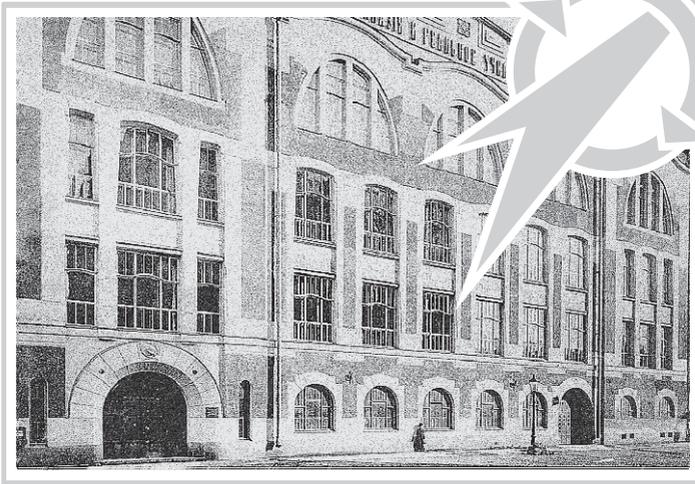


РОССИЙСКАЯ АКАДЕМИЯ НАУК
Отделение нанотехнологий и информационных технологий

САНКТ-ПЕТЕРБУРГСКИЙ
ИНСТИТУТ ИНФОРМАТИКИ И АВТОМАТИЗАЦИИ РАН

ТРУДЫ СПИИРАН

proceedings.spiiras.nw.ru



ВЫПУСК 4(47)



Санкт-Петербург
2016

18+

Труды СПИИРАН

Выпуск № 4(47), 2016

Научный, научно-образовательный, междисциплинарный журнал с базовой специализацией в области информатики, автоматизации и прикладной математики

Журнал основан в 2002 году

Учредитель и издатель

Федеральное государственное бюджетное учреждение науки
Санкт-Петербургский институт информатики и автоматизации Российской академии наук
(СПИИРАН)

Главный редактор

Р.М. Юсупов, чл.-корр. РАН, д-р техн. наук, проф., С-Петербург, РФ

Редакционная коллегия

А.А. Ашимов, академик национальной академии наук Республики Казахстан д-р техн. наук, проф., Алматы, Казахстан

С.Н. Баранов, д-р физ.-мат. наук, проф., С.-Петербург, РФ

Н.П. Веселкин, академик РАН, д-р мед. наук, проф., С.-Петербург, РФ

В.И. Городецкий, д-р техн. наук, проф., С.-Петербург, РФ

О.Ю. Гусихин, Ph.D., Диаборн, США

В. Делич, д-р техн. наук, проф., Нови-Сад, Сербия

А.Б. Долгий, Dr. Habil., проф., Сент-Этьен, Франция

М. Железны, Ph.D., доцент, Пльзень, Чешская республика

Д.А. Иванов, д-р экон. наук, проф., Берлин, Германия

И.А. Каляев, д-р техн. наук, профессор, член-корреспондент РАН, Таганрог, РФ

Г.А. Леонов, член-корр. РАН, д-р физ.-мат. наук, проф., С.-Петербург, РФ

К.П. Марков, Ph.D., доцент, Аизу, Япония

Ю.А. Меркурьев, академик Латвийской академии наук, Dr. Habil., проф., Рига, Латвия

Р.В. Мещеряков, д-р техн. наук, профессор, Томск, РФ

Н.А. Молдовян, д-р техн. наук, проф., С.-Петербург, РФ

В.Е. Павловский, д-р физ.-мат. наук, профессор, Москва, РФ

А.А. Петровский, д-р техн. наук, проф., Минск, Беларусь

В.А. Путилов, д-р техн. наук, проф., Апатиты, РФ

В.Х. Пшихопов, д-р техн. наук, профессор, Таганрог, РФ

А.Л. Ронжин (зам. главного редактора), д-р техн. наук, проф., С.-Петербург, РФ

А.И. Рудской, член-корр. РАН, д-р техн. наук, проф., С.-Петербург, РФ

В. Сгурев, академик Болгарской академии наук, д-р техн. наук, проф., София, Болгария

В.А. Скормин, Ph.D., проф., Бингемптон, США

А.В. Смирнов, д-р техн. наук, проф., С.-Петербург, РФ

Б.Я. Советов, академик РАО, д-р техн. наук, проф., С.-Петербург, РФ

В.А. Сойфер, член-корр. РАН, д-р техн. наук, проф., Самара, РФ

Б.В. Соколов, д-р техн. наук, проф., С.-Петербург, РФ

Л.В. Уткин, д-р техн. наук, проф., С.-Петербург, РФ

А.Л. Фрадков, д-р техн. наук, проф., С.-Петербург, РФ

Н.В. Хованов, д-р физ.-мат. наук, проф., С.-Петербург, РФ

Л.Б. Шереметов, д-р техн. наук, Мехико, Мексика

А.В. Язенин, д-р техн. наук, профессор, Тверь, РФ

Адрес редакции

191718, Санкт-Петербург, 14-я линия, д. 39,

e-mail: publ@iias.spb.su, сайт: <http://www.proceedings.spiiras.nw.ru/>

Подписано к печати 01.08.2016. Формат 60×90 1/16. Усл. печ. л. 15,0. Заказ № 282. Тираж 150 экз., цена свободная
Отпечатано в Редакционно-издательском центре ГУАП, 190000, Санкт-Петербург, Б. Морская, д. 67

Журнал зарегистрирован Федеральной службой по надзору в сфере связи и массовых коммуникаций,
свидетельство ПИ № ФС77-41695 от 19 августа 2010 г.
Подписной индекс 29393 по каталогу «Почта России»

Журнал входит в «Перечень ведущих рецензируемых научных журналов и изданий, в которых должны быть опубликованы основные научные результаты диссертации на соискание ученой степени доктора и кандидата наук»

© Федеральное государственное бюджетное учреждение науки

Санкт-Петербургский институт информатики и автоматизации Российской академии наук, 2016

Разрешается воспроизведение в прессе, а также сообщение в эфир или по кабелю опубликованных в составе печатного периодического издания-журнала «Труды СПИИРАН» статей по текущим экономическим, политическим, социальным и религиозным вопросам с обязательным указанием имени автора статьи и печатного периодического издания-журнала «Труды СПИИРАН»

SPIIRAS Proceedings

Issue № 4(47), 2016

Scientific, educational, and interdisciplinary journal primarily specialized
in computer science, automation, and applied mathematics

Trudy SPIIRAN ♦ Founded in 2002 ♦ Труды СПИИРАН

Founder and Publisher

Federal State Budget Institution of Science

St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences
(SPIIRAS)

Editor-in-Chief

R.M. Yusupov, Prof., Dr. Sci., Corr. Member of RAS, St. Petersburg, Russia

Editorial Board Members

A.A. Ashimov, Prof., Dr. Sci., Academician
of the National Academy of Sciences of the
Republic of Kazakhstan, Almaty, Kazakhstan
S.N. Baranov, Prof., Dr. Sci., St. Petersburg, Russia
N.P. Veselkin, Prof., Dr. Sci., Academician of RAS,
St. Petersburg, Russia
V.I. Gorodetski, Prof., Dr. Sci., St. Petersburg, Russia
O.Yu. Gusikhin, Ph. D., Dearborn, USA
V. Delic, Prof., Dr. Sci., Novi Sad, Serbia
A. Dolgui, Prof., Dr. Habil., St. Etienne, France
M. Zelezny, Assoc. Prof., Ph.D., Plzen, Czech
Republic
I.A. Kalyaev, Prof., Dr. Sci., Corr. Member of RAS,
Taganrog, Russia
D.A. Ivanov, Prof., Dr. Habil., Berlin, Germany
G.A. Leonov, Prof., Dr. Sci., Corr. Member of RAS,
St. Petersburg, Russia
K.P. Markov, Assoc. Prof., Ph.D., Aizu, Japan
Yu.A. Merkuriev, Prof., Dr. Habil., Academician
of the Latvian Academy of Sciences, Riga, Latvia
R.V. Meshcheryakov, Prof., Dr. Sci., Tomsk, Russia
N.A. Moldovian, Prof., Dr. Sci., St. Petersburg, Russia
V.E. Pavlovskiy, Prof., Dr. Sci., Moscow, Russia
A.A. Petrovsky, Prof., Dr. Sci., Minsk, Belarus

V.A. Putilov, Prof., Dr. Sci., Apatity, Russia
V.K. Pshikhopov, Prof., Dr. Sci., Taganrog, Russia
A.L. Ronzhin (Deputy Editor-in-Chief),
Prof., Dr. Sci., St. Petersburg, Russia
A.I. Rudskoi, Prof., Dr. Sci., Corr. Member of RAS,
St. Petersburg, Russia
V. Sgurev, Prof., Dr. Sci., Academician
of the Bulgarian academy of sciences, Sofia,
Bulgaria
V. Skormin, Prof., Ph.D., Binghamton, USA
A.V. Smirnov, Prof., Dr. Sci., St. Petersburg, Russia
B.Ya. Sovetov, Prof., Dr. Sci., Academician of RAE,
St. Petersburg, Russia
V.A. Soyfer, Prof., Dr. Sci., Corr. Member of RAS,
Samara, Russia
B.V. Sokolov, Prof., Dr. Sci., St. Petersburg, Russia
L.V. Utkin, Prof., Dr. Sci., St. Petersburg, Russia
A.L. Fradkov, Prof., Dr. Sci., St. Petersburg, Russia
N.V. Hovanov, Prof., Dr. Sci., St. Petersburg,
Russia
L.B. Sheremetov, Assoc. Prof., Dr. Sci., Mexico,
Mexico
A.V. Yazenin, Prof., Dr. Sci. Tver, Russia

Editorial Board's address

14-th line VO, 39, SPIIRAS, St. Petersburg, 199178, Russia,

e-mail: publ@iias.spb.su, web: <http://www.proceedings.spiiras.nw.ru/>

Signed to print 01.08.2016

Printed in Publishing center GUAP, 67, B. Morskaya, St. Petersburg, 190000, Russia

The journal is registered in Russian Federal Agency for Communications and Mass-Media Supervision,
certificate ПИ № ФС77-41695 dated August 19, 2010 r.

Subscription Index 29393, Russian Post Catalog

© Federal State Budget Institution of Science

St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, 2016

СОДЕРЖАНИЕ

Информационная безопасность

Федорченко А.В., Левшун Д.С., Чечулин А.А., Котенко И.В.
АНАЛИЗ МЕТОДОВ КОРРЕЛЯЦИИ СОБЫТИЙ БЕЗОПАСНОСТИ В СИЕМ-СИСТЕМАХ.
ЧАСТЬ 1 5

Гавришев А.А., Жук А.П., Осипов Д.Л.
АНАЛИЗ ТЕХНОЛОГИЙ ЗАЩИТЫ РАДИОКАНАЛА ОХРАННО-ПОЖАРНЫХ
СИГНАЛИЗАЦИЙ ОТ НЕСАНКЦИОНИРОВАННОГО ДОСТУПА 28

Васечкин Е.А., Таранов А.Б.
МОДЕЛЬ СИГНАЛОВ ПОБОЧНЫХ ЭЛЕКТРОМАГНИТНЫХ ИЗЛУЧЕНИЙ
ВИДЕОИНТЕРФЕЙСОВ 46

Методы управления и обработки информации

Кротов К.В.
МНОГОУРОВНЕВАЯ МОДЕЛЬ ПОСТРОЕНИЯ РАСПИСАНИЙ ОБРАБОТКИ ПАРТИЙ
ДАННЫХ В КОНВЕЙЕРНЫХ СИСТЕМАХ ПРИ ФОРМИРОВАНИИ КОМПЛЕКТОВ И
НАЛИЧИИ ОГРАНИЧЕНИЙ 65

Карпович С.Н.
МНОГОЗНАЧНАЯ КЛАССИФИКАЦИЯ ТЕКСТОВЫХ ДОКУМЕНТОВ С ИСПОЛЬЗОВАНИЕМ
ВЕРОЯТНОСТНОГО ТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ ML-PLSI 92

Бирюков Д.Н., Ломако А.Г., Жолус Р.Б.
ПОПОЛНЕНИЕ ОНТОЛОГИЧЕСКИХ СИСТЕМ ЗНАНИЙ НА ОСНОВЕ
МОДЕЛИРОВАНИЯ УМОЗАКЛЮЧЕНИЙ С УЧЕТОМ СЕМАНТИКИ РОЛЕЙ 105

Липатов А.А., Ушаков В.Н., Никитина М.В.
МЕТОДЫ ВЫЯВЛЕНИЯ ОДНОРОДНЫХ И НЕОДНОРОДНЫХ ГРУПП ОБЪЕКТОВ НА
ОСНОВЕ НЕОПРЕДЕЛЕННЫХ КАЧЕСТВЕННЫХ ДАННЫХ 130

Ильясов Б.Г., Герасимова И.Б., Карамзина А.Г., Суюргулова Ю.Р.
МОДЕЛЬ НАКОПЛЕНИЯ ЗНАНИЙ В НАУЧНОЙ ШКОЛЕ В РЕЗУЛЬТАТЕ
САМООРГАНИЗАЦИИ ПРОЦЕССА ОБМЕНА ИНФОРМАЦИИ 144

Фридман А.Я., Курбанов В.Г.
ИНФОРМАЦИОННАЯ ТЕХНОЛОГИЯ ИНТЕГРАЦИИ ПРОСТРАНСТВЕННЫХ ДАННЫХ В
СИТУАЦИОННУЮ СИСТЕМУ МОДЕЛИРОВАНИЯ 163

Бураков В.В., Волков В.Ф., Потрясаев С.А., Салухов В.И., Шедько Н.А.
ФОРМАЛИЗОВАННЫЕ ПОДХОДЫ К ИССЛЕДОВАНИЮ ТОЧНОСТИ МОДЕЛЕЙ
ЭКОНОМИЧЕСКИХ СИСТЕМ 192

Алгоритмы и программные средства

Шишигин Д.С.
К ВЫБОРУ ТЕХНОЛОГИИ ИНТЕГРАЦИИ ПРИКЛАДНОГО ПРОГРАММНОГО
ОБЕСПЕЧЕНИЯ С САПР 211

Л.В. Уткин, Ю.А. Жук
ПОЛНОГЕНОМНЫЙ ПОИСК АССОЦИАЦИЙ С ИСПОЛЬЗОВАНИЕМ МАТРИЦ ПАРНЫХ
СРАВНЕНИЙ 225

CONTENTS

Information Security

Fedorchenko A.V., Levshun D.S., Chechulin A.A., Kotenko I.V. AN ANALYSIS OF SECURITY EVENT CORRELATION TECHNIQUES IN SIEM-SYSTEMS. PART 1	5
Gavrishev A.A., Zhuk A.P., Osipov D.L. ANALYSIS OF PROTECTION TECHNOLOGIES RADIO FIRE ALARM SYSTEMS AGAINST UNAUTHORIZED ACCESS	28
Vasechkin E.A., Taranov A.B. MODEL SIGNALS STRAY ELECTROMAGNETIC RADIATION VIDEO INTERFACES	46

Methods of Information Processing and Management

Krotov K.V. A MULTILEVEL SCHEDULING MODEL FOR DATA BATCHES PROCESSING IN CONVEYOR SYSTEMS WHEN FORMING SETS AND IN THE PRESENCE OF RESTRICTION	65
Karpovich S.N. MULTI-LABEL CLASSIFICATION OF TEXT DOCUMENTS USING PROBABILISTIC TOPIC MODELING	93
Biryukov D.N., Lomako A.G., Zholus R.B. ONTOLOGICAL KNOWLEDGE SYSTEM COMPLETION BASED ON MODELING INFERENCES TAKING INTO ACCOUNT ROLE SEMANTICS	105
Lipatov A.A., Ushakov V.N., Nikitina M.V. METHODS OF DETECTING HOMOGENEOUS AND HETEROGENEOUS GROUPS OF OBJECTS BASED ON UNCERTAIN QUALITATIVE DATA	130
Ilyasov B.G., Gerasimova I.B., Karamzina A.G., Suyargulova Iu.R. A MODEL OF KNOWLEDGE ACCUMULATION IN THE SCIENTIFIC SCHOOL AS A RESULT OF SELF-ORGANIZATION OF INFORMATION EXCHANGE PROCESS	144
Fridman A.Ja., Kurbanov V.G. INFORMATION TECHNOLOGY TO INTEGRATE SPATIAL DATA INTO THE SITUATIONAL MODELLING SYSTEM	163
V.V. Burakov, V.F. Volkov, S.A. Potryasaev, V.I. Salukhov, N.A. Shedko METHOD FOR MODERNIZING INFORMATION SYSTEMS FOR ENVIRONMENTAL MONITORING BASED ON THEIR FUNCTIONALITY	191

Algorithms and Software

Shishigin D.S. DETERMINATE IDENTIFICATION OF RUSSIAN TEXT LETTER BIGRAMS	211
Utkin L.V., Zhuk Y.A. AXIOMS OF MULTICRITERIA OPTIMIZATION METHODS ON A FINITE SET OF ALTERNATIVES	225

А.В. ФЕДОРЧЕНКО, Д.С. ЛЕВШУН, А.А. ЧЕЧУЛИН, И.В. КОТЕНКО
**АНАЛИЗ МЕТОДОВ КОРРЕЛЯЦИИ СОБЫТИЙ
БЕЗОПАСНОСТИ В SIEM-СИСТЕМАХ. ЧАСТЬ 1**

Федорченко А.В., Левшун Д.С., Чечулин А.А., Котенко И.В. Анализ методов корреляции событий безопасности в SIEM-системах. Часть 1.

Аннотация. Статья посвящена анализу методов корреляции событий безопасности в системах управления информацией и событиями безопасности (SIEM-системах). Процесс корреляции событий безопасности рассматривается в виде многоуровневой иерархии этапов, цель каждого из которых заключается в выполнении определенных операций над обрабатываемыми данными безопасности. На основе результатов проведенного анализа в работе приводится описание каждого этапа процесса корреляции и схемы их взаимодействия.

Ключевые слова: процесс корреляция данных, события безопасности, анализ событий безопасности, системы оценки защищенности, SIEM-системы.

Fedorchenko A.V., Levshun D.S., Chechulin A.A., Kotenko I.V. An Analysis of Security Event Correlation Techniques in Siem-Systems. Part 1.

Abstract. The paper is devoted to the analysis of security event correlation methods in Security Information and Event Management (SIEM) systems. The correlation process is considered to be a multilevel hierarchy of stages. The goal of each stage consists in executing appropriate operations on security data being processed. Based on this analysis we outline each correlation stage and their interaction scheme.

Keywords: data correlation process; security event; security event analysis; computer network security evaluation systems; SIEM systems.

1. Введение. В настоящее время все больше внимания уделяется обеспечению безопасности информации как в крупных учреждениях и компаниях, так и в средних и малых организациях. Защищаемые объекты имеют различные уровни доступа, всевозможные варианты развертывания вычислительных сред и разнообразные топологии сетевого взаимодействия. Задача обеспечения безопасности с помощью универсальных средств обнаружения и предотвращения атак усложняется, в том числе за счет стремительного роста числа пользователей и разнообразия типов устройств, использования облачных технологий и многократного увеличения объема и скорости передачи и обработки информации. Одним из классов средств, позволяющих обеспечивать безопасность систем любого уровня и набора устройств, являются системы управления информацией и событиями безопасности (Security Information and Event Management, SIEM) [1, 2]. Преимущество данных решений заключается в гибкости применения и независимости от набора спецификаций и платформ для конечной защищаемой инфраструктуры.

Схема взаимодействия компонент SIEM-систем в общем виде всегда содержит модуль корреляции информации и событий безопасности, который является основополагающим элементом применяемых механизмов безопасности. Работа компонентов корреляции в широком смысле направлена на обнаружение атак, вредоносной активности и нарушений политики безопасности. В узком смысле компонент корреляции предназначен для поиска связей, зависимостей и причинно-следственных отношений между событиями безопасности и другой информацией безопасности. Данный компонент выполняет как функции корреляции событий, так и их пред- пост обработку в зависимости от конкретной реализации системы.

Учитывая необходимость оперирования событиями и информацией безопасности, компонент корреляции можно рассматривать с разных точек зрения. Корреляция представляется либо с точки зрения рассмотрения ее как процесса, либо с точки зрения множеств входных и выходных данных, преобразуемых внутри модуля корреляции. В первом случае корреляция представляет собой последовательность операций над событиями, специально определенную для получения конкретного решения и являющуюся непрерывной относительно работы всей системы. Во втором случае корреляция задается через множество типов событий [3], преобразуемых таким образом, что несколько событий могут образовывать одно более сложное событие и восприниматься системой как неделимое [4-6].

Стоит отметить, что в процесс корреляции можно представить с использованием двух основных типов операций (функций): (1) *комбинирования событий безопасности в одно мета-событие*; (2) *идентификации и удаления (или обозначения) ложных или бесполезных событий безопасности* [7]. Важными операциями корреляции также являются представление событий безопасности и обучение (переобучение) системы корреляции во время работы. Вместе с тем данные типы операций на разных этапах процесса корреляции выполняют отличающиеся по своему характеру действия.

Главным образом данная *работа посвящена* анализу методов корреляции за счет их детального изучения, выявления принципов работы основных компонентов модуля корреляции, классификации и последующего сравнения компонентов друг с другом по основным характеристикам. *Целью работы* является задание основополагающих этапов корреляции и определение достоинств и недостатков различных методов корреляции, используемых в мировой практике для реализаций современных SIEM-решений.

Данный анализ производился за счет изучения научно-технической литературы, включающей описания как отдельных методов корреляции и их общих обзоров, так и особенностей реализации компонентов корреляции в конкретных решениях открытых продуктов данного класса. *Новизна* статьи заключается в предложении собственной схемы этапов процесса корреляции и оценке применения в них различных методов. Качественное сравнение методов корреляции достигается за счет определения их общих и частных характеристик, на основе которых предлагается новая система классификации.

Результаты работы представлены в двух статьях. Данная статья является первой частью описания проводимых исследований, где корреляция событий безопасности рассматривается как процесс, разделенный на отдельные этапы. В разделе 2 приводятся релевантные работы по тематике описания процесса корреляции. В них рассматриваются прототипы систем корреляции предупреждений для систем обнаружения вторжений (СОВ). В данных работах также анализируются различные варианты выполнения базовых этапов в процессе корреляции. В разделе указываются преимущества и недостатки используемых подходов и схем. В разделе 3 описывается процесс корреляции событий безопасности с точки зрения оптимального позиционирования этапов работы модуля корреляции и распределения обработки данных по уровням. В данном разделе раскрываются задачи, возлагаемые на каждый этап, необходимость их использования и возможные способы реализации.

2. Релевантные работы. Тема представления процесса корреляции событий безопасности в виде последовательно выполняемых этапов активно исследуется на протяжении последних 20 лет. За это время были предложены различные методики обработки разнородных данных и преобразования в процессе корреляции низкоуровневых событий к высокоуровневым, а также рассмотрены возможные схемы, описывающие сам процесс корреляции. Вместе с развитием данной тематики публиковались работы, посвященные классификации этапов процесса корреляции событий безопасности [8-13].

В [8] предлагается разбиение процесса корреляции на выполняемые задачи. Выделяются следующие составляющие процесса: *сжатие (compression)*, *счет (count)*, *подавление (suppression)*, *логическая замена (boolean)* и *обобщение (generalization)*. Под *сжатием* подразумевается преобразование в одно событие безопасности нескольких одинаковых событий. *Счет* представляет собой замену похожих событий безопасности одним новым событием, а *подавление* — осуществление задержки обработки событий безопасности с низким

приоритетом до окончания обработки события безопасности с более высоким приоритетом. Процесс *логической замены* заключается в преобразовании некоторого множества событий в новое событие, удовлетворяющее определенному логическому шаблону. В результате выполнения *обобщения* производится перевод события безопасности к высокоуровневому представлению (суперклассу) для удовлетворения необходимой важности уведомления. Несомненным достоинством работы является возможность добавления статистических методов в описанную модель корреляции событий безопасности, несмотря на то, что она основана на строго детерминированных подходах. Недостатки представленной работы заключаются в отсутствии среди указанных задач элемента *предупреждения ошибок (fault prediction)* и элемента *предупреждения нарушений (preventive maintenance)*.

В [9] производится обзор работ в области корреляции предупреждений для систем обнаружения вторжений (Intrusion Detection System, IDS). В частности, рассматриваются этапы и операции процесса корреляции, описывается модель данных формата обмена сообщениями обнаружения вторжений (Intrusion Detection Message Exchange Format, IDMEF), а также приводится пример процесса корреляции для обнаружения типовой атаки. Процесс корреляции в данной работе условно делится на три этапа: (1) *предобработка (preprocessing)*; (2) *анализ предупреждений (alarm analysis)*; (3) *корреляция предупреждений (alarm correlation)*. На втором этапе выделяются такие методы и этапы процесса корреляции, как измерение схожих признаков (*similarity measures*), кластеризация, интеллектуальный анализ данных (*data mining*), удаление, редукция и слияние. Стоит отметить, что в результате выполнения каждого из трех этапов формируются простые события, мета-события и сценарии атак соответственно, а по окончании выполнения процесса корреляции формируется отчет.

В [10] выделяется шесть этапов процесса корреляции: *нормализация (normalization)*, *агрегация (aggregation)*, *корреляция (correlation)*, *отсевание ложных срабатываний (false alert reduction)*, *анализ стратегии атаки (attack strategy analysis)* и *приоритизация (prioritization)*. Описываются четыре основных метода корреляции: (1) на основе сценариев атак; (2) ориентированного на правила; (3) статистического и (4) временного. Главное отличие [10] от аналогичных работ заключается в точном связывании этапов процесса корреляции с используемыми в них конкретными методами. В данной работе также выделены отдельные группы методов корреляции.

В [11-13] авторы выделяют пять подходов к процессу корреляции событий безопасности, основанных на: (1) *подобии* (сходстве) (*similarity based*); (2) *предопределении сценариев атак* (*predefined attack scenarios based*); (3) *многоуровневых вычислениях* (*multi-stage*) на базе *предпосылок и последствий*; (4) *использовании множества источников информации* (*multiple information sources*); (5) *фильтрации* (*filter based*).

Первый подход заключается в вычислении величины подобия двух событий безопасности на основе атрибутов, ассоциируемых с этими событиями. События, величина подобия которых достаточно велика, группируются.

Второй подход заключается в объединении в последовательность связанных этапов проведения атак на основе заранее определенных шаблонов сценариев атак. Данный подход применяется для получения агрегированного и более высокоуровневого взгляда на угрозы безопасности.

Третий подход основывается на формировании сценариев атак путем связывания отдельных этапов их проведения при условии, что один из этапов является необходимым условием для проведения другого.

Четвертый подход направлен на приоритизирование и классификацию потоков событий безопасности в зависимости от источника информации о событии безопасности.

Пятый подход основан на удалении из процесса корреляции событий по заранее определенным правилам (фильтрам). Решение об удалении события из процесса корреляции принимается на основе значений одного или нескольких его атрибутов.

Похожая классификация приводится в [14]. Приведенная авторами модель корреляции событий безопасности состоит из двух частей: (1) подхода, *основанного на графах атак* (*an attack graph-based*) и (2) подхода, *основанного на подобии* (*similarity-based*). При этом первый подход используется для корреляции событий безопасности, вызванных известными атаками, а также для построения гипотез о вероятно необнаруженных или упущенных событиях безопасности. В свою очередь, второй подход применяется для корреляции событий безопасности, вызванных неизвестными атаками, а также для уточнения известных графов атак.

В [7, 15, 16] выделяются и описываются отдельные этапы и уровни процесса корреляции.

В [7] Крюгел, Валеур и Вигна раскрывают методы и подходы к корреляции предупреждений в зависимости от фазы процесса, указывают на их достоинства и недостатки, а также спорные моменты. В работе

процесс корреляции предупреждений представлен в виде этапов, которые преобразуют оповещения сенсоров в отчеты о вторжениях и направлены на разные аспекты процесса корреляции (рисунок 1).

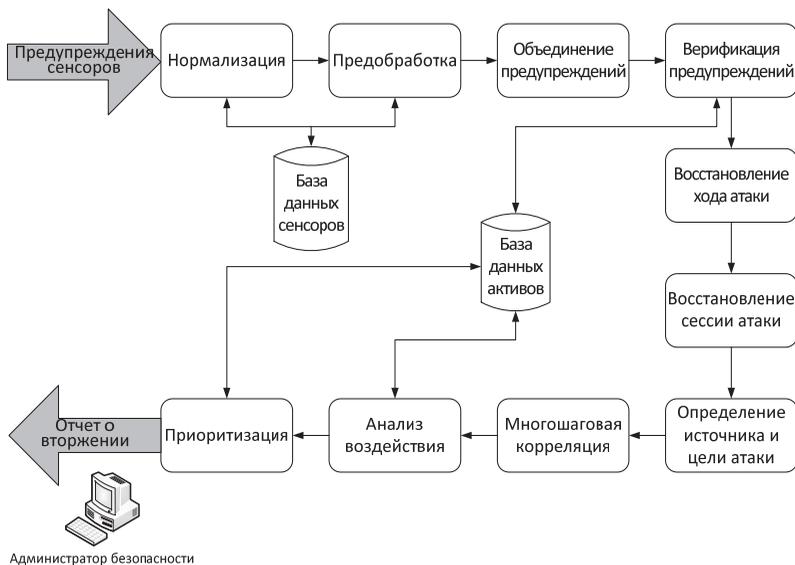


Рис. 1. Представление процесса корреляции предупреждений в [7]

Именно данная схема взята за основу во многих других работах (в том числе и в настоящей статье) для представления этапов процесса корреляции и их последовательности. В [7] рассматриваются методы корреляции предупреждений в системах обнаружения вторжений (Intrusion Detection System, IDS), которые также применимы в SIEM-системах, несмотря на то, что предупреждения являются только отдельным типом событий. Иными словами, за счет расширения типов обрабатываемых событий SIEM-системы расширяют системы обнаружения вторжений в рамках корреляции. В данной работе также производится структурное деление этапов процесса корреляции предупреждений в зависимости от их целей, таких как: сбор (*collection*), агрегация и верификация (*aggregation and verification*), анализ высокоуровневых структур (*high-level structures*), крупномасштабная корреляция (*large-scale correlation*), оценка (*evaluation*). В рамках детектирования атак приводятся следующие виды систем: (1) на основе злоупотреблений (*misuse-based*; в основу положена база знаний; атака идентифицируется при соответствии записи базы с параметрами входных

данных); (2) на основе аномалий (*anomaly-based*; текущее состояние сравнивается с эталонным с помощью оценки вероятности отклонения).

В [15] процесс корреляции предупреждений представлен в виде логических блоков. Авторы выделяют следующие блоки: *нормализация данных (Data Normalization Unit)*, *корреляция на основе фильтрации (Filter-based Correlation Unit)*, *редукция данных (Data Reduction Unit)*, *анализ намерений (Intention Recognition)* и *анализ воздействий (Impact Analysis)*.

Отличительной особенностью данной работы является представление модели процесса корреляции (рисунок 2), которая снижает количество обрабатываемых событий безопасности так рано, как это только возможно. Это осуществляется путем вывода из процесса корреляции незначимых или ложных событий безопасности еще на начальных этапах процесса. Авторы ввели дополнительный компонент для работы с некоррелируемыми данными. По итогам эксперимента на наборах данных DARPA 2000 авторам удалось добиться процента редукции 99,38 % (в среднем).

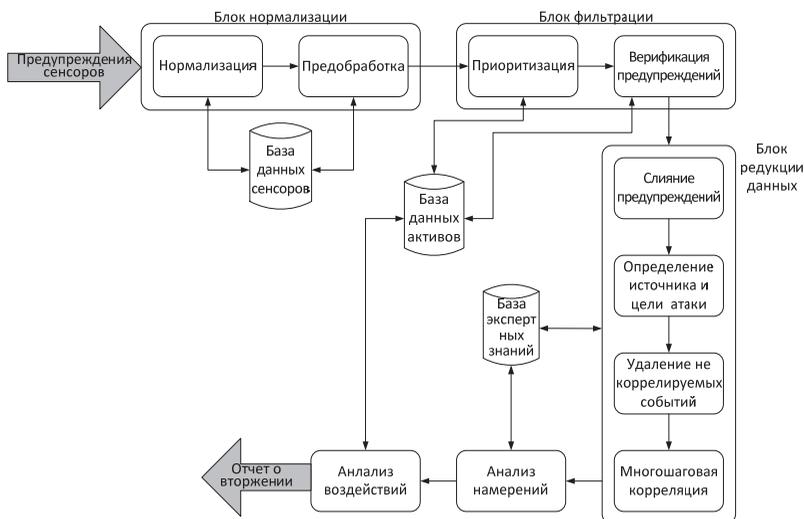


Рис. 2. Модель процесса корреляции событий безопасности [15]

Важно отметить, что данная модель не лишена недостатков. Во-первых, в блоке фильтрации этапы верификации и приоритезации безусловно будут обрабатывать в том числе и дубликаты событий безопасности, так как этап слияния предупреждений находится на более высоком уровне (блок редукции данных). Учитывая ресурсоемкость процесса проверки событий безопасности на подлинность, данное

решение подлежит дополнительному рассмотрению. Во-вторых, этап удаления из процесса корреляции событий безопасности данных, которые не могут быть коррелированы (*Удаление не коррелируемых событий*), не оставляет процессу права на ошибку. При увеличении среднего показателя редукции событий безопасности и облегчении дальнейшего анализа открытым остаётся вопрос гарантии того, что из процесса корреляции не удаляются важные события. И, в-третьих, задача модуля анализа воздействия заключается в исключении из процесса корреляции сценариев атак, влияние которых на инфраструктуру сети незначительно или невозможно, то есть в улучшении коэффициента редукции, при этом модуль анализа воздействия в блок редукции не входит.

В [16] предлагается разделение методов процесса корреляции событий безопасности по следующим уровням обработки данных (рисунок 3): (1) *первичные (сырые) данные (raw data)*; (2) *события (events)*; (3) *отчеты (reports)*. В зависимости от конкретного уровня, данные подвергаются соответствующей обработке для дальнейшего использования на более высоких уровнях, а в итоге — для принятия решения о возможных контрмерах и визуализации результатов.

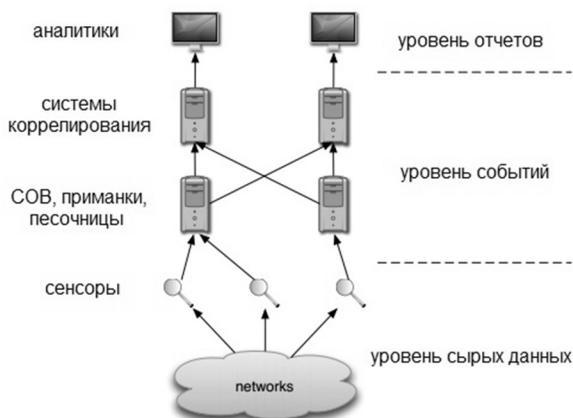


Рис. 3. Уровни процесса корреляции событий безопасности [16]

Уровень первичных данных состоит в основном из сетевых сенсоров, которые принимают необработанные данные и осуществляют их первичный анализ. Источниками данных для системы корреляции событий безопасности на данном уровне также могут быть системы журналирования различных сетевых приложений. На уровне первичных данных реализуются следующие процессы: *отбор пакетов (packet sampling)*, *вероятностный анализ (probabilistic analysis)*,

обнаружение аномальной активности (attack detection), обнаружение сканирования портов (detection of port scans), идентификация приложений (application identification) и анализ полезной нагрузки пакетов (payload analysis). На уровне событий выполняется локальный процесс корреляции событий безопасности и распределенный процесс корреляции событий безопасности, включая соблюдение конфиденциальности данных. На уровне отчетов специализированные приложения визуализируют результаты процесса корреляции наиболее подходящим образом. Уровень отчетов выполняет генерацию возможных активных контрмер и верификацию событий безопасности.

