаляются все дуги, инцидентные сущностям, не включенным в схему тиражирования.

Алгоритм создания копии при тиражировании данных

Операция создания копии по заданной схеме – это операция построения графа $G_{pk}(V_{pk}, E_{pk})$, где $V_{pk} \subset V_{ph}$, $E_{pk} \subset E_{ph}$. Учитывая то, что схема задает для копирования набор сущностей V_{sc} , становится возможным определить множество таблиц в этом графе

$$V_{pk} = \left\{ \bigcup_{e \in V_{sc}} MT(e) \cup S(e) \cup \left(\bigcup_{\substack{d \in V_{sc} \\ (e,d) \in E_{sc}}} Mid(e,d) \right) \right\}.$$

При этом в пакет тиражирования должны попасть объекты, имеющие отношение к начальному, -- множество $O(t, pk) \subset O(t, db)$ для каждой таблицы $t \in V_{pk}$.

Разработанный алгоритм выполняет рекурсивный обход графа схемы тиражирования в глубину. Перед запуском алгоритма необходимо указать корневой объект и в качестве входных параметров операции задать корневую сущность e_{start} и идентификатор корневого объекта $Id(o_{start})$. Если N – число вершин в графе, aK – общее число строк в таблицах, то одна и та же вершина может быть выбрана при обходе в худшем случае K раз, таким образом, рекурсивный вызов может быть выполнен $O(K \cdot N)$. Показано, что алгоритм создания копии (реплики) данных при тиражировании имеет сложность $O(K^3 N + KN^2)$.

Алгоритм восстановления данных при тиражировании

Операция восстановления данных из полученной копии подразумевает обновление данных, содержащихся в локальной БД подсистемы ИС, на основе реплики БД, созданной в другой подсистеме. Пакет тиражирования содержит множество таблиц $V_{pk} \subset V_{ph}$, заполненных строками $O(t, pk)$.

Необходимо организовать просмотр и анализ всех записей. При анализе очередной строки в пакете нужно искать запись в локальной БД, которая представляет тот же самый объект. В случае успешного поиска требуется обновить значения атрибутов, иначе – добавить объект в множество $O(t, db)$.

Правила сопоставления для каждой сущности задаются схемой тиражирования, по которой была создана резервная копия. Строки считаются *похожими*, если у них совпадают значения идентифицирующих атрибутов $IdentAt(e)$ с точностью $EqAc(e)$, где $e \in V_j$. При распознавании строк могут возникать *неопределенности*, связанные с неполнотой или ошибочностью данных. Их разрешение целесообразно возложить на пользователя. Если для объекта в пакете не был найден соответствующий объект в БД или их было найдено несколько, то этот экземпляр будет помещен в *промежуточный* (или *транзитный*) пакет вместе с результатами распознавания. Чтобы сохранить целостность данных, необходимо помещать в транзитный пакет и все дочерние объекты «сомнительного» экземпляра.

Будем считать, что граф $G_{pk}(V_{pk}, E_{pk})$ и граф логической модели $G_l(V_l, E_l)$ на узле-приемнике изоморфны соответствующим графам на узле-источнике (подсхема модели данных, образованная схемой тиражирования, одинакова в этих подсистемах ИС). Для этих условий сложность одной итерации алгоритма восстановления при обходе таблиц копии – $O(K \cdot A)$, где K – общее число строк в таблицах, A – число атрибутов. Поскольку в цикле просматриваются все строки таблицы, общая сложность процедуры восстановления равна $O(K^2 \cdot A)$.

Если требуется вручную устанавливать соответствие объектов из промежуточного пакета с объектами БД подсистемы-приемника, то при добавлении нового экземпляра объекта в БД по решению пользователя запускается специальный алгоритм восстановления данных из транзитного пакета.

Заключение

В работе предложен подход к тиражированию данных в системах, управляемых метаданными.

Современные системы управления базами данных (Oracle, Microsoft SQL Server и др.) содержат встроенные средства тиражирования данных, однако большинство из них имеют ограничения: репликация возможна только между БД одного формата, структуры таблицы-источника и таблицы-приемника должны быть одинаковы.

Предлагаемый подход снимает эти ограничения. Его можно сравнить с тиражированием моментальных снимков в Microsoft SQL Server, но использование моделей, метаданных различного уровня позволяет пользователю работать в терминах предметной области ИС, а также осуществлять гибкую настройку правил сопоставления различных экземпляров сущностей при их передаче из одной подсистемы ИС в другую.

В качестве инструментального средства разработки компонентов тиражирования CASE-системы METAS выбрана платформа Microsoft .NET, для доступа к данным в БД используется технология ADO.NET, данные в пакете тиражирования передаются в формате XML.

Благодарности

Статья частично финансирована из проекта **ITHEA XXI** Института Информационных теорий и Приложений FOI ITHEA и консорциума FOIBulgaria (www.ithea.org, www.foibg.com)

Библиографический список

- [1] Артемов Д.В. Microsoft SQL Server // Корпоративные базы данных'96: Материалы технической конференции / Центр информационных технологий. – Москва, 1996.
- [2] Барон Г., Ладьженский Г. Технология тиражирования данных в распределенных системах // Открытые системы. – 1994. №2. С. 17-22.
- [3] Сиколенко В.В. Поддержка распределенных систем в СУБД Oracle // Системы управления базами данных. – 1996. – № 4. – С. 27-35.
- [4] Armendariz-Inigo J.E., Juarez-Rodriguez J.R., de Mendivil J.R.G., Garitagoitia J.R., Munoz-Escori F.D., Irun-Briz L. Relaxed Approaches for Correct DB-Replication with SI Replicas // Software and Data Technologies: Proceedings of the Third International Conference, ICSOFT 2008. Porto, Portugal, July 22-24, 2008. – Springer-Verlag Berlin Heidelberg, 2009. P. 161-174.
- [5] Карауш А.С. Модель тиражирования библиографических баз данных // EL Pub2003: VIII Международная конференция по электронным публикациям / Институт вычислительных технологий СО РАН. Новосибирск, 2003.
- [6] Лядова Л.Н. Метамоделирование и многоуровневые метаданные как основа технологии создания адаптируемых информационных систем // International Book Series "Information Science & Computing", Number 4. Supplement to the International Journal "Information Technologies & Knowledge" Vol. 2, 2008. P. 125-132.
- [7] Стрелков М.А. Компонент тиражирования данных в системах, основанных на метаданных // Инженерный вестник / Научно-технический журнал общественного объединения «Белорусская инженерная академия» – Минск, 2006. – №1(21)/2.

Сведения об авторах

Лядова Людмила Николаевна – доцент кафедры информационных технологий в бизнесе Пермского филиала Государственного университета – Высшей школы экономики; 614070, Россия, г. Пермь, ул. Студенческая, 38; e-mail: LNLyadova@mail.ru.

Стрелков Михаил Александрович – студент магистратуры; Пермский государственный университет; 614990, Россия, г. Пермь, ул. Букирева, 15; e-mail: strelkopf@yandex.ru.

МУЛЬТИАГЕНТНЫЙ ПОДХОД К РЕШЕНИЮ ПРОБЛЕМЫ РАВНОМЕРНОГО РАСПРЕДЕЛЕНИЯ ВЫЧИСЛИТЕЛЬНОЙ НАГРУЗКИ

Александр Миков, Елена Замятина, Арсений Козлов

Аннотация: В работе рассматриваются различные подходы к решению проблемы балансировки во время распределенного имитационного эксперимента. Авторы предлагают решать эту проблему с помощью управляемой балансировки, которая позволяет при распределении нагрузки по вычислительным узлам использовать не только знания о состоянии вычислительной системы, но и знания о поведении конкретной имитационной модели. При этом предполагается использовать мультиагентный подход. Авторы приводят архитектуру мультиагентной системы балансировки, рассматривают вопросы реализации и результаты проведенных экспериментов.

Keywords: Распределенные вычисления; мультиагентная система; динамическая балансировка; имитационная модель.

ACM Classification Keywords: 1.6 Simulation and Modelling – 1.6.2 Simulation Languages; 1.2 Artificial Intelligence – 1.2.5 Programming Languages and Software – Expert system tools and techniques.

Введение

Метод имитационного моделирования является известным, признанным, а иногда, и единственным методом исследования сложных систем. Зачастую имитационный эксперимент требует больших вычислительных ресурсов. В этом случае целесообразно проводить распределенный имитационный эксперимент, используя ресурсы нескольких узлов вычислительной системы (ВС) (компьютерной сети, многопроцессорной ЭВМ и т.д.). Использование вычислительных ресурсов нескольких узлов вычислительной системы позволяет оптимизировать имитационный эксперимент по времени и по надежности (при выходе из строя одного или нескольких вычислительных узлов их функции берут на себя другие узлы). Распределенная имитационная модель (ИМ) представляет собой совокупность логических процессов LP_i , где $i = 1..n$, которые выполняются параллельно на различных узлах вычислительной системы (ВС) и обмениваются друг с другом сообщениями. Для синхронизации логических процессов в системах распределенного имитационного моделирования применяют специальные алгоритмы: консервативный, оптимистический [Fujimoto, 2003]. Оба этих алгоритма следят за тем, чтобы не возник парадокс времени – ситуация, когда процесс посылает сообщение со штампом времени меньшим, чем локальное время процесса, которому отсылают это сообщение. Однако выигрыши от параллельного выполнения распределенной имитационной модели могут быть сведены к нулю в случае, если возникнет дисбаланс загрузки вычислительных узлов.

Причина дисбаланса заключается в следующем:

- гетерогенность вычислительной системы (узлы ВС имеют разную производительность, а линии связи – разную пропускную способность);
- гетерогенность имитационной модели (некоторые процессы имеют гораздо большую интенсивность обменов, нежели другие, часть процессов оказывает большую нагрузку на вычислительные узлы, выполняя одно событие за другим, в то время как другие могут быть приостановлены, ожидая прихода того или иного сообщения);
- нагрузка, которую оказывают на вычислительные узлы сторонние приложения.

Таким образом, целесообразно корректировать возникающий дисбаланс. С этой целью обычно разрабатывают специальное программное обеспечение, которое следит за нагрузкой вычислительных узлов и восстанавливает равномерное распределение приложений по вычислительным узлам, выполняя перенос части компонентов приложений на другие, менее загруженные узлы. При этом программное обеспечение, выполняющее балансировку, следит за передачей сообщений по линиям связи. Нагрузка линий связи также должна быть сбалансирована.

Восстановление баланса нагрузки является хорошо известной задачей, которая получила множество решений. Разработано большое количество алгоритмов. Однако очень часто эти алгоритмы применимы только для конкретного приложения, для решения конкретной задачи. Существуют и решения, которые применимы для оптимизации распределенного имитационного эксперимента. Однако, несмотря на то, что разработчики подсистемы балансировки SPEEDES [Wilson, 1998] и Charm++ [Zheng, 2005] пытались разработать алгоритмы балансировки, которые могли бы адаптироваться к изменяющейся обстановке, к характерным особенностям той или иной имитационной модели, результаты экспериментов показали, что эффективности от применения этих алгоритмов им удалось добиться лишь в частных случаях. Действительно, найти универсальный алгоритм, который мог бы быть эффективным (обеспечить сокращение времени выполнения имитационного эксперимента) для любой имитационной модели практически невозможно.