В результате проведенного анализа была сформирована собственная модель процесса корреляции, в основу которой вошли необходимые с точки зрения авторов этапы обработки данных. Данные этапы были логически сгруппированы и связаны в зависимости от функциональной задачи каждого этапа. Описанные результаты раскрываются в разделе 3.

3. Процесс корреляции событий безопасности. Процесс корреляции событий безопасности является сложной задачей. Поэтому предлагается разбить ее на подзадачи с помощью декомпозиции. Применение декомпозиции к системе, реализующей процесс корреляции событий безопасности, позволит представить систему в виде простых функциональных модулей. Такой подход обеспечит рассмотрение каждого модуля независимо друг от друга. При этом важно четко определить функциональную нагрузку каждого модуля и порядок их взаимодействия.

Кроме того, предлагается применить многоуровневый подход к декомпозиции системы корреляции событий безопасности. Многоуровневый подход подразумевает разбиение на группы множества модулей системы корреляции, а также упорядочивание групп модулей по уровням, образующим иерархию. В соответствии с принципом иерархии, для каждого промежуточного уровня следует указать непосредственно примыкающие к нему соседний вышележащий и нижележащий уровни. При этом, с одной стороны, каждая группа модулей одного уровня должна быть сформирована таким образом, чтобы все модули этой группы для выполнения своих функций использовали результаты обработки событий безопасности, полученные на соседнем нижележащем или текущем уровне. С другой стороны, результаты обработки данных каждого модуля, отнесенного к некоторому уровню, могут быть переданы только модулям соседнего вышележащего уровня или среди модулей текущего уровня. Такой подход позволит определить задачи, выполняемые на каждом из

уровней, принципы их взаимодействия, а также типы данных, передаваемые между ними. Кроме того, декомпозиция системы, реализующей процесс корреляции, упрощает разработку, отладку и тестирование отдельных уровней или модулей.

Предлагаемое обобщенное представление системы корреляции событий безопасности приведено на рисунке 4.

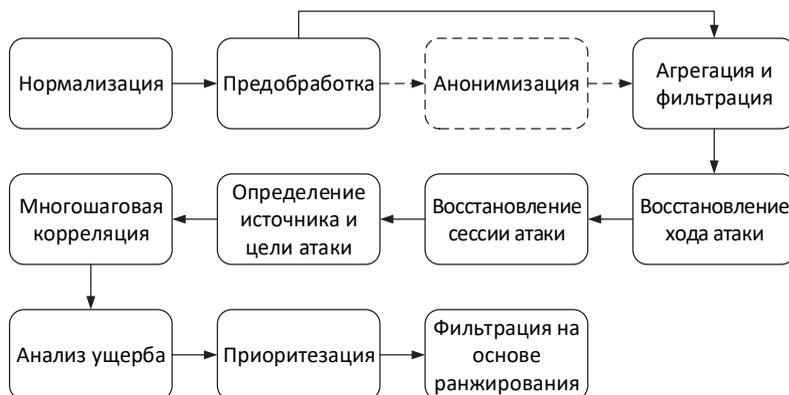


Рис. 4. Представление системы корреляции в виде модулей

В основу компонентов данной схемы легли этапы процесса корреляции, описанные в [7]. Отметим, что в рамках указанного представления данные, обработанные одним модулем, используются в качестве входных данных для следующего модуля. Тем не менее прохождение каждого события безопасности последовательно через одни и те же модули не является необходимым. Реальный процесс корреляции событий безопасности в зависимости от конкретной задачи может быть представлен гораздо более сложными схемами. Например, обработка потока событий безопасности может включать циклические операции между некоторыми модулями системы, а сами модули в некоторых ситуациях могут использоваться параллельно.

Важно отметить, что в рамках модульного представления системы корреляции событий безопасности не отражена реализация процесса верификации или проверки источников событий безопасности на подлинность. Если система корреляции событий безопасности будет получать в качестве входных данных события безопасности от любых источников, ничто не помешает злоумышленнику сгенерировать поток ошибок второго рода, притворившись одним из сенсоров. Наличие подобного потока событий значительно ухудшает качество процесса корреляции событий безопасности и может привести к обнаружению

сценариев атак, которых не существуют. Задача верификации источников событий безопасности ложится на этап сбора данных, поэтому в рамках системы корреляции событий безопасности каждое событие безопасности уже считается верифицированным, то есть полученным от разрешенного источника событий безопасности.

Далее модули системы корреляции событий рассматриваются более подробно.

Нормализация. Из-за того что источники данных могут поставлять информацию в разном формате, возникает необходимость преобразования формата каждого события безопасности в некоторый нормализованный формат, который был бы понятен всем модулям обработки. Данное преобразование, или *нормализация*, означает, что синтаксис и семантика события безопасности прозрачны и беспрепятственно определяемы.

Предобработка. После нормализации обработанные события безопасности нуждаются в дополнительной *предобработке*, так как часть источников может пропускать некоторые поля данных, важные для процесса корреляции (например, время начала, время окончания и источник события).

Анонимизация. Данный модуль системы корреляции необходим, если производится работа с событиями минимум от двух источников, между которыми не установлено доверительное отношение. Анонимизация применяется для удаления или сокрытия конфиденциальной (или важной с юридической точки зрения) информации из событий безопасности. Существует две операции данного модуля: *анонимизация* и *псевдоанонимизация* [17, 18]. Анонимизация препятствует восстановлению конфиденциальных данных, в то время как псевдоанонимизация — обратима, а значит, оригинальные данные могут быть восстановлены доверительной стороной. В общем случае желательно проводить псевдоанонимизацию, так как это позволяет получить доступ к оригинальной информации в ситуациях, когда необходим дальнейший анализ. Однако данное решение накладывает значительные вычислительные ограничения. Ключевой задачей модуля анонимизации является как сохранение необходимых свойств для анализа безопасности, так и способность их сокрытия от нежелательных сторон. Извлечение подобных свойств предполагает *деанонимизацию* (обратный процесс) данных, что возможно только при использовании псевдоанонимизирующих методов.

Агрегация и фильтрация. Задача модуля фильтрации и агрегации заключается в удалении из системы корреляции событий по заранее определенным правилам (фильтрам), и в объединении данных, которые возникли в результате независимого обнаружения одного и того же события различными источниками. Решение об удалении события из

системы корреляции принимается на основе значений одного или нескольких его атрибутов. Не прошедшие фильтрацию события безопасности больше в процессе корреляции не участвуют, а оставшиеся — переходят на этап агрегации. Решение об агрегировании двух событий безопасности принимается на основе содержащихся в них данных. При идентичности значений заранее определенных атрибутов событий, а также удовлетворении временных характеристик событий заданному интервалу, такие события безопасности агрегируют в одно мета-событие.

Восстановление хода атаки. Задача модуля восстановления хода атаки ограничена объединением событий безопасности, вызванных активностью одного злоумышленника по отношению к одной цели (рисунок 5). Восстановление хода атаки построено на объединении событий безопасности с совпадающими атрибутами цели и источника атаки, временные параметры которых попадают в заданный временной интервал. Требование к временным параметрам заключается в том, чтобы время окончания более раннего события, характерного для конкретной атаки, было достаточно и определенно близко ко времени старта другого события, продолжающего соответствующую атаку.

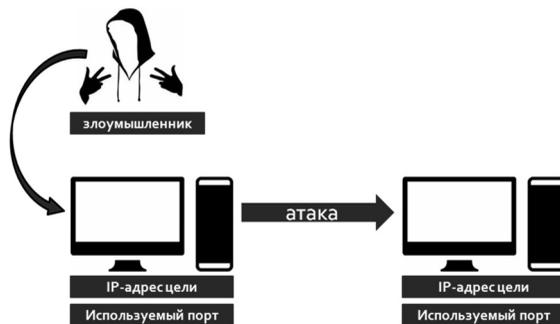


Рис. 5. Восстановление хода атаки

Восстановление сессии атаки. Цель данного модуля — поиск связи между сетевыми (*network-based*) и системными (*host-based*) событиями безопасности (рисунок 6). Это необходимо для объединения ряда событий безопасности, вызванных злоумышленником при тестировании различных эксплоитов против определенной программы или запуском одного и того же эксплоита несколько раз для подбора правильных значений определенных параметров (например, смещений и адресов памяти для переполнения буфера). Процесс поиска связи между событиями усложняется за счет отличающегося предоставления информации в сетевых и системных событиях. Сетевые сенсоры могут

предоставить информацию, характеризующую обнаруженные атаки, например, IP-адреса источника и цели, используемые порты. Данные, поступающие от системных сенсоров, с другой стороны, содержат информацию об объекте, который был атакован, и субъекте, которым была осуществлена данная атака. Именно обнаружение связи между сетевыми и системными событиями позволяет определить отдельную сессию производимой атаки.

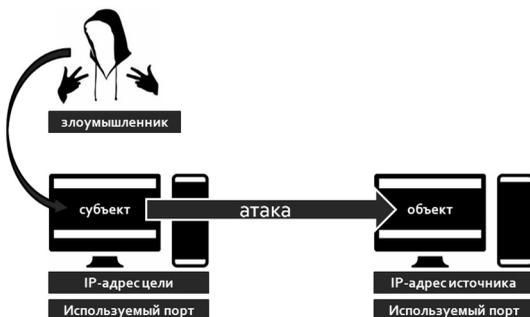


Рис. 6. Восстановление сессии атаки

Определение цели и источника атаки. Задача этого модуля — идентификация хостов, которые являются либо источником, либо целью обнаруживаемых атак. Данный модуль объединяет события безопасности, ассоциируемые с отдельным хостом, который атакует несколько жертв (сценарий *один-ко-многим* (*one2many*)) (рисунок 7а)), и с несколькими хостами, которые атакуют одну жертву (сценарий *многие-к-одному* (*many2one*)) (рисунок 7б)).

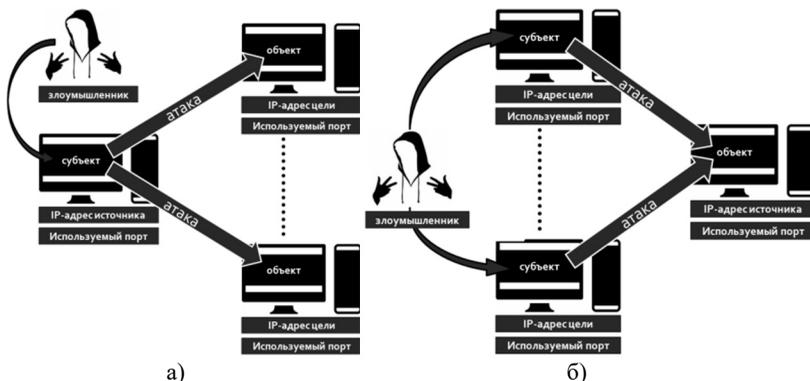


Рис. 7. Определение цели и источника атаки: а) сценарий один-ко-многим; б) сценарий многие-к-одному

Многошаговая корреляция. Модуль используется для распознавания сложных сценариев, которые состоят из нескольких отдельных атак (рисунок 8).

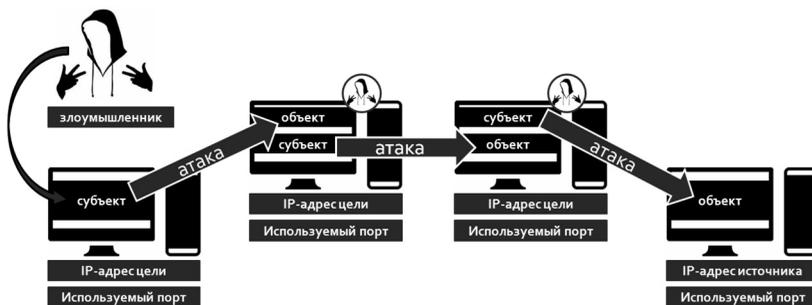


Рис. 8. Схема многошаговой корреляции

Обычно эти сценарии определяются с использованием той или иной формы экспертных знаний [19-21]. Модуль многошаговой корреляции также можно использовать для верификации высокоуровневых событий безопасности. При этом определяются сценарии атак, которые заведомо не имеют значения. Это позволяет удалить из процесса корреляции последовательности событий, которые коррелированы неверно. Например, при опросе сетевого окружения на предмет наличия какого-либо оборудования (принтера или сканера) последовательность событий по количеству запросов будет выглядеть как сканирование сети, однако реально данная последовательность незначительна и фактически не является атакующими действиями. К подобному роду ошибок также можно отнести действия приложений, использующих в работе пиринговые (peer-to-peer, P2P) сети.

Анализ ущерба. Помимо аналитической информации, полученной на предыдущих этапах, данный модуль использует стороннюю (не содержащуюся в событиях) информацию для анализа сценариев атак с точки зрения ущерба от их влияния на инфраструктуру сети или используемые ресурсы. На основе данных об ущербе модуль назначает более высокую степень важности сценариям атак, которые угрожают более значимым активам сети. Информация о сети и соответствующих ресурсах хранится в базе данных активов, которая содержит подробности об используемых сетевых сервисах, зависимостях между ними, а также их важности для функционирования сети.

Приоритизация. Модуль приоритизации должен учитывать политику безопасности и требования безопасности инфраструктуры, в которой развернута система корреляции. Фактически, модуль

ориентирован на пожелания пользователя, использующего систему корреляции. Поэтому его основной задачей является выделение сценариев атак в соответствии с их приоритетом для пользователя.

Фильтрация на основе ранжирования. Этот модуль используется для снижения количества рассматриваемых сценариев атак, критичность реализации которых мала для корректной работы защищаемой инфраструктуры. Данный модуль должен учитывать политику безопасности и требования безопасности инфраструктуры, в которой развернута система корреляции. Удаление из процесса корреляции сценариев атак с низким рангом (ущерб от влияния которых на анализируемую сеть отсутствует или незначителен) снижает количество ошибок второго рода, увеличивая точность работы процесса корреляции.

На следующем этапе исследования системы, реализующей процесс корреляции событий безопасности, была произведена группировка описанных модулей (рисунок 9).

Модуль нормализации преобразует события безопасности от разнородных источников в формат, понятный системе корреляции событий безопасности. Преобразование происходит без потерь и дополнений. Модуль предобработки дополняет нормализованные события безопасности значения атрибутов, которые могли быть пропущены источником события безопасности. Модуль анонимизации работает со значениями атрибутов предобработанных событий безопасности, преобразуя или удаляя их.

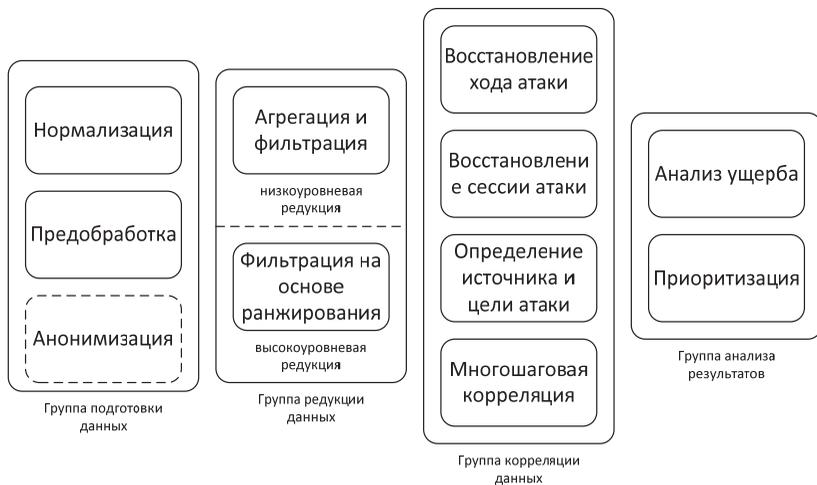


Рис. 9. Группировка модулей

Перечисленные выше модули были объединены в *группу подготовки данных*, которая непосредственно обрабатывает данные, полученные от разнородных источников, с целью их подготовки для модулей более высокого уровня.

Модули восстановления хода атаки, восстановления сессии атаки, определения цели и источника атаки и многошаговой корреляции в совокупности направлены на решение задачи поиска причинно-следственных связей между событиями. Данные модули выделены в *группу корреляции данных*.

Важной задачей процесса корреляции событий безопасности является редукция — операция уменьшения общего количества событий до необходимого и достаточного для проведения анализа, не превосходя вычислительные способности инфраструктуры [22, 23].

По признаку осуществления редукции общего потока информации, модули агрегации и фильтрации и фильтрации на основе ранжирования объединены в *группу редукции данных*. Стоит отметить, что модуль агрегации и фильтрации оперирует с событиями безопасности, в отличие от модуля фильтрации на основе ранжирования, который работает со сценариями атак. Поэтому условно разделим группу редукции данных на две части: *низкоуровневой редукции* и *высокоуровневой редукции*.

Группа высокоуровневой редукции нуждается в специальной информации, зависящей от ущерба при реализации атаки на активы инфраструктуры с учетом критичности потери данных активов. Данные результаты используются в модулях анализа ущерба и приоритизации, которые предлагается выделить в *группу анализа результатов*. При формировании из описанных групп обобщенной системы корреляции, с учетом разбиения на уровни, была получена следующая иерархическая схема (рисунок 10).

На первом уровне, оперирующем непосредственно с событиями безопасности от различных источников, находится группа подготовки данных. Нормализованные, преобразованные и при необходимости анонимизированные данные поступают на второй уровень, где целесообразно расположить группу низкоуровневой редукции для первичной агрегации и фильтрации данных. Дальнейшие операции для увеличения общего показателя редукции производятся над данными более высокого уровня (сложными событиями, сценариями атак). Необходимый переход от простых событий безопасности к сложным событиям, атакам и сценариям атак осуществляется на третьем уровне, который соответствует группе корреляции данных. После прохождения третьего уровня данные дополнительно обрабатываются на четвертом

уровне, который соответствует группе анализа результатов корреляции. После этого данные передаются на пятый уровень, где расположена группа высокоуровневой редукции, осуществляющая фильтрацию сценариев атак.

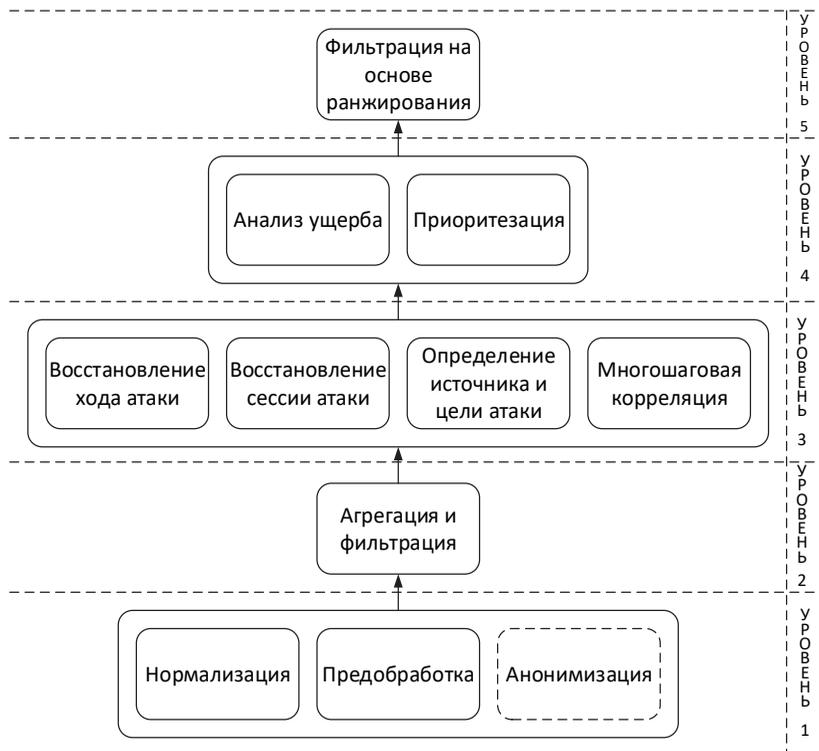


Рис. 10. Многоуровневая иерархия групп модулей

5. Заключение. В данной статье указано место и роль процесса корреляции в SIEM-системе. Описаны основные этапы корреляции, их роль и необходимость использования, а также произведен их анализ с точки зрения реализации системы оценки защищенности компьютерных инфраструктур. По итогам исследования получена четырехуровневая модель процесса корреляции с детализацией каждого уровня.

В результате проделанной работы были однозначно разделены этапы процесса корреляции. Это позволит в дальнейшем более детально исследовать этапы отдельно, а также выделить конкретные методы для эффективного применения на каждом из этапов. В следующей части описания проведенных исследований будут непосредственно

рассмотрены и классифицированы методы корреляции событий безопасности. Также будет произведена оценка применимости представленных методов на различных этапах процесса корреляции.

Литература

1. *Kotenko I.V., Chechulin A.A.* A Cyber Attack Modeling and Impact Assessment Framework // Proceedings of 5th International Conference on Cyber Conflict 2013 (CyCon 2013). 2013. pp. 119–142.
2. *Kotenko I.V., Polubelova O.V., Saenko I.V.* The Ontological Approach for SIEM Data Repository Implementation // 2012 IEEE International Conference on Green Computing and Communications (GreenCom). IEEE Computer Society. 2012. pp. 761–766.
3. *Liu G., Mok A.K., Yang E.J.* Composite Events for Network Event Correlation // IEEE/IFIP International Symposium on Integrated Network Management. 1999. pp. 247–260.
4. *Котенко И.В., Саенко И.Б., Полубелова О.В., Чечулин А.А.* Технологии управления информацией и событиями безопасности для защиты компьютерных сетей // Проблемы информационной безопасности. Компьютерные системы. 2012. № 2. С. 57–68.
5. *Котенко И.В., Саенко И.Б., Полубелова О.В., Чечулин А.А.* Применение технологии управления информацией и событиями безопасности для защиты информации в критически важных инфраструктурах // Труды СПИИРАН. 2012. Вып. 1 (20). С. 27–56.
6. *Котенко И.В., Саенко И.Б., Чечулин А.А.* Проактивное управление информацией и событиями безопасности в информационно-телекоммуникационных системах // Вопросы радиоэлектроники. 2014. Вып 3. № 1. С. 170–180.
7. *Kruegel C., Valeur F., Vigna G.* Intrusion Detection and Correlation: Challenges and Solutions // University of California, Santa Barbara, USA: Springer. 2005. pp. 29-33.
8. *Jakobson G., Weissman M.D.* Alarm correlation // IEEE Network. 1993. vol. 7(6). pp. 52–59.
9. *Zurutuza U., Uribeetxeberria R.* Intrusion Detection Alarm Correlation: A Survey // Proceedings of IADAT International Conference on Telecommunications and computer Networks. 2004. pp. 1–3.
10. *Sadoddin R., Ghorbani A.* Alert Correlation Survey: Framework and Techniques // Proceedings of 2006 International Conference on Privacy, Security and Trust: Bridge the Gap Between PST Technologies and Business Services (PST'06). 2006. Article no. 37.
11. *Dadkhah S., Shoja M.R.K., Taheri H.* Alert Correlation through a Multi Components Architecture // International Journal of Electrical and Computer Engineering (IJECE). 2013. vol. 3. no. 4. pp. 461–466.
12. *Elshoush H.T., Osman I.M.* Alert correlation in collaborative intelligent intrusion detection systems — A survey // Applied Soft Computing. 2011. pp. 4349–4365.
13. *Ning P., Xu D.* Correlation analysis of intrusion alerts // Intrusion Detection Systems: series Advances in Information Security. Springer. 2008. vol. 38. pp. 65–92.
14. *Ahmadinejad S.H., Jalili S., Abadi M.* A hybrid model for correlating alerts of known and unknown attack scenarios and updating attack graphs // Computer Networks. 2011. no. 55. pp. 2221–2240.
15. *Elshoushand H.T., Osman I.M.* An improved framework for intrusion alert correlation // Proceedings of World Congress on Engineering 2012 (WCE 2012). 2012. vol. 1. pp. 518–524.

16. *Limmer T., Dressler F.* Survey of event correlation techniques for attack detection in early warning systems. Tech report // University of Erlangen, Dept. of Computer Science. 2008. 37 p.
17. *Flegel U.* Pseudonymizing Unix Log Files // Infrastructure Security. 2002. LNCS 2437. pp. 162–179.
18. *Pang R., Paxson V.* A high-level programming environment for packet trace anonymization and transformation // Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications. 2003. pp. 339–351.
19. *Kotenko I.V., Chechulin A.A.* Computer Attack Modeling and Security Evaluation based on Attack Graphs // Proceedings of 7th International Conference on “Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications” (IDAACS’2013). 2013. pp. 614–619.
20. *Kotenko I.V., Chechulin A.A.* Fast Network Attack Modeling and Security Evaluation based on Attack Graphs // Journal of Cyber Security and Mobility. 2014. vol. 3. no. 1. pp. 27–46.
21. *Котенко И.В., Степашкин М.В., Дойникова Е.В.* Анализ защищенности автоматизированных систем с учетом социо-инженерных атак // Проблемы информационной безопасности. Компьютерные системы. 2011. № 3. С. 40–57.
22. *Файзуллин Р. Р., Васильев В. И.* Метод оценки защищенности сети передачи данных в системе мониторинга и управления событиями информационной безопасности на основе нечеткой логики // Вестник УГАТУ. 2013. Том 17. № 2(55). С. 150–156.
23. *Karlzen H.* An Analysis of Security Information and Event Management: The Use of SIEMs for Log Collection, Management and Analysis // Department of Computer Science and Engineering, University of Gothenburg. 2009. 45 p.

References

1. *Kotenko I.V., Chechulin A.A.* A Cyber Attack Modeling and Impact Assessment Framework. Proceedings of 5th International Conference on Cyber Conflict 2013 (CyCon 2013). 2013. pp. 119–142.
2. *Kotenko I.V., Polubelova O.V., Saenko I.V.* The Ontological Approach for SIEM Data Repository Implementation. 2012 IEEE International Conference on Green Computing and Communications (GreenCom). IEEE Computer Society. 2012. pp. 761–766.
3. *Liu G., Mok A.K., Yang E.J.* Composite Events for Network Event Correlation. IEEE/IFIP International Symposium on Integrated Network Management. 1999. pp. 247–260.
4. *Kotenko I.V., Saenko I.V., Polubelova O.V., Chechulin A.A.* [Methods for security event and information management for computer networks protection]. *Problemy informacionnoj bezopasnosti. Komp'yuternye sistemy – Problems of information security. Computer systems.* 2012. vol. 3. pp. 57–68 (In Russ.).
5. *Kotenko I.V., Saenko I.V., Polubelova O.V., Chechulin A.A.* [Application of the methods of security event and information management for information protection in the critical infrastructures]. *Trudy SPIIRAN–SPIIRAS Proceedings.* 2012. vol. 1 (20). pp. 27–56. (In Russ.).
6. *Kotenko I.V., Saenko I.V., Chechulin A.A.* [Proactive management of security information and events in the information and telecommunication systems]. *Voprosy radioelektroniki –Questions of Radioelectronics.* 2014. vol. 3. no. 1. pp. 170–180 (In Russ.).
7. *Kruegel C., Valeur F., Vigna G.* Intrusion Detection and Correlation: Challenges and Solutions. University of California, Santa Barbara, USA: Springer. 2005. pp. 29–33.

8. Jakobson G., Weissman M.D. Alarm correlation. *IEEE Network*. 1993. vol. 7(6). pp. 52–59.
9. Zurutuza U., Uribeetxeberria R. Intrusion Detection Alarm Correlation: A Survey. Proceedings of IADAT International Conference on Telecommunications and computer Networks. 2004. pp. 1–3.
10. Sadoddin R., Ghorbani A. Alert Correlation Survey: Framework and Techniques. Proceedings of 2006 International Conference on Privacy, Security and Trust: Bridge the Gap Between PST Technologies and Business Services (PST'06). 2006. Article no. 37.
11. Dadkhah S., Shoja M.R.K., Taheri H. Alert Correlation through a Multi Components Architecture. *International Journal of Electrical and Computer Engineering (IJECE)*. 2013. vol. 3. no. 4. pp. 461–466.
12. Elshoush H.T., Osman I.M. Alert correlation in collaborative intelligent intrusion detection systems — A survey. *Applied Soft Computing*. 2011. pp. 4349–4365.
13. Ning P., Xu D. Correlation analysis of intrusion alerts. *Intrusion Detection Systems: series Advances in Information Security*. Springer. 2008. vol. 38. pp. 65–92.
14. Ahmadinejad S.H., Jalili S., Abadi M. A hybrid model for correlating alerts of known and unknown attack scenarios and updating attack graphs. *Computer Networks*. 2011. vol. 55. pp. 2221–2240.
15. Elshoushand H.T., Osman I.M. An improved framework for intrusion alert correlation. Proceedings of World Congress on Engineering 2012 (WCE 2012). 2012. vol. 1. pp. 518–524.
16. Limmer T., Dressler F. Survey of event correlation techniques for attack detection in early warning systems. Tech report. University of Erlangen. Dept. of Computer Science. 2008. 37 p.
17. Flegel U. Pseudonymizing Unix Log Files. *Infrastructure Security*. 2002. LNCS 2437. pp. 162–179.
18. Pang R., Paxson V. A high-level programming environment for packet trace anonymization and transformation. Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications. 2003. pp. 339–351.
19. Kotenko I.V., Chechulin A.A. Computer Attack Modeling and Security Evaluation based on Attack Graphs. Proceedings of 7th International Conference on “Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications” (IDAACS'2013). 2013. pp. 614–619.
20. Kotenko I.V., Chechulin A.A. Fast Network Attack Modeling and Security Evaluation based on Attack Graphs. *Journal of Cyber Security and Mobility*. 2014. vol. 3. no. 1. pp. 27–46.
21. Kotenko I.V., Stepashkin M.V., Dojnikova E.V. [Security analysis of information systems taking into account social engineering attacks]. *Problemy informacionnoy bezopasnosti Kompyuternye sistemy – Problems of information security. Computer systems*. 2011. vol. 3. pp. 40–57 (In Russ.).
22. Fajzullin R.R., Vasil'ev V.I. [Protectability assessment method of a data-transmission network in security information and event management system on a basis of fuzzy logic]. *Vestnik UGATU – Proceedings USATU*. 2013. vol. 17. no. 2 (55). pp. 150–156 (In Russ.).
23. Karlzen H. An Analysis of Security Information and Event Management: The Use of SIEMs for Log Collection, Management and Analysis. Department of Computer Science and Engineering, University of Gothenburg. 2009. 45 p.

Федорченко Андрей Владимирович — младший научный сотрудник лаборатории проблем компьютерной безопасности, Федеральное государственное бюджетное

учреждение науки Санкт-Петербургского института информатики и автоматизации Российской академии наук (СПИИРАН). Область научных интересов: безопасность компьютерных сетей, обнаружение вторжений, вредоносные программы. Число научных публикаций — 14. fedorchenko@comsec.spb.ru, <http://comsec.spb.ru/ru/staff/fedorchenko>; 14-я линия В.О., 39, ком. 205, Санкт-Петербург, 199178; р.т.: +7-(812)-328-71-81.

Fedorchenko Andrey Vladimirovich — junior researcher of computer security problems laboratory, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Science (SPIIRAS). Research interests: computer network security, intrusion detection, malware. The number of publications — 14. fedorchenkoandrei28@rambler.ru, <http://comsec.spb.ru/ru/staff/fedorchenko>; 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone: +7-(812)-328-71-81.

Левшун Дмитрий Сергеевич — программист лаборатории проблем компьютерной безопасности, Федеральное государственное бюджетное учреждение науки Санкт-Петербургского института информатики и автоматизации Российской академии наук (СПИИРАН). Область научных интересов: компьютерная безопасность, защита встроенных устройств, системы киберфизической безопасности, безопасность распределённых систем, корреляция событий безопасности. Число научных публикаций — 5. levshun@comsec.spb.ru, <http://comsec.spb.ru/levshun>; 14-я линия В.О., 39, ком. 205, Санкт-Петербург, 199178; р.т.: +7-(812)-328-71-81.

Levshun Dmitry Sergeevich — software developer of computer security problems laboratory, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Science (SPIIRAS). Research interests: distributed system security, embedded devices, event correlation, cyber-physical security systems. The number of publications — 5. levshun@comsec.spb.ru, <http://comsec.spb.ru/levshun>; 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone: +7-(812)-328-71-81.

Чечулин Андрей Алексеевич — к-т техн. наук, старший научный сотрудник лаборатории проблем компьютерной безопасности, Федеральное государственное бюджетное учреждение науки Санкт-Петербургского института информатики и автоматизации Российской академии наук (СПИИРАН). Область научных интересов: безопасность компьютерных сетей, обнаружение вторжений, анализ сетевого трафика, анализ уязвимостей. Число научных публикаций — 150. andreych@bk.ru, <http://comsec.spb.ru/ru/staff/chechulin>; 14-я линия В.О., 39, ком. 205, Санкт-Петербург, 199178; р.т.: +7-(812)-328-71-81.

Chechulin Andrey Alexeevich — Ph.D., senior researcher of computer security problems laboratory, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Science (SPIIRAS). Research interests: computer network security, intrusion detection, analysis of the network traffic, vulnerability analysis. The number of publications — 150. andreych@bk.ru, <http://comsec.spb.ru/ru/staff/chechulin>; 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone: +7-(812)-328-71-81.

Котенко Игорь Витальевич — д-р техн. наук, профессор, заведующий лабораторией проблем компьютерной безопасности, Федеральное государственное бюджетное учреждение науки Санкт-Петербургского института информатики и автоматизации Российской академии наук (СПИИРАН). Область научных интересов: безопасность компьютерных сетей, в том числе управление политиками безопасности, разграничение доступа, аутентификация, анализ защищенности, обнаружение компьютерных атак, межсетевые экраны, защита от вирусов и сетевых червей, анализ и верификация

протоколов безопасности и систем защиты информации, защита программного обеспечения от взлома и управление цифровыми правами, технологии моделирования и визуализации для противодействия кибер-терроризму. Число научных публикаций — 450. ivkote@comsec.spb.ru, <http://www.comsec.spb.ru>; 14-я линия В.О., 39, Санкт-Петербург, 199178; p.t.: +7-(812)-328-71-81, Факс: +7(812)328-4450.

Kotenko Igor Vitalievich — Ph.D., Dr. Sci., professor, head of computer security problems Laboratory, St. Petersburg Institute for Informatics and Automation of Russian Academy of Sciences (SPIIRAS). Research interests: computer network security, including security policy management, access control, authentication, network security analysis, intrusion detection, firewalls, deception systems, malware protection, verification of security systems, digital right management, modeling, simulation and visualization technologies for counteraction to cyber terrorism. The number of publications — 450. ivkote@comsec.spb.ru, <http://www.comsec.spb.ru>; 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone: +7-(812)-328-71-81, Fax: +7(812)328-4450.