Предлагается хотя бы частично разрешить эту проблему, применив *управляемую балансировку* [Миков, 2007]. Управляемая балансировка предполагает, что программное обеспечение, применяемое для восстановления равномерной загрузки должно включать экспертный компонент, который на основании логического вывода предлагает решения для переноса избыточной нагрузки с перегруженного вычислительного узла на менее загруженный. Программные средства балансировки предназначены для системы распределенного имитационного моделирования Triad.Net. (Triad – система автоматизированного проектирования и имитационного моделирования вычислительных систем [Mikov, 1995], Triad.Net – распределенная ее версия [Миков, 2008], [Миков, 2009], для описания имитационной модели используют язык Triad).

Кроме того, для разработки оригинального алгоритма балансировки, учитывающего все особенности конкретного параллельного (распределенного) приложения, авторы предлагают специальные языковые средства. Эти языковые конструкции написаны на входном языке системы моделирования Triad.Net (язык Triad). Для того чтобы можно было сократить временные затраты на саму балансировку, авторы предлагают использовать мультиагентный подход, который позволяет применить децентрализованный алгоритм балансировки и тем самым сократить время на обмены между компонентами самой подсистемы балансировки. Архитектура мультиагентной подсистемы балансировки, протоколы и алгоритмы взаимодействия агентов будут описаны ниже.

Задача балансировки

Задача балансировки – это задача отображения неизоморфных связанных графов $B: TM \rightarrow NG$, где TM – множество графов моделей, а NG – множество графов конфигураций компьютерной сети. Граф G из NG , $G = (C, Ed)$, определяется множеством вычислительных узлов C и множеством ребер Ed , представляющих линии связи. Граф M из TM , задает имитационную модель. Имитационную модель (ИМ) M можно представить и как совокупность логических процессов $MP = \{LP_j\}$, $j = 1..n$, взаимодействующих между собой путем передачи сообщений.

Описание имитационной модели

Описание модели в системе Triad можно определить как $M = (STR, ROUT, MES)$, где STR – слой структур, $ROUT$ – слой рутин, MES – слой сообщений.

Слой структур представляет собой совокупность объектов, взаимодействующих друг с другом посредством посылки сообщений. Каждый объект имеет полюсы (входные и выходные), которые служат соответственно для приёма и передачи сообщений. Основа представления слоя структур – графы. В качестве вершин графа следует рассматривать отдельные объекты. Дуги графа определяют связи между объектами.

Объекты действуют по определённому алгоритму поведения, который описывают с помощью *рутины*. Рутинa представляет собой последовательность *событий* e_i , планирующих друг друга (E – множество событий; множество событий рутинa является частично упорядоченным в модельном времени). Выполнение события сопровождается изменением состояния объекта. Состояние объекта определяется значениями переменных рутинa. Таким образом, система имитации является событийно-ориентированной. Рутинa так же, как и объект, имеет входные и выходные полюса. Входные полюса служат соответственно для приёма сообщений, выходные полюса – для их передачи. В множестве событий рутинa выделено входное событие e_m . Все входные полюса рутинa обрабатываются входным событием. Обработка выходных полюсов осуществляется остальными событиями рутинa. Для передачи сообщения служит специальный оператор *out* (*out*<сообщение>*through*<имя полюса>). Совокупность рутин определяет *слой рутин ROUT*.

Слой сообщений (MES) предназначен для описания сообщений сложной структуры.

Система Triad реализована таким образом, что пользователю необязательно описывать все слои. Так, если возникает необходимость в исследовании структурных особенностей модели, то можно описать в модели только слой структур.

Алгоритмом имитации называют совокупность объектов, функционирующих по определённым сценариям, и синхронизирующий их алгоритм.

Алгоритм исследования

Для сбора, обработки и анализа имитационных моделей в системе Triad.Net существуют специальные объекты – *информационные процедуры* и *условия моделирования*. Информационные процедуры и условия моделирования реализуют алгоритм исследования модели. Информационные процедуры ведут наблюдение за элементами модели (событиями, переменными, входными и выходными полюсами), указанными пользователем. Если в какой-либо момент времени имитационного эксперимента пользователь решит, что следует установить наблюдение за другими элементами или выполнять иную обработку собираемой информации, он может сделать соответствующие указания, подключив к модели другой набор информационных процедур. Условия моделирования анализируют результат работы информационных процедур и определяют, выполнены ли условия завершения моделирования. Информационные процедуры и условия моделирования используют и для сбора информации о поведении модели, о ее характеристиках, и в системе балансировки для сбора информации о модели.

Для имитационной модели в Triad возможно использование операций над моделью (добавление и удаление вершины, добавление и удаление дуг и ребер и др.). Операции изменяют структуру имитационной модели во время имитационного прогона, а это в свою очередь требует перераспределения нагрузки на вычислительных узлах.

Кроме того, авторы разрабатывают систему имитации с удаленным доступом. Это предполагает балансировку запросов при обращении удаленных пользователей к имитационной модели. В настоящее время реализуется распределенная версия Triad.Net. Синхронизация объектов модели, располагающихся на разных вычислительных узлах, выполняется оптимистическим алгоритмом.

Разработка подсистемы управляемой балансировки (централизованный алгоритм)

Как уже говорилось ранее, управляемая балансировка Triad.Net предполагает наличие экспертного компонента. База знаний экспертного компонента включает знания исследователя о конкретной имитационной модели. К примеру, исследователь знает, что через 100 единиц модельного времени с начала моделирования два объекта имитационной модели, представленные логическими процессами LP_i и LP_j , должны интенсивно обмениваться информацией в течение некоторого временного интервала.

В этом случае целесообразно к моменту времени

$$t = \text{"начало моделирования"} + 100 \text{ ед. модельного времени}$$

перенести эти два процесса на один вычислительный узел или расположить их на соседних вычислительных узлах, если линия связи между ними имеет хорошую пропускную способность. При переносе нагрузки экспертный компонент использует правила, учитывающие знания о топологии ВС и ИМ, знания о нагрузке узлов, линий связи и о функционировании имитационной модели.

Таким образом, система балансировки в Triad.Net включает следующие компоненты: подсистему анализа и принятия решения, подсистему миграции, подсистему мониторинга имитационной модели, подсистему мониторинга вычислительной системы, базу знаний с правилами перераспределения нагрузки, редактор правил, механизм вывода и подсистему оценки качества оптимизации. *Подсистема мониторинга ВС* собирает информацию о текущем состоянии вычислительной системы, на которой выполняется имитационный эксперимент. *Подсистема мониторинга имитационной модели* собирает информацию о текущем состоянии ИМ (использует механизм информационных процедур). *Подсистема анализа* получает информацию от подсистем мониторинга и принимает решение о перераспределении нагрузки, после чего обращается к *экспертной системе* и получает от нее рекомендации о том, какие объекты имитационной модели следует перенести и на какие вычислительные узлы. Далее управление передают *подсистеме миграции*, которая и осуществляет перенос объектов. *Подсистема оценки качества оптимизации* определяет, насколько изменилось время имитационного прогона и условия этого прогона. Управление является централизованным, все правила хранятся в единой базе знаний, подсистема мониторинга вычислительной системы также располагается на одном из выделенных вычислительных узлов, взаимодействуя с остальными узлами. Централизованный алгоритм характеризуется достаточно ощутимыми временными затратами на коммуникацию.

Мультиагентная подсистема балансировки (децентрализованный алгоритм)

Динамическая система балансировки TriadBalance является *мультиагентной*, она состоит из *совокупности агентов разных типов*: агента-датчика вычислительного узла; агента-датчика имитационной модели; агента анализа; агента миграции; агента распределения. Агенты каждого типа действуют по своему сценарию для достижения цели, а вместе они реализуют балансировку распределенной имитационной модели.

Агент-датчик вычислительного узла постоянно собирает информацию о нагрузке вычислительного узла, о состоянии линий связи, о наличии свободной памяти. Этот агент взаимодействует с агентом анализа, используя доску объявлений.

Агент-датчик имитационной модели ведет наблюдение за объектами имитационной модели во время имитационного прогона, регистрируя интенсивность обмена между объектами, частоту выполнения тех или иных событий, частоту изменения, переменных и т.д. Агент-датчик имитационной модели использует механизм информационных процедур. Он взаимодействует с агентом распределения (опять же используя доску объявлений).

Агент анализа, взаимодействуя с агентом-датчиком вычислительного узла (эти агенты являются реактивными), принимает решение о необходимости перераспределения нагрузки. Он является когнитивным агентом и, принимая решения, использует правила из базы знаний агента. В этих правилах анализируются данные о работе аппаратуры: общее время мониторинга; загруженность процессора (в процентах); общее количество дескрипторов в системе; общее количество потоков в системе и т.д.

Агент распределения выполняет следующую роль: он должен решить, какой из объектов, если их несколько на одном узле, следует перенести на другой узел. Для решения этой проблемы агент должен располагать информацией, полученной от агента-датчика имитационной модели. Эту информацию агент распределения извлекает с доски объявлений. Агент распределения определяет, на какой вычислительный узел стоит перенести объект имитационной модели. С этой целью он обращается к агентам распределения вычислительных узлов-соседей.

Таким образом, алгоритм балансировки является *децентрализованным* и представляет собой последовательность шагов:

Агенты-датчики ВС (они располагаются на каждом вычислительном узле) регулярно на всем протяжении имитационного эксперимента собирают статистику о загруженности узлов и размещают ее в базе данных (доска объявлений). Каждый вычислительный узел располагает своей доской объявлений.

Агент анализа, сканируя доски объявлений, с помощью правил из базы знаний определяет, необходимо ли выполнять балансировку. Например, при превышении показателя загрузки вычислительного узла ($Node.Pwr > PwrLimit$) агент анализа принимает решение о необходимости выполнения балансировки. В этом случае агент анализа обращается к агенту распределения.

Агент распределения извлекает информацию из базы данных. Информация в базе данных появляется в результате работы агента-датчика имитационной модели. Для извлечения информации о модели используют механизм информационных процедур. В частности, агенты-датчики собирают информацию о частоте выполнения событий имитационных объектов, о частоте обменов между ними (частота появления сообщений на входных и выходных полюсах рутин). Таким образом, агент распределения на основании анализа данных на доске объявлений может сделать вывод о том имитационном объекте, который следует перенести на другой узел. Кроме того, он общается с агентами распределения, расположенными на соседних вычислительных узлах. Он сообщает о перегрузке, о том, что ему необходимо освободиться от некоторых своих имитационных объектов. Агенты распределения других вычислительных узлов извещают о своей нагрузке, пополняя базу данных агента распределения, корректируя правила в базе знаний этого агента. Во время сеанса взаимодействия агенты-распределения других узлов могут сообщить о том, что они незагружены. Далее агент распределения действует по правилам, которые хранятся в базе знаний. Для принятия решения об адресе целевого узла агент распределения, наряду с другими правилами, использует, в частности, правила топологических характеристик ИМ и ВС. При выборе целевого узла сообщение направляется агенту-миграции.