Поддержка исследований. Работа выполнена при финансовой поддержке РФФИ (проекты №14-07-00697, 14-07-00417, 15-07-07451, 16-37-00338, 16-29-09482 офи_м), при частичной поддержке бюджетных тем № 0073-2015-0004 и 0073-2015-0007, а также гранта РФФИ 15-11-30029 в СПИИРАН

Acknowledgements. This research is supported by RFBR (projects No. 14-07-00697, 14-07-00417, 15-07-07451, 16-37-00338, 16-29-09482 офи_м), in part by the budget (projects No. 0073-2015-0004 and 0073-2015-0007) and by the grant of RSF 15-11-30029 in SPIIRAS

РЕФЕРАТ

Федорченко А.В., Левшун Д.С., Чечулин А.А., Котенко И.В. **Анализ методов корреляции событий безопасности в SIEM-системах. Часть 1.**

Данная статья посвящена анализу различных схем, реализующих процесс корреляции событий безопасности в SIEM-системах. Исследуемая предметная область изучается в мировом сообществе на протяжении более чем двух десятилетий и в данном исследовании были учтены многие устоявшиеся аспекты корреляции. Рассмотрен ряд работ, направленных преимущественно на модульное представление процесса корреляции и применение различных подходов на его отдельных этапах.

В ходе проделанного анализа была получена и обоснована собственная схема процесса корреляции, состоящая из нескольких уровней и отдельных этапов на каждом из них, а также из установленных связей как между уровнями, так и между этапами. В работе подробно описывается каждый из этапов процесса корреляции, а также определенные для него функции.

SUMMARY

Fedorchenko A.V., Levshun D.S., Chechulin A.A., Kotenko I.V. **An Analysis of Security Event Correlation Techniques in Siem-Systems. Part 1.**

This paper provides an analysis of various schemes that can be applied for the security events correlation in SIEM systems. This field has been studied in the international community for over two decades, so in this paper we take into account established aspects of correlation. The paper contains a description of several techniques of information correlation that use a modular representation of correlation process and apply different approaches at separate stages of correlation.

During the investigation of correlation techniques the new scheme of correlation process was obtained and proved. This scheme contains several levels which, in their turn, consist of independent elements. This scheme also contains links between levels and between elements. In addition to the general scheme, the paper provides a description of each correlation stage and its specific functions.

А.А. ГАВРИШЕВ, А.П. ЖУК, Д.Л. ОСИПОВ
**АНАЛИЗ ТЕХНОЛОГИЙ ЗАЩИТЫ РАДИОКАНАЛА
ОХРАННО-ПОЖАРНЫХ СИГНАЛИЗАЦИЙ ОТ
НЕСАНКЦИОНИРОВАННОГО ДОСТУПА**

Гавришев А.А., Жук А.П., Осипов Д.Л. Анализ технологий защиты радиоканала охранно-пожарных сигнализаций от несанкционированного доступа.

Аннотация. В настоящее время наблюдается рост радиоканальных охранно-пожарных сигнализаций. Одним из основных недостатков является уязвимость канала связи к несанкционированному доступу. Авторами статьи на основе проведенного анализа и аппарата нечеткой логики определены наиболее защищенные беспроводные технологии охранно-пожарных сигнализаций с получением количественной оценки защищенности. К наиболее защищенным технологиям от комплексных угроз (просмотр, подмена, перехват, радиоэлектронное подавление) относятся технологии на основе шумоподобных сигналов (технологии на основе хаотических последовательностей и сверхширокополосных сигналов), а к наименее защищенным от данных угроз относятся криптографические методы (устройство имитозащиты). Показана необходимость дальнейших исследований, направленных на повышение защищенности радиоканала охранно-пожарных сигнализаций. Перспективным видится развитие систем с технологиями защиты радиоканала на основе шумоподобных сигналов (технологии на основе хаотических последовательностей и сверхширокополосных сигналов).

Ключевые слова: радиоканал, сигнализация, оценка защищенности, нечеткая логика.

Gavrishev A.A., Zhuk A.P., Osipov D.L. An Analysis of Technologies to Protect a Radio Channel of Fire Alarm Systems against Unauthorized Access.

Abstract. There is currently a growth of radio fire alarm systems. One of the main disadvantages of such systems is the vulnerability of the communication channel to unauthorized access. Based on the conducted analysis and the apparatus of fuzzy logic, the authors identified the most tamper-proof wireless technologies for fire alarm systems to give quantitative assessment of the security level. The most secure technologies to combat complex threats (view, substitution, interception, jamming) include technologies based on noise-like signals (technologies based on random sequences and ultra-wideband signals), while the least secure technologies include cryptographic techniques (simulation protection device). The necessity of further research aimed at enhancing security of the radio channel of fire alarm systems is shown. The development of systems with protection technologies based on noise-like signals is seen as promising.

Keywords: radio channel, alarm, security assessment, fuzzy logic.

1. Введение. В последние годы наблюдается бурный рост радиоканальных технических систем охраны, к которым относится беспроводная охранно-пожарная сигнализация (ОПС). Доля радиоканальных ОПС на рынке составляет сегодня от 15% до 25%, несмотря на то, что сегмент проводных систем имеет несколько десятилетий форы в своем развитии. Привлекательность радиоканального оборудования высокая, так как конкуренция здесь ниже, чем в сегменте проводных систем охраны, а рентабельность выше и находится на уровне 25–30%. Также беспроводные системы имеют такое неоспоримое преимущество, как сокращение времени монтажа [1].

Однако у радиоканальных ОПС есть существенный недостаток, который заключается в уязвимости беспроводного канала связи к несанкционированному доступу (НСД) со стороны третьих лиц. Данное обстоятельство может негативно сказаться на их работе и снизить уровень обеспечения безопасности, что является недопустимым фактором.

Целью данной статьи является сравнительный анализ количественных оценок защищенности беспроводных технологий ОПС от НСД на основе применения аппарата нечеткой логики.

2. Анализ технологий защиты радиоканала охранно-пожарных сигнализаций и оценка их защищенности. Согласно [2], одним из основных недостатков беспроводных ОПС является уязвимость канала связи к НСД. Как следствие, они подвержены угрозам НСД со стороны третьих лиц. Такими угрозами являются [3]: перехват, просмотр, подмена, радиоэлектронное подавление.

В настоящее время основными методами противодействия данным угрозам в беспроводных ОПС являются [4]:

- применение криптографических методов защиты;
- применение шумоподобных сигналов.

На основе данной классификации методов противодействия проведем первоначальный анализ технологий защиты радиоканала ОПС от НСД. Все способы (устройства) защиты радиоканала ОПС для краткости будут обозначаться литерой «Т» со сквозной нумерацией. В данной работе не рассматриваются беспроводные ОПС, в которых в качестве радиоканала используются мобильные сети.

Проанализируем криптографические методы защиты (КМЗ).

В работе [5] описывается беспроводная охранно-пожарная сигнализация на базе технологии IEEE 802.15.4, отличающаяся защищенностью трафика алгоритмом шифрования AES-128 (далее «Т1»). Следует отметить, что стандарт IEEE 802.15.4 обладает помехоустойчивостью по отношению к непреднамеренным помехам [6], при этом в [7] отмечено, что помехоустойчивость этого стандарта к преднамеренным помехам является невысокой.

В работе [8] описывается способ и устройство передачи извещений для централизованной охраны, который включает в себя использование имитовставки и скремблирования передаваемого сообщения (далее «Т2»).

В последнее время появился интерес к охранным роботам, которые могут патрулировать территории и помещения. Для этих целей они могут быть снабжены видеокамерами или датчикам, информация с которых поступает на центральный пульт управления по радиоканалу. В работе [9] предлагается использовать метод преобразования инфор-

мационных потоков, основанный на матричном умножении в поле GF(2) для защиты радиоканала охранного робототехнического комплекса (далее «ТЗ»).

Частным случаем применения КМЗ в радиоканале является имитозащита. Так, в патентах, основным прототипом которых является [10], предлагается метод защиты линии связи и оконечных датчиков систем охраны от навязывания ложных данных (имитозащищенности) в виде устройства имитозащиты, основанный на системе «свой–чужой» за счет использования в блоке контроля инициализирующего генератора первой псевдослучайной последовательности (ПСП-1), генератора второй ПСП (ПСП-2) и в каждом из датчиков генераторов ПСП-2, функции генерации которых идентичны функции генератора ПСП-2 блока контроля (далее «Т4»). Недостатком данного подхода является то обстоятельство [11], что сигналы, представленные в данном устройстве ПСП, циркулируют по открытым линиям связи без защиты от НСД, и, как следствие, они могут быть перехвачены, подавлены помехами или может быть раскрыта их структура.

Таким образом, КМЗ защищают радиоканал от подмены и просмотра трафика, однако уязвимы для перехвата и подавления помехами. Кроме того, данные алгоритмы обладают продолжительным временем выполнения команд шифрования-расшифрования, что сказывается на производительности беспроводных ОПС.

В настоящее время большое развитие получают идеи использования для защиты радиоканала шумоподобных сигналов (ШПС).

Анализ методов защиты радиоканала ОПС на основе ШПС, показал, что их целесообразно разделить на следующие группы:

- технологии защиты радиоканала ОПС на основе передачи сигналов на частотно-временных позициях;
- технологии защиты радиоканала ОПС на основе псевдослучайной перестройки рабочей частоты;
- технологии защиты радиоканала ОПС на основе фазоманипулированных сигналов;
- технологии защиты радиоканала ОПС на основе сверхширокополосных сигналов;
- технология защиты радиоканала ОПС на основе хаотических последовательностей.

Рассмотрим технологии защиты радиоканала ОПС на основе передачи сигналов на частотно-временных позициях. В работе [4] отмечается, что в настоящее время в радиоканальных охранно-пожарных системах используется передача сигналов на случайных частотно-временных позициях (ЧВП). В системах с ЧВП для обеспечения не-

предсказуемости случайных позиций используются генераторы случайных чисел (ГСЧ). По такому принципу реализовано устройство, предложенное в работе [12] (далее «Т5»). К его недостатку можно отнести требование дополнительного частотно-временного ресурса [13]. Есть и другие способы выбора позиций в ЧВП. Так в работе [14] ЧВП выбираются с помощью индивидуального уникального ключа, который присвоен каждому охраняемому объекту (далее «Т6»). Одним из недостатков данного устройства является затруднительность смены уникального ключа в случае компрометации охраняемого объекта. В работе [15] описывается метод передачи сообщений с помощью скачкообразной передачи частоты на центральное приемное устройство, которое хранит указатели на будущие частоты и время для каждого датчика, динамически обновляемые, а также ID каждого из датчиков (далее «Т7»). Основным недостатком данного устройства является малое допустимое количество частотно-временных позиций [13]. В работе [16] описывается способ радиосвязи между охраняемыми объектами и пунктом централизованной охраны (далее «Т8»). Для борьбы с НСД используются «прыгающие» частоты, которые передаются в псевдослучайное время и на произвольной несущей частоте.

Технология защиты на основе передачи сигналов на ЧВП обеспечивает защиту радиоканала от НСД за счет непредсказуемости позиции в каждый момент времени. Однако общим недостатком подобных систем может стать недостаточная криптостойкость ГСЧ и ПСП в силу ограниченности периода генерации случайных чисел.

Рассмотрим технологии защиты радиоканала ОПС на основе псевдослучайной перестройки рабочей частоты (ППРЧ). В работе [17] описывается радиоканальный комплекс охраны (далее «Т9»). Радиопередатчики выполнены по технологии «прыгающих частот», в соответствии с которой каждый выход в эфир передатчика возможен на одной из 1024 заранее запрограммированных частот связи. Каждый передатчик имеет свой псевдослучайный алгоритм скачков частоты. Кроме того, в данном комплексе применяется помехоустойчивое кодирование, исправляющее большинство ошибок. В работе [18] описывается система тревожной сигнализации, обладающая защищенностью от преднамеренных помех, за счет использования «прыгающих» частот, которые, за счет псевдослучайного алгоритма, выбираются случайно (далее «Т10»). В работе [19] изменение частоты происходит по определенному псевдослучайному закону, например, с помощью ГСЧ, что обеспечивает плавность, а не скачкообразность (далее «Т11»).

Рассмотрим технологии защиты радиоканала ОПС на основе фазоманипулированных сигналов (ФМС). В работе [20] предлагается

устройство охранной сигнализации, в котором в качестве сигналов используются псевдослучайные ФМС (далее «Т12»). Данное устройство отличается сложностью конструкции, ограниченным ансамблем шумоподобных сигналов в силу ограниченности периода генерации ПСП. В работе [21] предлагается устройство охранной сигнализации, предназначенное для оповещения об изменении местоположения контролируемого объекта (далее «Т13»). В качестве сигналов используются так же ФМС. Данное устройство отличается низкой помехоустойчивостью.

Обратимся к технологии защиты радиоканала ОПС на основе сверхширокополосных сигналов (СШПС). В работе [22] рассматривается способ и устройство скрытой передачи извещений охранной сигнализации (далее «Т14»). В качестве сигналов используются бинарные СШПС с низкой спектральной плотностью. Основным недостатком данного устройства является возможность постановки помех злоумышленником [23]. В работе [24] описывается система охранной радиосвязи с СШПС на ЧВП, защищенная от средств радиоэлектронной борьбы криптостойкой расстановкой пакетов (далее «Т15»). Однако в [25] отмечается, что данная реализация технологии с СШПС не является совершенной.

Рассмотрим технологию защиты радиоканала ОПС на основе хаотических последовательностей. В работе [26] предлагается альтернативный подход для защиты радиоканала, заключающийся в использовании в блоке контроля и оконечном датчике перезаписываемых накопителей хаотических последовательностей (ХП). Данный подход обладает увеличенной защищенностью радиоканала от встраивания и перехвата информации за счет использования перезаписываемых накопителей ХП, позволяющих уменьшить вероятность перехвата и подмены ПСП, циркулирующих между датчиком и блоком контроля [11] (далее «Т16»). Кроме того, в данном подходе реализована возможность проверки имитозащищенности оконечного датчика. Одним из основных недостатков данного устройства является необходимость периодической перезаписи ХП в накопителях для повышения защищенности радиоканала. Следует заметить, что в литературе встречаются исследования по защищенности систем с хаотическими сигналами, доказывающие их повышенную защищенность от НСД. Так в работах [27, 28] обосновывается значительное преимущество перед другими видами ШПС по показателю структурной скрытности. Работы [29, 30] обосновывают значительную защищенность данных систем от подмены, перехвата и подавления помехами.

Таким образом, технологии с ШПС обеспечивают комплексную защиту от основных видов НСД (просмотр, подмена, перехват и по-

давление помехами). К общим недостаткам большинства рассмотренных систем можно отнести сложность приемо-передающей аппаратуры и ограниченный ансамбль используемых шумоподобных сигналов.

2.1. Анализ технологий защиты радиоканала, применяемых в охранно-пожарных сигнализациях, с позиции угроз несанкционированного доступа. КМЗ обеспечивают защиту только от просмотра трафика и подмены. Однако они недостаточно эффективны при перехвате и подавлении трафика помехами. Использование ШПС обеспечивает защиту от просмотра, подмены, перехвата и подавления трафика помехами. Тем не менее представленные выше технологии использования ШПС не являются в равной степени эффективными против угроз информационной безопасности. Краткий анализ методов НСД к системам с ШПС (технологии защиты радиоканала на основе ЧВП, ППРЧ, ФМС, СШПС, ХП) проведен авторами в работе [31].

Таким образом, использование ШПС позволяет в значительной мере повысить защищенность беспроводных ОПС по сравнению с КМЗ. Однако сами ШПС должны отвечать следующим критериям [32]: обладать максимальным уровнем априорной неопределенности используемых параметров сигналов и обладать большим числом сигнальных конструкций.

2.2. Количественная оценка защищенности беспроводных охранно-пожарных сигнализаций. Эффективность применения того или иного способа защиты радиоканала от НСД должна определяться с помощью адекватной методики, позволяющей выяснить уровень защищенности сигнализации с помощью количественной и качественной оценки. В настоящее время известно множество различных методик оценки защищенности. В работах [33-36] проводится анализ методик оценок защищенности. Некоторые из них приведены в таблице 1. Обобщенные выводы по ним показывают, что в настоящее время не существует совершенных методик оценки защищенности беспроводных ОПС от НСД. Кроме недостатков, присущих каждой из методик в отдельности (достаточно громоздкий математический аппарат, отсутствие правила перевода количественных показателей в качественные, невозможность применения к беспроводным ОПС), у них существует общая отрицательная черта — оценка защищенности зачастую не несет комплексного характера и направлена на одну или несколько угроз в области информационной безопасности (ИБ), в то время как злоумышленники применяют весь арсенал противоправных действий для достижения своих целей. Еще одним недостатком (на этот раз присущим зарубежным методам оценки защищенности) является наличие технологий «двойного назначения», что может повлечь за собой потенциальные трудности с их использованием на территории России.

Таблица 1. Методики оценок защищенности

Авторы	Область применения	Используемый математический аппарат	Полученный результат
Авраменко В.С., Козленко А.В. (2010)	Автоматизированные системы	Коэффициент готовности из теории надежности	Модель количественной оценки
Бондарь И.В., Золотарев В.В., Попов А.М. (2010)	Информационные системы	Математическая статистика и теория графов	Методика оценки защищенности на основе международных стандартов
Политов М.С., Мельников А.В. (2008)	Информационные системы	Экстраполяционное вероятностное прогнозирование	Оценка защищенности с повышенной достоверностью
Karabacak B., Sogukpinar I. (2004)	Информационные системы	Теория вероятности	Количественная оценка на основе опросной модели
Fu S., Liu Z., Zhou H., Liu W., Li B. (2015)	Информационные системы	Теория информации	Количественная и качественная оценка рисков в сфере ИБ
Goel S., Chen V. (2005)	Информационные системы	Матричные вычисления	Количественная и качественная методика оценки рисков ИБ
Боговик А.В., Игнатов В.В. (2006)	Радиотехнические системы специального назначения	Теория вероятности и теория надежности	Методики оценок защищенности систем связи, функционирующих в экстремальных условиях
Литвиненко В.П. (2009)	Радиотехнические системы	Теория вероятности	Оценки различных видов скрытности (энергетическая, структурная)
Бабкин А.Н., Эсауленко А.В. (2012)	Радиотехнические системы	Теория вероятности	Методика определения параметров радиоканала, обеспечения выполнения заданного критерия эффективности
Щербаков В.Б., Ермаков С.А. (2010)	Беспроводная сеть стандарта IEEE 802.11	Нечеткая логика	Методика оценки рисков ИБ

По мнению авторов, в настоящее время оценку защищенности беспроводных ОПС от НСД целесообразно проводить на основе понятий нечеткой логики, позволяющих наглядно и просто представить количественную и качественную оценки защищенности [37, 38].

Для этого в работе [33] была предложена методика оценки защищенности беспроводных сигнализаций, основанная на аппарате нечеткой логики (рисунок 1). Данная методика позволяет получить оценку защищенности беспроводных сигнализаций с учетом обобщенного *At*-уровня атаки на них и с учетом обобщенного *P*-уровня защиты от этих атак. Проведем количественную оценку защищенности рассмотренных выше беспроводных систем ОПС на основе данной методики.

Введем следующие обозначения: «ОН» — очень низкий, «Н» — низкий, «С» — средний, «В» — высокий, «ОВ» — очень высокий. В таблице 2 приведены нечеткие значения переменных и соответствующие им численные значения. Следует заметить, что «очень низкому» уровню атаки соответствует численное значение «1», а «очень низкому» уровню защиты — численное значение «5».

Таблица 2. Соответствие нечетких переменных числовым значениям

Численное значение	Нечеткий параметр	
	<i>P</i> -уровень защиты	<i>At</i> -уровень атаки
1	ОВ	ОН
2	В	Н
3	С	С
4	Н	В
5	ОН	ОВ

Сначала проведем вычисления для беспроводных ОПС, использующих КМЗ (таблица 3). Для краткости условимся обозначать угрозу «просмотр» как «У1», угрозу «подмена» как «У2», угрозу «перехват» как «У3», угрозу «подавление помехами» как «У4».

Таблица 3. Уровни защиты и уровни атаки КМЗ

Угрозы	<i>P</i> -уровень защиты				<i>At</i> -уровень атаки			
	<i>T1</i>	<i>T2</i>	<i>T3</i>	<i>T4</i>	<i>T1</i>	<i>T2</i>	<i>T3</i>	<i>T4</i>
У1	3	3	2	5	4	4	4	5
У2	3	2	2	2	4	4	4	4
У3	4	5	5	5	5	5	5	5
У4	4	5	5	5	5	5	5	5

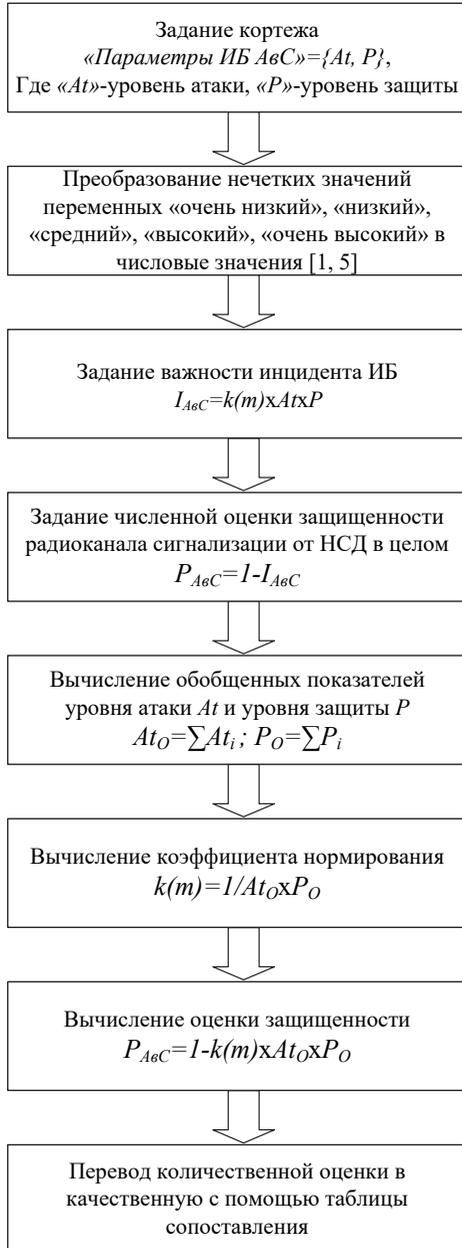


Рис. 1. Этапы методики оценки защищенности

В качестве наглядного примера проведем расчет для устройства «Т4». На основе таблицы 3 и 5 этапа методики оценки защищенности (рисунок 1) вычислим обобщенные показатели уровня защиты и уровня атаки, которые равны $P_o=17$ и $At_o=19$. При этом нормирующий коэффициент $k(m)$, при максимальных значениях $P_o=20$ и $At_o=20$, $k(m)=0,0025$. Следует заметить, что вычисление для других устройств (способов) защиты радиоканала ОПС проводятся аналогично. Окончательные расчеты оценок защищенности будут приведены ниже.

Проведем вычисления для беспроводных ОПС, использующих технологии защиты радиоканала на основе ШПС.

Рассмотрим технологии защиты радиоканала ОПС на основе ЧВП (таблица 4).

Таблица 4. Уровни защиты и уровни атаки технологий на основе ЧВП

Угрозы	P-уровень защиты				At-уровень атаки			
	T5	T6	T7	T8	T5	T6	T7	T8
У1	2	3	3	3	4	4	5	5
У2	2	4	2	3	4	5	4	4
У3	3	3	3	3	4	4	4	4
У4	3	3	3	3	4	5	4	4

Рассмотрим технологию защиты радиоканала ОПС на основе ШПРЧ (таблица 5).

Таблица 5. Уровни защиты и уровни атаки технологий на основе ШПРЧ

Угрозы	P-уровень защиты			At-уровень атаки		
	T9	T10	T11	T9	T10	T11
У1	3	3	3	4	4	4
У2	3	3	3	4	4	4
У3	3	4	3	5	5	5
У4	4	4	3	5	5	5

Рассмотрим технологии защиты радиоканала ОПС на основе ФМС (таблица 6).

Таблица 6. Уровни защиты и уровни атаки технологий на основе ФМС

Угрозы	P-уровень защиты		At-уровень атаки	
	T12	T13	T12	T13
У1	3	3	4	4
У2	2	2	4	4
У3	3	3	5	5
У4	3	4	5	5

Рассмотрим технологии защиты радиоканала ОПС на основе СШПС (таблица 7).

Таблица 7. Уровни защиты и уровни атаки технологий на основе СШПС

Угрозы	<i>P</i> -уровень защиты		<i>At</i> -уровень атаки	
	<i>T14</i>	<i>T15</i>	<i>T14</i>	<i>T15</i>
У1	2	2	4	4
У2	2	2	4	4
У3	2	2	4	4
У4	3	3	5	4

Рассмотрим технологию защиты радиоканала ОПС на основе ХП (таблица 8).

Таблица 8. Уровни защиты и уровни атаки технологии на основе ХП

Угрозы	<i>P</i> -уровень защиты	<i>At</i> -уровень атаки
	<i>T16</i>	<i>T16</i>
У1	2	4
У2	2	4
У3	2	4
У4	2	4

Далее проведем количественную оценку защищенности рассмотренных технологий защиты радиоканала ОПС. Количественная оценка защищенности вычисляется по следующей формуле [33]:

$$P_{\text{авс}} = 1 - k(m) \times At_o \times P_o. \quad (1)$$

Формула (1) представлена на рисунке 1 на предпоследнем этапе методики. Стоит заметить, что количественная оценка находится в диапазоне [0; 1]. Далее количественная оценка переводится в качественную оценку (таблица 9) [33].

Таблица 9. Сопоставление количественных и качественных оценок защищенности

Значение количественной оценки защищенности	Значение качественной оценки защищенности
$0 \leq P_{\text{авс}} < 0,2$	Очень низкая
$0,2 \leq P_{\text{авс}} < 0,4$	Низкая
$0,4 \leq P_{\text{авс}} < 0,6$	Средняя
$0,6 \leq P_{\text{авс}} < 0,8$	Высокая
$0,8 \leq P_{\text{авс}} < 1$	Очень высокая

Составим на основе количественных оценок защищенности ранжированный список технологий защиты радиоканала ОПС (таблица 10).

Таблица 10. Оценка технологий защиты радиоканала ОПС

№	Устройство (способ)	Метод защиты радиоканала	Количественная оценка защищенности	Качественная оценка защищенности
1	T16	ХП	0,6800	Высокая
2	T15	СШПС	0,6400	Высокая
3	T14	СШПС	0,6175	Высокая
4	T5	ЧВП	0,6000	Высокая
5	T7	ЧВП	0,5325	Средняя
6	T12	ФМС	0,5050	Средняя
7	T8	ЧВП	0,4900	Средняя
8	T11	ППРЧ	0,4600	Средняя
9	T13	ФМС	0,4600	Средняя
10	T6	ЧВП	0,4150	Средняя
11	T9	ППРЧ	0,4150	Средняя
12	T10	ППРЧ	0,3700	Низкая
13	T1	КМЗ	0,3700	Низкая
14	T3	КМЗ	0,3700	Низкая
15	T2	КМЗ	0,3250	Низкая
16	T4	КМЗ	0,1925	Очень низкая

3. Заключение. Таким образом, в данной работе был проведен анализ технологий защиты радиоканала ОПС от НСД, которыми в настоящее время являются КМЗ и ШПС. КМЗ, хоть и защищают радиоканал от просмотра и подмены, однако малоэффективны против комплексных угроз (перехват, просмотр, подмена и радиоэлектронное подавление). Использование ШПС (технологии защиты радиоканала на основе ЧВП, ППРЧ, ФМС, СШПС, ХП) обеспечивает защиту радиоканала от перечисленного комплекса угроз.

Сравнительный анализ количественных оценок защищенности беспроводных охранно-пожарных сигнализаций, выполненный на основе аппарата нечеткой логики, показал, что наиболее защищенными от комплексных угроз являются системы с технологиями защиты радиоканала ОПС на основе ШПС (технологии на основе ХП («Т16») и СШПС («Т15»)), которые имеют количественные показатели 0,6800 и 0,6400, а менее всего защищенными — технология с КМЗ («Т4»), которая имеет количественный показатель 0,1925.

Следует заметить, что не одна из количественных оценок защищенности устройств (способов) защиты радиоканала ОПС не близка к нижней границе значения количественной оценки защищенности,

$R_{авс}=0,8$ и соответствующей «очень высокому» значению качественной оценки защищенности (таблица 9).

Таким образом, необходимы дальнейшие исследования, направленные на разработку новых устройств (способов) повышения защищенности радиоканала ОПС. Перспективным видится развития систем с технологиями защиты радиоканала на основе ШПС (технологии на основе ХП и СШПС).

Литература

1. Радиоканальные системы охраны. URL: <http://www.aktivsb.ru/info860.html> (дата обращения: 01.06.2015).
2. *Цыбенко Л.В.* Анализ устройств радиоохранной сигнализации // Омский научный вестник. 2007. № 1(52). С. 94–96.
3. *Моисеев В.С., Козар А.Н., Дятчин В.В.* Информационная безопасность автоматизированных систем специального назначения // Казань: Отечество. 2006. 384 с.
4. *Брауде-Золотарев Ю.* Алгоритмы безопасности радиоканалов // Алгоритм безопасности. 2013. № 1. С. 64–66.
5. *Антонюк О.И., Лисенко О.М., Розумный В.Г., Туру А.Г.* Система пожечно-охранной сигнализации: корисна модель 10842 // Патент Украины № 200507552. 2005. 6 с.
6. *Варгаузин В.А.* Радиосети для сбора данных от сенсоров, мониторинга и управления на основе стандарта IEEE 802.15.4 // ТелеМультиМедия. 2005. № 6. С. 23–27.
7. *Пименов П.Н., Мырова О.Л.* Эффективность воздействия сверхкороткого электромагнитного импульса на широкополосные системы радиосвязи // Технологии ЭМС. 2015. № 1(52). С. 17–21.
8. *Рунов Ю.А.* Способ передачи извещений для систем централизованной охраны // Евразийский патент № 019227. 2014. 5 с.
9. *Скуратов В.В.* Матричное умножение над полем GF(2) в защите беспроводных каналов систем управления робототехническим комплексом // Информационно-управляющие системы. 2013. № 4. С. 88–90.
10. *Лепешкин О.М. и др.* Устройство для имитозащиты контролируемых объектов // Патент РФ № 2310236. 2007. 9 с.
11. *Жук А.П., Гавришев А.А.* Альтернативный подход повышения структурной скрытности сигналов-переносчиков устройства имитозащиты контролируемых объектов // Спецтехника и связь. 2015. № 2. С. 59–63.
12. *Косарев С.А., Брауде-Золотарев Ю.М.* Способ радиосвязи охраняемых объектов и центра охраны // Патент РФ № 2295778. 2007. 11 с.
13. *Василевский В.В., Завьялов С.А.* Способ передачи извещений для систем централизованной охраны // Омский научный вестник. 2010. № 2. С. 203–206.
14. *Молдаванов А.В., Максимов О.Н., Тимчук А.А., Юшин А.И.* Способ контроля состояния охраняемого объекта // Патент РФ № 2221279. 2004. 6 с.
15. *Partyka A.* Transmission of urgent messages in frequency hopping system for intermittent transmission // Patent US 6870875. 2005. 22 p.
16. *Грибок В.П., Косарев С.А., Райгородский Ю.В., Шентовецкий А.Ю.* Способ радиосвязи между охраняемыми объектами и пунктом централизованной охраны // Патент РФ № 2351066. 2009. 23 с.
17. *Федяев С.Л., Максимов В.С., Федяев Ю.С.* Радиоканальный сигнализационный комплекс охраны // Патент РФ № 111938. 2011. 5 с.
18. *Герасимчук А.Н., Косарев С.А.* Система тревожной сигнализации для обслуживания компактной группы объектов недвижимости // Патент РФ № 95882. 2010. 11 с.
19. *Василевский В.В., Завьялов С.А.* Способ передачи извещений для систем централизованной охраны // Патент РФ № 2371775. 2009. 13 с.

20. *Алиев Э.А., Магомедов Д.А., Карагишиев У.Д.* Радиосистема охраны на шумоподобных сигналах // Патент РФ № 2103742. 1998. 5 с.
21. *Леньшин В.П., Рихтер С.Г.* Устройство охранной сигнализации // Патент РФ № 2234135. 2004. 4 с.
22. *Гарбацевич В.А., Конейкин В.В.* Способ дистанционной передачи сообщения и устройства для его осуществления // Патент РФ № 2081456. 1997. 5 с.
23. *Шилов А.В.* Способ передачи сообщения о срабатывании охранной сигнализации // Патент РФ № 2199151. 2003. 4 с.
24. *Руднев А.Н., Брауде-Золотарев Ю.М., Давыдов Ю.Л., Косарев С.А.* Система радиосвязи технических средств охраны // Доклады VI МНТК «Перспективные технологии в средствах передачи информации». 2005. С. 57–59.
25. *Брауде-Золотарев Ю.* О выборе наилучших сверхширокополосных сигналов // Технологии и средства связи. 2014. № 1. С. 54–57.
26. *Осипов Д.Л., Жук А.П., Гавришев А.А.* Устройство имитозащиты контролируемых объектов с повышенной структурной скрытностью сигналов-переносчиков // Патент РФ № 2560824. 2015. 15 с.
27. *Корчинский В.В.* Оценка структурной скрытности сигнальных конструкций на основе хаотических сигналов в системах передачи конфиденциальной информации // Наукові праці ОНАЗ ім. О.С. Попова. 2012. № 1. С. 77–81.
28. *Сиващенко С.И.* Скрытность радиосистем со сложными и хаотическими сигналами // Системи управління, навігації та зв'язку. 2009. № 3(11). С. 56–58.
29. *Жук А.П., Баркетов С.В., Сазонов В.В.* Вариант помехоустойчивой хаотической системы передачи информации // Информационное противодействие угрозам терроризма. 2010. № 15. С. 126–130.
30. *Баркетов С.В. и др.* Когерентная система передачи информации хаотическими сигналами // Патент РФ № 2326500. 2008. 6 с.
31. *Гавришев А.А.* К вопросу о несанкционированном доступе к беспроводным системам связи на основе шумоподобных сигналов // Сборник тезисов международной научно-практической конференции «Пожаротушение: проблемы, технологии, инновации». 2016. С. 218–220.
32. *Горохов С.М., Захарченко Н.В., Корчинский В.В.* Критерии эффективности скрытых методов передачи // Цифрові технології. 2012. № 12. С. 147–150.
33. *Гавришев А.А., Бурмистров В.А., Осипов Д.Л.* Оценка защищенности беспроводной сигнализации от несанкционированного доступа на основе понятий нечеткой логики // Прикладная информатика. 2015. Вып. 10. № 4(58). С. 62–69.
34. *Жук А.П., Осипов Д.Л., Гавришев А.А.* Анализ методов оценки защищенности беспроводной сигнализации // Сборник трудов III Международной научно-практической конференции «Информационная безопасность в свете Стратегии Казахстан-2050». 2015. С. 139–144.
35. *Бабкин А.Н., Эсауленко А.В.* Эффективность функционирования радиоканала в системах безопасности // Вестник Воронежского института МВД России. 2012. № 4. С. 90–91.
36. *Щербаков В.Б., Ермаков С.А.* Безопасность беспроводных сетей: стандарт IEEE 802.11 // М.: РадиоСофт. 2010. 255 с.
37. *Заде Л.* Понятие лингвистической переменной и его применение к принятию приближенных решений // М.: Мир. 1976. 163 с.
38. *Файзуллин Р.Р., Васильев В.И.* Метод оценки защищенности сети передачи данных в системе мониторинга и управления событиями информационной безопасности на основе нечеткой логики // Вестник УГАТУ. 2013. Вып. 17. № 2(55). С. 150–156.

References

1. Radiokanal'nye sistemy okhrany [Radiochannel protection systems]. Available at: <http://www.aktivs.ru/info860.html> (accessed: 01.06.2015). (In Russ.)

2. Tsybenko L.V. [The analysis of devices of radio security alarm system]. *Omskii nauchnyi vestnik – Omsk Scientific Bulletin*. 2007. vol. 1(52). pp. 94–96. (In Russ.).
3. Moiseev V.S., Kozar A.N., Dyatchin V.V. *Informatsionnaya bezopasnost' avtomatizirovannykh sistem spetsial'nogo naznacheniya*. [Information security of the automated systems of special purpose]. Kazan: Otechestvo Publ. 2006. 384 p. (In Russ.).
4. Braude-Zolotarev Yu. [Safety radio's algorithms]. *Algoritm bezopasnosti – Safety algorithm*. 2013. vol. 1. pp. 64–66 (In Russ.).
5. Antonjuk O.I., Lysenko O.M., Rozumnyj V.G., Turu A.G. [The system of fire alarm]. Patent Ua. no. 10842. 2005. 6 p. (In Ukr.).
6. Vargauzin V.A. [Radio network to collect data from the sensors, monitoring and control based on the IEEE 802.15.4 standard]. *TeleMul'tiMediya – TeleMultiMedia*. 2005. vol. 6. pp. 23–27 (In Russ.).
7. Pimenov P.N., Myrova O.L. [The impact of ultra short electromagnetic pulses for reducing the quality of service broadband communication systems]. *Tekhnologii elektromagnitnoi sovместимости – Technology of electromagnetic compatibility*. 2015. vol. 1(52). pp. 17–21 (In Russ.).
8. Runov Yu.A. *Sposob peredachi izvешhenij dlja sistem centralizovannoj ohrany* [A method of transmitting notices to central security systems]. Eurasian patent. no. 019227. 2014. 5 p. (In Russ.).
9. Skuratov V.V. [Matrix multiplication above a GF(2) field to protect wireless channels of robotic complex control]. *Informatsionno-upravliaiushchie sistemy – Information and Control Systems*. 2013. vol. 4. pp. 88–90 (In Russ.).
10. Lepeshkin O.M. et al. *Ustrojstvo dlja imitozashhity kontroliruemykh ob'ekto* [Device for imitation protection of objects being monitored]. Patent RF. no. 2310236. 2007. 9 p. (In Russ.).
11. Zhuk A.P., Gavrishiev A.A. [Alternative approach of increased structural stealth signal-carrying device simulation protection of the controlled objects]. *Spetstekhnika i svyaz' – Specialized machinery and communication*. 2015. vol. 2. pp. 59–63 (In Russ.).
12. Kosarev S.A., Braude-Zolotarev Yu.M. *Sposob radiosvyazi ohranjaemykh ob'ektov i centra ohrany* [Method of provision of radiocommunication between guided objects and center of guidance]. Patent RF. no. 2295778. 2007. 11 p. (In Russ.).
13. Vasilevskii V.V., Zav'yalov S.A. [Data transmission method for centralized wireless security systems]. *Omskii nauchnyi vestnik – Omsk Scientific Bulletin*. 2010. vol. 2. pp. 203–206 (In Russ.).
14. Moldavanov A.V., Maksimov O.N., Timchuk A.A., Yushin A.I. *Sposob kontrolja sostojanija ohranjaemogo ob'ekta* [A method of monitoring the status of the protected object]. Patent RF. no. 2221279. 2004. 6 p. (In Russ.).
15. Andrzej Partyka. Transmission of urgent messages in frequency hopping system for intermittent transmission. Patent US. no. 6870875. 2005. 22 p.
16. Gribov V.P., Kosarev S.A., Raigorodskii Yu.V., Sheptovetskii A.Yu. *Sposob radiosvyazi mezhdu ohranjaemyimi ob'ektami i punktom centralizovannoj ohrany* [Wireless method of communication between guarded objects and centralised guarding station]. Patent RF. no. 2351066. 2009. 23 p. (In Russ.).
17. Fedyayev S.L., Maksimov V.S., Fedyayev Yu.S. *Radiokanal'nyj signalizacionnyj kompleks ohrany* [Radiochannel signalling system protection]. The utility model RF. no. 111938. 2011. 5 p. (In Russ.).
18. Gerasimchuk A.N., Kosarev S.A. *Sistema trevozhnoj signalizacii dlja obsluzhivaniya kompaktnoj gruppy ob'ektov nedvizhimosti* [Alarm system to service the compact group of real estate]. The utility model RF. no. 95882. 2010. 11 p. (In Russ.).
19. Vasilevskii V.V., Zav'yalov S.A. *Sposob peredachi izvешhenij dlja sistem centralizovannoj ohrany* [Method for transfer of notices for centralized security systems]. Patent RF. no. 2371775. 2009. 13 p. (In Russ.).

20. Aliev E.A., Magomedov D.A., Karagishiev U.D. *Radiosistema ohrany na shumopodobnyh signalah* [Protection radio system with spread spectrum signals]. Patent RF. no. 2103742. 1998. 5 p. (In Russ.).
21. Len'shin V.P., Rikhter S.G. *Ustrojstvo ohrannoj signalizacii* [An alarm device]. Patent RF. no. 2234135. 2004. 4 p. (In Russ.).
22. Garbatsevich V.A., Kopeikin V.V. *Sposob distancionnoj peredachi soobshhenija i ustrojstvo dlja ego osushchestvlenija* [Transmitting messages method and device for its implementation]. Patent RF. no. 2081456. 1997. 5 p. (In Russ.).
23. Shilov A.V. *Sposob peredachi soobshhenija o srabatyvanii ohrannoj signalizacii* [A method of transmitting messages triggering the alarm]. Patent RF. no. 2199151. 2003. 4 p. (In Russ.).
24. Rudnev A.N., Braude-Zolotarev Yu.M., Davydov Yu.L., Kosarev S.A. [The radio communication system of protection's systems]. *Doklady VI Mezhdunarodnoi nauchno-tehnicheskoi konferentsii "Perspektivnye tekhnologii v sredstvakh peredachi informatsii"* [Proceedings of the VI International scientific-technical conference "Advanced technologies in information transfer"]. Vladimir. Russia. 2005. pp. 57–59 (In Russ.).
25. Braude-Zolotarev Yu. [Advantages of UWB signals]. *Tekhnologii i sredstva svyazi – Communication Technologies & Equipment*. 2014. vol. 1. pp. 54–57 (In Russ.).
26. Osipov D.L., Zhuk A.P., Gavrishev A.A. *Ustrojstvo imitozashhity kontroliruemym ob'ektom s povyshennoj strukturnoj skrytnost'ju signalov-perenoschikov* [Apparatus for protection against imitation of controlled objects with high structural security of carrier signals]. Patent RF. no. 2560824. 2015. 15 p. (In Russ.).
27. Korchinskii V.V. [Evaluation of structural stealth signal designs based on chaotic signals in the transmission systems of confidential information]. *Naukovi praci ONAZ im. O.S. Popova – Proceedings of the O.S. Popov ONAT*. 2012. vol. 1. pp. 77–81 (In Russ.).
28. Sivashchenko S.I. [Secrecy of radio system with difficult and chaotic signals]. *Sistemy upravlinnja, navigacii' ta zv'jazku – Systems of control, navigation and communication*. 2009. vol. 3(11). pp. 56–58 (In Russ.).
29. Zhuk A.P., Barketov S.V., Sazonov V.V. [Option chaotic noise immunity data transmission system]. *Informatsionnoe protivodeistvie ugrozam terrorizma – Information counteraction to the terrorism threats*. 2010. vol. 15. pp. 126–130 (In Russ.).
30. Barketov S.V., Zhuk A.P., Sazonov V.V., Avdeenko S.I., Zhuk E.P., Lokhov V.I., Golub' J.S. *Kogerentnaja sistema peredachi informacii haoticheskimi signalami* [Coherent data transmission system using random signals]. Patent RF. no. 2326500. 2008. 6 p. (In Russ.).
31. Gavrishev A.A. [About unauthorized access to wireless communications systems based on noise-like signals]. *Sbornik tezisev mezhdunarodnoi nauchno-prakticheskoi konferentsii "Pozharotushente: problemy, tekhnologii, innovatsii"* [Abstracts of the International scientific-practical conference "Fire fighting: Issues, Technologies, Innovations"]. Moscow. Russia. 2016. pp. 218–220 (In Russ.).
32. Gorokhov S.M., Zakharchenko N.V., Korchinskii V.V. [Criteria of efficiency of hidden communication techniques]. *Cyfrovi tehnologii' – Digital Technologies*. 2012. vol. 12. pp. 147–150 (In Russ.).
33. Gavrishev A.A., Burmistrov V.A., Osipov D.L. [Assessment the security of wireless alarm from unauthorized access based on the concepts of fuzzy logic]. *Prikladnaya informatika – Journal of Applied Informatics*. 2015. vol. 10. vol. 4(58). pp. 62–69 (In Russ.).
34. Zhuk A.P., Osipov D.L., Gavrishev A.A. [Security analysis methods for assessing wireless alarm]. *Sbornik trudov III Mezhdunarodnoi nauchno-prakticheskoi konferentsii "Informatsionnaya bezopasnost' v svete Strategii Kazakhstan-2050"* [Proceedings of the III International Scientific and Practical Conference "Information security from the Kazakhstan-2050 Strategy"]. Kazakhstan. 2015. pp. 139–144 (In Russ.).
35. Babkin A.N., Esaulenko A.V. [Efficiency of functioning of the radio channel in safety systems]. *Vestnik voronezhskogo instituta MVD Rossii – Vestnik of Voronezh Institute of the Ministry of Interior of Russia*. 2012. vol. 4. pp. 90–91 (In Russ.).

36. Shcherbakov V.B., Ermakov S.A. Bezopasnost' besprovodnyh setej: standart IEEE 802.11 [Wireless Security standard IEEE 802.11]. Moscow: RadioSoft Publ. 2010. 255 p. (In Russ.).
37. Zade L. The concept of a linguistic variable and its application to approximate reasoning. New York. American Elsevier Publ. 1973. (Rus. ed.: Zade L. Ponjatie lingvisticheskoj peremennoj i ego primenenie k prinjatiju priblizhennyh reshenij. Moscow: Mir Publ. 1976. 163 p.)
38. Fayzullin R.R., Vasilyev V.I. [Protectability assessment method of a data-transmission network in security information and event management system on a basis of fuzzy logic]. *Vestnik UGATU – Scientific journal of USATU*. 2013. vol. 13. no. 2(55). pp. 150–156 (In Russ.).

Гавришев Алексей Андреевич — аспирант кафедры организации и технологии защиты информации института информационных технологий и телекоммуникаций, ФГАОУ ВПО Северо-Кавказский федеральный университет (СКФУ). Область научных интересов: имитозащита объектов информационных систем, математические и информационные модели защиты информации. Число научных публикаций — 12. alexxx.2008@inbox.ru; ул. Пушкина, 1, Ставрополь, 355009; п.т.: +7(8652)95-68-08, Факс: +7(8652)95-68-03.

Gavrishev Aleksey Andreevich — Ph.D. Student of organization and technology of information protection department of information technologies and telecommunications institute, North-Caucasus Federal University (NCFU). Research interests: simulation protection of objects of information systems, mathematical and information models of information security. The number of publications — 12. alexxx.2008@inbox.ru; 1, Pushkin Street, Stavropol, 355009, Russia; office phone: +7(8652)95-68-08, Fax: +7(8652)95-68-03.

Жук Александр Павлович — к-т техн. наук, профессор, профессор кафедры организации и технологии защиты информации института информационных технологий и телекоммуникаций, ФГАОУ ВПО Северо-Кавказский федеральный университет (СКФУ). Область научных интересов: теория и практика построения защищённых телекоммуникационных и навигационных систем, имитозащита объектов информационных систем. Число научных публикаций — 250. alekszhuk@mail.ru; ул. Пушкина, 1, Ставрополь, 355009; п.т.: +7(8652)95-68-08, Факс: +7(8652)95-68-03.

Zhuk Aleksandr Pavlovich — Ph.D., professor, professor of organization and technology of information protection department of information technologies and telecommunications institute, North-Caucasus Federal University (NCFU). Research interests: theory and practice of construction of protected telecommunications and navigation systems, simulation protection of objects of information systems. The number of publications — 250. alekszhuk@mail.ru; 1, Pushkin Street, Stavropol, 355009, Russia; office phone: +7(8652)95-68-08, Fax: +7(8652)95-68-03.

Осипов Дмитрий Леонидович — к-т техн. наук, доцент, доцент кафедры прикладной математики и компьютерной безопасности института информационных технологий и телекоммуникаций, ФГАОУ ВПО Северо-Кавказский федеральный университет (СКФУ). Область научных интересов: технологии программирования, имитозащита объектов информационных систем. Число научных публикаций — 60. DmtrOsipov@Yandex.ru; ул. Пушкина, 1, Ставрополь, 355009; п.т.: +7(8652)95-68-08, Факс: +7(8652)95-68-03.

Osipov Dmitriy Leonidovich — Ph.D., associate professor, associate professor of applied mathematics and computer technologies department of information technologies and telecommunications institute, North-Caucasus Federal University (NCFU). Research interests: programming technology, simulation protection of objects of information systems. The number of publications — 60. DmtrOsipov@Yandex.ru; 1, Pushkin Street, Stavropol, 355009, Russia; office phone: +7(8652)95-68-08, Fax: +7(8652)95-68-03.

РЕФЕРАТ

Гавришев А.А., Жук А.П., Осипов Д.Л. **Анализ технологий защиты радиоканала охранно-пожарных сигнализаций от несанкционированного доступа.**

В данной работе авторами рассматриваются вопросы защиты радиоканала охранно-пожарных сигнализаций от несанкционированного доступа. Проведенный анализ технологий защиты радиоканала охранно-пожарных сигнализаций подтвердил, что базовыми методами защиты являются криптографические методы защиты и шумоподобные сигналы. Отмечено, что среди технологий с шумоподобными сигналами выделяются технологии защиты радиоканала на основе передачи сигналов на частотно-временных позициях, на основе псевдослучайной перестройки рабочей частоты, на основе фазоманипулированных сигналов, на основе сверхширокополосных сигналов и на основе хаотических последовательностей. На основе проведенного анализа с помощью аппарата нечеткой логики производится количественная оценка защищенности беспроводных охранно-пожарных сигнализаций с составлением ранжированного списка. В результате установлено, что наиболее защищенными технологиями от комплексных угроз (просмотр, подмена, перехват и подавление) являются технологии на основе шумоподобных сигналов (технологии на основе хаотических последовательностей и сверхширокополосных сигналов), а наименее защищенными — технология с криптографическими методами защиты (устройство имитозащиты). Показана необходимость дальнейших исследований, направленных на повышение защищенности радиоканала охранно-пожарных сигнализаций. Перспективным видится развития систем с технологиями защиты радиоканала на основе шумоподобных сигналов (технологии на основе хаотических последовательностей и сверхширокополосных сигналов).

SUMMARY

Gavrishev A.A., Zhuk A.P., Osipov D.L., **An Analysis of Technologies to Protect a Radio Channel of Fire Alarm Systems against Unauthorized Access.**

This paper considers the issues of protecting the radio channel of fire alarm systems from unauthorized access. The conducted analysis of technologies to protect the radio channel of fire alarms has confirmed that the basic protection methods are cryptographic methods and noise-like signals. It is noted that the technologies of noise-like signals include protection technologies based on: signal transmission on time-frequency positions; frequency-hopping spread spectrum (FHSS); phase modulation signals; ultra-wideband signals; and random sequences. According to the conducted analysis, as well as using the fuzzy logic apparatus, we quantitatively assessed the security of wireless fire alarms and made the ranked list. As a result, it was found out that the most secure technologies to combat complex threats (view, substitution, interception and jamming) are technologies based on noise-like signals (random sequences and ultra-wideband signals), while the least protected technologies appear to be cryptographic methods. The necessity of further research aimed at enhancing security of the radio channel of fire alarm systems is shown. The development of systems with protection technologies based on noise-like signals is seen as promising.

Е.А. ВАСЕЧКИН, А.Б. ТАРАНОВ

**МОДЕЛЬ СИГНАЛОВ ПОБОЧНЫХ ЭЛЕКТРОМАГНИТНЫХ
ИЗЛУЧЕНИЙ ВИДЕОИНТЕРФЕЙСОВ**

Васечкин Е.А., Таранов А.Б. Модель сигналов побочных электромагнитных излучений видеоинтерфейсов.

Аннотация. Измерение энергии сигналов побочных электромагнитных излучений является актуальной задачей при рассмотрении вопросов электромагнитной совместимости и защиты информации. При построении видеосистем средств вычислительной техники используются параллельные линии передачи, побочные электромагнитные излучение от которых взаимодействуют. В статье представлены особенности схемотехнической реализации видеосистем и модели излучения учитывающие их. Помимо этого приведена модель информационных сигналов видеоинтерфейсов, учитывающая компенсационные свойства дифференциальных линий.

Ключевые слова: побочное электромагнитное излучение, дифференциальная линия, видеоинтерфейс, синфазная составляющая тока.

Vasechkin E.A., Taranov A.B. A Model of Signals of Side Electromagnetic Radiation of Video Interfaces.

Abstract. Energy measurement of side electromagnetic radiation signals is an important task in addressing issues of electromagnetic compatibility and information security. When constructing video systems of computer equipment, parallel transmission lines are used, side electromagnetic radiations from which interact. The article presents the characteristics of circuit realization of video systems and radiation models considering them. In addition, we present a model of information signals of video interfaces, which takes into account the compensatory properties of differential lines in a video system with a DVI interface.

Keywords: side electromagnetic radiation, differential line, video interface, in-phase component of the current.

1. Введение. В условиях бурного развития средств обработки и передачи информации, на фоне возрастания возможностей средств технической разведки и глобальной информатизации общества все более значимой становится необходимость решения проблем защиты информации. Анализ [1-5] показал, что для средств вычислительной техники (СВТ) одним из наиболее опасных является канал утечки за счет побочных электромагнитных излучений (ПЭМИ). В качестве основного источника ПЭМИ СВТ выступает видеосистема [6]. Для оценки степени защищенности СВТ должны проводиться измерения энергии ПЭМИ, учитывающие особенности аппаратной реализации узлов и блоков СВТ. Существующие модели, описывающие ПЭМИ от видеосистемы, не учитывают этих особенностей, а значит, взаимодействию ПЭМИ от нескольких информационных линий не будет уделено должного внимания. Поэтому актуальна задача создания модели сигналов ПЭМИ видеосистемы, с помощью которой возможно исследовать взаимодействие ПЭМИ от ее информационных линий между собой. Стоит отметить, что анализ современной отечественной и ино-

странной литературы по данной тематике затруднен в силу закрытости данной тематики исследований.

2. Проводные интерфейсы СВТ. Особенностью построения проводных интерфейсов начиная с 1995 года является использование дифференциального способа передачи информации по симметричным линиям [7]. Суть этого способа заключается в том, что на оба проводника линии подаются разноточные по модулю, но противоположные по знаку сигналы. Сигналы, одновременно передаваемые по проводникам дифференциальной линии, называют комплементарными. Например, для интерфейса *DVI* общая схема дифференциальной линии в [8], представлена следующим образом:

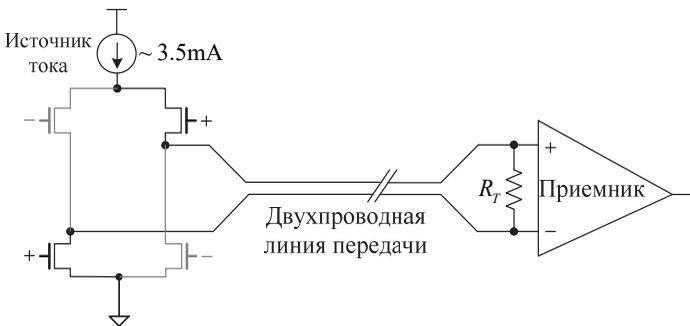


Рис. 1. Общая схема дифференциальной линии

На рисунке 1 R_T — сопротивление согласующего резистора. На прямой вход операционного усилителя поступает сигнал U^+ , а сигнал U^- на инверсный вход. На выходе приемника будет сигнал $U_{out} = U^+ - U^-$. Сигналы, протекающие в дифференциальной линии, схематично представлены на рисунке 2.

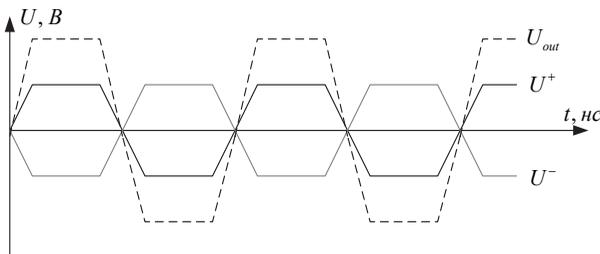


Рис. 2. Сигналы в дифференциальной линии

Использование дифференциального способа передачи имеет следующие преимущества [9]:

1. Повышение помехоустойчивости приемника дифференциальной линии за счет того, что помехи, воздействующие на оба проводника дифференциальной линии, вычитаются на входе приемника.

2. Снижение напряженности ПЭМИ в пространстве за счет того, что токи, возникающие в проводниках, противоположны по знаку, следовательно, и поля, возникающие в пространстве, будут компенсировать друг друга.

3. Существующие модели излучения от проводных интерфейсов. Простейшими моделями, описывающими излучение от линии передачи сигнала, являются модели излучения элементарного диполя и рамочной антенны [10]. Ограничением на применение данных моделей является длина линии, она не должна превосходить $0.1 \cdot \lambda$, где λ — наименьшее значение волны сигнала передаваемого по физической линии [11]. В качестве допущения используется тот факт, что сигналы распространяются по аддитивной модели Гаусса. Связанно это с тем, что при проведении измерений многолучевость распространения излучения исключается за счет использования безэховых камер, отстройки по максимальному уровню излучения и использованию направленных антенн [12].

Взаимная ориентация векторов \vec{E} и \vec{H} в поле диполя Герца представлена на рисунке 3.

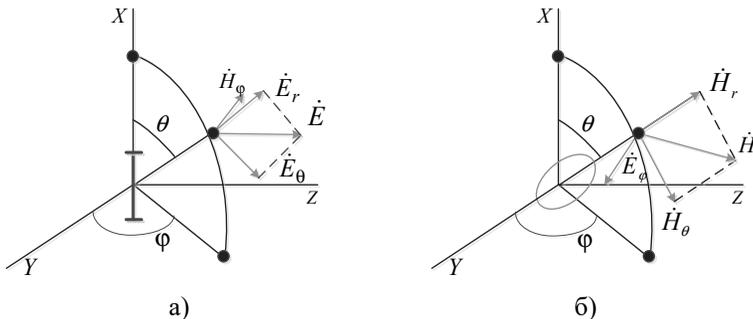


Рис. 3. Взаимная ориентация векторов \vec{E} и \vec{H} в поле диполя Герца (а), рамочной антенны (б)

Значение векторов напряженности электромагнитного поля вокруг диполя Герца длиной l имеет следующий вид:

$$\dot{H}_\varphi(\omega, r) = \frac{\dot{I}(\omega)l}{4\pi} \left[j \frac{k}{r} + \frac{1}{r^2} \right] \sin \theta e^{-jkr}, \quad (1)$$

$$\dot{E}_r(\omega, r) = \frac{k}{\omega \epsilon} \frac{\dot{I}(\omega) l}{2\pi} \left[\frac{1}{r^2} - j \frac{1}{kr^3} \right] \cos \theta e^{-jkr}, \quad (2)$$

$$\dot{E}_\theta(\omega, r) = \frac{k}{\omega \epsilon} \frac{\dot{I}(\omega) l}{4\pi} \left[j \frac{k}{r} + \frac{1}{r^2} - j \frac{1}{kr^3} \right] \sin \theta e^{-jkr}, \quad (3)$$

где $k = \omega / c$ — волновое число; c — скорость света; $\dot{I}(\omega)$ — комплексное значение амплитуды тока, проходящего по линии на частоте ω ; r — расстояние до точки наблюдения; ϵ — диэлектрическая проницаемость среды распространения электромагнитной волны; θ, φ — значение сферических координат точки наблюдения.

Значение векторов напряженности электромагнитного поля вокруг рамочной антенны радиусом R :

$$\dot{E}_\varphi(\omega, r) = -j \frac{\omega \mu}{k} \frac{\dot{I}(\omega) a}{4\pi} \left[j \frac{k}{r} + \frac{1}{r^2} \right] \sin \theta e^{-jkr}, \quad (4)$$

$$\dot{H}_r(\omega, r) = j \frac{\dot{I}(\omega) a}{2\pi} \left[\frac{1}{r^2} - j \frac{1}{kr^3} \right] \cos \theta e^{-jkr}, \quad (5)$$

$$\dot{H}_\theta(\omega, r) = j \frac{\dot{I}(\omega) a}{4\pi} \left[j \frac{k}{r} + \frac{1}{r^2} - j \frac{1}{kr^3} \right] \sin \theta e^{-jkr}, \quad (6)$$

где $a = k\pi R^2$ — площадь рамочной антенны; μ — магнитная проницаемость среды распространения электромагнитной волны.

Для высокоскоростных линий передачи характерно превосходство в несколько раз длины линии над длинной электромагнитной волны, передаваемой вдоль кабеля [13]. В этих случаях более адекватным является описание излучения моделью антенны бегущей волны (рисунок 4), которую можно представить в виде суперпозиции n диполей длиной dL . В каждом из диполей комплексное значение силы тока отличается от предыдущего в силу особенностей распространения электромагнитных волн вдоль направляющих систем.

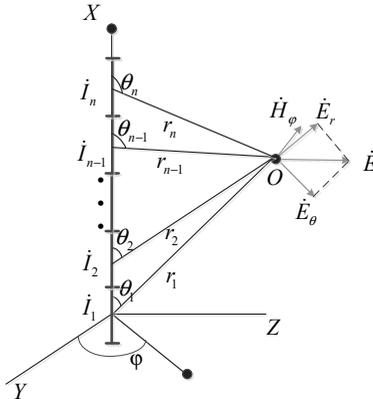


Рис. 4. Модель антенны бегущей волны

Выражения для составляющих напряженностей магнитного и электрического полей имеют вид:

$$\dot{H}_\varphi(\omega, r) = \sum_{n=1}^N \frac{\dot{I}_n(\omega)l}{4\pi} \left[j \frac{k}{r_n} + \frac{1}{r_n^2} \right] \cdot \sin \theta_n \cdot e^{-jk r_n}, \quad (7)$$

$$\dot{E}_r(\omega, r) = \sum_{n=1}^N \frac{k}{\omega \epsilon} \frac{\dot{I}_n(\omega)l}{2\pi} \left[\frac{1}{r_n^2} - j \frac{1}{k r_n^3} \right] \cdot \sin \theta_n \cdot e^{-jk r_n}, \quad (8)$$

$$\dot{E}_\theta(\omega, r) = \sum_{n=1}^N \frac{k}{\omega \epsilon} \frac{\dot{I}_n(\omega)l}{4\pi} \left[j \frac{k}{r_n} + \frac{1}{r_n^2} - j \frac{1}{k r_n^3} \right] \cdot \sin \theta_n \cdot e^{-jk r_n}. \quad (9)$$

Недостатком представленных моделей является то, что они не учитывают, что передача сигналов в современных интерфейсах осуществляется по дифференциальным линиям.

4. Модель излучения от дифференциальной линии. Дифференциальную линию можно представить в виде совокупности двух параллельных диполей (рисунок 5), находящихся на расстоянии Δr друг от друга или двух рамочных антенн, находящихся в параллельных плоскостях на том же расстоянии:

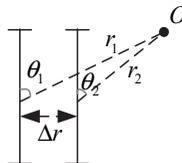


Рис. 5. Расположение двух диполей и точки наблюдения

В случаях, когда точка измерения находится на расстоянии $r \gg \Delta r$, углы $\theta_1 = \theta_2 = \theta$. Рассмотрим излучение двух диполей Герца, для случая, когда расстояние до точки наблюдения для каждого из них намного больше расстояния между ними $r_1 \gg \Delta r$ и $r_2 \gg \Delta r$. Значение напряженностей магнитного и электрического полей имеют вид:

$$\dot{H}_\varphi(\omega, r) = \frac{l}{4\pi} \cdot \sin \theta \cdot \left[\dot{I}_1(\omega) \cdot \left(j \frac{k}{r_1} + \frac{1}{r_1^2} \right) \cdot e^{-jk r_1} + \dot{I}_2(\omega) \cdot \left(j \frac{k}{r_2} + \frac{1}{r_2^2} \right) \cdot e^{-jk r_2} \right], \quad (10)$$

$$\dot{E}_r(\omega, r) = \frac{k \cdot l}{2\pi\omega\epsilon} \cos \theta \left[\dot{I}_1(\omega) \left(\frac{1}{r_1^2} - j \frac{1}{k r_1^3} \right) e^{-jk r_1} + \dot{I}_2(\omega) \left(\frac{1}{r_2^2} - j \frac{1}{k r_2^3} \right) e^{-jk r_2} \right], \quad (11)$$

$$\dot{E}_\theta(\omega, r) = \frac{k \cdot l}{4\pi\omega\epsilon} \sin \theta \left[\dot{I}_1(\omega) \left(j \frac{k}{r_1} + \frac{1}{r_1^2} - j \frac{1}{k r_1^3} \right) e^{-jk r_1} + \dot{I}_2(\omega) \left(j \frac{k}{r_2} + \frac{1}{r_2^2} - j \frac{1}{k r_2^3} \right) e^{-jk r_2} \right], \quad (12)$$

где r_1, r_2 — расстояние до точки наблюдения от каждого из диполей; $\dot{I}_1(\omega), \dot{I}_2(\omega)$ — комплексное значение амплитуд тока на частоте ω , проходящей по различным проводникам дифференциальной линии.

При этом максимальное значение разности расстояний от точки наблюдения до диполей будет равно Δr .

$$\max(|r_2 - r_1|) = \Delta r. \quad (13)$$

Существует необходимость изучения влияния разности расстояний от диполей до точки наблюдения. Для этого предположим, что значение амплитуд токов на каждой частоте одинаково в каждом диполе, а также параметры самих диполей одинаковы. Тогда отношения составляющих напряженности полей в одной и той же точке наблюдения для диполей имеет вид:

$$\frac{\dot{H}_\varphi(\omega, r_1)}{\dot{H}_\varphi(\omega, r_2)} = \left(\frac{jk r_1 + 1}{jk r_2 + 1} \right) \cdot \left(\frac{r_2}{r_1} \right)^2 \cdot e^{jk(r_2 - r_1)}, \quad (14)$$

$$\frac{\dot{E}_r(\omega, r_1)}{\dot{E}_r(\omega, r_2)} = \left(\frac{k r_1 - j}{k r_2 - j} \right) \cdot \left(\frac{r_2}{r_1} \right)^3 \cdot e^{jk(r_2 - r_1)}, \quad (15)$$

$$\frac{\dot{E}_\theta(\omega, r_1)}{\dot{E}_\theta(\omega, r_2)} = \frac{jk^2 r_1^2 + k r_1 - j}{jk^2 r_2^2 + k r_2 - j} \cdot \left(\frac{r_2}{r_1} \right)^3 \cdot e^{jk(r_2 - r_1)}. \quad (16)$$

Преобразуем выражение для случая максимальной разности расстояний, т.е. $r_1 = r_2 + \Delta r$.

$$\frac{\dot{H}_\varphi(\omega, r_1)}{\dot{H}_\varphi(\omega, r_2)} = \left(1 + \frac{jk \cdot \Delta r}{jkr_2 + 1}\right) \cdot \left(\frac{r_2}{r_2 + \Delta r}\right)^2 \cdot e^{j \cdot k \cdot \Delta r}, \quad (17)$$

$$\frac{\dot{E}_r(\omega, r_1)}{\dot{E}_r(\omega, r_2)} = \left(1 + \frac{k \cdot \Delta r}{kr_2 - j}\right) \cdot \left(\frac{r_2}{r_2 + \Delta r}\right)^3 \cdot e^{j \cdot k \cdot \Delta r}, \quad (18)$$

$$\frac{\dot{E}_\theta(\omega, r_1)}{\dot{E}_\theta(\omega, r_2)} = \frac{jk^2(r_2 + \Delta r)^2 + k(r_2 + \Delta r)_1 - j}{jk^2 r_2^2 + kr_2 - j} \cdot \left(\frac{r_2}{r_2 + \Delta r}\right)^3 \cdot e^{j \cdot k \cdot \Delta r}. \quad (19)$$

В реальном кабеле максимальное расстояние между проводниками дифференциальной линии будет менее $\Delta r = 0.005$ м. Таким образом, можно считать, что $\Delta r \rightarrow 0$, значит, значения выражений (14-16) в правой части будут $\rightarrow 1$. Следовательно, разностью расстояний от каждого из диполей до точки измерения можно пренебречь. С использованием этого допущения выражения (10-12) преобразуются к следующему виду:

$$\dot{H}_\varphi(\omega, r) = \frac{l}{4\pi} \cdot \sin \theta \cdot \left(j \frac{k}{r} + \frac{1}{r^2}\right) \cdot e^{-jkr} \cdot (\dot{I}_1(\omega) + \dot{I}_2(\omega)), \quad (20)$$

$$\dot{E}_r(\omega, r) = \frac{k \cdot l}{2\pi\omega\epsilon} \cos \theta \left(\frac{1}{r_1^2} - j \frac{1}{kr_1^3}\right) \cdot e^{-jkr_1} \cdot (\dot{I}_1(\omega) + \dot{I}_2(\omega)), \quad (21)$$

$$\dot{E}_\theta(\omega, r) = \frac{k \cdot l}{4\pi\omega\epsilon} \cdot \sin \theta \cdot \left(j \frac{k}{r} + \frac{1}{r^2} - j \frac{1}{kr^3}\right) \cdot e^{-jkr} \cdot (\dot{I}_1(\omega) + \dot{I}_2(\omega)). \quad (22)$$

Поскольку преобразование Фурье является линейным [14], то для интервала наблюдения длительностью T_H можно записать следующее выражение:

$$\dot{I}_1(\omega) + \dot{I}_2(\omega) = \frac{1}{\sqrt{2\pi}} \int_0^{T_H} (I_1(t) + I_2(t)) \cdot e^{-j\omega t} dt. \quad (23)$$

Из выражений (20-23) видно, что напряженность магнитного и электрического полей и значение синфазной составляющей тока диффе-

ренциальной линии $I_c(t) = I_1(t) + I_2(t)$ имеют прямо пропорциональную зависимость между собой. Аналогичная зависимость получается и для двух рамочных антенн, находящихся в параллельных плоскостях.

Высокоскоростную дифференциальную линию передачи можно описать как 2 параллельные антенны бегущей волны, находящиеся на расстоянии Δr . В этом случае для каждой двух параллельных диполей будут выполняться зависимости, полученные выше. Таким образом, напряженности поля ПЭМИ дифференциальной линии, описываемой с использованием антенны бегущей волны, будут также прямо пропорционально зависеть от синфазной составляющей тока передаваемой по этой линии.

В [9] рассматривалось представление сигналов дифференциальной линии в виде двух составляющих:

– синфазная составляющая $I_c(t) = \frac{I_1(t) + I_2(t)}{2}$;

– дифференциальная составляющая $I_d(t) = \frac{I_1(t) - I_2(t)}{2}$.