Агент миграции, получив запрос от агента распределения, выполняет непосредственно перенос выбранных объектов имитационной модели на выбранные целевые узлы сети.

Таким образом, выполняется *децентрализованный алгоритм балансировки*. Действительно, нет единого управления процессом балансировки, группа агентов управляет балансировкой на собственном узле, учитывая информацию только соседних вычислительных узлов.

Для повышения гибкости мультиагентной системы балансировки, адаптируемости когнитивных агентов к изменяющимся условиям используют два типа метаправил: это *метаправила, определяющие структуру правил* (их используют агенты распределения и анализа для распознавания нужных правил) и *метаправила, определяющие, как можно изменить правила*.

Второй вид метаправил определяет данные и правила, а также критерии, на основании которых будут изменяться правила балансировки когнитивных агентов. Так для агента анализа определены правила, примеры которых приведены ниже:

- *IF %Выражение1% ЗНАК %Выражение2% THEN ОТДАТЬ* (Если значения *Выражения1* и *Выражения2* удовлетворяют условию, то необходимо перенести часть объектов модели имитации с данного узла на другие).
- *IF %Выражение1% ЗНАК %Выражение2% THEN ПРИНЯТЬ* (Если значения *Выражения1* и *Выражения2* удовлетворяют условию, то необходимо перенести часть объектов модели имитации с других узлов на данный).
- *IF %Выражение% BETWEEN %Выражение1% AND %Выражение2% THEN ПРИНЯТЬ* (Если значение *Выражения* находится между значениями выражений *Выражение1* и *Выражение2*, то необходимо перенести часть объектов модели имитации с данного узла на другие).

Агент анализа для вычисления выражений использует информацию, размещенную на доске объявлений, агрегированную по узлу. Напротив, агент распределения использует детализированную информацию для выбора объектов модели имитации. В системе определены следующие типы структур правил для агента распределения: *IF %Выражение1% ЗНАК %Выражение2% THEN ОТДАТЬ*; *IF %Выражение% BETWEEN %Выражение1% AND %Выражение2% THEN ОТДАТЬ*.

В процессе имитационного эксперимента может сложиться ситуация, при которой некоторые из определенных пользователем правил не будут выполняться и, следовательно, уменьшится эффективность динамической балансировки, основанной на правилах. В таком случае система балансировки должна сама следить за состоянием правил и корректировать их, удалять или добавлять новые, в зависимости от сложившейся ситуации в процессе имитационного эксперимента. Для того чтобы агенты могли правильно понимать, в каких случаях необходимо корректировать, добавлять или удалять правила, в системе введены метаправила, определяющие правила изменения правил, например: «*Если частота выполнения правила больше (>, <, <=, >=, ==) заданного значения, то необходимо изменить правило*»; «*Если правило не выполняется никогда, то необходимо изменить правило*»; «*Если правило не может быть выполнено, то необходимо изменить правило*» и т.п. Таким образом, система балансировки для каждого правила ведет статистику по эффективности работы правил, на основе которой определяется необходимость их изменений.

Реализация подсистемы мультиагентной балансировки и результаты экспериментов

Настоящая версия мультиагентной подсистемы динамической балансировки PSU.HPC.TriadBalance распределённой имитационной модели реализована на кластере из 8 вычислительных узлов (рис. 1), работающих под управлением операционной системы (ОС) Windows HPC Server 2008 (далее WinHPC), которая является специализированной кластерной версией (High Performance Computing) операционной системы Windows Server 2008. Подсистема разработана с использованием .Net Framework 3.5 и пакета

сборку HPC Rack 2008, позволяющего управлять распределённым выполнением приложений в используемой авторами ОС. Вычислительным узлом кластера может считаться как отдельный компьютер, так и серверная стойка. Выделяются узлы трех типов: головной узел (Head Node), вычислительный узел (Compute Node) и узел брокера WCF (WCF Broker Node). Головной узел может также выполнять и другие функции, т.е. быть вычислительным узлом или узлом брокера WCF. Для управления кластерными вычислениями использовался «HPC Job Manager».

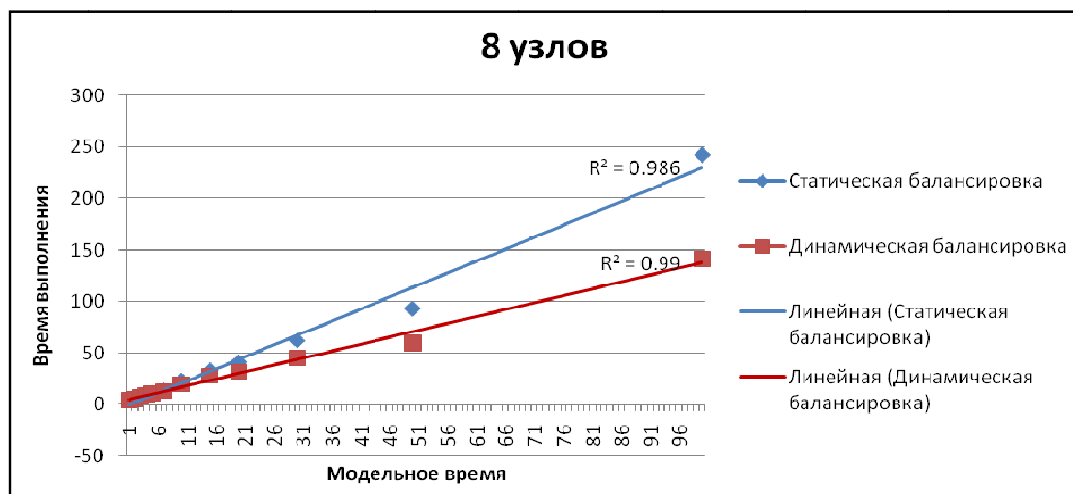


Рис. 1. Результаты прогона модели «Клиент-сервер» на 8 вычислительных узлах кластера

Перед запуском системы выполнялась *предварительная статическая балансировка модели* (начальное размещение объектов модели на вычислительных узлах сети). Следующий шаг – *настройка системы*. Пользователь, основываясь на знаниях о модели (он знает, как должна работать модель), модифицирует правила балансировки. Система балансировки включает *редакторы правил и метаправил*. Правила формируются в виде XML-файла. На основании этих правил агенты принимают решения о переносе объектов модели с одного вычислительного узла сети на другой.

В систему балансировки включена *подсистема визуализации*, с помощью которой можно определить, на каких узлах располагаются объекты имитационной модели. Текущее расположение объектов имитационной модели на вычислительных узлах можно получить в любой момент времени имитационного эксперимента по запросу пользователя. Кроме того, подсистема визуализации позволяет оценить качество оптимизации, отображая в удобных формах загрузку процессоров (и других показателей о функционировании модели и вычислительных узлов во время имитационного прогона).

Для проведения экспериментов была разработана имитационная модель "Клиент-Сервер". В результате экспериментов удалось получить данные, которые представлены на рис.1. Результаты показали, что время, затраченное на имитационный эксперимент при использовании подсистемы мультиагентной балансировки, действительно снижается. Применение подсистемы становится более эффективным при больших значениях системного времени моделирования.

Заключение

В работе рассмотрены вопросы реализации динамической балансировки, предложен централизованный и децентрализованный алгоритмы. Децентрализованный алгоритм основан на мультиагентном подходе. В работе рассматривается также архитектура подсистем управляемой балансировки, реализующих оба алгоритма. Для подсистемы балансировки, основанной на мультиагентном подходе, приведены перечень агентов, их назначение, алгоритмы их взаимодействия и особенности реализации. Кроме того, приводятся

результаты прогона модели «Клиент-Сервер», которые показывают выигрыш во времени при проведении распределенного имитационного эксперимента с применением мультиагентного подхода.

Благодарности

Статья частично финансирована из проекта **ITHEA XXI** Института Информационных теорий и Приложений FOI ITHEA и консорциума FOIBulgaria (www.ithea.org, www.foibg.com)

Библиографический список

- [Fujimoto, 2003] Fujimoto R.M. Distributed Simulation Systems // Proceedings of the 2003 Winter Simulation Conference S. Chick, P. J. Sanchez, D. Ferrin, and D. J. Morrice, eds. Pp. 124-34.
- [Wilson,1998] Wilson L.F. and Shen W. Experiments In Load Migration And Dynamic Load Balancing In Speedes // Proceedings of the 1998 Winter Simulation Conference. D.J. Medeiros, E.F. Watson, J.S. Carson and M.S. Manivannan, eds. Washington. 1998. Pp.483-490.
- [Zheng, 2005] Zheng G. Achieving High Performance on Extremely Large Parallel Machines: Performance Prediction and Load Balancing // Ph.D. Thesis, Department of Computer Science. University of Illinois at Urbana-Champaign, 2005,165 p. [<http://charm.cs.uiuc.edu/>].
- [Миков, 2007] Миков А.И., Замятина Е.Б., Осмехин К.А. Динамическое распределение объектов имитационной модели, основанное на знаниях // Proceedings of XIII International Conference "Knowledge-Dialogue-Solution" (KDS), ITHEA, Sofia, 2007. Vol.2., Pp. 618-624.
- [Mikov, 1995] Mikov A.I. Simulation and Design of Hardware and Software with Triad //Proc.2nd Intl.Conf. on Electronic Hardware Description Languages. Las Vegas, USA, 1995. Pp. 15-20.
- [Миков, 2008] Миков А.И., Замятина Е.Б. Технология имитационного моделирования больших систем // Труды Всероссийской научной конференции «Научный сервис в сети Интернет» – М.: Изд-во МГУ, 2008. С.199-204.
- [Миков, 2009] Миков А.И., Замятина Е.Б., Козлов А.А. Оптимизация параллельных вычислений с применением мультиагентной балансировки // Труды международной научной конференции «Параллельные Вычислительные Технологии». Нижний Новгород – Челябинск, Изд. ЮУрГУ, 2009. С. 599-604.

Сведения об авторах

Александр Миков – Кубанский государственный университет, профессор, заведующий кафедрой «Вычислительные технологии»; Россия, г. Краснодар, ул. Аксайская, 40/1-28; e-mail: alexander_mikov@mail.ru.

Елена Замятина – Пермский государственный университет, доцент кафедры математического обеспечения вычислительных систем; Россия, г. Пермь, 614017, ул. Тургенева, 33–40; e-mail: e_zamyatina@mail.ru.

Арсений Козлов – Пермский государственный университет, студент кафедры математического обеспечения вычислительных систем; Россия, г. Пермь, ул. Левченко, 6-87; e-mail: Arko87@yandex.ru.

ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ – ЭКОЛОГИЧЕСКИЙ ПОДХОД К ИССЛЕДОВАНИЯМ

Максим Луцкий, Николай Сидоров

Abstract: *The software ecology as a new branch of the software engineering is presented. Three directions of software ecology decisions (green software, sustainable development, ecosystems) are looked.*

Keywords: *software, software engineering, software ecosystem, reverse engineering.*

ACM Classification Keywords: *software, software engineering, distribution, maintence and enchainment .*

Введение

В статье рассматривается применение экологического подхода к исследованию программного обеспечения и формулируются основные положения нового раздела инженерии программного обеспечения – экологии программного обеспечения.

Можно указать три причины для применения экологических исследований к программному обеспечению. Первые две причины связаны с концепцией устойчивого развития и определяются влиянием программных продуктов и их производств на окружающую среду. Третья причина вызвана необходимостью наблюдений за программным обеспечением как за организованной системой в контексте реального мира.

Первая причина, основывается на предположении, что программные продукты и их производство в индустрии программного обеспечения, как в других инженерных отраслях влияют на окружающую среду и следовательно на биосферу [1, 2]. Это влияние может быть вредным и неуправляемым.

Вторая причина, основывается на парадоксе развития научно-технического прогресса, суть которого заключается в том, что приходится осуществлять деятельность по ликвидации результатов деятельности. При этом, в любой отрасли деятельности человека, и в индустрии программного обеспечения также, понятна и разумна постановка задачи – утилизировать как можно больше из того, что стало не нужно [3, 4].

Третья причина основывается на наблюдении, что эффективное планирование развития и обслуживания программных продуктов требует понимания их места в реальном мире, с учетом их взаимодействий с реальным миром, внутри продукта, и внутри элементов продукта [5].

Применение экологического подхода в программном обеспечении

Под экологическим подходом будем понимать методологию исследований, в основе которой лежит рассмотрение объекта, как части окружающей среды, как правило в форме экосистемы, с учетом устойчивого развития.

Распространение общих принципов и требований экологичности на процессы жизненного цикла программного обеспечения объясняется тем, что в настоящий момент программное обеспечение является продуктом инженерной технической деятельности [6]. Индустрия программного обеспечения потребляет значительные ресурсы, а программные продукты, функционируют во всех хозяйственных отраслях, поэтому прямо или косвенно индустрия и продукты оказывают значительное влияние на окружающую среду [3]. Кроме того, программное обеспечение потенциально опасных технических объектов, в случае его неправильного функционирования, может быть причиной прямых экологических катастроф.

В контексте концепции устойчивого развития [2, 7], программное обеспечение является его активом вследствие того, что оно продукт производственной деятельности человека. Известно, что производственная деятельность человечества все чаще вступает в противоречие с процессами, поддерживающими устойчивый круговорот в биосфере. В. Вернадский указывал на необходимость решения задачи перехода от стихийных взаимодействий человека и биосферы к сознательным, которые превращают биосферу в ноосферу и обеспечивают устойчивое развитие [8]. Сейчас, эта задача вновь актуальна в контексте программы совместных действий, в интересах устойчивого развития, и программное обеспечение играет важную роль в ее решении вследствие того, что оно является следующим:

- - основой всех информационных и коммуникационных технологий;
- - наукоемким образованием;
- - продуктом коллективной, все чаще глобальной деятельности.

В контексте экологического подхода программное обеспечение рассматривается, как технический объект, взаимодействующий с окружающей средой. Такой аспект рассмотрения относится к той части экологии, которая иногда обозначается «относящейся к окружающей среде» (environmental) [7]. Экологизация программного обеспечения фактически делает отрасль разработки программных продуктов объектом исследований инженерной экологии техносферы [3, 9]. Основные цели такой экологизации – это сохранение природных ресурсов и защита окружающей среды. Достижение целей требует постановки и решения соответствующих задач.

Применение экологических исследований в индустрии программного обеспечения показывает, что их распространение идет в трех основных направлениях:

- первое, связано с применением к программному обеспечению общих принципов и требований экологичности производства и использования технических объектов [10, 11];
- второе, связано с реализацией ресурсосберегающих и безотходных производств программного обеспечения [4, 12, 13];
- третье, связано с применением к программному обеспечению экологических исследований, рассматривая его как экосистему [14, 15].

Экологическое производство и использование программного обеспечения

Наиболее ярким проявлением первого направления является GreenSoftware – «Зеленое» программное обеспечение [10, 11]. В рамках этого направления основной акцент делается на использовании программного обеспечения, как средства прямого или косвенного уменьшения вредного влияния на окружающую среду. Например, использование электронной почты вместо бумажной, видеоконференций вместо поездок. Кроме того, рассматриваются влияния на окружающую среду при производстве и эксплуатации программного обеспечения, например, эффективное энергопотребление, уменьшение выделения CO и CO₂ в атмосферу и разогрева оборудования, других подобных аспектов взаимодействия между окружающей средой и вычислительными средствами на всех этапах жизненного цикла программного обеспечения. В рамках GreenSoftware важным также является минимизация использования основных (объем памяти, время и быстродействие процессора) и дополнительных (объем внешней памяти, распределение каналов) вычислительных ресурсов.

Наиболее яркими примерами такого подхода являются стратегия корпорации IBM по разработке GreenSoftware [10] и стратегия GreenAsus компании Asus [11].

При разработке программного обеспечения корпорация IBM предлагает следующие подходы, большую часть из которых она использует в качестве стратегии IBM - GreenSoftwareStrategy, при разработке собственного «зеленого» программного обеспечения:

- сокращение деловых поездок, используя on-line коммуникации;
- загрузка недостаточно загруженных серверов, для уменьшения потребления энергии и рабочей площади;
- планирование графика выполнения рабочей нагрузки во время умеренной нагрузки, для того, чтобы использовать энергию меньшей стоимости;
- эффективное управление ресурсами жизненных циклов;
- оптимизация разных приложений для уменьшения ресурсов и энергии, которые потребляются;
- объединение и укрупнение ресурсов для уменьшения необходимой площади и упрощения вычислительной инфраструктуры;
- оптимизация нагревания, вентиляции и кондиционирования воздуха, на рабочих местах для сокращения потребления энергии;
- эффективное управление хранением данных;
- сокращение потребления энергии при уменьшении рабочей нагрузки;
- оптимизация бизнес-процессов для сокращения потребления энергии и операционных затрат;
- сокращение использования бумажных носителей путем введения в бизнес-процессы электронных документов.

Стратегия GreenAsus была сформулирована в 2000 году и делится на четыре части. Первая, Green-Design, должна обеспечивать создание и внедрение в производство «зеленых» компонентов, которые состоят из легко восстанавливаемых материалов и эффективно утилизируются. Вторая, Green-Manufacturing обеспечивает разработку и внедрение экономичных и экологичных технологий на производстве. Например, компания полностью отказалась от использования свинца, а до 2009 года собиралась прекратить применение галогенов. Третья часть, Green-Procurement, регламентирует логистику товаров, использование материалов и компонентов от внешних производителей. Кроме этого, стимулируются поставщики на пересмотр своего отношения к проблемам экологии. Четвертая часть, Green-Marketing решает задачи утилизации старых компьютеров. Компания рассматривает последнюю часть важной и в 2010 году собирается начинать свою деятельность по переработке электроники в Украине.

Другие компании начинают решать задачи этого направления. Например, Intel также пересматривает свое отношение к проблемам экологии, разрабатывая новый стандарт, который минимизирует вредное влияние на окружающую среду на всех этапах производства.

Ресурсосберегающее и безотходное производство программного обеспечения

Ресурсосберегающее производство – это производство и реализация продуктов с минимальным расходом вещества и энергии на всех этапах производственного цикла и с наименьшим воздействием на человека и природные системы. Основой ресурсосберегающего производства являются ресурсосберегающие технологии [9].

Рассмотрим существующие ресурсы инженерии программного обеспечения (принципы, процессы, конструкции), которые отвечают экологическому подходу и могут использоваться как основа для

построения ресурсосберегающих и безотходных технологий разработки и сопровождения программного обеспечения (табл. 1).

Таблица 1. Методы, принципы, конструкции экологического программного обеспечения

| № п/п | Методы | Подпрограммное (процедурное) программирование | Объектно-ориентированное программирование | Модульное программирование |
|-------|------------------------------------|--|---|------------------------------------|
| 1 | Конструкция | Подпрограмма (закрытая, открытая) | Класс (C++, C #) | Модуль (Модуля 2,3), пакет (Ada) |
| 2 | Принципы реализации конструкций | Параметризация | Полиморфизм, наследование | Раздельная компиляция |
| 3 | Принципы использования конструкций | Процедурная абстракция абстракция выполнения, шаблонирование | Повторное и многократное использование (унификация) | Принципы использования конструкций |
| 4 | Фундаментальная абстракция | Абстрактный тип данных | | |

Первой такой конструкцией была подпрограмма, которую рассматривали как средство сокращения текста программы. Гибкость конструкции обеспечивала параметризация. Вскоре взгляд на подпрограмму начал изменяться, и её стали рассматривать, как средство структурной организации программ. Приспособление программы к ожидаемым изменениям постановки задачи или сопровождения сводилось к замене одного или нескольких тел подпрограмм или к изменению значений некоторых фактических параметров в обращении к подпрограммам. Поэтому, применялись подпрограммы как процедурные абстракции, абстракции выполнения (закрытые подпрограммы) или шаблоны (открытые подпрограммы). В качестве метода применения подпрограммы использовалось подпрограммное или процедурное программирование. Позднее, следуя стремлению улучшить конструкцию для реализации абстрактных типов данных были созданы модуль и класс.

В 1984 году, при развертывании исследований связанных с повторными использованными программного обеспечения обращалось внимание на то, что с годами количество новых применений вычислительных машин уменьшается [16, 17]. Это свидетельствовало о том, что сопровождаемое программное обеспечение, которое впоследствии стали называть наследуемым, содержало опыт, и его следовало повторно использовать при создании нового программного обеспечения. С учетом этого жизненный цикл программного обеспечения был расширен дополнительными процессами (рис. 3). С одной стороны, он был дополнен доменным анализом. Его цель, путем анализа опыта, накопленного в домене строить повторно используемые решения, для применения в разработке нового программного обеспечения [18]. С другой стороны, в жизненный цикл, в контексте фазы ликвидации были введены процессы, которые связаны с утилизацией программного обеспечения (рис.4, таб. 2) [13].

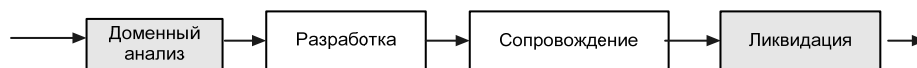


Рис. 3 Расширение жизненного цикла программного обеспечения

Утилизация включает три процесса – повторное использование, восстановление и переработку наследуемого программного обеспечения, а все неутраченные компоненты уничтожаются (рис. 4).