Дифференциальная составляющая тока показывает абсолютно сбалансированную часть токов, протекающих по проводникам дифференциальной линии. Излучение от такой составляющей определяется разностью расстояний от точки измерений до этих проводников, и будет заметно на небольших расстояниях [11]. Как отмечалось выше, синфазная составляющая тока является основной причиной появления ПЭМИ. Поскольку $I_1(t)$ и $I_2(t)$ комплементарны, то синфазная составляющая тока $I_c(t)$ отличается от постоянной составляющей только в моменты передачи по дифференциальной линии фронтов и срезов импульсов. Основными причинами возникновения синфазной составляющей являются:

- задержка по времени между сигналами в дифференциальной линии, обусловленная особенностями передатчика и сигнального тракта;
- различие между фронтами и срезами импульсов в проводниках дифференциальной линии;
- различием характеристик передающих трактов сигналов дифференциальной линии.

По результатам приведенных рассуждений можно сделать вывод о прямой зависимости энергии электромагнитных излучений от значения синфазной составляющей тока.

5. Модель излучения от нескольких дифференциальных линий. Одним из наиболее распространенных видеointерфейсов в настоящее время является *Digital Video Interface (DVI)* [15]. Как и у

большинства видеointерфейсов передача различных составляющих цветов пикселя, выводимого на экран, осуществляется по параллельным линиям. Учитывая эту особенность видеointерфейсов, модель сигналов побочных электромагнитных излучений от них можно представить в виде совокупности двух моделей:

- модель информационных сигналов видеointерфейсов, учитывающая компенсационные свойства дифференциальных линий;
- модель излучения от нескольких дифференциальных линий в видеосистеме.

Взаимосвязь этих моделей представлена на рисунке 6.



Рис. 6. Взаимосвязь моделей, описывающих излучение видеосистемы

На рисунке 6 $D_K^R[0,7]$, $D_K^G[0,7]$, $D_K^B[0,7]$ — значения последовательностей, соответствующих 8-битному представлению уровней яркости базовых цветов пикселей по модели *RedGreenBlue(RGB)*, на k этапе работы; $Q_K^R[0,9]$, $Q_K^G[0,9]$, $Q_K^B[0,9]$, $Q_K^{\text{sinxp}}[0,9]$ — значения передаваемых по интерфейсным линиям последовательностей, на k этапе работы модели информационных сигналов видеointерфейсов; $I_C^R(t)$, $I_C^G(t)$, $I_C^B(t)$, $I_C^{\text{sinxp}}(t)$ — непрерывные значения синфазных составляющих тока, соответствующих передаче последовательностей $Q_K^R[0,9]$, $Q_K^G[0,9]$, $Q_K^B[0,9]$, $Q_K^{\text{sinxp}}[0,9]$ по интерфейсным линиям видеосистемы; $\dot{E}(\omega, r)$, $\dot{H}(\omega, r)$ — комплексное значение напряженностей электрического и магнитного полей на частоте ω и расстоянии до точки измерения r .

Блок измерения синфазных составляющих тока необходим для согласования дискретных и непрерывных данных, используемых в моделях. Стоит отметить, что измерения синфазных составляющих тока по всем дифференциальным линиям должно происходить одновременно.

Рассмотрим сначала модель излучения от нескольких дифференциальных линий. Проводники параллельных дифференциальных линий видеосистемы находятся на расстояниях соизмеримых с расстоянием между проводниками одной дифференциальной линии. Испол-

зую рассуждения, указанные выше, можно показать, что излучение от нескольких дифференциальных линий подобно излучению от проводников одной дифференциальной линии. Таким образом, синфазные составляющие в каждой линии будут оказывать воздействие на суммарный сигнал ПЭМИ от этих линий. Реальная модель излучения ПЭМИ представляет собой совокупность взаимосвязанных диполей, и может быть описана достаточно сложной моделью распределенного диполя. Выбор точки измерения, а также использование различных по свойствам антенн вносит дополнительную случайность в результаты измерений. Поэтому для оценки ПЭМИ нескольких дифференциальных каналов может быть использована упрощенная модель:

$$\dot{H}(\omega, r) = K_H(\omega, r) \cdot \left(I_C^R(\omega) + I_C^G(\omega) + I_C^B(\omega) + I_C^{cunxp}(\omega) \right), \quad (24)$$

$$\dot{E}(\omega, r) = K_E(\omega, r) \cdot \left(I_C^R(\omega) + I_C^G(\omega) + I_C^B(\omega) + I_C^{cunxp}(\omega) \right), \quad (25)$$

где $K_H(\omega, r)$, $K_E(\omega, r)$ — пространственно-частотные характеристики излучений от дифференциальных линий; $I_C^R(\omega)$, $I_C^G(\omega)$, $I_C^B(\omega)$, $I_C^{cunxp}(\omega)$ — комплексные значения амплитуд синфазных составляющих токов, протекающих в линиях видеосистемы на частоте ω .

Представленная модель указывает на зависимость уровней напряженностей электрических и магнитного полей от суммарного значения синфазных составляющих токов, протекающих в нескольких дифференциальных линиях. И в дальнейшем используется для исследования взаимного влияния ПЭМИ нескольких дифференциальных линий.

Поскольку электромагнитное излучение возникает только в моменты изменения тока или напряжения в линии [6], то для сигналов цифровых видеointерфейсов СВТ ПЭМИ будут возникать только в моменты передачи фронтов и срезов сигналов, передаваемых по интерфейсной линии. Максимальный джиттер между сигналами, передаваемыми по параллельным интерфейсным линиям, составляет не более $0,25 \cdot T_0$ (где T_0 — время передачи одного бита информации) [8]. С учетом этих двух особенностей можно сделать вывод о взаимном влиянии ПЭМИ от параллельных интерфейсных линий видеосистемы только в моменты передачи по ним фронтов и срезов сигналов.

Модель информационных сигналов видеointерфейсов учитывает моменты совпадений знакоперемен в последовательностях, передаваемых по дифференциальным линиям. Это позволяет оценить степень взаимодействия ПЭМИ от нескольких дифференциальных линий. При этом ПЭМИ могут как компенсировать, так и усиливать друг друга.

Под степенью взаимодействия понимается коэффициент K , принимающий значение 1 в случае совпадения во всех последовательностях моментов знакоперемен, значение 0 в обратной ситуации.

Структурная схема модели представлена на рисунке 7. Разрабатываемая модель является динамической, детерминированной, дискретной [16].

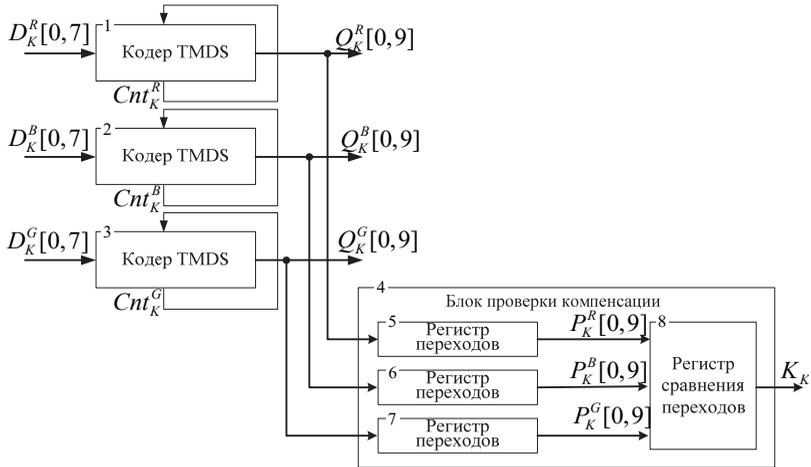


Рис. 7. Параметры модели информационных сигналов видеоинтерфейса DVI

В соответствии со структурной схемой модели входными параметрами является множество входных 8-битных последовательностей $D_K^R [0,7] \in A$, $D_K^B [0,7] \in A$, $D_K^G [0,7] \in A$, поступающих на вход кодера *Transition Minimized Differential Signaling (TMDS)*, алгоритм работы которого указан в [17]. Множество A — множество всех возможных 8-битных комбинаций на входе кодера *TMDS*.

В качестве выходных параметров выступают:

- множество 10-битных выходных последовательностей на выходе кодера *TMDS* $Q_K^R [0,9] \in B$, $Q_K^B [0,9] \in B$, $Q_K^G [0,9] \in B$, где B — множество разрешенных комбинаций на выходе кодера;

- коэффициент $K_K \in [0,1]$, учитывающий взаимное влияние информационных дифференциальных линий друг на друга.

Внутренними управляемыми параметрами модели является начальное состояние кодера *TMDS* $Cnt_0^R \in C$, $Cnt_0^B \in C$, $Cnt_0^G \in C$ где

$C = \{-8, -6, -4, -2, 0, 2, 4, 6, 8\}$ — множество возможных состояний кодера *TMDS*.

Внутренними неуправляемыми параметрами являются:

– множество 10-битных последовательностей с выхода регистров перехода: $P_K^R [0,9] \in D$, $P_K^B [0,9] \in D$, $P_K^G [0,9] \in D$, где D — множество возможных комбинаций на выходе регистров перехода для выходных сигналов с кодера *TMDS*;

– множество состояний кодера *TMDS* на k -шаге работы $Cnt_K^R \in C$, $Cnt_K^B \in C$, $Cnt_K^G \in C$.

Один шаг работы модели равен одному шагу работы кодера *TMDS*.

Работу кодера *TMDS* можно описать с использованием инъективной функции $f: A \times C \rightarrow B \times C$. Инъективные свойства функции определяются однозначностью кодирования информации и однозначностью определения выходного состояния при известной входной последовательности бит и известном состоянии кодера на предыдущем шаге работы модели. Модель работы кодера *TMDS* имеет следующий вид:

$$(Q_K [0,9], Cnt_K) = f(D_K [0,7], Cnt_{K-1}) \quad (26)$$

где $D_K [0,7]$ — значение 8-битной последовательности на входе кодера *TMDS*; $Q_K [0,9]$ — значение 10-битной последовательности на входе кодера *TMDS*; Cnt_K , Cnt_{K-1} — значение состояний кодера *TMDS* на k и $k-1$ шаге работы.

Работу регистра переходов можно описать с использованием инъекции $h: B \rightarrow D$. Логическая функция h описывается следующим выражением:

$$h(Q_K [0,9], Q_{K-1} [9]) = [Q_K [0] \oplus Q_{K-1} [9], Q_K [1] \oplus Q_K [0], \dots, Q_K [9] \oplus Q_K [8]]. \quad (27)$$

Таким образом, функция преобразует выходную последовательность кодера *TMDS* $Q_K [0;9]$ в последовательность, в которой обозначены моменты переходов импульсов. Модель работы регистра переходов можно описать следующим образом:

$$P_K [0,9] = h(Q_K [0,9]). \quad (28)$$

Работу регистра сравнения переходов можно описать с использованием логической функции g . Функцию g необходима для определения единиц встречающихся во всех трех последовательностях

только один раз. Другими словами эта функция, выделяет переходы, встречающиеся только в одной из трех последовательностях $Q_K [0, 9]$. Функция g для i элементов последовательностей $P_K^R [0, 9], P_K^B [0, 9], P_K^G [0, 9]$ имеет следующий вид:

$$g(P_K^R [i], P_K^B [i], P_K^G [i]) = \begin{cases} 1, & \text{если } w([P_K^R [i], P_K^B [i], P_K^G [i]]) = 1 \\ 0, & \text{если } w([P_K^R [i], P_K^B [i], P_K^G [i]]) \neq 1 \end{cases} \quad (29)$$

где $w([P_K^R [i], P_K^B [i], P_K^G [i]])$ — вес по Хеммингу вектора $[P_K^R [i], P_K^B [i], P_K^G [i]]$.

С использованием этой функции вектор коэффициента K_K на k шаге работы имеет следующий вид:

$$K_K = \frac{g(P_K^R [0], P_K^B [0], P_K^G [0]) + \dots + g(P_K^R [9], P_K^B [9], P_K^G [9])}{10}. \quad (30)$$

Для нахождения значений коэффициента K для нескольких шагов работы модели необходимо найти среднее арифметическое этого коэффициента:

$$K_{1,N} = \frac{1}{N} \sum_{i=1}^N K_i \quad (31)$$

Значение коэффициента $K_{1,N} = 0$ говорит о том, что во всех 3 линиях видеосистемы на всех N этапах передаются последовательности с совпадающими моментами знакоперемен. При таком значении коэффициента, изменяя соотношение фаз между сигналами, передаваемыми по параллельным линиям, наблюдается изменение энергии суммарного ПЭМИ. Для сигналов с большими значениями коэффициента $K_{1,N}$ значение энергии будет варьироваться меньше, при изменении соотношений фаз между сигналами в параллельной линии. В случае если $K_{1,N} = 1$, за счет изменения соотношений фаз между сигналами в параллельных линиях невозможно добиться изменения суммарного уровня ПЭМИ. Коэффициент $K_{1,N} \in [0, 1]$ позволяет выбрать из всего множества возможных входных последовательностей, соответствующих цветам пикселей, выводимых на экран, множество входных последовательностей, при прохождении которых по

интерфейсным линиям ПЭМИ от отдельных дифференциальных линий будет оказывать максимальное воздействие друг на друга.

Для оценки взаимного влияния дифференциальных линий друг на друга был проведен эксперимент. В ходе которого измерялся уровень ПЭМИ от видеосистемы в частотной области при различных комбинациях сигналов, передаваемых по интерфейсным линиям. При каждом измерении изменяется соотношение фаз между сигналами, передаваемыми по параллельным интерфейсным линиям видеосистемы. При этом использовались периодические сигналы, поскольку они позволяют регистрировать ПЭМИ узкополосными приборами. Последовательность пикселей серого цвета, соответствующая уровню яркости 33 по составляющим в модели *RGB*, на выходе кодера будет иметь вид $Q_k^1 [0,9] = [0,0,0,0,0,1,1,1,1,1]$. При выводе на экран монотонного изображения с таким цветом пикселей, на физическом уровне по всем трем составляющим будет передаваться периодические сигналы. При выводе на экран уровня яркости цвета 223 на физическом уровне будет образовываться последовательность $Q_k^2 [0,9] = [1,1,1,1,1,0,0,0,0,0]$. Последовательности $Q_k^1 [0,9]$ и $Q_k^2 [0,9]$ инверсны между собой и периодичны с одной и той же скважностью, а также имеют значение коэффициента $K_{1,N} = 0$. Это позволяет оценить взаимное влияние дифференциальных линий друг на друга. Результаты измерения и условия проведения эксперимента приведены в таблице 1.

Таблица 1. Результаты измерений ПЭВМ In Win и монитора LG E2241

Условия проведения эксперимента: разрешение экрана монитора 1920*1080, расстояние от антенны до ПЭВМ $L=1$ м, частота измерений $f=148,51$ МГц, полоса пропускания $RBW=10$ кГц		
RGB комбинации	Уровень сигнал +шум, дБ/мкВ	Уровень шума, дБ/мкВ
33/33/33	38,1	6
33/33/223	40,0	
33/223/33	37,3	
223/33/33	39,1	
223/223/33	40,9	
223/223/223	46,4	
33/223/223	42,8	
223/33/223	44,6	

Как видно из результатов измерений, разность между максимальным и минимальным уровнем сигнала ПЭМИ составляет 9,1 дБ. Отличие по уровню излучений для комбинаций 33/33/33 и 223/223/223 связано с влиянием ПЭМИ от линии синхронизации. Таким образом, в

эксперименте показано, что изменяя соотношение фаз между сигналами в параллельной линии можно добиться изменения суммарного излучения от видеосистемы.

3. Заключение. В результате проведенных исследований разработан модель информационных сигналов видеосистемы, учитывающая компенсационные свойства дифференциальных линий. Эта модель позволяет выделить последовательности, при передаче которых по интерфейсным линиям образующиеся ПЭМИ будут оказывать друг на друга взаимное влияние. Результаты исследования этого влияния в ходе эксперимента показывают, что суммарное ПЭМИ видеосистемы может быть снижено в несколько раз.

Данные исследования в будущем позволят более объективно оценивать суммарный уровень ПЭМИ от видеосистемы за счет учета взаимного влияния дифференциальных линий. Кроме того, представленные модели могут быть использованы в задачах повышения защищенности информации обрабатываемой СВТ за счет снижения суммарного излучения от видеосистемы.

Литература

1. *Зайцев А.П., Шелупанов А.А., Мецераков Р.В.* и др. Технические средства и методы защиты информации: учеб. для вузов // М.: ООО «Издательство Машиностроение». 2009. 508 с.
2. *Халаян Д. Б.* Защита информации. Вас подслушивают? Защищайтесь! // М.: НОУ ШО «Баярд». 2004. 432 с.
3. *Хорев А. А.* Защита информации от утечки по техническим каналам. Часть 1. Технические каналы утечки информации : учеб. пособие // М. : Гостехкомиссия России. 1998. 320 с.
4. *Бузов Г.А., Калинин С.В., Кондратьев А.В.* Защита от утечки информации по техническим каналам : учеб. пособие // М.: Горячая линия – Телеком. 2005. 416 с.
5. *Хорев А. А.* Оценка возможности побочных электромагнитных излучений и наводок // Доклады ТУСУРа. 2014. № 2(32). С. 207–213.
6. *Kuhn G.* Electromagnetic Eavesdropping Risks of Flat-Panel Displays // 4th Workshop on Privacy Enhancing Technologies. 2004. pp. 20.
7. *Уэйкерли Дж.Ф.* Проектирование цифровых устройств // М.: Постмкет. 2003. 544 с.
8. Digital Display Working Group (DDWG). Digital Visual Interface (DVI) revision 1.0. 1999. URL:http://www.cs.unc.edu/~stc/FAQs/Video/dvi_spec-V1_0.pdf. (дата обращения: 20.03.2016).
9. *Джонсон Г.В., Грэхем М.* Высокоскоростная передача цифровых данных. Высший курс черной магии // М.: Вильямс. 2005. 1024 с.
10. *Гольдштейн Л.Д., Зернов Н.В.* Электромагнитные поля и волны // М.: «Советское радио». 1971. 664 с.
11. *Горбунова А.А., Коновалюк М.А., Баев А.Б.* Модель электромагнитных излучений проводных интерфейсов передачи данных средств вычислительной техники // Специальная техника. 2013. № 5. С. 26–36.

12. Гуляйкин Д.А., Тарусов В.А., Шалагинов В.А. Модель системы измерения побочных электромагнитных излучений // Вестник РГРТУ. 2014. № 1. С. 62–68.
13. Сазонов Д. М. Антенны и устройства СВЧ // М.: Высшая школа. 1988. 432 с.
14. Зюко А.Г., Кловский Д.Д., Коржик В.И. и др. Теория электрической связи: учебник для вузов / под ред. Д.Д. Кловского // М.: Радио и связь. 1999. 432 с.
15. Колесниченко О.В., Шишигин И.В., Соломенчук В.Г. Аппаратные средства РС // СПб.: БХВ-Петербург. 2010. 800 с.
16. Советов Б. Я., Яковлев С. А. Моделирование систем: учеб. для вузов // М.: Высшая школа. 2001. 343 с.
17. National Semiconductor. LVDS Owner's Manual revision 2.0. 2000. URL: http://www.gaw.ru/pdf/interface/LVDS/LOM_2.pdf. (дата обращения: 20.03.2016).

References

1. Zaicev A.P. *Tekhnicheskie sredstva i metody zaschity informacii: Uchebnik dlya vuzov* [Hardware and methods of information security: tutorial]. М.: ООО «Izdatel'stvo Mashinostroeniye». 2009. 508 p. (In Russ.).
2. Halyapin D.B. *Zaschita informacii. Vas podslushivayut? Zashischaites'!* [Information security. Did you overhear? Defend yourself!]. М.: NOU ShO «Bayard». 2004. 432 p. (In Russ.).
3. Horev A.A. *Zaschita informacii ot utechki po tekhnicheskim kanalam. Chast'1. Tekhnicheskie kanaly utechki informacii: ucheb. posobie* [Information protection against leakage via technical channels. Part 1. Technical channels of information leakage: tutorial]. М.: Gostehkomissiya Rossii. 1998. 320 p.
4. Buzov G. A., Kalinin S. V., Kondrat'ev A. V. *Zaschita ot utechki informacii po tekhnicheskim kanalam: ucheb. posobie* [Protection against leakage of information through technical channels: tutorial] . М.: Goryachaya liniya – Telekom. 2005. 416 p. (In Russ.).
5. Horev A. A. [Evaluation of the possibility of side electromagnetic radiation and interference]. *Doklady TUSUR – Reports TUSUR*. 2014. vol. 2(32). pp. 207–213. (In Russ.).
6. Kuhn, G. *Electromagnetic Eavesdropping Risks of Flat-Panel Displays*. 4th Workshop on Privacy Enhancing Technologies. 2004. pp. 20.
7. Ueykerli J.F. *Proektirovanie cifrovih ustroystv* [Designing digital devices]. М.: Postmarket. 2003. 544 p. (In Russ.).
8. Digital Display Working Group (DDWG). *Digital Visual Interface (DVI) revision1.0*. April 1999. Available at: http://www.cs.unc.edu/~stc/FAQs/Video/dvi_spec_V1_0.pdf (accessed 20.03.2016).
9. Johnson G.V., Grjehem M. *Visokoskorostnaia peredacha cifrovih dannih. Visshii kyrs chernoy magii* [High speed transmission of digital data. The highest rate of black magic]. М.: Graham-Williams. 2005. 1024 p. (In Russ.).
10. Goldshtein L.D., Zernov N.V. *Electomagnitnie polia I volni* [Electromagnetic fields and waves]. М.: "Soviet Radio". 1971. 664 p. (In Russ.).
11. Gorbunova A.A., Konovaliyk M.A., Baev A.B. [Model EMI wired data interfaces of computer technology]. *Special'naja tehnika – Special equipment*. 2013. vol. 5. pp. 26–36. (In Russ.).
12. Gulyaykin D.A., Tarusov V.A., Shalaginov V.A. [Model measurement system side electromagnetic radiation]. *Vestnik RSREU – Bulletin RSREU*. 2014. vol. 1. pp. 62–68. (In Russ.).
13. Sazonov D.M. *Anteni i ustroystva SVCH* [Antennas and microwave devices]. М.: Higher School. 1988. 432 p. (In Russ.).

14. Zyuko A.G., Kloovsky D.D., Korzhik V.I. et. al. *Teoria elektricheskoi svyazi: uchebnik dlja vuzov* [The theory of electric communication: textbook for universities. Ed. by D.D. Kloovsky]. M.: Radio and Communications. 1999. 432 p. (In Russ.).
15. Kolesnichenko O.V., Shishigin I.V., Solomenchuk V.G. *Apparatnie sredstva PC* [Hardware PC] SPb.: BHV-Petersburg. 2010. 800 p. (In Russ.).
16. Sovetov B.J., Yakovlev S.A. *Modelirovanie sistem: ucheb. dlja vuzov* [Modeling systems: Proc. for high schools]. M.: Higher School. 2001. 343 p. (In Russ.).
17. Natinal Semiconductor. LVDS Owner's Manual revision 2.0. 2000. Available at: http://www.gaw.ru/pdf/interface/LVDS/LOM_2.pdf. (accessed 20.03.2016).

Васечкин Евгений Александрович — сотрудник, Академия Федеральной службы охраны Российской Федерации. Область научных интересов: техническая защита информации. Число научных публикаций — 6. djonv87@inbox.ru; Приборостроительная, 35, Орел, 302034; р.т.: +7(4862)549533.

Vasechkin Eugene Aleksandrovich — researcher, The Academy of Federal Security Guard Service of the Russian Federation. Research interests: technical protection of information. The number of publications — 6. djonv87@inbox.ru; 35, Priborostroitelnaya Street, Orel, 302034, Russia; office phone: +7(4862)549533.

Таранов Алексей Борисович — сотрудник, Академия Федеральной службы охраны Российской Федерации. Область научных интересов: техническая защита информации. Число научных публикаций — 5. taalbo@yandex.ru; Приборостроительная, 35, Орел, 302034; р.т.: +7(4862)549533.

Taranov Alexei Borisovich — researcher, The Academy of Federal Security Guard Service of the Russian Federation. Research interests: technical protection of information. The number of publications — 5. taalbo@yandex.ru; 35, Priborostroitelnaya Street, Orel, 302034, Russia; office phone: +7(4862)549533.

РЕФЕРАТ

Васечкин Е.А., Таранов А.Б. **Модель сигналов побочных электромагнитных излучений видеоинтерфейсов.**

В результате развития средств обработки и передачи информации, а также возрастания возможностей средств технической разведки все более значимой становится задача решения проблем защиты информации. При этом для средств вычислительной техники наиболее опасным является канал утечки информации за счет побочных электромагнитных излучений. Наиболее уязвимой к утечке по этому каналу является видеосистема. Поэтому существует важная задача, связанная с оценкой энергии побочных электромагнитных излучений от видеосистемы.

Анализ аппаратной реализации оборудования современных видеоинтерфейсов позволил выделить следующие особенности:

- передача сигналов по ним осуществляется с использованием дифференциальных линий;

- информация о цвете пикселей передается по параллельным линиям.

Детальное изучение этих особенностей позволяет выдвинуть гипотезу о взаимодействии побочных электромагнитных излучений от параллельных дифференциальных линий в видеосистеме.

В статье показана зависимость напряженностей электрических и магнитных полей побочного электромагнитного излучения от нескольких дифференциальных линий. Также приведена модель информационных сигналов видеоинтерфейса, учитывающая компенсационные свойства дифференциальных линий. Использование этой модели позволяет выделить последовательности, при передаче которых по интерфейсным линиям будет наблюдаться взаимное влияние побочных электромагнитных излучений от параллельных дифференциальных линий.

По результату проведенного эксперимента можно сделать вывод о необходимости учета взаимного влияния дифференциальных линий в задачах оценки защищенности средств вычислительной техники.

SUMMARY

Vasechkin E.A., Taranov A.B. **A Model of Signals of Side Electromagnetic Radiation of Video Interfaces.**

With the development of means of information processing and transmission, as well as with the increasing opportunities of technical intelligence tools, the problem of information security becomes more vital. For computer equipment, the channel of information leakage due to side electromagnetic radiation poses a serious threat. A video system is the most vulnerable to leakage through this channel. Therefore, there is an important problem related to the energy assessment of side electromagnetic radiation from the video system.

The analysis of hardware implementation of modern video interfaces equipment allowed us to identify the following features:

- signaling is carried out using the differential transmission lines;
- information about pixels color is transmitted along parallel lines.

A detailed study of these features allows us to put forward a hypothesis about the interaction of side electromagnetic radiation from parallel differential lines in the video system.

The article shows the dependence of the electric and magnetic fields of side electromagnetic radiation on several differential lines. We also present a model of video information signals, which takes into account the compensatory properties of differential lines. Using this model allows one to select such sequences, during transmission of which via the interface lines there will be a mutual influence of side electromagnetic radiation from parallel differential lines.

The results of the conducted experiment allowed us to conclude about the necessity to consider the mutual influence of differential lines when evaluating computer equipment security.

К.В. КРОТОВ

**МНОГОУРОВНЕВАЯ МОДЕЛЬ ПОСТРОЕНИЯ РАСПИСАНИЙ
ОБРАБОТКИ ПАРТИЙ ДАННЫХ В КОНВЕЙЕРНЫХ
СИСТЕМАХ ПРИ ФОРМИРОВАНИИ КОМПЛЕКТОВ И
НАЛИЧИИ ОГРАНИЧЕНИЙ**

Кротов К.В. Многоуровневая модель построения расписаний обработки партий данных в конвейерных системах при формировании комплектов и наличии ограничений.

Аннотация. Рассматривается задача построения расписаний обработки партий данных различных типов в конвейерной системе при наличии ограничений на время ее функционирования и условия формирования комплектов из результатов обработки. В статье исследуется первый этап решения указанной задачи, связанный с обоснованием многоуровневой модели принятия решений по составам партий данных, составам групп партий, обрабатываемых в течение заданных интервалов времени функционирования системы, расписаний обработки партий данных в группах с учетом условия формирования из результатов обработки комплектов разных типов.

Ключевые слова: модель многоуровневого программирования, партии данных, группы партий данных, комплекты, формируемые из результатов обработки.

Krotov K.V. A Multilevel Scheduling Model for Data Batches Processing in Conveyor Systems when Forming Sets and in the Presence of Restrictions.

Abstract. The paper considers the task of making schedules for processing data of various types in the conveyor system in the presence of time restrictions on its operation and under condition of forming sets from the processing results. The paper is dedicated to the first stage of solving the above problem. This stage is associated with the rationale of a multilevel model of making decisions on the composition of data batches; composition of groups of batches being processed within the specified time intervals of the system; scheduling of processing of data batches in groups taking into account the conditions of forming sets of various types from the processing results.

Keywords: multilevel programming model, data batches, groups of data batches, sets formed from processing results.

1. Введение. Конвейеризация программ предполагает, что за каждым сегментом конвейера закреплены для выполнения определенные их части [1]. Введем в рассмотрение следующие обозначения: i — идентификатор типа данных, обрабатываемых в системе, n — количество типов обрабатываемых данных ($i = \overline{1, n}$), n^i — количество данных i -го типа, которые должны быть обработаны. Данные i -го типа обрабатываются соответствующей им программой, тогда i — идентификатор программы, выполняемой в составе конвейера, обрабатывающей данные i -го типа. Однократное выполнение конвейеризированной программы i -го типа обеспечивает обработку одного элемента данных i -го типа. Если значение

$n^i \geq 1$ ($i = \overline{1, n}$), тогда обрабатывающая программа должна быть выполнена в конвейерной системе n^i раз. Программы, выполняющие обработку данных, находятся в оперативной памяти каждого из сегментов конвейера. Производительность всех сегментов конвейера является одинаковой, объемы вычислений, выполняемые в соответствии с программным кодом, реализующим обработку данных i -ых типов ($i = \overline{1, n}$) на каждом сегменте, различны. Тогда функционирование каждого сегмента конвейера при обработке данных характеризуется длительностями выполнения им операций с данными каждого i -го типа. Т. к. являются различными длительности выполнения программ на соответствующих сегментах, тогда может быть сформировано расписание обработки данных, представляющее собой порядок запуска программ на выполнение. В результате управление вычислительным процессом в конвейерных системах предполагает определение порядка запуска программ обработки данных на выполнение.

Постановка задачи предполагает задание значений $n^i > 1$ ($i = \overline{1, n}$). Однотипные данные могут быть объединены для обработки в партии. Партия — это совокупность данных одного типа, которые обрабатываются без переналадки (перенастройки) конвейера. Переналадка (перенастройка) сегмента конвейера представляет собой завершение активного состояния программы, выполнявшей обработку данных, и активизацию программы, которая будет реализовывать обработку данных в соответствии с сформированным расписанием. Характеристиками партии являются: тип i и количество данных в ней. Партия является фиксированной, если в нее входят все данные i -го типа. Если партия содержит не все n^i данных, в этом случае в системе обрабатывается ни одна, а несколько партий данных этого типа.

Постановка задачи также предполагает задание ограничений время функционирования системы при обработке данных. Т. е. задается количество интервалов времени функционирования системы при обработке данных и значения длительностей этих интервалов. Формируемые партии данных будут обработаны в течение этих интервалов времени. Совокупность партий, обрабатываемых в течение одного интервала времени функционирования системы, называется группой партий. Группы партий формируются в количестве, соответствующем задаваемому количеству интервалов времени функционирования системы.

Особенностью постановки задачи является наличие условия формирования комплектов разных типов из результатов обработки (составы комплектов разных типов являются заданными). Комплект данных (результатов обработки) — это совокупность результатов обработки разных типов, входящих в него в заданном количестве. То есть составы комплектов являются заданными и не могут быть изменены. Комплекты заданных составов формируются из данных, обрабатываемых в группах партий (обрабатываемых в течение заданных интервалов времени функционирования системы). При формировании комплектов предполагается, что данные всех n типов в определенном количестве входят в комплекты каждого типа.

В соответствии с выполненными рассуждениями управление вычислительным процессом в конвейерной системе предполагает: 1) определение составов партий данных, обрабатываемых в системе; 2) определение составов групп партий (совокупности партий), обрабатываемых в течение интервалов времени функционирования системы, количество и длительности которых задаются в качестве исходных данных решаемой задачи; 3) определение порядка обработки партий данных, входящих в каждую из групп партий (расписаний обработки партий данных). Для решения поставленной задачи определения составов партий, групп партий и порядков обработки партий применены методы, реализующие локальную оптимизацию решений (это связано с доказанным свойством NP-полноты рассматриваемой задачи). Поэтому эффективными решениями по составам партий, групп партий и порядков обработки партий будут являться те решения, которые при реализации локального поиска позволяют получить минимальные значения введенных критериев оптимизации (локально эффективные решения).

2. Анализ публикаций. Развитие современных методов построения расписаний обработки партий представлено в работах [2-11]. В [2] выполнена классификация задач управления обработкой партий (построения расписаний обработки партий). Задачи различаются по виду процесса обработки (непрерывный либо дискретный), способам представления времени моделирования (непрерывное либо дискретное), способам формирования партий и т.д. Управление обработкой партий предполагает построение расписаний для фиксированных партий, определение количества и размера партий до реализации процедуры построения расписаний (алгоритм определения составов партий никак не связан с характеристиками оборудования и процессом обработки, составы партий определяются без связи с построением расписаний),

определение размеров партий совместно с решением задачи построения расписаний их обработки. В [2] рассматривается задача формирования партий и распределения их по обрабатывающим приборам при дискретном времени моделирования для непрерывного производства. При ее решении использован аппарат частично целочисленного линейного программирования (ЧЦЛП). Большая размерность модели (наличие тридцати одного ограничения), использование методов ЧЦЛП, задание ограниченного количества обрабатывающих приборов (ограниченного количества ресурсов) делают затруднительным определение решения для задач большой размерности за ограниченное время. Решение задач обработки партий рассматривается в работе [3], однако указанные там методы построения расписаний предполагают наличие фиксированных партий данных при их обработке на ограниченное количество приборов (задача определения оптимальных составов партий в этой работе не рассматривается). В работах [4, 5] рассматривается задача управления обработкой партий в непрерывном (химическом) производстве. Под партиями в этом случае подразумевается объемы материалов, участвующих в процессе производства (объемы партий определяются способом производства требуемого вида продукта, количеством конечного продукта, определяемым в соответствии со спросом). Таким образом, в [4, 5] реализуется распределение обработки партий (размер которых не оптимизируется) материалов различных видов, обеспечивающих выпуск продуктов разных типов, по параллельно действующим машинам. При этом рассматриваются задачи ограниченной размерности (2 продукта, 2-3 прибора в обрабатывающей конвейерной системе). В работе [6] рассматривается решение задачи среднесрочного планирования выпуска продукции при ограниченном количестве ресурсов и последующего составления расписаний обработки сформированных при планировании партий. При этом под партией подразумевается совокупность изделий одного типа, выпуск которых закрепляется за определенным предприятием или производственным участком. Размеры партий определяются в соответствии с заказами на производство и директивными сроками их выпуска. Модель оптимизации составов партий учитывает только стоимостные параметры выпуска продукции, но не учитывает временные характеристики и особенности технологических процессов. На основе полученного решения по распределению заказов по производственным участкам (сформированным составам партий изделий) решается задача выделения для них ресурсов с целью обработки. Совместно задача планирования (определения составов

партий) и управления выпуском (формирования расписаний обработки партий) в работе не решается. В [6] использованы модель большой размерности и аппарат ЧЦП, что ограничивает размерность решаемой задачи. В работе [7] решается задача определения количества и составов партий единичных (разнотипных) требований, обрабатываемых на одном приборе, с заданными директивными сроками обработки и стоимостью доставки партий. Составы партий разнотипных требований формируются с учетом директивных сроков. Задача предполагает наличие одного обрабатывающего прибора, для которого отсутствуют простои при обработке партий. Тогда формирование партий предполагает и одновременное автоматическое построение расписания их обработки (т.е. вопрос оптимизации использования ресурсов обрабатывающих приборов в данной работе не рассматривается). В результате решается только задача определения составов партий на основе заданных директивных сроков окончания обработки требований. Аналогичная задача формирования партий требований в соответствии с их директивными сроками завершения, обрабатываемых на параллельных машинах, рассматривается в [8]. Объединение требований в партии реализуется с использованием эвристической процедуры в соответствии со значениями параметров начала обработки и директивными сроками окончания обработки требований. Так как обработка двух партий на одной машине (приборе) не может пересекаться, тогда сроки начала и окончания обработки партий используются для их распределения по параллельным машинам (расписание вытекает из сформированных составов партий). В работе [9] решается задача планирования производства полупроводников, предполагающая совместное формирование составов партий обрабатываемых пластин разных типов и расписаний обработки партий в конвейерной системе с параллельно действующими машинами. При определении локально оптимальных решений по составам партий и расписаниям их обработки используется один обобщенный критерий, в котором совместно учитываются эти решения. Особенностью аппарата формирования составов партий является использование метода отжига. Реализуется поиск в окрестностях текущего локально эффективного решения, для формирования новых решений, входящих в окрестности, введены эвристические правила, оперирующие со случайно выбираемыми партиями, изменяющие как составы партий, так и расписания их обработки (изменение позиции партии, изменение обрабатывающего прибора для выполнения операции, создание новых партий). На основе сформированного совместного решения по составам партий и

расписаний их обработки выполняется оценка эффективности полученного расписания с использованием дизъюнктивного графа. Предложенный в [9] подход позволяет реализовывать стохастический поиск локально оптимальных решений. Планированию производственного процесса посвящена работа [10], в которой решается задача распределения заказов на производство продукции по сменным заданиям, распределения заданий по партиям и формирование расписаний обработки партий. Для определения составов сменных заданий используется эвристическая процедура, для определения составов партий разработана имитационная процедура, позволяющая моделировать прохождение партиями конвейерной системы. Определение эффективных составов партий предполагает задание параметров имитационной модели, соответствующих размерам партий изделий каждого типа, и проведение моделирования. В соответствии с результатами моделирования выбираются значения параметров размеров партий, обеспечивающие минимальное значение критерия. После того как с использованием моделирования определены оптимальные размеры партий, выполняется формирование расписаний их обработки. Работа [11] реализует решение задачи определения составов партий компонент, из которых выполняется формирование элементов (аналог формирования комплектов из обработанных в системе изделий). Для формирования партий и расписаний их обработки введена оптимизационная модель, являющаяся многопараметрической и многоиндексной. Определение решений по количеству и составам партий осуществляется в работе путем полного перебора возможных значений этих параметров. Расписание для полученного решения формируется посредством использования эвристической процедуры. В итоге при большой размерности задачи (значительное число типов компонент и количество компонент каждого типа) прямой перебор при формировании партий не обеспечивает решение поставленной задачи за ограниченное время.