Рис. 4 Содержание фазы ликвидации

Таблица 2 Процессы утилизации

| Процессы утилизации | Общая характеристика процессов утилизации | Методы инженерии программного обеспечения | Средства инженерии программного обеспечения |
|-------------------------|--|--|--|
| Повторное использование | Создания компонентов, хранение и применение компонентов | Классификация, хранение, поиск | Среда программирования, мониторинговые системы, экспертные системы |
| Восстановление | Анализ программного обеспечения, преобразование программного обеспечения на одном уровне | Синтаксического семантического анализа прямой и обратной инженерии | Анализаторы преобразователи абстрактные экстракторы |
| Переработка | Преобразование программного обеспечения на разных уровнях представления. Анализ программного обеспечения | Синтаксического и семантического анализа, реверсивной инженерии | CASE, преобразователи, абстрактор, экстрактор. |

Реализация принципа повторного использования привела к появлению понятия реверсивной инженерии программного обеспечения, соединение которой с прямой инженерией дало понятие реинженерии и позволило в экологическом аспекте рассматривать разработку и сопровождение программного обеспечения, как его круговорот (рис. 5) [4] (подкова – «horseshoe» [5]).

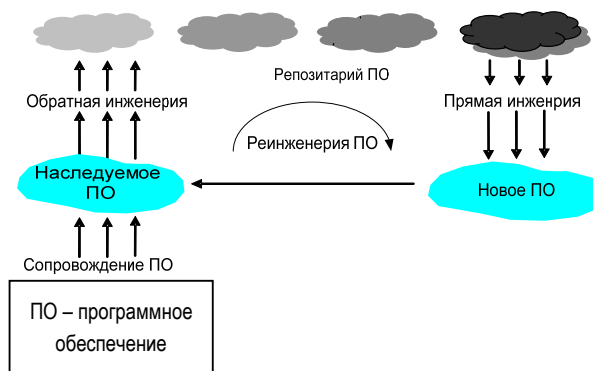


Рис. 5 «Круговорот» программного обеспечения

В доменном анализе и утилизации программного обеспечения сложилось несколько способов создания (табл. 3) и применения (рис. 6) [4] компонентов повторного использования.

Таблица 3. Создание компонентов программного обеспечения

| Характеристика | Способы создания компонентов программного обеспечения | | |
|------------------------------|---|------------------------------|---|
| | Утилитарный | Индуктивный | Дедуктивный |
| Характер способа | Экстенсивный | Интенсивный | Интенсивный |
| Метафора | «Свалка» | «Фабрика компонентов» | «Высокая технология» |
| Тип компонентов | Пассивные | Большие пассивные | Большие активные |
| Техника применения | Композиционная | Композиционная и адаптивная | Адаптивная и композиционная |
| Метод разработки компонентов | Утилизация | Доменный анализ, снизу вверх | Структурное проектирование, сверху вниз |
| Тип применения | Случайный | Случайный систематический | систематический |
| Средства переработки | Не нужны | Нужные | Не нужны |

| | | | |
|------------------------------------|---------------|--------------|---------------|
| и возобновления | | | |
| Мощность инфраструктуры применения | Высокая | Высокая | Низкая |
| Формы повторного использования | «Черный ящик» | «Белый ящик» | «Черный ящик» |

Индуктивный и дедуктивный способы создания компонентов впоследствии позволили превратить компоненты повторного использования в компоненты многократного использования, которые сейчас составляют основу компонентной разработки программного обеспечения.

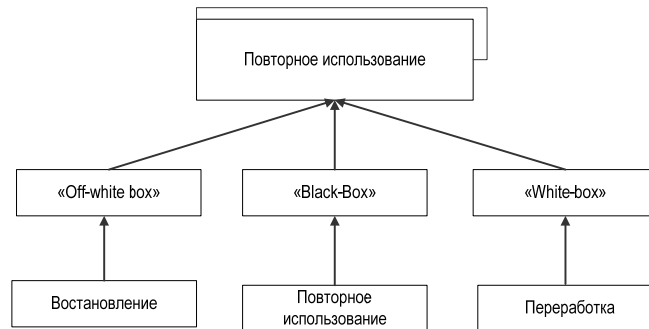


Рис. 6 Аспекты применения принципов повторного использования

Появились также две парадигмы применения компонентов: адаптивная и сборочная. В соответствии с адаптивной парадигмой – применяется модель программного обеспечения, похожая на фрейм, где компоненты заполняют слоты. В соответствии с сборочной парадигмой – компоненты объединяются программным кодом, как «клеем».

Таким образом, видно, что в инженерии программного обеспечения есть принципы, процессы и конструкции и принципы, которые могут использоваться при построении безотходных, ресурсосберегающих технологий и продуктов.

Программное обеспечение как экосистема

В работе [5] относительно программного обеспечения показано следующее:

- изменение (развитие) – непереносимое свойство программ, обусловленное наличием обратных связей и связанное с законами эволюции программ;
- метасистема, в рамках которой развивается программа включает продукты деятельности, процессы и организацию, содержит большое количество обратных связей, стабилизирующих внутренних механизмов, влияющих на процессы планирования, управления повышения их эффективности;
- эффективное планирование и обслуживание программы требует понимания ее места в метасистеме, а также взаимодействий как между элементами, так и внутри них.

При этом, программы, о которых идет речь, по классификации SPE являются E – программами, а метасистема, – это их внешняя среда – реальный мир [5] .

Рассматривая программные системы в контексте реального мира как экосистемы с учетом рассмотренных выше, особенностей к их исследованиям целесообразно применить экологический

подход. Тогда, одна из основных задач исследований будет состоять в сборе, накоплении, систематизации и анализе информации о количественном характере взаимосвязей внутри программной системы, между программной системой и внешней средой для получения следующих результатов (рис. 7):

- оценка качества исследуемых программных систем – как экосистем;
- влияние причин, изменений компонентов, источников и факторов воздействия;
- прогноз устойчивости программных систем как экосистем и допустимости изменений.

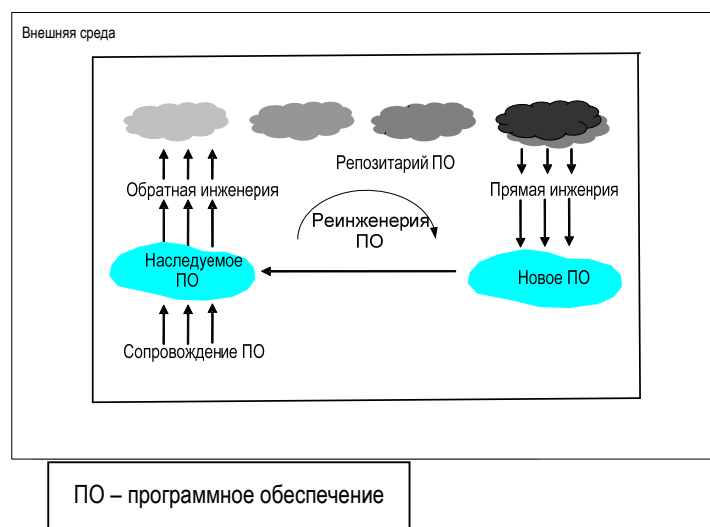


Рис. 7 Е – программа в реальном мире

Таким образом, для выполнения исследований должен быть организован мониторинг программной системы в контексте внешней среды, а исследование программной системы как экосистемы должно осуществляться на основе системного подхода, например, в следующих аспектах:

- морфологическом – исследуется устройство системы;
- функциональном – исследуются функции компонентов системы, обычно в терминах входных, выходных и управляющих параметров, возмущающих воздействий и параметров состояния.

Принимая во внимание тот факт, что исследуется наследуемая программная система, важную роль в экологическом исследовании будет играть реверсивная инженерия.

Заключение

Таким образом, принимая во внимание представление направленные экологических исследований программного обеспечения, в целом, видимо, можно говорить об экологии программного обеспечения как разделе инженерии программного обеспечения, хотя и понимая, что наиболее полно это касается «энвайроиментологии», как основы рационального использования ресурсов в целях устойчивого развития.

Практическим результатом экологических исследований было бы создание научно-обоснованных технологий создания и применения программного обеспечения, основанных на оптимальном использовании ресурсов, в том числе и природных, обеспечивающих их поддержку, восстановление и контролируемое устойчивое развитие.

Такой результат может быть получен только путем широкой пропаганды экологического подхода как в среде разработчиков, так и пользователей программного обеспечения.

Библиография

- [Баландин, Вернадский, 1999] Р.К. Баландин, В.Н. Вернадский: жизнь, мысль, бессмертие. М., «Знание», 1979.
- [Сидоров, 2006] М.О.Сидоров. Экология программного обеспечения. – Материалы Всеукраинской конференции аспирантов и студентов «Инженерия программного обеспечения» – К.: НАУ, 2006. – С. 41 - 48
- [Мазур, Молданов, 2006] И.И. Мазур, О.И. Молданов. Курс инженерной [экологии: - М.: Высш.шк., 1999.-447 с] |.
- [Сидоров, Шарепа, 1990] Н.А. Сидоров, А.Н. Шарепа. Средство для утилизации программного обеспечения // УсиМ.- 1990.- №5.- с. 50-54.
- [Lehman, Belady,1985] М.М. Lehman, L.A. Belady. Program Evolution.- Academic Press.- 1985.- 532 p.
- [Sommerville, 2001] I. Sommerville. Software Engineering.- Addition- Wesley.- 2001.- 625 p.
- [Угольницкий, 2002] Г.Л. Угольницкий. Иерархические управления устойчивым развитием социальных организаций. – Общественная наука и современность. - №3. – 2002. – с. 133 – 140.
- [Вернадский, 1994] В.Н. Вернадский. Несколько слов о неосфере. – Успехи современной биологии. - №18. – вып. – 2. –с. 113 – 120.
- [2000] Словарь терминов по экологии.- 2000 г.
- [IBM Software, 2008] IBM Software: A green strategy for your entire organization. IBM Software for a greener world June. 2008. NY 10589. U.S.A. Produced in the United| States| of| America|. May| 2008.
- [Турнов, 2008] Н. Турнов. Зеленый свет| для ASUS| // PC| WEEK/UE.- #24|(93).- 2008.
- [Сидоров, 1989] Н.А. Сидоров. Повторное использование| программного| обеспечения// Кибернетика. – 1989. - №3 – с.46-51
- [Сидоров, 1994] Н.А. Сидоров. Утилизация программного обеспечения экономический аспект // Кибернетика и системный анализ.- 1994.- №3.- с. 151-166.
- [Messershmitt] D.G. Messershmitt , C. Szyperski. Software Ecosystems: Understanding an Indispensable Technology and Industry. – MIT press. – 2003. – 233 p.
- [Lungu] M.F. Lungu. Reverse Engineering Software Ecosystems. – Doct. Diss. – USI.- 2009. – 208 p.
- [Boehm, 1987] B.W. Boehm.| Improving| Software| Productivity| // Computer|, v. 20, n. 9, 1987, г. 43-57
- [Biggerstaff, Foreword, 1984] T. Biggerstaff|. Foreword| //IEEE Trans|, on| Software| Engeneer|. v. 10, n. 5, - в 1984 г. 474-476
- [Prieto-Diaz], 1987] R. Prieto-Diaz| Domian|. Analysis| For| Reusability| // Compsac| 87 Oct|., 7-9.-1987.- г. 23-30

Информацияобавторах



Maksim Lyzkiy – *the 1-st deputy of rector, National aviation university, Cand. Of Sc. (techn.).*

Major Field of Scientific Research: Software engineering, information technology, education.