Таким образом, решение задачи определения составов партий и построения расписаний их обработки реализуется путем привлечения: 1) аппарата ЧЦЛП (однако при большой размерности задачи получение ее решения за ограниченное время является затруднительным, решения по составам партий формируются без учета расписаний их обработки); 2) методов формирования партий с учетом директивных сроков окончания обработки входящих в них требований (однако формирование групп партий, обрабатываемых в течение заданных интервалов времени, с использованием данных

методов является затруднительным); 3) эвристических процедур и правил (однако применение правил не позволяет получить решения, приближающиеся к оптимальным). В соответствии с этим разработка моделей и методов определения оптимальных составов партий, групп партий и расписаний их обработки является актуальной задачей.

3. Постановка цели и задач научного исследования. Цель работы состоит в совершенствовании методов построения расписаний обработки партий данных в конвейерных системах. Совершенствование методов связано с применением иерархического подхода к решению задачи определения расписаний обработки партий в течение задаваемых интервалов времени функционирования системы. Применение иерархического подхода предусматривает обоснование моделей и методов определения решений на каждом из уровней, на которых функционируют алгоритмы, реализующие формирование составов партий, составов групп партий, обрабатываемых в течение интервалов времени. В основу разрабатываемых методов положен подход к определению эффективных решений, связанный с их локальной оптимизацией (используется подход, связанный с реализацией поиска локально эффективных решений).

4. Основное содержание работы. Рассматриваемая задача является задачей с полной информацией, все параметры, характеризующие обрабатываемые данные (типы данных, количество данных, длительности обработки данных различных типов, составы комплектов различных типов и т.д.) и функционирующую систему (количество сегментов, дисциплина обработки партий, длительности интервалов времени функционирования системы и т.д.) являются заданными. Т.к. через i обозначен идентификатор типа данных, обрабатываемых в системе, тогда через d_i обозначим момент времени поступления в систему данных каждого i -го типа ($i = \overline{1, n}$). Для всех программ моменты времени их загрузки в систему и моменты поступления данных на обработку одинаковы, при этом $d_i = 0$. Обозначим через l индекс сегмента конвейера, осуществляющего выполнение l -й части программы, при этом $l = \overline{1, L}$. Каждым сегментом конвейера выполняются вычисления, соответствующие назначенной для него части программы. Дисциплина обслуживания выполняемых в системе программ предполагает прохождение данными, которые они обрабатывают, всех сегментов конвейера, при этом прерывание обработки данных является недопустимым. Выполнение на каждом l -м сегменте назначенной ему части i -го

программы характеризуется параметром длительности обработки данных, однозначно соответствующей объему вычислений.

Для классификации решаемой задачи построения расписаний выполнены следующие рассуждения. В соответствии с [12] задача составления расписаний может быть представлена в следующем виде: $\alpha|\beta|\gamma$, где параметр α соответствует классу (типу) обрабатываемой системы, β соответствует типу решаемой задачи и характеризует ее особенности, γ – вид критерия оптимизации для определения эффективного решения. Рассматривается конвейерная система, которая в [3] и [12] определена как *FlowShop*. Тогда в качестве α указывается F . Решаемая в системе задача обработки данных предполагает наличие ограничений на ресурсы, тогда в качестве характеристики β задачи указывается *res* [12]. Также решение задачи предполагает объединение обрабатываемых данных в партии. В этом случае в соответствии с [3] в качестве дополнительной характеристики решаемой задачи указывается *batch*. Критерий, в соответствии с которым определяется эффективное расписание, учитывает особенности метода определения порядка обработки партий, реализующего жадный подход. Метод предполагает, что в последовательности обработки партий, порядок выполнения операций с которыми оптимизирован на предыдущих шагах алгоритма, добавляется новая партия данных и определяется ее эффективное местоположение в этих последовательностях. Критерий эффективности решения по порядку обработки партий данных сформирован с точки зрения условия реализации внутренней цели функционирования системы, которая обуславливает необходимость эффективного использования оборудования системы (эффективное использование оборудования системы связано с минимизацией простоев сегментов конвейера при реализации обработки партий). Для классификации задачи в рассмотрение введено обозначение простоя (ожидания) отдельного l -го сегмента конвейера при обработке им партий в виде $T_l^{ож}$. Критерий, используемый при определении локально эффективного расписания обработки партий, учитывает суммарные простои всех сегментов конвейера при реализации операций. Для классификации задачи критерий обозначен в виде

$\sum_{l=1}^L T_l^{ож}$. Тогда рассматриваемая задача с точки зрения [3], [12] может

быть охарактеризована как: $F | res, batch | \sum_{l=1}^L T_l^{ож}$.

Особенностью решаемой задачи является необходимость формирования комплектов из результатов обработки данных. Комплект результатов обработки данных представляет собой совокупность данных, при этом количество данных каждого типа, входящих в комплект одного типа, является заданным. Через g обозначим идентификатор типа комплекта, который формируется из результатов обработки данных в конвейерной системе, а через $g^{ком}$ — общее количество типов формируемых комплектов. Через w_{ig} обозначим количество данных i -го типа, которые должны входить в один комплект g -го типа. В рассмотрение введена матрица (W) , значения элементов w_{ig} которой равны числу данных i -го типа, входящих в один комплект g -го типа.

Обработка партий данных разных типов выполняется с учетом ограничений на время функционирования системы, тогда для идентификации интервалов времени в рассмотрение введены обозначения: t^z — интервал времени, в течение которого может быть реализована обработка партий данных; Z — количество этих интервалов ($z = \overline{1, Z}$). Обработка партий данных при наличии ограничений (интервалов t^z , $z = \overline{1, Z}$) должна быть реализована с точки зрения внешней цели функционирования системы, для достижения которой количество сформированных комплектов должно быть maximизировано. То есть обработка данных направлена на осуществление операций на всех сегментах конвейера с данными разных типов таким образом, чтобы из результатов обработки могло быть сформировано максимальное количество комплектов. Формирование партий данных i -ых типов ($i = \overline{1, n}$) вместо фиксированных партий позволяет получить более эффективное решение (с точки зрения внешней цели функционирования системы — максимизации количества комплектов). При этом должно быть определено как количество партий данных, так и их составы. Так как обработка партий данных i -ых типов ($i = \overline{1, n}$) реализуется в течение заданных интервалов времени функционирования системы

t^z ($z = \overline{1, Z}$), тогда на основе решения по количеству и составам партий формируются группы партий, каждая из которых обрабатывается в течение одного интервала t^z ($z = \overline{1, Z}$).

Входными данными для системы построения расписаний обработки партий являются: типы и количество данных, обрабатываемых в системе; значения интервалов времени t^z функционирования системы при обработке данных; количество Z интервалов времени; матрица (W) , элемент w_{ig} которой равен количеству данных i -го типа, которые должны входить в один формируемый комплект g -ого типа ($i = \overline{1, n}$; $g = \overline{1, g^{KOM}}$). Выходными решениями являются: количество и составы партий данных i -ых типов ($i = \overline{1, n}$); составы групп партий, обрабатываемых системой в течение заданных интервалов времени t^z , определяемые в соответствии с условием формирования из обрабатываемых в группах данных максимального количества комплектов; расписания обработки партий данных каждой группы. В итоге обобщенная цель функционирования системы представлена как совокупность иерархически упорядоченных целей подсистем (функций, решаемых на иерархически упорядоченных уровнях системы построения расписаний обработки партий). Формирование решений на уровнях системы построения расписаний обработки партий осуществляется следующим образом: первый уровень — решения по количеству и составам партий, второй уровень — решения по составам групп партий; третий уровень — решения по порядку обработки входящих в группы партий на сегментах конвейера. Выполненная декомпозиция обобщенной функции системы на совокупность иерархически упорядоченных подфункций позволяет рассмотреть задачу как задачу теории иерархических игр [13, 14].

С целью формирования модели иерархической игры для построения расписаний обработки партий в рассмотрение введены следующие обозначения: m_i — количество партий данных i -го типа ($i = \overline{1, n}$), формируемых на первом уровне принятия решений, при $i = \overline{1, n}$ элементы m_i образуют вектор (M) ; (A) — матрица, элемент a_{ih} которой — это количество данных i -го типа в h -ой партии ($h \leq m_i$), размерность матрицы (A) — $n \times h^{max}$, где $h^{max} = \max_{1 \leq i \leq n} (m_i)$. Решение, формируемое на первом уровне системы (количество и составы партии)

имеет вид: $[(M), (A)]$. Через N^z обозначим группу партий, обрабатываемых в течение одного из интервалов t^z ($z = \overline{1, Z}$). Партии данных i -го типа могут входить в различные группы партий N^z ($z = \overline{1, Z}$). Через m_i^z обозначим количество партий данных i -го типа в группе партий N^z , через $(A)_i^z$ — вектор количества данных i -го типа в m_i^z партиях в группе N^z . Партии данных i -го типа, входящие в группу N^z , определены с использованием набора: $[i, m_i^z, (A)_i^z]$. Группа N^z — совокупность наборов: $N^z = \{[i, m_i^z, (A)_i^z]_k \mid k = \overline{1, k_z}\}^z$, где k_z — количество типов данных, партии которых входят в N^z . При $m_i^z = 0$ набор параметров $[i, m_i^z, (A)_i^z]$ не входит в группу партий N^z . Решение, формируемое на втором уровне системы — совокупность групп партий, имеет вид: $\{N^z \mid (z = \overline{1, Z})\}$.

Расписание обработки партий группы N^z обозначено как π^z , оно представляет собой совокупность последовательностей π^l запуска партий на обработку на каждом l -ом сегменте конвейера ($l = \overline{1, L}$). Расписание π^z имеет вид: $\pi^z = \{\pi^1, \pi^2, \pi^3, \dots, \pi^L\}^z$. Расписание обработки партий π^z формируется в предположении, что порядок обработки партий является одинаковым на всех L сегментах конвейера. Для формализации вида последовательностей π^l расписания π^z в рассмотрении введена матрица порядка обработки партий в системе $(P)^z$; элемент $p_{ij}^z = 1$, если партия данных i -го типа занимает в последовательности π^l j -ю позицию, $p_{ij}^z = 0$ в противном случае, размерность матрицы $k_z \times n_p^z$, где k_z — количество типов данных в партиях группы N^z , n_p^z — количество партий в последовательностях π^l расписания π^z . Порядок обработки партий группы на всех сегментах одинаков, поэтому достаточно определения одной матрицы порядка $(P)^z$. В рассмотрение введена матрица $(R)^z$ — матрица количества данных i -ых типов в партиях, занимающих в последовательности π^l j -е позиции (элемент r_{ij}^z равен

количеству данных i -го типа в партии, занимающей j -ю позицию в π^l , размер матрицы $k_z \times n_p^z$). Решение, формируемое на нижнем уровне иерархии, имеет вид: $\{f(P)^z, (R)^z | z = \overline{1, Z}\}$.

С точки зрения аппарата теории иерархических игр каждым игроком принимается решение, соответствующее функции (подцели) отдельного уровня. Игрок на верхнем уровне принимает решение по составам партий данных, игрок на втором уровне принимает решение по составам групп партий, игрок на третьем уровне принимает решение по расписанию обработки партий данных в конвейерной системе.

В соответствии с функциями уровней системы между ними выполняется обмен информацией следующего вида: 1) на вход первого уровня подаются: количество n типов обрабатываемых данных; количество n^i ($i = \overline{1, n}$) данных каждого типа, обрабатываемых в системе; матрица (W) составов комплектов; с выхода уровня передаются составы партий данных i -ых типов ($i = \overline{1, n}$) — решение $[f(M), (A)]$; 2) на вход второго уровня — составы партий данных; с выхода уровня — сформированные с учетом интервалов времени t^z группы партий данных N^z (решение $\{N^z | (z = \overline{1, Z})\}$); 3) на вход третьего уровня — решение $\{N^z | (z = \overline{1, Z})\}$; с выхода уровня — сформированные расписания π^z ($z = \overline{1, Z}$). При распределении совокупности партий данных i -ых типов (представленной в решении $[f(M), (A)]$) по группам партий N^z ($z = \overline{1, Z}$) состав партий не изменяется (значения m_i и a_{ih} , поступившие с верхнего уровня, изменены быть не могут). В общем виде модель иерархической игры в соответствии с введенными обозначениями имеет следующий вид [14, 15]:

1) первый уровень:

$$f_1([M, A], \{N^{z*} | z = \overline{1, Z}\}) \rightarrow \min(\max); \quad (1)$$

2) второй уровень:

$$f_2(\{N^z | z = \overline{1, Z}\}, \{[(P)^z, (R)^z]^* | z = \overline{1, Z}\}) \rightarrow \min(\max); \quad (2)$$

3) третий уровень:

$$f_3(\{N^z | z = \overline{1, Z}\}, \{[(P)^z, (R)^z] | z = \overline{1, Z}\}) \rightarrow \min(\max). \quad (3)$$

Комплекты формируются только из данных, входящих в партии, включенные в соответствующие группы (партий, которые будут обработаны в соответствии с построенными расписаниями). Поэтому для текущих составов партии наилучшим решением по составам групп партий будет являться решение, которое обеспечивает максимальное количество формируемых комплектов. Данное условие соответствует внешней цели функционирования системы, в соответствии с которой необходимо реализовать выпуск большего количества комплектов в течение заданных временных интервалов. Тогда для определения эффективности решений на втором уровне в качестве критерия используется количество комплектов, сформированных из данных, обработанных в составе групп партий. Решение $\{ N^{z*} | (z = \overline{I, Z}) \}$ соответствует максимальному количеству комплектов разных типов, которые могут быть сформированы из данных, обработанных в группах.

Количество данных, обработанных в течение интервалов t^z , зависит от количества и составов сформированных партий (составы групп партий зависят от количества и составов партий). Решение по порядку обработки партий в каждой из групп является зависящим от составов партий данных в этой группе. Таким образом, определение количества и составов партий данных различных типов, распределение партий по группам, построение расписаний обработки партий групп с учетом ограничений на длительность интервалов t^z должно обеспечить формирование максимального количества комплектов.

Для формирования решения по составам комплектов данных на основе решения по группам $\{ N^z | (z = \overline{I, Z}) \}$ в рассмотрение введена матрица (N^{KOM}) , значения элементов g -го столбца которой соответствуют количествам данных i -ых типов, включаемых в комплекты g -го типа (элемент n_{ig}^{KOM} соответствует количеству данных i -го типа, которые будут входить в состав комплектов g -го типа). Для идентификации количества комплектов, сформированных из данных, обработанных в группах $N^z (z = \overline{I, Z})$, в рассмотрение введен вектор N^K , значение g -го элемента которого соответствует количеству комплектов g -го типа (элемент n_g^K — количество комплектов g -го типа, сформированных из данных, обработанных в группах $N^z (z = \overline{I, Z})$). Для реализации алгоритма определения составов комплектов введен вектор N^{gz} , элемент n_i^{gz} — это количество данных

i -го типа, обрабатываемых в группах партий $N^z (z = \overline{1, Z})$, которые могут быть распределены по комплектам g^{KOM} типов. Перед началом реализации алгоритма распределения данных по комплектам элемент

$$n_i^{gz} \text{ определяется следующим образом: } n_i^{gz} = \sum_{z=1}^Z \sum_{h=1}^{m_i^z} (a_h)_i^z.$$

Алгоритм формирования комплектов предполагает, что за одну итерацию данные каждого типа в количестве $n_i^{gz} (i = \overline{1, n})$ равномерно распределяются по комплектам всех g^{KOM} типов. При формировании комплекта некоторого g -го типа для всех n типов данных выполняется проверка условий $n_i^{gz} \geq w_{ig} (i = \overline{1, n})$. Если для каждого i -го типа данных ($i = \overline{1, n}$) введенное условие выполняется, тогда в комплект рассматриваемого g -го типа выделяется по w_{ig} данных каждого типа. В этом случае количество комплектов рассматриваемого g -го типа увеличивается: $n_g^K = n_g^K + 1$, выполняется изменение значений n_i^{gz} для каждого i -го типа данных ($i = \overline{1, n}$): $n_i^{gz} = n_i^{gz} - w_{ig}$, количество данных i -ых типов, использованных при формировании комплекта g -го типа изменяется: $n_{ig}^{KOM} = n_{ig}^{KOM} + w_{ig} (i = \overline{1, n})$. На следующем шаге реализуется формирование одного экземпляра комплекта $(g+1)$ -го типа и т.д. Итерации алгоритма продолжаются до тех пор, пока хотя бы для одного из типов комплектов и всех n типов данных выполняется условие $n_i^{gz} \geq w_{ig} (i = \overline{1, n})$. После распределения данных i -ых типов в количестве $n_i^{gz} (i = \overline{1, n})$ по комплектам количество

комплектов различных типов вычисляется выражением $\sum_{g=1}^{g^{KOM}} n_g^K$.

Количество данных i -го типа, вошедших во все сформированные комплекты определяется следующим образом: $\sum_{g=1}^{g^{KOM}} n_g^K \cdot w_{ig}$. Количество

данных i -го типа, обрабатываемых в группах партий $N^z (z = \overline{1, Z})$, но не включаемых в составы ни одного из сформированных комплектов,

определяется выражением: $\sum_{z=1}^Z \sum_{h=1}^{m_i^z} (a_h)_i^z - \sum_{g=1}^{g^{KOM}} n_g^k \cdot w_{ig}$. Общее количество данных различных типов, обрабатываемых в группах партий, но не включаемых в комплекты, вычисляется выражением:

$$\sum_{i=1}^n \left(\sum_{z=1}^Z \sum_{h=1}^{m_i^z} (a_h)_i^z - \sum_{g=1}^{g^{KOM}} n_g^k \cdot w_{ig} \right).$$

Для определения вида модели вычислительного процесса обработки партий данных, включенных в группы N^z ($z = \overline{1, Z}$) (при наличии ограничений на длительность интервалов t^z), введены обозначения: t_i^l — длительность обработки данных i -го типа на l -ом сегменте конвейера ($l = \overline{1, L}$); t_{ik}^l — длительность переналадки l -го сегмента с обработки данных i -го типа на обработку данных k -го типа; t_{ii}^l — длительность первоначальной наладки l -го сегмента на обработку данных i -го типа; t_{ji}^{nl} — момент времени начала обработки партии данных i -го типа, занимающей в последовательности π^l j -ю позицию; $(t_{ji}^{nl})^z$ — матрица моментов времени начала обработки партий данных i -ых типов, занимающих в π^l j -е позиции (для группы партий N^z); t_{jq}^{ol} — момент времени начала обработки данных с q -ым порядковым номером в партии, занимающей j -ю позицию в последовательностях π^l ; $(t_{jq}^{ol})^z$ — матрица моментов времени начала обработки q -ых данных в партии, занимающей в π^l j -ю позицию ($q = \overline{1, n_j^z}, n_j^z = \sum_{h=1}^{n^z} r_{hj}^z$, где n_j^z — количество данных в партии, входящей в группу N^z , занимающей j -ю позицию в π^l). С использованием элементов матриц $(P)^z$ и $(t_{jq}^{ol})^z$ элементы матрицы $(t_{ji}^{nl})^z$ определяются следующим образом: $[t_{ji}^{nl}]^z = p_{ij}^z \cdot [t_{jl}^{ol}]^z$, где $i = \overline{1, k_z}, j = \overline{1, n_p^z}$, $[t_{jl}^{ol}]^z$ — момент времени начала обработки первых данных в партии, занимающей j -ю позицию в π^l . В рассмотрение введена матрица переналадок (t_{ik}^l) ,

элементы t_{ik}^l которой соответствуют длительностям переналадки сегментов с обработки данных i -го типа на обработку данных k -го типа, элементы t_{ii}^l — длительности первоначальной наладки сегментов на обработку данных i -ых типов. Выполним дальнейшие рассуждения по обоснованию модели вычислительного процесса обработки партий данных в конвейерной системе для одной из групп N^z (опуская при этом индекс группы z).

Формализация выражений для определения параметров модели вычислительного процесса t_{ji}^{nl} , t_{qj}^{ol} ($i = \overline{1, k_z}, j = \overline{1, n_p}^l, l = \overline{1, L}$) выполнена следующим образом. Для первого сегмента выражения для t_{jq}^{ol} формируются следующим образом: 1) если t_{11}^{ol} — момент времени начала обработки данных ($q=1$) в партии, занимающей первую позицию в последовательности π^l , t_{1i}^{nl} — момент времени начала обработки партии данных i -го типа, занимающей в π^l первую позицию, тогда $t_{11}^{ol} = t_{1i}^{nl} = t_{ii}^l$, 2) выражение $\sum_{h=1}^{k_z} t_{hh}^l \cdot p_{hl}$ определяет интервал первоначальной наладки первого сегмента на обработку данных i -го типа; $(q-1) \sum_{h=1}^{k_z} t_h^l \cdot p_{hl}$ — время обработки предыдущих $(q-1)$ -их данных в этой партии, $(t)^l$ — вектор длительностей обработки на первом приборе данных (t_i^l — элемент вектора), тогда значения t_{1q}^{ol} моментов времени начала обработки любых данных в первой партии в π^l определяется выражением вида:

$$t_{1q}^{ol} = \sum_{h=1}^{k_z} t_{hh}^l \cdot p_{hl} + (q-1) \sum_{h=1}^{k_z} t_h^l \cdot p_{hl}; \quad (4)$$

3) если t_{i,i_2}^{lnep} — интервал переналадки первого сегмента конвейера с обработки данных i -го типа (первая позиция партии в π^l) на обработку данных другого типа (вторая позиция партии в π^l), t_{21}^{ol} — момент времени начала обработки данных в первой позиции во второй партии (t_{2i}^{nl} — начало обработки этой партии данных i -го типа в π^l

($t_{2i}^{nl} = t_{2i}^{0l}$), тогда с учетом (4) значения t_{2i}^{0l} (t_{2i}^{nl}) определяются выражением вида:

$$t_{2i}^{nl} = t_{2i}^{0l} = \sum_{h=1}^{k_z} t_{hh}^l \cdot p_{hl}^l + \sum_{h=1}^{k_z} t_h^l \cdot r_{hl}^l + t_{i_1 i_2}^{Inep}, \quad (5)$$

где значение $t_{i_1 i_2}^{Inep}$ определяется следующим образом:

$$t_{i_1 i_2}^{Inep} = t_{i_r i_r'}^l, \text{ где } \begin{cases} i_r = i \mid p_{i_l}^l = l, i = \overline{1, k_z} \\ i_r' = j \mid p_{j_2}^l = l, j = \overline{1, k_z} \end{cases}.$$

Выражение $\sum_{h=1}^{k_z} t_{hh}^l \cdot p_{hl}^l + \sum_{h=1}^{k_z} t_h^l \cdot r_{hl}^l + t_{i_1 i_2}^{Inep}$ позволяет определить момент времени начала обработки второй в π^l партии данных, выражение $(q-1) \sum_{h=1}^{k_z} t_h^l \cdot p_{h2}^l$ — длительность обработки данных, предшествующих данным в q -ой позиции в этой партии, тогда время начала обработки q -ых данных i -го типа в партии, занимающей вторую позицию в π^l , определяется следующим образом:

$$t_{2q}^{0l} = \sum_{h=1}^{k_z} t_{hh}^l \cdot p_{hl}^l + \sum_{h=1}^{k_z} t_h^l \cdot r_{hl}^l + t_{i_1 i_2}^{Inep} + (q-1) \sum_{h=1}^{k_z} t_h^l \cdot p_{h2}^l, \quad q = \overline{1, n_2^z}. \quad (6)$$

По аналогии с (5), (6) сформированы выражения для вычисления моментов времени начала обработки третьей, четвертой партий и данных в q -ых позициях в них:

$$\begin{aligned} t_{3i}^{nl} &= \sum_{p=1}^{k_z} t_{hh}^l \cdot p_{hl}^l + \sum_{h=1}^{k_z} t_h^l \cdot r_{hl}^l + t_{i_1 i_2}^{Inep} + \sum_{h=1}^{k_z} t_h^l \cdot r_{h2}^l + t_{i_2 i_3}^{Inep}; \\ t_{3q}^{0l} &= \sum_{h=1}^{k_z} t_{hh}^l \cdot p_{hl}^l + \sum_{h=1}^{k_z} t_h^l \cdot r_{hl}^l + t_{i_1 i_2}^{Inep} + \sum_{h=1}^{k_z} t_h^l \cdot r_{h2}^l + t_{i_2 i_3}^{Inep} + (q-1) \sum_{h=1}^{k_z} t_h^l \cdot p_{h3}^l; \\ t_{4i}^{nl} &= \sum_{h=1}^{k_z} t_{hh}^l \cdot p_{hl}^l + \sum_{h=1}^{k_z} t_h^l \cdot r_{hl}^l + t_{i_1 i_2}^{Inep} + \sum_{h=1}^{k_z} t_h^l \cdot r_{h2}^l + t_{i_2 i_3}^{Inep} + \sum_{h=1}^{k_z} t_h^l \cdot r_{h3}^l + t_{i_3 i_4}^{Inep}; \\ t_{4q}^{0l} &= \sum_{h=1}^{k_z} t_{hh}^l \cdot p_{hl}^l + \sum_{h=1}^{k_z} t_h^l \cdot r_{hl}^l + t_{i_1 i_2}^{Inep} + \sum_{h=1}^{k_z} t_h^l \cdot r_{h2}^l + t_{i_2 i_3}^{Inep} + \\ &+ \sum_{h=1}^{k_z} t_h^l \cdot r_{h3}^l + t_{i_3 i_4}^{Inep} + (q-1) \sum_{h=1}^{k_z} t_h^l \cdot p_{h4}^l. \end{aligned}$$

Тогда моменты времени начала обработки партии данных i -го типа, занимающей j -ю позицию в π^l , моменты времени начала обработки q -ых данных в партии в j -й позиции в π^l , определяются выражениями вида:

$$t_{ji}^l = \sum_{h=1}^{k_z} t_{hh}^l \cdot p_{hl}^l + \sum_{f=1}^{j-1} \sum_{h=1}^{k_z} t_h^l \cdot r_{hf}^l + \sum_{h=1}^{j-1} t_{i_h t_{h+1}}^{l nep},$$

$$t_{jq}^{0l} = \sum_{h=1}^{k_z} t_{hh}^l \cdot p_{hl}^l + \sum_{f=1}^{j-1} \sum_{h=1}^{k_z} t_h^l \cdot r_{hf}^l + \sum_{h=1}^{j-1} t_{i_h t_{h+1}}^{l nep} + (q-l) \sum_{h=1}^{k_z} t_h^l \cdot p_{hj}^l.$$

Выражение для t_{jq}^{0l} ($q = \overline{1, n_j^z}$, при $l > 1$) сформированы следующим образом. Для $l=2$ и $j=1$ имеем:

$$t_{11}^{02} = t_{1i}^{n_2} = \max\left(\sum_{h=1}^{k_z} t_{hh}^2 \cdot p_{h1}; t_{11}^{01} + \sum_{h=1}^{k_z} t_h^1 \cdot p_{h1}\right);$$

$$t_{12}^{02} = \max\left(t_{11}^{02} + \sum_{h=1}^{k_z} t_h^2 \cdot p_{h1}; t_{12}^{01} + \sum_{h=1}^{k_z} t_h^1 \cdot p_{h1}\right);$$

$$t_{1n_1}^{02} = \max\left(t_{1, n_1-1}^{02} + \sum_{h=1}^{k_z} t_h^2 \cdot p_{h1}; t_{1n_1}^{01} + \sum_{h=1}^{k_z} t_h^1 \cdot p_{h1}\right), \text{ где } n_1^z = \sum_{i=1}^{k_z} r_{i1}^z.$$

Выражения для определения t_{2q}^{02} ($q = \overline{1, n_2^z}$, $n_2^z = \sum_{h=1}^{k_z} r_{h2}^z$) имеют вид:

$$t_{21}^{02} = t_{2i}^{n_2} = \max\left(t_{1n_1}^{02} + \sum_{h=1}^{k_z} t_h^2 \cdot p_{h1} + t_{i_1 i_2}^{2 nep}; t_{21}^{01} + \sum_{h=1}^{k_z} t_h^1 \cdot p_{h,2}\right);$$

$$t_{22}^{02} = \max\left(t_{21}^{02} + \sum_{h=1}^{k_z} t_h^2 \cdot p_{h2}; t_{22}^{01} + \sum_{h=1}^{k_z} t_h^1 \cdot p_{h2}\right);$$

$$t_{2n_2}^{02} = \max\left(t_{2, n_2-1}^{02} + \sum_{h=1}^{k_z} t_h^2 \cdot p_{h2}; t_{2, n_2}^{01} + \sum_{h=1}^{k_z} t_h^1 \cdot p_{h2}\right).$$

Вид выражений для t_{3i}^{02} и t_{3q}^{02} :

$$t_{31}^{02} = t_{3i}^{n_2} = \max\left(t_{2n_2}^{02} + \sum_{h=1}^{k_z} t_h^2 \cdot p_{h2} + t_{i_2 i_3}^{2 nep}; t_{31}^{01} + \sum_{h=1}^{k_z} t_h^1 \cdot p_{h3}\right);$$

$$t_{3q}^{02} = \max\left(t_{3, q-1}^{02} + \sum_{h=1}^{k_z} t_h^2 \cdot p_{h3}; t_{3q}^{01} + \sum_{h=1}^{k_z} t_h^1 \cdot p_{h3}\right), \text{ где } q = \overline{2, n_3^z}, n_3^z = \sum_{h=1}^{k_z} r_{h3}^z.$$

Сформированы обобщенные выражения для t_{jq}^{0l} (с учетом особенностей для t_{ll}^{0l} и t_{jl}^{0l} , $l = \overline{2, L}$, $q = \overline{1, n_j^z}$, $n_j^z = \sum_{h=1}^{k_z} r_{hj}^z$) в виде:

$$t_{ll}^{0l} = t_{li}^{nl} = \max\left(\sum_{h=1}^{k_z} t_h^l \cdot p_{hl}; t_{ll}^{0l-1} + \sum_{h=1}^{k_z} t_h^{l-1} \cdot p_{hl}\right);$$

$$t_{jl}^{0l} = t_{ji}^{nl} = \max\left(t_{j-1, n_{j-1}}^{0l} + \sum_{h=1}^{k_z} t_h^l \cdot p_{h, j-1} + t_{i_{j-1}^l}^{lnep}; t_{jl}^{0l-1} + \sum_{h=1}^{k_z} t_h^{l-1} \cdot p_{hj}\right);$$

$$t_{jq}^{0l} = \max\left(t_{jq-1}^{0l} + \sum_{h=1}^{k_z} t_h^l \cdot p_{hj}; t_{jq}^{0l-1} + \sum_{h=1}^{k_z} t_h^{l-1} \cdot p_{hj}\right).$$

Полученные выражения представляют собой модель вычислительного процесса обработки партий данных i -ых типов ($i = \overline{1, n}$) на l -ых сегментах конвейера ($l = \overline{1, L}$). Метод построения расписаний обработки партий данных на третьем уровне системы реализует жадный подход к оптимизации. Он предполагает добавление текущей рассматриваемой партии в конец последовательностей π^l ($l = \overline{1, L}$), в которых на предыдущих шагах алгоритма были размещены партии данных, и определение эффективного местоположения этой партии в π^l (определение для рассматриваемой партии позиций в π^l ($l = \overline{1, L}$), которые обеспечивают локально оптимальное решение). Так как в задаче заданы ограничения на интервалы времени функционирования системы, учитываемым ресурсом является время работы системы. В этом случае решение по порядку обработки партий должно быть построено таким образом, чтобы временной ресурс системы использовался в полной мере. Тогда критерий эффективности расписаний обработки партий соответствует внутренней цели функционирования системы, определяющей необходимость минимизации простоев оборудования при обработке. При определении текущей эффективной позиции рассматриваемой партии в π^l критерий характеризует простои сегментов конвейера при обработке текущего количества партий, находящихся в π^l .