Nikolay Sidorov – *head of software engineering department, National aviation university, Doct. of sc. (techn.).*

Major Fields of Scientific Research: Software engineering, software education.

КОНЦЕПЦИЯ ПОСТРОЕНИЯ И МЕТОДЫ РЕАЛИЗАЦИИ АВТОМАТИЗИРОВАННОЙ СИСТЕМЫ ОПЕРАТИВНО-ДИСПЕТЧЕРСКОГО УПРАВЛЕНИЯ ГАЗОТРАНСПОРТНОЙ СИСТЕМОЙ УКРАИНЫ

Юрий Пономарев, Богдан Клюк, Виктор Борисенко

Abstract: в работе определены основные концептуальные принципы построения, методы разработки и средства реализации автоматизированной системы оперативно-диспетчерского управления (АСОДУ) газотранспортной системой (ГТС) Украины. В качестве современного программно-технического ядра (каркаса) АСОДУ ГТС был обосновано выбран новый класс систем оперативного управления – MES (). Важной особенностью систем класса MES является наличие единого реляционного хранилища оперативной информации, собираемой множеством систем нижнего уровня типа SCADA и передаваемой по современным корпоративным каналам передачи данных.

ACM Classification Keywords: J/1 Administrative data processing

Введение

Разработка и внедрение корпоративных систем управления бизнес-процессами ориентировано на высокий уровень информационной поддержки принятия управленческих решений, что позволяет обеспечить высокие уровни управляемости, прибыльности и, в конечном итоге, капитализации компании. Перспективным направлением повышения качества и эффективности работы таких систем является использование при их построении современных интеллектуальных компонентов и технологий.

1. Интеллектуальное структурно-инвариантное объектно-ориентированное ядро интегрированного распределенного банка данных

В качестве базовой интеллектуальной информационной компоненты корпоративной системы управления бизнес-процессами было разработано и внедрено в государственной газотранспортной компании "Укртрансгаз" интеллектуальное структурно-инвариантное объектно-ориентированное Ядро (ИЯ) интегрированного распределенного банка данных (ИРБД). Работа проводилась в отраслевом институте транспорта газа (г. Харьков, Украина).

Одной из наиболее важных технических характеристик ИЯ ИРБД, существенно повышающих эффективность сопровождения и развития системы, является его *унифицированная стандартизованная саморасширяемая структура*, которая позволяет включать (изменять) метаинформацию о новых сущностях (объектах предметной области) и их атрибутах без добавления новых таблиц, столбцов и отношений между ними. Способность к структурно-инвариантному информационному саморасширению является весьма важным интеллектуальным механизмом, характерным, например, для естественных биологических систем, обладающих достаточно развитым мозгом. Для реализации данной возможности ИЯ ИРБД содержит в едином структурированном виде, как метаданные, так и фактографическую информацию, разделяемую несколькими распределенными приложениями.

Область метаданных является реализацией формализованного представления в реляционной базе данных структурно-семантических знаний о предметной области и интеллектуальной корпоративной системе управления бизнес-процессами. Основу метаданных составляют описание классов объектов, их типов (категорий), типовых (базовых) отношений между классами, различных иерархий функционально-семантического группирования классов, формируемых разработчиками и расширяемых системными

администраторами по мере развития ИРБД. К метаданным относится и метасистемная служебная информация о конечных пользователях системы, их привилегиях и правах доступа к данным ИРБД, а также информация о функциональных подсистемах корпоративной системы управления бизнес-процессами, реализующих их приложениях и разработчиках. Метаданные вносятся в ИРБД на этапе разработки или развития системы.

Фактографические данные описывают конкретные статические или динамические свойства элементов объекта автоматизации либо являются нормативно-справочной информацией, необходимой для единообразного представления нормативных характеристик объекта. Фактографические данные вносятся в ядро ИРБД как на этапах разработки, внедрения, так и во время эксплуатации системы.

Понятия классов объектов и их свойств лежат в основе предлагаемого подхода, т.к. на их основе строится вся система знаний об объектах автоматизации. Класс объекта является базовым унифицированным *информационным семантическим шаблоном*, в котором представлен набор общих свойств, а также данные о типовой внутренней структуре конкретных сущностей (экземпляров, объектов), относящихся к данному классу. Метаинформация структурно-категорного типа, содержащаяся в классе, определяет общие особенности подмножества его экземпляров и правила построения отношений между ними, а также связей между экземплярами одного либо разных классов.

Для каждого класса определен расширяемый набор свойств и, возможно, типовых состояний объектов класса, а также множество допустимых семантических отношений объектов данного класса с другими объектами.

Свойства классов по видам делятся на *нормативные, статические и динамические*. *Нормативные* свойства определяются для промышленных изделий и представляют собой номинальные характеристики всех экземпляров объектов данного типа (марок). Значения нормативных свойств класса задаются для каждой марки промышленного изделия, относящейся к этому же классу. *Статические* свойства представляют неизменяемые во времени или редко изменяемые характеристики экземпляра объекта, такие, например, как километр расположения крана на трубопроводе. Значения статических свойств задаются для каждого экземпляра объекта отдельно. *Динамические* свойства отражают режимные характеристики экземпляра объекта, изменяющиеся во времени, например выходной ток катодной установки. Значения динамических свойств задаются для каждого экземпляра объекта и имеют связь с описанием его состояния, в результате перехода объекта в которое они были зарегистрированы. Каждое состояние характеризуется информацией о дате и времени перехода объекта в это состояние и выхода из него. Состояния объектов могут быть объединены причинно-следственными связями и сгруппированы для отражения детализации сложных состояний. Одной из важнейших характеристик свойства является тип данных, который определяется в унифицированном универсальном формате, не привязанном к особенностям конкретной СУБД.

Среди классов выделено несколько семантических категорий (понятия, технологические объекты, организационные объекты и др.). Категория (тип), к которому принадлежит класс, определяет допустимое подмножество видов свойств, которыми он может характеризоваться.

Разработчик может в виде дерева определить иерархическое группирование классов объектов, которые будут использованы в его приложениях, по какому-либо признаку (например, функциональному). Вершины дерева представляют собой семантические группы классов и классы объектов (конечные вершины), а ветви - отношения вида "тип - подтип". Корневой вершине такого дерева присваивается наименование всей иерархической семантической группы. Иерархическое семантическое дерево классов в интеллектуальной корпоративной системе управления бизнес-процессами используется, например, для

упрощения задания сложных составных условий фильтрации или организации сложного концептуального поиска информации.

Кроме того, для классов допустимы *отношения наследования* (для объектов в данной реализации они не предусмотрены). Отношение наследования устанавливается в том случае, если возникает необходимость уже существующий класс разделить на два или более классов в связи появлением новых свойств, которые присущи только некоторому подмножеству объектов данного класса. В этом случае уже существующий класс (родительский) автоматически становится абстрактным с сохранением у него общих для всех объектов свойств, а новыми свойствами, присущими только части объектов класса, наделяется новый наследованный подкласс.

2. Интеллектуальная адаптивная технология синхронизации данных

Все известные способы и системы синхронизации (репликации) данных между распределенными серверами баз данных, встроенные в современные промышленные СУБД, имеют ряд существенных недостатков, основными среди которых являются следующие:

- неудобная, сложная, негибкая настройка механизмов распределенной репликации данных;
- ненадежная работа встроенных систем репликации при наличии сегментов корпоративной сети с невысокой пропускной способностью (в газотранспортной системе – это удаленные пункты сбора данных с устройств линейной телемеханики, газораспределительные станции и т.п.);
- неэффективная работа встроенных механизмов репликации со сложными, гибкими распределенными базами данных, имеющими в составе развитые концептуально-семантические метаданные (к ним и относится ИЯ ИРБД. [Борисенко, 2004]).

В целях преодоления вышеуказанных недостатков в отраслевом институте транспорта газа была разработана, апробирована и внедрена интеллектуальная технология репликации (синхронизации) данных (ИТРД).

Разработанная технология имеет адаптивный характер и базируется на процедурном методе представления знаний, необходимых для организации эффективного, надежного и качественного обмена данными между удаленно расположенными информационными серверными узлами ИРБД..

Предлагаемая технология базируется на следующих основных принципах реализации:

- изменения данных должны распространяться по древовидной структуре сначала на подчиненные информационные узлы (серверы), а потом на вышестоящий узел. Информация обо всех узлах ИРБД и о подключении их друг к другу, должна храниться в специальной таблице ядра ИРБД на каждом узле;
- изменения данных ИРБД, выполненные на конкретном узле должны накапливаться в промежуточном буфере обмена (при помощи триггеров регистрации), после чего передаваться на другие узлы, подключенные к данному серверу;
- промежуток времени, за который выполняется попытка передачи изменений на смежный сервер, должен задаваться администратором ИРБД в настройках информации. При этом для каждого смежного сервера этот промежуток может быть задан отдельно;
- при записи изменений текущего сервера в буфер обмена необходимо помечать данные, относящиеся к одному непрерывному логическому блоку информации маркерами начала и конца;

- после того, как изменения текущего сервера будут переданы на все смежные серверы, данные из буфера обмена должны быть удалены и сохранены в архиве. Время хранения данных в архиве должно задаваться адаптивными параметрами настройки;

- передача данных на удаленный сервер должна производиться также в промежуточный буфер обмена. Каждая удачно переданная строка буфера обмена должна подтверждаться операцией commit;

- данные буфера, переданные с других серверов, могут быть развернуты только в том случае, если были выявлены непрерывные логические блоки (найлены маркеры начала и конца блоков);

- должен быть реализован контроль работы системы синхронизации данных путем введения обратной семантико-событийной связи по выбранному каналу (удаленный режим контроля).

- должна быть реализована возможность удаленного управления процессом репликации по выбранному каналу (удаленный командный режим);

- должна существовать возможность семантической пометки блока изменений как требующего подтверждения при успешной передаче с указанием почтового ящика, на который потом должны прийти письма-подтверждения со всех серверов. Письма должен высылать почтовый робот удаленного управления серверами ИАСУ.

- должна существовать гибкая система разграничения прав пользователей на внесение изменений в различные объекты ИРБД; Каждый пользователь должен иметь права изменять свое подмножество данных, определенное системным интегратором или разработчиком;

- права на изменение, добавление и удаление данных должны быть двух видов - права с полной ответственностью и права с ограниченной ответственностью. Права с полной ответственностью присваиваются пользователям на непересекающиеся подмножества данных (интеллектуальная процедура раздачи прав должна это строго контролировать). Права с ограниченной ответственностью могут быть даны на пересекающиеся подмножества;

- должна быть обеспечена возможность запрета репликации данных об объектах заданных классов;

- должна существовать интеллектуальная процедура передачи изменений структуры объектов ИРБД только заданного типа в рамках общих потоков данных.