В этом случае критерий эффективности расписания обработки партий на нижнем уровне учитывает: а) время простоя сегментов в

ожидании начала обработки партий данных (с учетом интервалов наладки, переналадки и последующего ожидания); б) время простоя сегментов в ожидании готовности данных при их обработке внутри партий. Простой l -го сегмента в ожидании обработки первой в π^l ($l = \overline{1, L}$) партии равен значению t_{l1}^{0l} , суммарное время простоя сегментов в ожидании начала обработки партий в последовательностях π^l определяется выражением $\sum_{l=1}^L t_{l1}^{0l}$. Интервал

простоя l -го сегмента в ожидании начала обработки следующей партии после окончания обработки предыдущей определяется выражением вида: $t_{j1}^{0l} - \left[t_{j-1, n_{j-1}^z}^{0l} + \sum_{h=1}^{k_z} t_h^l \cdot p_{h, j-1} \right]$, где $j > 1$,

$n_{j-1}^z = \sum_{h=1}^{k_z} r_{h, j-1}^z$ — количество данных в предшествующей $(j-1)$ -ой партии. Суммарный простой l -го сегмента в ожидании начала обработки всех j -ых партий ($j = \overline{1, n_p^z}$) определен следующим образом:

$$\sum_{j=2}^{n_p^z} \left[t_{j1}^{0l} - \left[t_{j-1, n_{j-1}^z}^{0l} + \sum_{h=1}^{k_z} t_h^l \cdot p_{h, j-1} \right] \right].$$

В этом случае суммарной простоя всех L сегментов в ожидании начала обработки партий на них (простой между партиями для всех L сегментов) определяется выражением:

$$\sum_{l=1}^L \sum_{j=2}^{n_p^z} \left[t_{j1}^{0l} - \left[t_{j-1, n_{j-1}^z}^{0l} + \sum_{h=1}^{k_z} t_h^l \cdot p_{h, j-1} \right] \right]. \quad (7)$$

Простой l -го сегмента в ожидании готовности к обработке данных, занимающих q -ю позицию в j -ой партии в последовательности π^l , определяется выражением вида: $t_{jq}^{0l} - \left[t_{j, q-1}^{0l} + \sum_{h=1}^{k_z} t_h^l \cdot p_{hj} \right]$, где $j = \overline{1, n_p^z}$. Это выражение соответствует интервалу между данными в q -ой и $(q-1)$ -ой позициях в j -ой партии в π^l . Тогда суммарный простой

l -го прибора в ожидании готовности к обработке всех данных в j -ой партии в π^l вычисляется с использованием выражения:

$$\sum_{q=2}^{n_j^z} \left[t_{jq}^{0l} - \left[t_{j,q-1}^{0l} + \sum_{h=1}^{k_z} t_h^l \cdot p_{hj} \right] \right], \quad (8)$$

где q – номер позиции данных в j -ой партии в π^l , n_j^z — число данных в этой j -ой партии, $n_j^z = \sum_{h=1}^{k_z} r_{hj}$. На основе (8) общий простой l -го сегмента в ожидании готовности к обработке данных внутри всех партий в π^l определяется выражением вида:

$$\sum_{j=1}^{n_p^z} \sum_{q=2}^{n_j^z} \left[t_{jq}^{0l} - \left[t_{j,q-1}^{0l} + \sum_{h=1}^{k_z} t_h^l \cdot p_{hj} \right] \right].$$

Суммарный простой всех L сегментов конвейера в ожидании готовности данных внутри партий вычисляется выражением:

$$\sum_{l=2}^L \sum_{j=1}^{n_p^z} \sum_{q=2}^{n_j^z} \left[t_{jq}^{0l} - \left[t_{j,q-1}^{0l} + \sum_{h=1}^{k_z} t_h^l \cdot p_{hj} \right] \right]. \quad (9)$$

При оценивании эффективности решения на нижнем уровне иерархии для текущего количества партий, добавленных в π^l ($l = \overline{1, L}$), учитывается общий простой сегментов конвейера при определении порядка обработки этих партий, и должны быть учтены: выражение $\sum_{l=1}^L t_{l1}^{0l}$ и формулы (7) и (9).

Рассматриваемая задача предполагает ограничение времени функционирования системы, тогда при формировании комплектов могут быть использованы только данные, партии которых будут обработаны в группах N^z ($z = \overline{1, Z}$). Максимизация количества комплектов, сформированных из данных, обработанных в составе групп партий, отвечает внешней цели функционирования системы. Тогда для достижения этой цели решение по составам групп $\{N^z | z = \overline{1, Z}\}$ должно анализироваться с точки зрения количества комплектов, которые могут быть сформировано из данных, входящих

в партии, включенные в группы N^z ($z = \overline{I, Z}$). Таким образом, на втором уровне иерархии системы на основе решения по составам групп партий $\{N^z | z = \overline{I, Z}\}$ необходимо определить решение по составам комплектов данных, формируемых из результатов их обработки в группах партий. Общее количество комплектов, сформированных из данных, партии которых включены в группы, определяется выражением, введенным ранее: $\sum_{g=1}^{g^{ком}} n_g^k$. Приведенное

выражение использовано в качестве критерия эффективности решений $\{N^z | z = \overline{I, Z}\}$ по составам групп партий.

На первом уровне иерархии принимается решение по составам партий данных в виде $[(M), (A)]$. В соответствии с видом модели (1) критерий эффективности решений на первом уровне должен учитывать как решение $[(M), (A)]$, формируемое на этом уровне, так и эффективное решение по составам групп партий, формируемое на втором уровне в виде $\{N^{z*} | z = \overline{I, Z}\}$. Решение $\{N^{z*} | z = \overline{I, Z}\}$ характеризуется количеством сформированных комплектов разных типов — вектором N^k (решение $\{N^{z*} | z = \overline{I, Z}\}$ обеспечивает максимальное количество комплектов). Тогда при определении значения критерия f_1 , характеризующего решение $[(M), (A)]$ на первом уровне, используется само это решение и решение по количеству комплектов каждого типа — вектор N^k , характеризующий эффективное решение по составам групп партий $\{N^{z*} | z = \overline{I, Z}\}$.

Входными параметрами для решаемой задачи являются: количество n типов данных, обрабатываемых в системе, количество данных каждого типа n^i ($i = \overline{1, n}$), длительности интервалов времени t^z ($z = \overline{I, Z}$) функционирования системы при обработке данных. Предполагается, что все данные, подаваемые на вход системы для обработки, должны быть использованы при формировании комплектов, то есть все данные, поданные на вход, обрабатываются системой и из них формируются комплекты. Тогда внешняя цель функционирования системы предполагает, что все данные, поданные на вход конвейера, обрабатываются и используются при формировании комплектов. В этом случае решение $[(M), (A)]$ охарактеризовано количеством данных, которые должны быть

обработаны в соответствии с задаваемыми входными характеристиками и включены в комплекты, но в соответствии с решением N^k в комплекты не вошли. В соответствии с решением $[(M), (A)]$ количество данных i -го типа, которое должно быть

обработано в системе определяется выражением вида: $\sum_{h=1}^{m_i} a_{ih}$, где

a_{ih} — элемент матрицы (A) (количество данных i -го типа в h -ой партии ($h = \overline{1, m_i}$)), m_i — количество партий данных i -го типа. Общее количество данных различных типов, которые должны быть

обработаны в системе, определяется выражением $\sum_{i=1}^n \sum_{h=1}^{m_i} a_{ih}$. Количество

данных различных типов, используемых при формировании комплектов, определяется выражением $\sum_{i=1}^n \sum_{g=1}^{g^{KOM}} n_{ig}^{KOM} \cdot w_{ig}$. Тогда

количество данных, которые должны быть обработаны в соответствии с решением $[(M), (A)]$, но не использованы при формировании комплектов, определяется выражением вида:

$$\sum_{i=1}^n \sum_{h=1}^{m_i} a_{ih} - \sum_{i=1}^n \sum_{g=1}^{g^{KOM}} n_{ig}^{KOM} \cdot w_{ig}.$$

Полученное выражение позволяет охарактеризовать достижение внешней цели функционирования системы при формировании составов партий данных.

Для задания индекса последней партии в последовательности π^L использовано обозначение n_p^z (где n_p^z — количество партий, входящих в группу N^z). Если через n_j^z обозначено количество данных в партии, входящей в группу N^z и занимающей j -ю позицию в последовательности π^L , тогда при $j = n_p^z$ через $(t_{j, n_j^z}^{0L})^z$ обозначен момент времени начала обработки последних данных в партии с индексом n_p^z на L -ом сегменте конвейера, а окончание обработки этой партии определяется выражением $(t_{j, n_j^z}^{0L})^z + \sum_{i=1}^{k_z} (t_i^L)^z \cdot p_{i,j}^z$ (при $j = n_p^z$).

При задании значений t^z ($z = \overline{1, Z}$) ограничение на время обработки партий данных, входящих в группы N^z , имеет вид:

$[(t_{j,n_j}^{0L})^z + \sum_{i=1}^{k_z} (t_i^L)^z \cdot p_{i,j}^z] \leq t^z$ при условии $j = n_p^z$ ($z = \overline{1, Z}$). В

соответствии с выполненными рассуждениями модель многоуровневого программирования для определения эффективных составов партий, составов групп партий и расписаний обработки партий имеет вид (с учетом (1-3)):

$$1) \text{ первый уровень: } \min f_1, f_1 = \sum_{i=1}^n \sum_{h=1}^{m_i} a_{ih} - \sum_{i=1}^n \sum_{g=1}^{g^{KOM}} n_{ig}^{KOM} \cdot w_{ig};$$

$$2) \text{ второй уровень: } \max f_2, f_2 = \sum_{g=1}^{g^{KOM}} n_g^k;$$

$$3) \text{ третий уровень: } \min f_3^z \text{ (} z = \overline{1, Z} \text{), где:}$$

$$f_3^z = \sum_{l=2}^L [t_{l1}^{0l}]^z + \sum_{l=1}^L \sum_{j=2}^{n_p^z} \left[[t_{jl}^{0l}]^z - \left[[t_{j-1,n_{j-1}}^{0l}]^z + \sum_{h=1}^{k_z} [t_h^l]^z \cdot p_{h,j-1}^z \right] \right] + \\ + \sum_{l=2}^L \sum_{j=1}^{n_p^z} \sum_{q=2}^{n_j^z} \left[[t_{jq}^{0l}]^z - \left[[t_{j,q-1}^{0l}]^z + \sum_{h=1}^{k_z} [t_h^l]^z \cdot p_{hj}^z \right] \right].$$

4) ограничения на длительность реализации обработки партий в N^z

$$(z = \overline{1, Z}): \left[(t_{j,n_j}^{0L})^z + \sum_{i=1}^{k_z} (t_i^L)^z \cdot p_{i,n_j}^z \right] \leq t^z \text{ при } j = n_p^z (z = \overline{1, Z}).$$

Таким образом, модель многоуровневого программирования представлена в виде совокупности критериев оптимизации на каждом уровне иерархии принятия решений и ограничения на время функционирования системы при обработке данных.

5. Заключение. Результатом выполненных исследований является сформированная многоуровневая модель построения комплексных расписаний обработки партий с формированием из обработанных данных комплектов различных типов (при ограничении на время реализации выполнения операций с партиями в группах). Использование сформированной модели позволяет определить: 1) эффективные составы партий (количество партий, число данных в них), для которых общее количество данных, не вошедших в состав комплектов разных типов, будет минимальным; 2) эффективных составов групп партий, обрабатываемых в течение установленных

интервалов t^z , для которых общее количество сформированных комплектов будет максимальным; 3) эффективных порядков обработки партий на сегментах конвейера. Дальнейшие исследования направлены на разработку методов построения комплексных расписаний обработки партий данных при формировании комплектов и наличии ограничений на длительности интервалов реализации операций с партиями в группах, соответствующих сформированной модели.

Литература

1. *Хьюз К., Хьюз Т.* Параллельное и распределенное программирование на C++ // М.: Вильямс. 2004. 672 с.
2. *Mendez C.A. et al.* State-of-the-art review of optimization methods for short-term scheduling of batch processes // *Computers and Chemical Engineering*. 2006. vol. 30. pp. 913–946.
3. *Ковалев М.М.* Модели и методы календарного планирования: Курс лекций // Минск: Изд-во БГУ. 2004. 63 с.
4. *Tan Y., Huangi W., Sun Y., Yue Y.* Comparative Study of Different Approaches to Solve Batch Process Sheduling and Optimisation Problems // *Proceedings of the 18th International Conference on Automation & Computing*. Loughborough University. Leicestershire. UK. 2012. pp. 424–444.
5. *Adonyi R., Romero J., Puigjaner L., Friedler F.* Incorporating heat integration in batch process scheduling // *Applied Thermal Engineering*. 2003. vol. 23. pp. 1743–1762.
6. *Kreipl S.* Planning and Scheduling in supply chains: An Overview of Issues in Practice // *Production and Operations Management*. 2004. vol. 17. no. 1. pp. 77–92.
7. *Steiner G., Zhang R.* Minimizing the weight number of late jobs with Batch setup times and delivery costs on a single machine // *Vienna: Itech Education and Published*. 2007. 436 p.
8. *Koehler F. Khuller S.* Optimal Batch Schedules for Parrallel Machines // *Algorithms and Data Structures: 13th International Symposium*. Berlin: Springer. 2013. pp. 475–486.
9. *Yugma C., et al.* Batching and Scheduling Algoritm for the diffusion Area in Semiconductor Manufacturing // *International Journal of Production Research*. 2012. vol. 50. no. 8. pp. 2118–2132.
10. *Chandra P.* Managing Batch Processors to reduce lead time in a semiconductor packaging line // *International Jornal of Production Research*. 1997. vol. 35(3). pp. 611–633.
11. *Surjandari I., Rachman A., Dhini A.* The Batch Sheduling Model for Dynamic multiitem, Multilevel Production in an assembly Job-Shop with Parrallel Machines // *International Journal of Technology*. 2015. vol. 1. pp. 84–96.
12. *Lawler E.L. et al.* Sequencing and scheduling: Algorithms and complexity // *Department of Operations Research, Statistics, and System Theory [BS]*. 1989. vol. R 8909. pp. 1–70.
13. *Петросян Л.А., Зенкевич Н.А., Семина Е.А.* Теория игр // М.: Высшая школа. 1999. 300 с.
14. *Гермейер Ю.Б.* Игры с непротивоположными интересами // М.: Наука. 1976. 327 с.

References

1. Hughes C., Hughes T. Parallel and distributed programming using C++. Addison-Wesley Professional. 2004. 720 p. (Russ. ed.: H'juz K., H'juz T. *Parallel'noe i raspredeleennoe programirovanie na C++*. M.: Vil'jams. 2004. 672 p.).
2. Mendez C.A. et al. State-of-the-art review of optimization methods for short-term scheduling of batch processes. *Computers and Chemical Engineering*. 2006. vol. 30. pp. 913–946.
3. Kovalev M.M. *Modeli i metody kalendarnogo planirovanija: Kurs lekcij* [Models and methods of scheduling. A course of lectures]. Minsk: Izd-vo BGU. 2004. 63 p. (In Russ.).
4. Tan Y. Huangi W., Sun Y., Yue Y. Comparative Study of Different Approaches to Solve Batch Process Sheduling and Optimisation Problems. Proceedings of the 18th International Conference on Automation & Computing. Loughborough University. Leicestershire. UK. 2012. pp. 424–444.
5. Adonyi R., Romero J., Puigjaner L., Friedler F. Incorporating heat integration in batch process scheduling. *Applied Thermal Engineering*. 2003. vol. 23. pp. 1743–1762.
6. Kreipl S. Planning and Scheduling in supply chains: An Overview of Issues in Practice. *Production and Operations Management*. 2004. vol. 17. no. 1. pp. 77–92.
7. Steiner G., Zhang R. Minimizing the weight number of late jobs with Batch setup times and delivery costs on a single machine. Vienna: Itech Education and Published. 2007. 436 p.
8. Koehler F. Khuller S. Optimal Batch Schedules for Parrallel Machines. Algorithms and Data Structures: 13th International Symposium. Berlin: Springer. 2013. pp. 475–486.
9. Yugma C., et al. Batching and Scheduling Algorithm for the diffusion Area in Semiconductor Manufacturing. *International Journal of Production Research*. 2012. vol. 50. no. 8. pp. 2118–2132.
10. Chandra P. Managing Batch Processors to reduce lead time in a semiconductor packaging line. *International Journal of Production Research*. 1997. vol. 35(3). pp. 611–633.
11. Surjandari I., Rachman A., Dhini A. The Batch Sheduling Model for Dynamic multiitem, Multilevel Production in an assembly Job-Shop with Parrallel Machines. *International Journal of Technology*. 2015. vol. 1. pp. 84–96.
12. Lawler E.L. et al. Sequencing and scheduling: Algorithms and complexity // Department of Operations Research, Statistics, and System Theory [BS]. 1989. vol. R 8909. pp. 1–70.
13. Petrosyan L.A., Zenkevich N.A., Semina E.A. *Teorija igr* [Game theory]. M.: Vysshaja shkola. 1999. 300 p. (In Russ.).
14. Germeier Y.B.: *Igry s neprotivopolozhnyimi interesami* [Game with no conflicting interests]. M: Nauka. 1976. 327 p. (In Russ.).

Кротов Кирилл Викторович — к-т техн. наук, доцент кафедры информационных систем, ФГАОУ ВО Севастопольский Государственный университет (СевГУ). Область научных интересов: теория расписаний, методы локальной оптимизации решений, многоуровневое программирование. Число научных публикаций — 70. krotov_k1@mail.ru; ул. Университетская, 33, Севастополь, 299053; р.т.: +7(978)730-38-19.

Krotov Kirill Victorovich — Ph.D., associate professor of information systems department, Sevastopol State University. Research interests: theory of schedules, methods of local optimization solutions, multi-level programming. The number of publications — 70. krotov_k1@mail.ru; 33, Universitetskaja str., Sevastopol, 299053; office phone: +7(978)730-38-19.

РЕФЕРАТ

***Кротов К.В.* Многоуровневая модель построения расписаний обработки партий данных в конвейерных системах при формировании комплектов и наличии ограничений.**

В настоящее время актуальными являются задачи управления вычислительными процессами в конвейерных системах. Одной из задач управления вычислительными процессами является составление расписаний обработки партий данных в конвейерной системе при наличии ограничений на интервалы времени ее функционирования и условия формирования комплектов из результатов обработки. В работе выполнена декомпозиция обобщенной цели функционирования системы построения расписаний обработки партий на совокупность взаимосвязанных иерархически упорядоченных подцелей, предполагающих определение составов партий, составов групп партий и расписаний обработки партий. В соответствии с выполненной декомпозицией и сформированными подцелями функционирования подсистем применен аппарат многоуровневого программирования для решения обобщенной задачи построения расписаний обработки партий данных. В работе реализуется обоснование многоуровневой модели (модели иерархической игры) построения комплексных расписаний обработки партий данных в конвейерной системе: обосновываются критерии оптимизации решений на каждом из уровней иерархии и ограничения на время обработки партий в системе.

SUMMARY

***Krotov K.V.* A Multilevel Scheduling Model for Data Batches Processing in Conveyor Systems when Forming Sets and in the Presence of Restriction.**

Nowadays computational process control in conveyor systems is a topical issue. Scheduling of data processing in a conveyor system with time restrictions on its operation and under condition of the formation of sets from processing results is one of the computing process management tasks. We have performed decomposition of the generalized goal of functioning of a scheduling system for data processing into a set of interrelated hierarchically arranged sub-goals that involve determining composition of data batches, groups of data batches and processing scheduling. A multilevel programming device for solving the generalized task of processing scheduling was applied in accordance with the performed decomposition and sub-goals of subsystems operation. The paper substantiates the use of the multilevel model (hierarchical game model) for complex scheduling of data processing in the conveyor system, i.e. the criteria of optimizing solutions at each hierarchical level and time restrictions on data batches processing in the system.

С.Н. КАРПОВИЧ
**МНОГОЗНАЧНАЯ КЛАССИФИКАЦИЯ ТЕКСТОВЫХ
ДОКУМЕНТОВ С ИСПОЛЬЗОВАНИЕМ ВЕРОЯТНОСТНОГО
ТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ ml-PLSI**

Карпович С.Н. Многозначная классификация текстовых документов с использованием вероятностного тематического моделирования ml-PLSI.

Аннотация. В работе рассмотрен подход к многозначной классификации текстовых документов на основе вероятностного тематического моделирования. На базе корпуса SCTM-ru построена тематическая модель методом обучения с учителем, приведен алгоритм многозначной классификации. Описан состав программного прототипа, реализующего предложенный подход.

Ключевые слова: многозначная (нечеткая) классификация, обучение с учителем, тематическое моделирование, обработка текста на естественном языке.

Карпович S.N. Multi-Label Classification of Text Documents using Probabilistic Topic Model ml-PLSI.

Abstract. In this paper, we describe an approach to multi-label classification of text documents based on probabilistic topic modeling. On the basis of SCTM-ru a topic model has been built with the help of supervised learning. A multi-label classification algorithm is presented. We propose tools for multi-label classification implementing this approach.

Keywords: multi-label classification, supervised learning, topic model, natural language processing.

1. Введение. В настоящее время количество информации, которое создается в текстовом электронном виде колоссально, год от года объем этой информации увеличивается. Лингвисты, маркетологи и аналитики нуждаются в инструментах автоматической обработки текстов на естественном языке. Одним из направлений обработки текстов является категоризация. Категоризация выполняется различными алгоритмами классификации и кластеризации. Задачи классификации волновали исследователей с середины XIX века, в работе [1] дан обзор ряда алгоритмов классификации 1969–1980 гг. Под задачей кластеризации документов понимают задачу разбиения заданной выборки текстовых документов на непересекающиеся подмножества — кластеры, так чтобы каждый кластер состоял из похожих документов, а документы разных кластеров существенно отличались. Под задачей классификации документов понимают задачу отнесения документа к одной из нескольких категорий (классов, тем) на основании содержания документа. Классификация относится к задачам обучения с учителем, когда алгоритм классификации сначала обучается на размеченных документах, а затем классифицирует новые документы.

Большое количество исследований алгоритмов кластеризации и классификации связаны с определением одной категории, к которой

может принадлежать документ. В реальном мире чаще бывает, что один и тот же документ может быть отнесен к нескольким категориям. Например, статья или новость про футбольный матч может быть отнесена к категориям: спорт, футбол, спортивные соревнования, городские мероприятия. Поэтому особенно актуальны методы и алгоритмы многозначной (нечеткой) классификации (multi-label classification) и мягкой кластеризации.

Цель данной работы — предложить подход к многозначной классификации с использованием методик вероятностного тематического моделирования. Создать тематическую модель на размеченных данных и предсказать категории, к которым относится новый документ, фраза или слово.

2. Обзор существующих алгоритмов многозначной классификации. Multi-label classification — не имеет устоявшегося русскоязычного термина, в литературе встречаются многозначная классификация и нечеткая классификация. В этой работе мы используем термин многозначная классификация. В машинном обучении многозначная классификация представляет собой вариант задач классификации, в которой к каждому документу (классифицируемому объекту) должны быть определены несколько меток. Не следует путать многозначную классификацию с многоклассовой классификацией, цель которой определить один класс из более чем двух классов кандидатов. В работе [2] представлен обзор алгоритмов многозначной классификации. Существует два основных метода для решения задач многозначной классификации: методы преобразования проблемы и метод адаптации. Метод преобразования проблемы трансформирует задачу в набор двоичных классификационных задач. Методы адаптации выполняют классификацию множества меток классов, решают задачу в ее полном виде.

Для решения задачи многозначной классификации используют адаптированные версии алгоритмов классификации, такие как: Boosting (AdaBoost), k-ближайших соседей, деревья решений, ядерные методы, SVM, нейронные сети. Метрики оценки качества многозначной классификации отличаются от обычной классификации в силу особенности задачи.

Тематическое моделирование — это способ построения тематической модели коллекции текстовых документов. Тематическая модель задает отношение между темами и документами в корпусе текстов. В обзоре [3] рассмотрены пять основных классов вероятностных тематических моделей: базовые, учитывающие отношения между документами, учитывающие отношения между

словами, темпоральные, обучаемые с учителем. Тематические модели задают мягкую кластеризацию слов и документов по кластерам-темам, означающую, что слово или документ могут быть отнесены сразу к нескольким темам с различными вероятностями. В результате синонимы с большой вероятностью будут отнесены к одной теме, так как часто употребляются в рамках одних и тех же контекстов, а омонимы попадут в разные, так как их контексты различаются. Тематические модели, как правило, основаны на гипотезе «мешка слов» и «мешка документов», т.е. порядок документов в коллекции не имеет значения и порядок слов в документе не имеет значения.

В работе [4] описан алгоритм тематической модели классификации под названием Label-LDA. В основе работы алгоритма лежит базовый алгоритм LDA, векторы документов и векторы тем порождаются распределением Дирихле. Основан он на двух сильных ограничениях, темы отождествляются с классами, предполагается, что для каждого документа точно известно множество всех классов, к которым он относится. Аналогичные ограничения касаются алгоритма Flat LDA, описанного в работе [5]. Для задачи классификации несбалансированных классов в этой работе был предложен алгоритм Prior-LDA, использующий частотную регуляризацию. По утверждению авторов работы [5] Flat-LDA, Prior-LDA являются частными случаями более общего алгоритма Dependency LDA, в котором предлагается моделировать классы документов через распределение тем документов и вводится новая неизвестная матрица класс-тема. В работе [6] предложен подход к многозначной классификации методом LDA с использованием знаний толпы под названием ML-PA-LDA-C. Используется информация не только о присутствующем классе, но и об отсутствующем применяется для построения модели по зашумленным размеченным данным. В работе устранено одно ограничение Label-LDA, предполагается, что точно не известно множество всех классов, к которым принадлежит документ.

3. Многозначная классификация с использованием Вероятностного тематического моделирования (multi-label Probabilistic Latent Semantic Indexing — ml-PLSI). Модели, разработанные на основе латентного размещения Дирихле (LDA), как указано в работе [7], не имеют сильных лингвистических обоснований. При этом классическая модель вероятностного латентно семантического анализа PLSA [8] не связана с какими-либо параметрическими априорными распределениями.

Пусть D — множество текстовых документов, W — словарь терминов. Каждый документ $d \in D$ представляет собой последовательность терминов n_d терминов (w_1, \dots, w_{n_d}) из словаря W .

С учетом гипотезы условной независимости $p(w|d, t) = p(w|t)$ по формуле полной вероятности получаем вероятностную модель порождения документа d :

$$p(w|d) = \sum_{t \in T} p(w|d, t) p(t|d) = \sum_{t \in T} p(w|t) p(t|d) = \sum_{t \in T} \varphi_{wt} \theta_{td},$$

Для вычисления φ_{wt} и θ_{td} используется EM-алгоритм.

Согласно вероятностному тематическому моделированию, впервые предложенному в работе [8], вероятностная модель появления пары «документ-слово» может быть записана тремя эквивалентными способами:

$$\begin{aligned} p(d, w) &= \sum_{t \in T} p(t) p(w|t) p(d|t) = \sum_{t \in T} p(d) p(w|t) p(t|d) = \\ &= \sum_{t \in T} p(w) p(t|w) p(d|t), \end{aligned}$$

где: T — множество тем;

$p(t)$ — неизвестное априорное распределение тем в коллекции;

$p(d)$ — априорное распределение на множестве документов,

эмпирическая оценка $p(d) = \frac{n_d}{n}$, где $n_d = \sum_{d \in D} n_d$ — суммарная длина

всех документов, а n_d — длина документа в словах;

$p(w)$ — априорное распределение на множестве слов,

эмпирическая оценка $p(w) = \frac{n_w}{n}$, где n_w — число вхождений слова w во все документы.

Если мы отождествим понятие темы тематической модели и категории документа, учтем, что задача построения тематической модели имеет бесконечно много решений [9], то сможем построить один из вариантов тематической модели, обучившись на размеченном корпусе. Построенная на данных предположениях тематическая модель зависит от качества выбранной для обучения коллекции. Например, категории в корпусе SCTM-ru проставлены авторами новостей, перед авторами не стояла задача указать все категории, которые только возможно для каждой новости, поэтому часть

документов не получили полный набор категорий, даже если они этого заслуживали. При этом объем корпуса позволяет предположить, что в своем большинстве авторы использовали наиболее характерные категории. Основываясь на знании толпы, мы можем рассчитывать вероятностную оценку отнесения слова к категории:

$$p(w|d) = \sum_{t \in T} p(w|t) p(t|d) = \sum_{t \in T} p(w|c) p(c|d) = \sum_{t \in T} \varphi_{wt} \theta_{td}.$$

Построенная таким образом тематическая модель может быть одним из множества решений задачи тематического моделирования. В работе [10] рассмотрен алгоритм создания тематической модели методом обучения с учителем.

Для уменьшения размерности векторного пространства рекомендуется все слова в корпусе привести к нормальной словоформе. Если документов в корпусе немного и алгоритм будет выполнен за конечное время, то этого делать необязательно, т.к. словоформа также может стать важной информацией для определения категории документа, слова в конкретной словоформе чаще могут встречаться в документах, принадлежащих одной категории. На основании семиологии, [11] слово в знаковой системе наделено смыслом и является частью языка. Поэтому в данной работе мы не приводим слова в корпусе к нормальной словоформе.

Алгоритм 1. Алгоритм построения тематической модели методом обучения с учителем:

Вход: коллекция документов D с указанием категорий C .

Выход: распределения $p(w|c)$, $p(d|c)$, $p(w|d)$:

1. Для всех $d \in D, w \in d$:

а. Считаем $p(w|d) = n_{dw} / n$.

2. Для всех $c \in C, d \in D$:

а. Считаем $p(d|c) = n_c / n$.

3. Для всех $w \in W, c \in C$:

а. Считаем $p(w|c) = \varphi_{wt} = n_{dwc} / n_{dw}$.

Обучить тематическую модель по размеченным данным — это значит рассчитать матрицы «слово-документ», «документ-категория» и «слово-категория» для каждого слова из коллекции документов. На первом шаге рассчитываем матрицу «слово-документ». Значения матрицы — это количество повторений слова в документе. На втором

шаге рассчитываем матрицу «документ-категория». Для этого для каждого документа в корпусе получаем список категорий, к каждой категории документ может быть отнесен не более одного раза, поэтому значения матрицы «документ-категория» — это единицы в том случае, если документ связан с категорией, и нуль, если такой связи нет. На третьем завершающем шаге рассчитываем матрицу «слово-категория». Значения матрицы — это вероятность встретить слово в этой категории. Как мы ранее отмечали, разметка документов категориями, содержит ошибки. Эти ошибки можно разделить на два вида:

1. Слово, которое имеет отношение к определенной категории, редко встречается в корпусе поэтому связь между словом и категорией не установилась.

2. Документ отмечен какой-либо категорией по ошибке.

В данной работе для уменьшения влияния ошибок на результат предлагаем использовать регуляризацию.

В результате обучения тематической модели мы получили три матрицы, по которым мы можем узнать, с какой вероятностью то или иное слово относится к теме, из каких слов формируется тема и к каким темам относится каждый документ. По матрице «слово-документ» можем восстановить каждый документ в формате мешка слов. По матрице «слово-категория» можем оценить с какой вероятностью то или иное слово относится к категории, можем рассчитать вероятность отнесения нескольких слов к категориям, для этого достаточно просуммировать вероятность отнесения каждого слова к категории $\sum_{w \in d} p(w|c)$. По матрице «документ-категория»

можем получить список всех документов, которые связаны с категориями. Для примера в таблице 1 приведены наиболее характерные и наиболее часто встречаемые слова для трех категорий: спорт, происшествия и политика.

Таблица 1. Пример слов характерных для категорий

Спорт	Происшествия	Политика
Зимнему	Даги	Реймер
Виндсёрфинг	Дагов	Халип
Перепёлкин	Гоа	Гольман
Трофименко	Моди	Минтимер
Заезда	Зингер	Муртаза
Кайтингу	Ксанте	Рахимова
Полумарафона	Ока	Дарькина
Фонак	Реанимация	Спутникам
Гимнаст	Стропила	Хамитов

Стоит заметить, что эти слова однозначно характеризуют категорию, к которой относятся, но не являются часто употребляемыми во всем корпусе; зачастую это имена собственные и фамилии либо редкие по написанию слова с буквой «ё».

4. Многозначная классификация. Переходя к следующему шагу, мы имеем тематическую модель, в которой определены отношения между словами и категориями. Для предсказания категорий нового документа, выполняем Алгоритм 2

Алгоритм 2. Многозначная классификация на базе вероятностного тематического моделирования:

Вход: Тематическая модель, новый документ d_{new} .

Выход: Список предсказанных категорий:

1. Для всех $w \in d_{new}$:

а. $\sum_{w \in d} p(w|c)$.

2. Возвращаем список категорий по убыванию суммы вероятностных оценок.

Следует учитывать, что слова, которые есть в новом документе и отсутствуют в корпусе, на котором обучалась тематическая модель, не будут учтены при предсказании категорий.

Для того чтобы отобрать только наиболее релевантные категории используем регуляризацию, а именно:

– Регуляризация по документам. Если слово встречается в большом количестве документов, то оно перестает быть для нас информативным, вероятность отнесения документа к теме по этому слову может быть сведена к нулю.

– Регуляризация по темам. Если слово встречается в большом количестве категорий, то оно перестает быть для нас информативным, вероятность отнесения слова к категории по этому слову может быть сведена к нулю.

5. Эксперимент с корпусом SCTM-ru. В качестве данных для исследования используем корпус SCTM-ru [12], созданный специально для тестирования задач тематического моделирования. Источником данных корпуса является международный новостной сайт «Русские Викиновости». Корпус SCTM-ru состоит из 7000 документов, 185 авторов, почти 12000 уникальных категорий. События, описанные в документах, распределены по более чем 2000 уникальным датам, с

ноября 2005 года по июнь 2014 года. В корпусе SCTM-ru 2400000 словоупотреблений, состоящих только из русских букв. Словарный состав корпуса — 150600 уникальных словоформ, 59000 уникальных лемм.

Каждая новость содержит указанные автором категории. У автора новости не стояла задача перечислить все категории, к которым новость может иметь отношение, тем не менее указанные категории дают весомые основания полагать, что новость сильно связана с темой этих категорий. Обычно новость определена автором к нескольким категориям, поэтому мы рассматривали алгоритмы многозначной классификации. Для каждого документа мы не знаем точного множества всех категорий.

В корпусе SCTM-ru часть категорий авторы используют редко, 19284 категориями размечено менее 50 новостей из корпуса. Так как эти категории не представляют большой ценности для нашей тематической модели и могут создавать дополнительные трудности при прогнозировании категорий новых документов, мы не будем их учитывать при построении тематической модели. Категории, которыми размечены более 50 новостей — 230 штук, далее их называем рабочие категории или просто категории, их мы используем для обучения тематической модели. Всего документов, размеченных рабочими категориями 6428. Для обучения модели используем 5000 новостей. Чтобы отсеять не информативные слова, междометия и предлоги, создаем словарь стоп-слов, в который вошло 151 слово.

Для разработки прототипа программы используется дистрибутив Anaconda, язык разработки Python и модули для машинного обучения pandas, numpy, scikit-learn. Векторизуем документы, создаем матрицу $p(w|d)$. Затем создаем вектора категорий и создаем матрицу $p(d|c)$. Затем рассчитываем матрицу $p(w|c)$. Таким образом, мы обучили тематическую модель на размеченных данных корпуса SCTM-ru.

6. Оценка качества полученной тематической модели.

Наиболее известным критерием является коэффициент неопределенности, используемый для оценки качества различных моделей. Коэффициент неопределенности представляет собой меру несоответствия модели $p(w|d)$ словам w , наблюдаемым в документах коллекции, и определяется через логарифм правдоподобия:

$$Perplexity(D) = \exp\left(-\frac{1}{n} \sum_{d \in D} \sum_{w \in d} n_{dw} \log p(w|d)\right),$$

где n — число всех рассматриваемых слов в текстовой коллекции, D — множество всех документов в коллекции, n_{dw} — частотность слова w в документе d , $p(w|d)$ — вероятность появления слова w в документе d . Чем меньше значение коэффициента неопределенности, тем лучше модель предсказывает появление слов в документах коллекции. Коэффициент неопределенности полученной тематической модели 0,11. В работе [13] отмечено, что коэффициент, вычисленный по той же самой коллекции, дает оптимистически заниженную оценку. Для предсказания тематических категорий используем незатронутые в обучении новости.