Предлагаемая интеллектуальная адаптивная система синхронизации данных реализует четыре основных режима функционирования:

— *автоматический широковещательный режим* (изменения затрагивают все сервера ИРБД и передаются от сервера к серверу в реальном масштабе времени автоматически);

— *автоматический режим с адресными метками* (изменения пишутся в историю изменений с указанием адресов серверов, на которых они должны быть развернуты);

— *принудительный широковещательный режим* (изменения передаются всем серверам автоматически, но данные разворачиваются по команде интегратора);

— *принудительный режим с адресными метками* (изменения разворачиваются по команде интегратора на заданных серверах).

Основные интеллектуальные программные средства поддержки синхронизации данных ИРБД реализованы в виде настраиваемого параметризованного пакета хранимых процедур СУБД Oracle и удаленной службы (интеллектуального агента) обслуживания узлов ИРБД. Интеллектуальный агент реализует все описанные выше алгоритмы синхронизации данных, а также выполняет ряд

дополнительных системных функций, требующих периодического вызова и самовосстановления после сбоев при возникновении нештатных ситуаций, таких как: перезагрузка сервера, сбой питания, обрыв канала передачи данных, критическая перезагрузка ресурсов сервера, временная нехватка оперативной памяти и т.п..

В качестве дополнительного модуля была разработана служебная программа, которая предназначена для удаленного выполнения SQL-запросов к информационным серверным узлам, а также для получения и просмотра результатов их выполнения в асинхронном режиме. Она может быть также использована для удаленного обновления и перекомпиляции хранимых процедур, триггеров, пакетов и других объектов ORACLE одновременно на нескольких узлах. Кроме того, был разработан web-базируемый программный модуль для администратора узлов ИРБД, который выполняет следующие основные функции:

- отображение в удобном для пользователя виде значений контролируемых параметров на основе информации последнего опроса;
- сигнализация о наличии аварийных значений параметров при помощи цветового выделения соответствующего аварийного узла и параметра;
- изменение значений управляющего набора параметров и отсылку соответствующего SQL-запроса;
- отображение статуса параметра (дата последнего опроса, наличие и характер ошибки опроса);
- добавление новых контролируемых и управляющих параметров.

Выводы

В работе было представлено описание двух современных интеллектуальных технологий, используемых для построения базовых информационно-процедурных компонент корпоративной системы управления бизнес-процессами.

Опыт разработки и внедрения рассмотренных интеллектуальных технологий показывает, что данное направление развития современных систем управления бизнес-процессами является актуальным, практически целесообразным и весьма перспективным подходом к решению более общей проблемы существенного повышения уровня капитализации, эффективности и рентабельности современных крупномасштабных промышленных предприятий.

Библиография

[Борисенко, 2004] В. Борисенко, Т. Борисенко. Процедурно-семантическая модель представления знаний и ее реализация в виде адаптивного объектно-ориентированного ядра корпоративной системы управления газотранспортной системой. - В кн. "Трубопроводные системы энергетики: Управление развитием и функционированием". - Новосибирск: Наука, 2004, с. 312-321.

Информация об авторах

Юрий Пономарев – заместитель директора по научно-исследовательским работам, Институт транспорта газа, маршала Конева, 16, г. Харьков, 61004, Украина;
e-mail: ponomarev@asugaz.kharkov.com

Виктор Борисенко – ученый секретарь Совета института, начальник отдела, Институт транспорта газа, маршала Конева, 16, г. Харьков, 61004, Украина; e-mail: vborisenko.nipi@naftogaz.net

Татьяна Борисенко – ведущий инженер-программист, Институт транспорта газа, маршала Конева, 16, г. Харьков, 61004, Украина; e-mail: tborisenko.nipi@naftogaz.net

МОДЕЛИРОВАНИЕ И АНАЛИЗ ФАКТОВ И СВЯЗЕЙ МЕЖДУ НИМИ

Павел Мальцев

Аннотация: В данной работе предложен подход к моделированию и анализу фактов, записи о которых имеются в хранилище данных, и связей между ними. Описаны основные положения формальной математической теории, лежащей в основе предлагаемого подхода. Программные компоненты, реализующие предложенный подход, будут включены в программный комплекс ViP, предназначенный для многомерного анализа данных, получаемых из гетерогенных источников, и позволяет упростить разработку приложений Business Intelligence. Кроме того, планируется использовать описанный подход при разработке исследовательского портала "Инновационное развитие регионов".

Keywords: исчисление фактов, Business Intelligence, BI, бизнес-анализ, Data Mining, Reporting, системы поддержки принятия решений, DSS, информационно-аналитические системы.

ACM Classification Keywords: H.4 Information Systems Applications: H.4.2 Types of Systems – Decision support (e.g., MIS).

Введение

В настоящее время ведётся разработка исследовательского портала, основной целью создания которого является интеграция информации об инновационной активности регионов, создание и апробация различных моделей инновационного развития, организация коллективной работы исследователей.

Проект создания портала "Инновационное развитие регионов" ставит задачу разработки информационно-аналитической системы, реализующей сбор, хранение, представление и анализ данных об инновационной активности регионов. При разработке портала особенно важно иметь инструмент, позволяющий извлечь максимум новых знаний из всего богатства данных, накопленных в хранилище портала. В настоящее время при решении подобных задач широко применяются методы интеллектуального анализа данных.

Пользователь портала, решая поставленные задачи, может задать вопрос: "Как связаны между собой факты, записи о которых имеются в хранилище данных?". Современные методы интеллектуального анализа данных позволяют определить лишь характер ранее выявленной связи, таким образом, само наличие некой связи между событиями должно быть установлено заранее. Например: при поиске ассоциаций, следует явно указать, между какими событиями требуется искать ассоциации. Гораздо полезнее и удобнее для пользователя иметь средство, которое позволяло бы *обнаруживать новые связи* или *проверять гипотезы исследователя о наличии связей между фактами*.

В работе представлен подход к моделированию и анализу фактов, записи о которых имеются в хранилище данных, и связей между ними. Этот подход планируется реализовать в программном комплексе ViP. Данный программный комплекс используется в качестве платформы при разработке портала «Инновационное развитие регионов» (информацию о комплексе ViP можно получить в [1]). В статье детально описаны основы формальной математической теории лежащей в основе предлагаемого подхода.

Общие сведения о предлагаемом подходе

Часто мы сталкиваемся с задачей исследования динамики развития той или иной системы. Пусть мы можем фиксировать отдельные статические состояния исследуемой нами системы. Чем характеризуются данные статические состояния? Очевидно, что они характеризуются какими-либо постоянными

значениями параметров исследуемой системы. Что же заставляет систему менять своё состояние? Очевидно, что система меняет своё состояние вследствие возникновения неких событий (внутренних или внешних по отношению системе). Следует понимать, что возникновение того или иного события определяет дальнейшее развитие всей системы.

Знания о том, как те или иные события влияют на развитие системы могут помочь в решении задачи прогнозирования дальнейшего развития системы. Кроме того, обладая данными знаниями можно прогнозировать возникновение событий в будущем.

В каждый момент времени в любой системе может происходить огромное количество событий, но не все события «полезны» для анализа. Поэтому введём понятие факта как существенного с точки зрения исследователя события. *Фактом* называется *значимое для анализа событие*.

В современных информационных системах накоплено великое множество записей о свершившихся событиях – фактах. Особый интерес представляют не столько сами факты, сколько связи между ними. Но, к сожалению, информация о связях между фактами в явном виде в информационных системах содержится редко. Умение проводить качественный анализ связей между фактами позволит извлечь новые, скрытые знания о причинно-следственных связях, что позволит решать задачу прогнозирования на более высоком уровне.

Выделим три типа связей между фактами: структурные, семантические, неявные. Рассмотрим характерные черты каждого типа связей:

1. *Структурные связи*. Информация о структурных связях заложена в структуре самой базы данных (БД), т.е. они явно выделены. Приведём пример явной связи. Пусть в БД содержатся данные о величине государственного финансирования тех или иных исследований и величине привлеченных инвестиций на выполнение данных исследований. Связь между фактами изменения величины государственного финансирования и размеров частных инвестиций является примером явной связи, т.к. заложена в структуру самой БД. Структурные связи не вызывают затруднения при анализе, т.к. они всегда очевидны. Более того, информационные системы всегда разрабатываются с учётом этих связей. Очевидность данных связей делает их менее ценными для решения поставленной задачи.
2. *Семантические связи*. В отличие от структурных эти связи менее явны, т.к. заложены в сами данные, в их семантику. Например: связь между ростом размера государственного финансирования исследований и ростом средней заработной платы исследователей явно в базе данных не выделена, но данная связь известна, очевидна и определяется самими данными.
3. *Неявные связи*. Знания о неявных связях являются самыми ценными для исследователей. Отличительной чертой неявных связей является то, что они неизвестны и ни в каком виде не выделены в данных, содержащихся в БД.

Поиск неявных связей и является целью предлагаемого подхода. Суть описываемого подхода – в *автоматическом построении математической модели фактов отдельной предметной области*, на основе накопленной статистики. Важную роль здесь играют так называемые *статистические шаблоны*. Данные шаблоны позволяют идентифицировать основные структуры, такие как следствия, обобщения, совокупности и т.д., то есть *структуры модели фактов*. Важной особенностью является то, что база статистических шаблонов *может редактироваться*.

Исследователь здесь обладает широким полем для эксперимента. Для описания моделей фактов предлагается формальная математическая теория – «*исчисление фактов*». По своей сути модель фактов представляет *онтологию*. Данная модель может быть впоследствии дополнена или отредактирована пользователем (экспертом), а внесённые пользователем изменения могут быть проверены на имеющихся

данных. Таким образом, описанный подход позволяет не только *автоматически строить онтологию фактов*, но и *осуществить поддержку эксперта в проверке его гипотез*.

Описание подхода следует начать с определения понятия «исчисление фактов», которое служит основой представленного подхода.

Исчисление фактов

Каждый факт характеризуется набором атрибутов, комбинация значений которых однозначно идентифицирует факт среди остальных. Для разных фактов наборы их атрибутов могут отличаться, но каждый факт обязательно обладает атрибутами пространства и времени.

Атрибут пространства определяет точку (область) пространства в которой произошло событие (факт).

Атрибут времени определяет момент времени, в который произошло событие (факт).

Основой для проводимого анализа являются данные из хранилища данных. Как известно, сущностным свойством хранилищ данных является *поддержка хронологии*. Таким образом, исходные данные уже содержат атрибут времени, чего нельзя сказать об атрибуте пространства. Если данные не содержат атрибута пространства, то будем считать, что все факты происходят в одной точке пространства.

Будем называть *структурой факта* набор атрибутов данного факта. Для каждого атрибута должно быть определено множество допустимых значений. Данное множество и будет характеризовать атрибут. Будем обозначать атрибуты строчными буквами латинского алфавита, при необходимости будем использовать индексы.

Атрибут времени будем обозначать t , а атрибут пространства – g .

Множество допустимых значений атрибута x будем обозначать $D(x)$. Если $D(x)$ конечно, то будем говорить, что x имеет *категориальный тип*. Будем считать, что атрибуты x_1 и x_2 одинаковы, если $D(x_1) = D(x_2)$, то есть множество допустимых значений должно отражать смысл атрибута, и здесь в первую очередь важен состав множества, а не его мощность.

Таким образом, факт представляется совокупностью атрибутов с зафиксированными значениями.

Факты будем обозначать строчными буквами греческого алфавита: $\alpha, \beta, \gamma, \dots$. Для обозначения структуры факта, т.е. набора его атрибутов, будем использовать следующее обозначение:

$$\alpha = (t, g, x_1, x_2, \dots, x_n),$$

где $t, g, x_1, x_2, \dots, x_n$ – атрибуты факта α . Таким образом, *фактом* будем называть совокупность n атрибутов. Будем полагать, что n – число из множества натуральных чисел.

Будем говорить, что наборы атрибутов (x_1, x_2, \dots, x_n) и (y_1, y_2, \dots, y_m) подобны, если выполняются следующие условия:

- 1) $n = m$;
- 2) $(\forall x_i \mid i \in \overline{1, n})(\exists j \in \overline{1, m}) : D(x_i) = D(y_j)$.

Введём следующие обозначения:

- 1) G – множество всех фактов;
- 2) $\alpha \text{ like } \beta$ (будем читать, как “ α подобен β ”) – *отношение подобия фактов*: будем говорить, что два факта подобны, если их наборы атрибутов подобны.

Следует заметить, что отношение подобия обладают свойствами транзитивности и коммутативности.

Далее приведём основное свойство структуры факта:

$$(\forall \alpha = (x_1, x_2, \dots, x_n) \mid \alpha \in G, n \in N)(\forall i, j \in \overline{1, n} \mid i \neq j) : D(x_i) \neq D(x_j).$$

Свойство транзитивности отношения подобия. Если $\alpha \text{ like } \beta$ и $\beta \text{ like } \gamma$, то $\alpha \text{ like } \gamma$.

Доказательство:

Пусть: $\alpha = (x_1, x_2, \dots, x_n)$, $\beta = (y_1, y_2, \dots, y_m)$, $\gamma = (z_1, z_2, \dots, z_k)$.

$\alpha \text{ like } \beta \Rightarrow n = m$, $\beta \text{ like } \gamma \Rightarrow m = k$, Таким образом $n = k$ (*).

$$\alpha \text{ like } \beta \Rightarrow (\forall x_i \mid i \in \overline{1, n})(\exists j \in \overline{1, m}) : D(x_i) = D(y_j) (**)$$

$$\beta \text{ like } \gamma \Rightarrow (\forall y_j \mid j \in \overline{1, m})(\exists p \in \overline{1, k}) : D(y_j) = D(z_p) (***)$$

$$\Rightarrow (\forall x_i \mid i \in \overline{1, n})(\exists p \in \overline{1, k}) : D(x_i) = D(z_p) \Rightarrow \alpha \text{ like } \gamma \blacksquare$$

Свойство коммутативности отношения подобия. Если $\alpha \text{ like } \beta$, то $\beta \text{ like } \alpha$.

Доказательство:

Пусть $\alpha = (x_1, x_2, \dots, x_n)$, $\beta = (y_1, y_2, \dots, y_m)$.

Предположим обратное:

$$\begin{aligned} \blacksquare \quad \overline{\beta \text{ like } \alpha} &\Rightarrow (\exists y_j \mid j \in \overline{1, m})(\forall i \in \overline{1, n}) : D(x_i) \neq D(y_j) \xrightarrow{\alpha \text{ like } \beta} \\ &\xrightarrow{\alpha \text{ like } \beta} (\exists i, j \in \overline{1, n})(\exists y_p \mid p \in \overline{1, m}) : D(x_i) = D(y_p) = D(x_j) \Rightarrow (\exists i, j \in \overline{1, n}) : D(x_i) = D(x_j) \end{aligned}$$

Но это противоречит основному свойству структуры факта. Таким образом, мы пришли к противоречию, а значит, наше предположение было неверным. Тем самым мы доказали свойство коммутативности отношения подобия. ■

Введём понятие класса фактов.

Множество фактов A будем называть *классом фактов*, если для него верно

$$(\forall \alpha, \beta \in A \mid \alpha \neq \beta)(\forall \gamma \in G \setminus A) : (\alpha \text{ like } \beta) \& \overline{(\alpha \text{ like } \gamma)}.$$

Таким образом, *класс фактов* – это множество подобных фактов, такое, что любой факт, не принадлежащий классу, не подобен ни одному факту из данного класса.

Классы фактов будем обозначать заглавными буквами греческого алфавита. Следует заметить, что каждому кубу в хранилище данных соответствует определённый класс фактов [2].

К сожалению, мы не всегда можем точно сказать, свершился ли факт к некоторому моменту времени t , тем более, если данный момент ещё не настал. Чаще всего, можно лишь *оценить вероятность* того, что тот или иной факт свершился (свершится) к установленному моменту. Вероятность свершения факта α к некоторому моменту времени t будем обозначать следующим образом: $P_t(\alpha)$, $\alpha \in G$.

Предметом наших исследований являются не отдельные факты, а *причинно-следственные связи между ними*, поэтому нам важно будет знать, какова вероятность свершения некоторого факта β при условии

свершения факта α . Такую *условную вероятность* свершения факта будем обозначать следующим образом: $P_\alpha(\beta)$, $\alpha, \beta \in G$.

Теперь можно перейти к рассмотрению *отношения следствия*. В основе предлагаемого подхода лежит предположение о том, что все факты связаны причинно-следственными связями. Другими словами: у любого факта α_1 есть причина – факт α_0 . Положим также, что существует один и только один факт, не имеющий причин, будем обозначать его o . Отношение следствия будем обозначать следующим образом: $\alpha \rightarrow \beta$, где α – причина, а β – следствие.

Введём следующие аксиомы отношения следствия:

$$\begin{aligned} 1^\circ & (\forall \alpha \in G \mid \alpha \neq o) : o \rightarrow \alpha \\ 2^\circ & \alpha \rightarrow \beta \Rightarrow \overline{\beta \rightarrow \alpha} \quad \alpha, \beta \in G \\ 3^\circ & \alpha \rightarrow \beta, \beta \rightarrow \gamma \quad \alpha, \beta, \gamma \in G \\ 4^\circ & (t', x_1, \dots, x_n) \rightarrow (t'', y_1, \dots, y_m) \Rightarrow t' \leq t'' \end{aligned}$$

Следует сделать замечание относительно последней аксиомы. Мы не будем описывать множество допустимых значений атрибута времени и, тем более, не будем определять отношение “меньше либо равно” на данном множестве. Важно здесь понимать, что факт-следствие всегда свершается не ранее факта-причины.

Исследователь при проведении анализа не самой сложной системы может столкнуться с огромным количеством фактов. Анализ связей между отдельными фактами не поможет извлечь из данных каких-либо ценных знаний. Исследователю важно иметь *механизмы обобщения и группировки фактов*.

Группировкой фактов будем называть способ объединения фактов под символом одного факта.

Первым примером группировки является *совокупность* фактов. Представим ситуацию, когда мы можем представить факт в виде группы более простых фактов. Например: факт α подразумевает наступление фактов: $\alpha_1, \alpha_2, \dots, \alpha_n$. Будем говорить, что α является совокупностью фактов $\alpha_1, \alpha_2, \dots, \alpha_n$.

Совокупность будем обозначать как *произведение*:

$$\alpha = \alpha_1 \times \alpha_2 \times \dots \times \alpha_n \text{ или } \alpha = \prod_{i=1}^n \alpha_i.$$

Факты, представимые в виде совокупности, будем называть *составными*, а факты, которые нельзя представить в виде совокупности, будем называть *атомарными* фактами.

Теперь представим себе ситуацию, когда под знаком некоторого факта β будем понимать факт свершения хотябы одного факта из некоторой группы $\beta_1, \beta_2, \dots, \beta_n$.

Такую группировку будем называть *суммой фактов* и обозначать следующим образом:

$$\beta = \beta_1 + \beta_2 + \dots + \beta_n \text{ или } \beta = \sum_{i=1}^n \beta_i.$$

Группу фактов будем называть *независимыми* друг от друга, если любая пара фактов из данной группы не связана отношением следствия.

Для *независимой группы* фактов $\alpha_1, \alpha_2, \dots, \alpha_n$ справедливы следующие свойства:

$$1^\circ P\left(\prod_{i=0}^n \alpha_i\right) = \prod_{i=0}^n P(\alpha_i)$$

$$2^\circ P\left(\sum_{i=0}^n \alpha_i\right) = \sum_{i=0}^n P(\alpha_i)$$

Будем называть *обобщением* способ объединения некоторой группы фактов под знаком одного факта таким образом, что факт, построенный в результате обобщения, обладает общими для группы свойствами.

- Пусть факт α является результатом обобщения группы фактов $\alpha_1, \alpha_2, \dots, \alpha_n$, данное утверждение будем записывать следующим образом:

$$\{\alpha_1, \alpha_2, \dots, \alpha_n\} \text{ is } \alpha \text{ или } \alpha_i \text{ is } \alpha \quad i = \overline{1, n}.$$

Для обобщения справедливы следующие *основные свойства обобщения*:

$$1^\circ (x_1 = c_1, x_2 = c_2, \dots, x_m = c_m, x_{m+1}, \dots, x_n) \text{ is } (x_1 = c_1, x_2 = c_2, \dots, x_m = c_m)$$

$$2^\circ (\forall \alpha | \alpha \rightarrow \beta; \gamma \text{ is } \beta) : \alpha \rightarrow \gamma$$

$$3^\circ (\forall \beta | \alpha \rightarrow \beta; \gamma \text{ is } \alpha) : \gamma \rightarrow \beta$$

Обобщение является мощным инструментом анализа, которое позволяет отсечь второстепенные свойства фактов и заглянуть глубоко в их суть.

Заключение

В данной работе рассмотрены базовые положения формальной математической теории, которая лежит в основе подхода к моделированию и анализу фактов и причинно-следственных связей между фактами. Описанный подход – основа для разработки программных средств, которые обеспечивают исследователей возможностями обнаруживать связи между фактами или проверять гипотезы о наличии связей. Программные средства проходят апробацию при создании исследовательского портала «Инновационное развитие регионов».

Благодарности

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

Библиография

- [1] Мальцев П. Моделирование многомерных данных в системе METAS BI-PLATFORM // Advanced Studies in Software and Knowledge Engineering: International Book Series / Sofia, 2008. С. 173-180.
- [2] Мальцев П.А., Лядова Л.Н. Формализация многомерной модели данных // Математика программных систем: Межвузовский сб. науч. тр. / Перм. ун-т. Пермь, 2006. С. 74-87.

Сведения об авторе

Павел Мальцев – Пермский государственный университет, аспирант кафедры математического обеспечения вычислительных систем; Россия, г. Пермь, 614990, ул. Букирева, 15; e-mail:

pavel.maltsev@mail.ru.