Используем следующие метрики оценки качества многозначной классификации:

1. Функция потерь Хэмминга для ошибочных предсказаний.
2. Количество документов, в которых первая предсказанная категория, совпала с любой из указанных автором новости.
3. Процент верных предсказаний из первых 50 предсказанных категорий.

Результаты оценки качества приведены в таблице 2.

Таблица 2. Оценка качества многозначной классификации

Метрика качества	Результат (100 новостей)	Результат (500 новостей)
Функция потерь Хэмминга	0,18	0,17
Первая предсказанная (точность)	82%	84%
Процент верных предсказаний из 50 первых категорий (полнота)	71%	72%

Построенная тематическая модель позволяет определить наиболее вероятные категории не только для текстового документа, но и для отдельной фразы или слова. Например, для фразы «выборы президента России» десять первых категорий отображены на рисунке 1, для слова «футбол» на рисунке 2 показаны десять наиболее вероятных категорий.

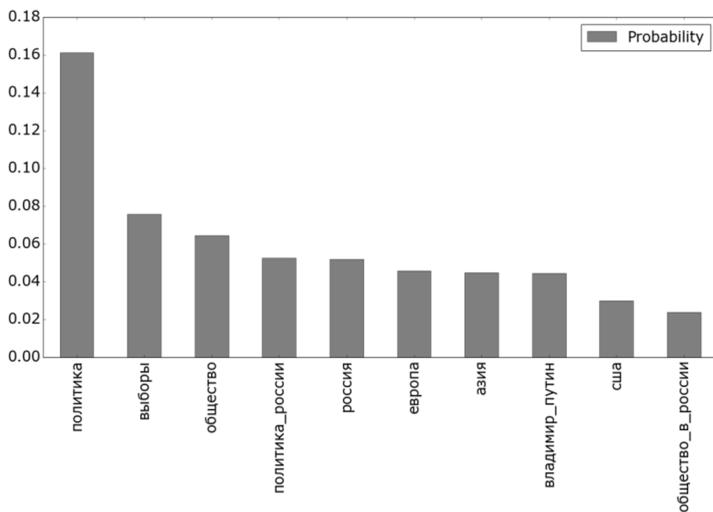


Рис. 1. Наиболее вероятные категории для фразы «выборы президента России»

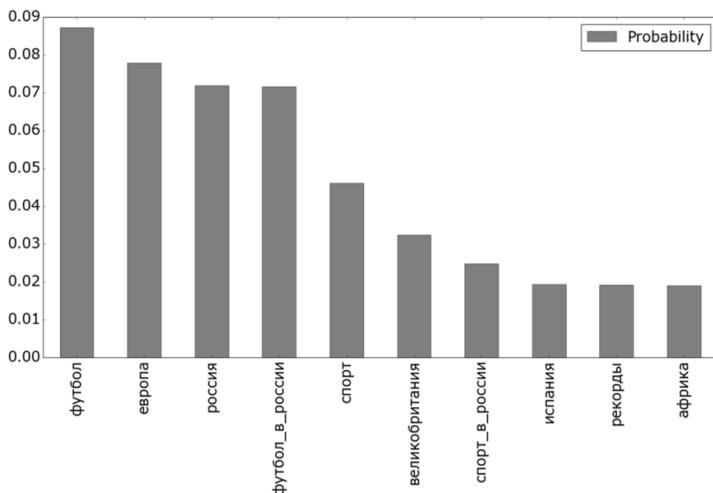


Рис. 2. Наиболее вероятные категории для слова «футбол»

6. Заключение. В результате проделанной работы разработан подход к обучению тематических моделей с учителем (supervised Probabilistic Latent Semantic Indexing — PLSI). Предложен метод многозначной классификации текстовых документов с использованием обученной тематической модели ml-PLSI. Проведенные эксперименты на корпусе SCTM-ги демонстрируют

перспективность использования тематического моделирования в задачах многозначной классификации. Поставленные в работе цели были достигнуты.

Вероятностная тематическая модель может быть использована в задачах ассоциативной классификации [14] в комбинации с другими алгоритмами классификации, а также может быть решателем в алгоритмах коллективного распознавания, описанных в работе [15].

Далее будут продолжены исследования возможностей многозначной классификации методом вероятностного тематического моделирования, будет проверена гипотеза обучения модели новыми словами за счет уже имеющейся информации о связях слов и категорий. Эксперимент с программной частью метода доступен на <<https://github.com/cimsweb/mlPLSI>>.

Литература

1. Журавлёв Ю.И., и др. Задачи распознавания и классификации со стандартной обучающей информацией // Журнал вычислительной математики и математической физики. 1980. Вып. 20. № 5. С. 1294–1309.
2. Tsoumakas G., Katakis I. Multi-label classification: an overview // International Journal of Data Warehousing & Mining. 2007. vol. 3(3). pp. 1–13.
3. Daud A. et al. Knowledge discovery through directed probabilistic topic models: a survey // Frontiers of computer science in China. 2010. vol. 4. no. 2. pp. 280–301.
4. Ramage D., Hall D., Nallapati R., Manning C. D. Labeled lda: a supervised topic model for credit attribution in multi-labeled corpora // Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing. USA. 2009. vol. 1. pp. 248–256.
5. Rubin T.N., Chambers A., Smyth P., Steyvers M. Statistical topic models for multilabel document classification // Machine Learning. 2012. vol. 88. no. 1–2. pp. 157–208.
6. Padmanabhan D. et al. Topic Model Based Multi-Label Classification from the Crowd // arXiv preprint arXiv:1604.00783. 2016.
7. Воронцов К.В., Потапенко А.А. Модификация EM-алгоритма для вероятностного тематического моделирования // Машинное обучение и анализ данных. 2013. Вып. 1. № 6. С. 657–686.
8. Hoffman T. Probabilistic Latent Semantic Indexing // Proceedings of the Twenty-Second Annual International SIGIR Conference on Research and Development in Information Retrieval. 1999. pp. 50–57.
9. Воронцов К.В., Потапенко А.А. Регуляризация, робастность и разреженность вероятностных тематических моделей // Компьютерные исследования и моделирование. 2012. Вып. 4. № 4. С. 693–706.
10. Blei D., McAuliffe J. Supervised topic models // Advances in neural information processing systems. 2008. vol. 20. pp. 121–128.
11. Плохотнюк В.С. Аксиоматизация семиологии и научный статус семиотики // Terra economicus. 2010. Вып. 8(4). С. 124–132.
12. Карпович С.Н. Русскоязычный корпус SCTM-RU для построения тематических моделей // Труды СПИИРАН. 2015. Вып. 39. С. 123–142.
13. Blei D., Ng A., Jordan M. Latent Dirichlet Allocation // Journal of Machine Learning Research // MIT Press. 2003. vol. 3(Jan). pp. 993–1002.
14. Городецкий В.И., Тушканова О.Н. Ассоциативная классификация: аналитический обзор. Часть 1 // Труды СПИИРАН. 2015. Вып. 1(38). С. 183–203.
15. Городецкий В.И., Тушканова О.Н. Ассоциативная классификация: аналитический обзор. Часть 2. // Труды СПИИРАН. 2015. Вып. 2(39). С. 212–240.

16. *Городецкий В.И., Серебряков С.В.* Методы и алгоритмы коллективного распознавания // Труды СПИИРАН. 2006. №3. С 139–171.

Reference

1. Zhuravlev Yu.I., et. al. [Recognition and classification problems with standard training information]. *Zh.-vychisl.-matem.-i-matem.-fiz. – Comput. Math. Math. Phys.* 1980. vol. 20. no. 5. pp. 1294–1309. (In Russ.).
2. Tsoumakas, Grigorios; Katakis, Ioannis. Multi-label classification: an overview. *International Journal of Data Warehousing & Mining*. 2007. vol. 3(3). pp. 1–13.
3. Daud A. et al. Knowledge discovery through directed probabilistic topic models: a survey // *Frontiers of computer science in China*. 2010. vol. 4. no. 2. pp. 280–301.
4. Ramage D., Hall D., Nallapati R., Manning C. D. Labeled lda: a supervised topic model for credit attribution in multi-labeled corpora. *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*. USA. 2009. vol. 1. pp. 248–256.
5. Rubin T.N., Chambers A., Smyth P., Steyvers M. Statistical topic models for multilabel document classification. *Machine Learning*. 2012. vol. 88. no. 1–2. pp. 157–208.
6. Padmanabhan D. et al. Topic Model Based Multi-Label Classification from the Crowd. arXiv preprint arXiv:1604.00783. 2016.
7. Vorontsov K.V., Potapenko A.A. [EM-like algorithms for probabilistic topic modeling]. *Mashinnoe obuchenie i analiz dannyh – Machine Learning and Data Mining*. 2013. vol. 1. no. 6. pp. 657–686. (In Russ.).
8. Hoffman T. Probabilistic Latent Semantic Indexing. *Proceedings of the Twenty-Second Annual International SIGIR Conference on Research and Development in Information Retrieval*. 1999. pp. 50–57.
9. Vorontsov K.V., Potapenko A.A. [Regularization, robustness and sparsity of probabilistic topic models]. *Kompyuternyye issledovaniya i modelirovaniye — Computer research and modeling*. 2012. vol. 4 no. 4. pp. 693–706. (In Russ.).
10. Blei D., McAuliffe. J. Supervised topic models. *Advances in neural information processing systems*. 2008. vol. 20. pp. 121–128.
11. Plokhotnuk V.S. [Axiomatization of semiology and scientific status of semiotics]. *Terra economicus*. 2010. vol. 8(4). pp. 124–132. (In Russ.).
12. Karpovich S.N. [The Russian language text corpus for testing algorithms of topic model]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2015. vol. 39. pp 123–142. (In Russ.).
13. Blei D., Ng A., Jordan M. Latent Dirichlet Allocation. *Journal of Machine Learning Research*. 2003. vol. 3. pp. 993–1002.
14. Gorodetsky V.I., Tushkanova O.N. [Associative classification: analytical overview. Part 1]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2015. vol. 1(38). pp. 183–203. (In Russ.).
15. Gorodetsky V.I., Tushkanova O.N. [Associative classification: analytical overview. Part 2.]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2015. vol. 2(39). pp 212–240. (In Russ.).
16. Gorodetsky V.I., Serebryakov S.V. [Methods and algorithms of the collective recognition: a survey]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2006. vol. 3. pp. 139–171. (In Russ.).

Карпович Сергей Николаевич — руководитель направления поисковой оптимизации, ООО "Рамблер Интернет Холдинг". Область научных интересов: тематическое моделирование, обработка текстов на естественном языке, data mining. Число научных публикаций — 1. cims@yandex.ru, <http://www.cims.ru>; Варшавское ш., 9, стр. 1, БЦ «Даниловская мануфактура», корпус «Ряды Солдатенкова», Москва, 117105; р.т.: +7(495)7851700.

Karpovich Sergey Nikolaevich — head of search engine optimization direction, Rambler Internet Holding LLC. Research interests: topic model, natural language processing, classification, clustering, data mining. The number of publications — 1. cims@yandex.ru, <http://www.cims.ru>; 9, Varshavskoe sh., str. 1, BC «Danilovskaja manufaktura», k. «Rjady Soldatenkova», 117105, Moscow; office phone: +7(495)7851700.

РЕФЕРАТ

Карпович С.Н. **Многозначная классификация текстовых документов с использованием вероятностного тематического моделирования ml-PLSI.**

В статье рассмотрен подход к задаче многозначной классификации текстовых документов с использованием методик вероятностного тематического моделирования. Большое количество исследований алгоритмов классификации и кластеризации, связаны с определением одной категории, к которой может принадлежать документ. Зачастую один документ может быть отнесен к нескольким категориям. Обозначена актуальность задачи. Проведен обзор существующих алгоритмов многозначной классификации.

Описана программная реализация алгоритма многозначной классификации. Методом обучения с учителем построена тематическая модель. Приведены оценки качества классификации, представлен пример предсказания возможных категорий для фразы.

Предложенный подход продемонстрировал свою эффективность. Вероятностные оценки отнесения документа к категории позволяют использовать его в задачах коллективного распознавания и в задачах ассоциативной классификации. Далее будут продолжены исследования возможностей многозначной классификации методом вероятностного тематического моделирования.

SUMMARY

Karpovich S.N. **Multi-Label Classification of Text Documents using Probabilistic Topic Model ml-PLSI.**

The paper considers an approach to multi-label classification of text documents based on probabilistic topic modeling. A large number of studies of clustering and classification algorithms are related to determination of one label to which a document may belong. Often one document can be relevant to many labels. The significance of this task is shown. A comparative analysis of multi-label classification algorithms has been conducted.

The article describes technology tools for multi-label classification. A topic model has been built with the help of supervised learning. Evaluation of classification quality is given, and an example of the prediction of possible categories for the phrase is presented. The developed approach has proven to be efficient. Probabilistic estimations of categorizing a document allow using it for collective recognition and associative classification tasks. We will continue our studies on the opportunities of multi-label classification by the method of probabilistic topic modeling.

Д.Н. БИРЮКОВ, А.Г. ЛОМАКО, Р.Б. ЖОЛУС
**ПОПОЛНЕНИЕ ОНТОЛОГИЧЕСКИХ СИСТЕМ ЗНАНИЙ НА
ОСНОВЕ МОДЕЛИРОВАНИЯ УМОЗАКЛЮЧЕНИЙ С УЧЕТОМ
СЕМАНТИКИ РОЛЕЙ**

Бирюков Д.Н., Ломако А.Г., Жолус Р.Б. Пополнение онтологических систем знаний на основе моделирования умозаключений с учетом семантики ролей.

Аннотация. В статье рассмотрен вопрос, связанный с автоматическим пополнением онтологии ролями и концептами, формируемыми интеллектуальной системой, при предоставлении ей новых фактов. Осуществление указанных предвычислений позволяет повысить информационное содержание онтологии на этапе предварительной обработки потока поступающих данных.

Ключевые слова: интеллектуальная система, система знаний, онтология, правдоподобные умозаключения.

Biryukov D.N., Lomako A.G., Zholus R.B. Ontological Knowledge System Completion Based on Modeling Inferences Taking into Account Role Semantics.

Abstract. The article considers the issue of automatic completion of ontology with roles and concepts formed by an intelligent system in the provision of new facts. Implementation of specified calculations allows increasing ontology information content during data stream preprocessing.

Keywords: intelligent system, knowledge system, ontology; plausible inferences.

1. Введение. В работе [1] предлагается наделить интеллектуальную систему (ИС) [2, 3] возможностью концептуально-логических преобразований над данными и знаниями в процессе эволюции ее многофункциональной памяти. В этой связи требуется найти возможности логического расширения, конструктивной типизации и смысловой агрегации потока поступающих данных, фактов и знаний с уже накопленными знаниями, представленными в онтологических слоях семантической памяти [4]. Последние хранятся в виде совокупности взаимосвязанных «Семантических Звеньев» (СЗ — два концепта, связанных ролью) [5], основные типы которых представлены в таблице 1, где $Fn: [X \rightarrow Y]$ — обозначение отношений между концептами типа X и концептами типа Y . Семантика используемых отношений представлена в поле «Описание СЗ» таблицы 1.

Вся совокупность СЗ представляет собой обобщенную онтологию тех предметных областей, в которых функционирует ИС. При этом абсолютно все концепты, входящие в различные СЗ и представленные в онтологии, должны быть иерархически связаны и принадлежать одному из частично упорядоченных множеств в базе {"Объекты" — "О", "Свойства" — "Р", "Действия" — "А"} [5].

При поступлении на вход ИС нового факта, ее система знаний посредством формально-логических умозаключений специального вида увязывает поступившие данные с уже построенной ранее онтологией.

Таблица 1. Основные типы семантических звеньев для онтологий произвольных предметных областей

Обозначение СЗ	Описание СЗ
$F11: [P \rightarrow P]$	Свойство аппроксимирует Свойство
$F12: [O \rightarrow O]$	Объект аппроксимирует Объект
$F13: [A \rightarrow A]$	Действие аппроксимирует Действие
$F2: [O \rightarrow P]$	Объект обладает Свойством
$F3: [P \rightarrow A]$	Свойство порождает способность к Действию
$F4: [A \rightarrow P]$	Действие пригодно для воздействия на Свойство
$F5: [A \rightarrow A]$	Действие следует за Действием
$F6: [O \rightarrow O]$	Объект состоит из Объекта
$F7: [A \rightarrow A]$	Действие состоит из Действия
$F8: [O \rightarrow O]$	Объект Производит Объект

2. Силлогистика как основа моделирования умозаключений.

Выбран подход к моделированию правдоподобных умозаключений над данными, представленными в онтологии, основанный на применении расширенной силлогистики.

В силлогистике исследуются различного рода логические отношения между категорическими атрибутивными высказываниями. В составе категорических атрибутивных высказываний выделяют кванторные слова, предикцирующие связки и термины. В каждом категорическом атрибутивном высказывании имеется два термина: *субъект* (α) — термин, обозначающий те предметы, о которых в высказывании что-либо утверждается или отрицается, и *предикат* (β) — термин, обозначающий то, что предикцируется, утверждается или отрицается об этих предметах.

Категорические атрибутивные высказывания по количеству делятся на единичные и множественные. Среди множественных выделяют общие (\forall) и частные (\exists) высказывания. По качеству рассматриваемые высказывания делятся на утвердительные и отрицательные.

Для осуществления классификации объектов, свойств, которыми обладают классифицируемые объекты (субъекты), а также действий (процессов) достаточно воспользоваться возможностями традиционной силлогистики, которая подразумевает, что все термины категорических атрибутивных высказываний при их интерпретации на не-

котором универсуме обязательно должны оказаться знаками таких свойств (классов), которые являются непустыми и неуниверсальными (в противном случае невозможно осуществить классификацию).

Известные законы силлогистики [6] (например: «*Всякий S есть P*», « $\models \neg$ (*Всякий S есть P* & *Некоторый S не есть P*)» и т.п., где « \models » — знак логического следования) вполне пригодны для их аксиоматического задания в системе знаний. Однако применение только этих законов, открывает лишь незначительную часть возможностей расширения знаний системы. Наибольший интерес представляют умозаключения, в которых от наличия некоторых отношений между терминами α и β и терминами γ и β , фиксируемых в посылках A_1 и A_2 , приходят к заключению о наличии определенного отношения между терминами α и γ . Такие умозаключения называют простым категорическим силлогизмом.

Общий термин β , содержащийся в посылках A_1 и A_2 в двухпосылочном умозаключении вида: $A_1, A_2 \vdash B$, связывает посылки и опосредует следование из них заключения B .

При формулировании простых категорических силлогизмов *меньшим* термином называют тот термин, который является субъектом заключения, *большим* — тот, который является предикатом заключения, а тот, который является общим термином — называют *средним* термином [6]. Посылку, содержащую меньший термин, называют *меньшей* посылкой, а посылку с большим термином — *большой* посылкой. Если помещать большую посылку на первое место, а меньшую — на второе (или под ней), то все простые категорические силлогизмы можно разделить по следующим четырем фигурам [6] (рисунок 1):

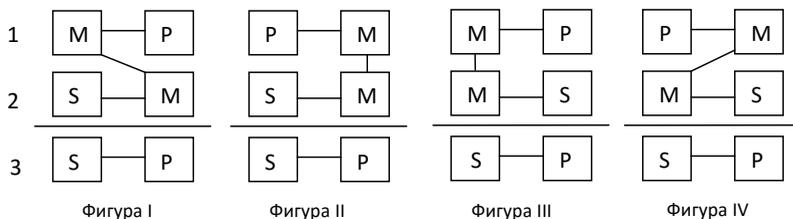


Рис. 1. Фигуры силлогизмов

На рисунке 1 цифрой «1» обозначается большая посылка, цифрой «2» — меньшая посылка, а цифрой «3» — заключение; «S» — меньший термин, «P» — больший, а «M» — средний термин («S» и «P» называют еще крайними терминами).

Если в фигуре указать тип высказываний (всего их четыре: *a*, *i*, *e*, *o* [6]), стоящих на местах посылок и заключения, то можно получить разновидность силлогизма по конкретной фигуре. Разновидности силлогизмов называются *модусами фигур*. В каждой фигуре имеется 64 модуса, т.е. по всем четырем фигурам можно построить 256 модусов из которых только 24 — правильных (правильными модусами называются такие модусы фигур, для которых между посылками и заключением существует отношение логического следования). Следует отметить, что проверять правильность силлогизмов можно как в рамках исчисления предикатов, так и в рамках логики предикатов. В то же время при реализации механизма осуществления правдоподобных умозаключений необходимо учитывать ограничения, накладываемые семантикой применяемых ролей (см. выше или [5]). В предложенной модели представления знаний присутствуют, например, такие роли как: *F5* (Действие следует за Действием), *F6* (Объект состоит из Объекта) и *F7* (Действие состоит из Действия), к которым нельзя применить правила, описываемые указанными выше фигурами.

При рассмотрении семантических конструкций, построенных на основе СЗ, содержащих роли «*Состоит из...*», следует учитывать, что в этом случае осуществляется описание процедуры членения предметов [6] (объектов, процессов), а не ограничения объема понятий, что влечет к невозможности прямого использования модусов силлогистики.

3. Моделирование умозаключений с учетом семантики ролей, используемых при построении онтологии. Для дальнейшего описания порядка моделирования рассуждений, осуществляемых на основе умозаключений с учетом семантики ролей, предлагается ввести понятие простейшей семантической конструкции (ПСК), которая предполагает возможность осуществления умозаключения. ПСК — пара семантических звеньев, имеющих в своем составе один общий концепт.

Можно утверждать, что рассмотрение правил построения сколь угодно сложных онтологий, а также правил осуществления правдоподобных умозаключений над знаниями, представленными в виде онтологии (путем применения семантических звеньев, типы которых приведены выше и в [5]), можно свести к рассмотрению правил построения простейших семантических конструкций, а также к рассмотрению правил осуществления правдоподобных умозаключений над ними (так как именно пара СЗ может участвовать в порождении новых знаний). Если учесть, что каждое СЗ характеризуется двумя концептами и ролью между ними, а также то, что если два СЗ взаимодействуют, то они должны иметь один общий концепт (случай, когда у двух звеньев попарно совпадают концепты, предлагается не рассматривать, так как он

не предполагает возможности осуществления логического вывода), тогда можно построить таблицу с возможными тройками концептов. Поскольку для построения онтологии предложено использовать всего три типа концептов («Объекты», «Свойства» и «Действия» [5, 7]), то и таблица возможных их комбинаций, состоящая из троек концептов, тоже будет небольшой — всего 27 вариантов. При этом каждый из вариантов требует отдельного рассмотрения, так как необходимо рассмотреть все возможные (допустимые) роли и их направления, а также комбинации кванторов. Фрагмент подобной таблицы приведен в таблице 2, где «К-т» — «Концепт», «О» – Объект, $F6$ и $F12$ — соответствующие типы СЗ, а « \rightarrow » — направление Роли в СЗ.

При построении таблицы учитывались ниже приведенные положения.

1. Для обозначения «*Всякий A есть B* » (формальное обозначение: «*Роль*(\forall Концепт_1, Концепт_2)»), предлагается использовать стрелку с «закрашенным» указателем (рисунок 2):



Рис. 2. Пример отображения конструкции типа: «*Всякий A есть B* »

Для обозначения «*Некоторый A есть B* » (формальное обозначение: «*Роль*(\exists Концепт_1, Концепт_2)») предлагается использовать стрелку с «полым» указателем (рисунок 3):



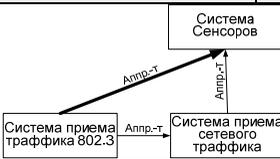
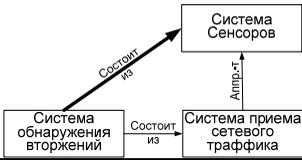
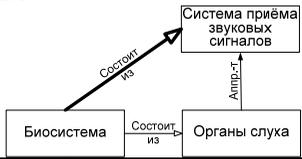
Рис. 3. Пример отображения конструкции типа: «*Некоторый A есть B* »

2. Отношение аппроксимации позволяет осуществлять деление определенного класса концептов на непересекающиеся подклассы относительно заранее выбранного основания, при этом каждый элемент подкласса должен аппроксимировать любой элемент класса, в противном случае классификация задана неверно. Ввиду этого следует считать, что случаи: $F11(\exists P_1, P_2)$, $F12(\exists O_1, O_2)$ и $F13(\exists A_1, A_2)$ недопустимы, ибо подобные отношения не могут встретиться при работе с базой знаний в корректно работающей ИС (на основе предложенных моделей).

3. В исходных данных, в силу введенных выше типов семантических звеньев, не могут присутствовать некоторые звенья, например: «О А», «Р О» и др.

В последующем при необходимости можно расширить перечень семантических звеньев. Добавление и удаление типов СЗ приведет к необходимости дополнения «таблицы логического вывода» (таблица 2).

Таблица 2. Фрагмент «таблицы логического вывода»

№	К-т №1	К-т №2	К-т №3	Примечание
	Направление ребра		Направление ребра	
1	O	O	O	
1.1	→		→	
1.1.1	F12		F12	
1.1.1.1	$O_1 \xrightarrow{(\forall)F12} O_2 \xrightarrow{(\forall)F12} O_3$			$F12(\forall O_1, O_2) \& F12(\forall O_2, O_3) \Rightarrow F12(\forall O_1, O_3)$  <p>Согласно определению, отношение аппроксимации транзитивно (из $x \sqsubseteq y \sqsubseteq z$ следует $x \sqsubseteq z$), а аппроксимируемые и аппроксимирующие классы — совместимы.</p>
1.1.1.2	$O_1 \xrightarrow{(\exists)F12} O_2 \xrightarrow{(\forall)F12} O_3$			$(\exists)F12$ — недопустимо
1.1.1.3	$O_1 \xrightarrow{(\forall)F12} O_2 \xrightarrow{(\exists)F12} O_3$			$(\exists)F12$ — недопустимо
1.1.1.4	$O_1 \xrightarrow{(\exists)F12} O_2 \xrightarrow{(\exists)F12} O_3$			$(\exists)F12$ — недопустимо
1.1.2	F6		F12	
1.1.2.1	$O_1 \xrightarrow{(\forall)F6} O_2 \xrightarrow{(\forall)F12} O_3$			$F6(\forall O_1, O_2) \& F12(\forall O_2, O_3) \Rightarrow F6(\forall O_1, O_3)$ <p><i>Утверждение:</i> Если $F6(\forall O_1, O_2) \& F12(\forall O_2, O_3)$ то $F6(\forall O_1, O_3)$.</p> <p>1. $WoO_2(o) \subseteq WoO_3(o)$, так как в данной теории T: $T \models \forall o(O_2(o) \supset O_3(o))$ и $T \not\models \forall o(O_3(o) \supset O_2(o))$ — согласно аксиоме аппроксимации;</p> <p>2. $(O_1 \xrightarrow{(\forall)F6} O_2) \sqsubseteq (O_1 \xrightarrow{(\forall)F6} O_3)$.</p> 
1.1.2.2	$O_1 \xrightarrow{(\exists)F6} O_2 \xrightarrow{(\forall)F12} O_3$			$F6(\exists O_1, O_2) \& F12(\forall O_2, O_3) \Rightarrow F6(\exists O_1, O_3)$ <p>Аналогично п.п. 1.1.2.1</p> 
1.1.2.3	$O_1 \xrightarrow{(\forall)F6} O_2 \xrightarrow{(\exists)F12} O_3$			$(\exists)F12$ — недопустимо
1.1.2.4	$O_1 \xrightarrow{(\exists)F6} O_2 \xrightarrow{(\exists)F12} O_3$			$(\exists)F12$ — недопустимо

4. Результатом взаимодействия двух семантических звеньев может стать конструкция, которая не предполагает образования никакого нового семантического звена (рисунок 4):

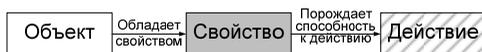


Рис. 4. Пример семантической конструкции, не предполагающей вывод нового семантического звена

Невозможность сформировать новое СЗ в приведенном примере обусловлена тем, что изначально среди базовых ролей не задано ни одной возможной роли между концептами «Объект» и «Действие».

5. В случае, если результатом взаимодействия двух семантических звеньев становится конструкция, которая потенциально может образовать новое семантическое звено (ввиду наличия подобных ролей), но логически такое звено не выводимо, то для обозначения таких конструкций предлагается использовать обозначение \emptyset . Пример подобных конструкций:

$$\begin{aligned}
 F6(\forall O_1, O_2) \& F12(\forall O_3, O_2) \Rightarrow \emptyset, \\
 F6(\forall O_1, O_2) \& F2(\forall O_2, P) \Rightarrow \emptyset, \\
 F2(\exists O, P_1) \& F11(\forall P_2, P_1) \Rightarrow \emptyset \text{ и т.п.}
 \end{aligned}$$

Подобное обозначение также предлагается использовать при описании результатов вывода для типов конструкций подобных той, которая рассмотрена на рисунке 4.

4. Порождение концептов в онтологической базе знаний. Одной из примечательных и важных особенностей предлагаемой системы знаний является ее способность к порождению концептов и включению их в уже имеющуюся онтологию.

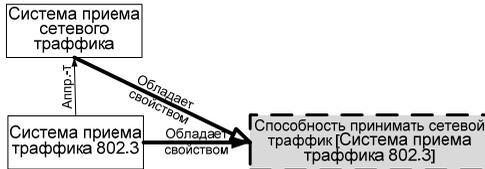
Так, например, если ИС необходимо вычислить значение функции в точке, находящейся вне области определения, то в качестве значения функции в этой точке порождается новый синтаксический объект, отражающий факт вызова в заданной точке. Этот объект в работе [8] предложено определять как строку, полностью совпадающую с текстом вызова. Такой способ доопределения функций назван *задержкой вычислений* функции или просто *задержкой функции*. Значение, получающееся в результате такой задержки, предложено называть *интенционалом функции* (для краткости — просто *интенционалом*). Пример одного из простейших правил порождения интенционалов приведен в таблице 3. Проведенный анализ широко применяемых в настоящее время машин логического вывода (FaCT++, HermiT, Pellet и др.) и дескрипционных логик, реализованных в них [17-28], показал, что они неспособны автоматически формировать новые понятия, именовать и встраивать их в уже имеющуюся онтологию.

Таблица 3. Пример правила, порождающего интенционал

№	К-т №1	К-т №2	К-т №3	Примечание
	Направление ребра		Направление ребра	
2	О	О	Р	
2.1	→		→	
2.1.1	$F12$		$F2$	
2.1.1.1	$O_1 \xrightarrow{(\forall)F12} O_2 \xrightarrow{(\forall)F2} P$			$F12(\forall O_1, O_2) \& F2(\forall O_2, P) \Rightarrow$ а) $F2(\forall O_1, P)$; в) $F11(\forall P[O_1], P)$; б) $\exists P[O_1]$; г) $F2(\forall O_1, P[O_1])$; д) $F2(\exists O_2, P[O_1])$.
<p>Утверждение_1: Если $F12(\forall O_1, O_2) \& F2(\forall O_2, P)$ то $F2(\forall O_1, P)$.</p> <p>Доказательство_1:</p> <p>– пусть для $\forall o \in O_2$ родом является U (универсум – н.в.г. решетки O), тогда признак $P(o)$ можно считать видовым отличием, так как других данных, на основании которых можно было бы выделить $\forall o \in O_2$ из U нет;</p> <p>– исходя из того, что $O_1 \xrightarrow{(\forall)F12} O_2$, следует: $\forall o(o \in O_1 \supset o \in O_2)$, т.е. понятие $oO_1(o)$ включается в понятие $oO_2(o)$,</p> <p>– следовательно, признак $P(o)$ является характерным для $\forall o \in O_1$.</p> <div style="text-align: center;"> </div> <p>Утверждение_2: Если $F12(\forall O_1, O_2) \& F2(\forall O_2, P)$ то</p> <p>а) существует $P[O_1]$, где содержание понятия $pP[O_1](p)$ больше чем $pP(p)$, при этом $WpP(p) = WpP[O_2](p)$ и $WpP[O_1](p) \subseteq WpP[O_2](p)$, а следовательно, имеет место отношение $F11(\forall P[O_1], P)$;</p> <p>б) $F2(\forall O_1, P[O_1])$.</p> <div style="text-align: center;"> </div> <p>$P[O_1]$ – является примером интенционального расширения, применяемого для пополнения базы знаний путем расширения области определения.</p> <p>Утверждение_3: Если $F12(\forall O_1, O_2) \& F2(\forall O_1, P[O_1])$ то $F2(\exists O_2, P[O_1])$.</p> <p>Доказательство:</p>				

– пусть $\forall o \in O_1$ признак $P[O_1](o)$ является видовым отличием, так как других данных, на основании которых можно было бы выделить $\forall o \in O_1$ из O_2 нет;

– исходя из того, что $O_1 \xrightarrow{(\forall)F^{12}} O_2$, следует: $\forall o(o \in O_1 \supset o \in O_2)$, т.е. понятие $oO_1(o)$ включается в понятие $oO_2(o)$, а само понятие O_1 содержит данные о том, что элементы, входящие в O_1 , а следовательно и в O_2 (т.к. $\forall o(o \in O_1 \supset o \in O_2)$) обладают свойством, входящим в $P[O_1]$, то можно утверждать, что признак $P[O_1](o)$ является характерным хотя бы для некоторых $o \in O_2$.



Таким образом, получив на вход два СЗ:



– любая «Система приема сетевого трафика» обладает свойством «Способность принимать сетевой трафик»;

– всякая «Система приема трафика 802.3» является подклассом «Систем приема сетевого трафика»;

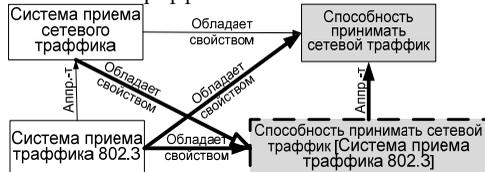
– система знаний создаст один концепт, «Способность принимать сетевой трафик [Система приема трафика 802.3]», и сформирует четыре СЗ:

– всякая «Система приема трафика 802.3» обладает свойством «Способность принимать сетевой трафик [Система приема трафика 802.3]»;

– свойство «Способность принимать сетевой трафик [Система приема трафика 802.3]» является подклассом (аппроксимирует) свойства свойством «Способность принимать сетевой трафик»;

– некоторые «Системы приема сетевого трафика» обладают свойством «Способность принимать сетевой трафик [Система приема трафика 802.3]», т.е. обладают способностями, присущими «Системам приема трафика 802.3»;

– все «Системы приема трафика 802.3» обладают свойством «Способность принимать сетевой трафик»:



Положительный эффект от вычислений над расширенной областью особенно заметен при моделировании многоэтапных процессов (например, многоэтапных атакующих воздействий и способов защиты от них). При этом следует отметить, что в ходе осуществления задержки вычислений осуществляется порождение нового типа данных, представленного в виде интенционала.

Изначально концепты произвольной предметной области представляются в онтологии в виде простых объектных термов, имена же концептов, полученных в ходе интенционального расширения, относятся к классу аппликативных объектных термов. Существенным свойством имен аппликативных объектных термов является то, что в них фиксируется их собственная «история» появления. Указанные термы далее можно использовать при построении спецификаций моделируемых процессов.

5. Пример, демонстрирующий некоторые возможности разработанной машины логического вывода, реализованной в ИС. Приведенные выше положения, связанные с порождением интенционалов, полезно пояснить на конкретном примере.

Пусть в памяти ИС имеется информация о том, что:

- 1) имеется «APM» (как класс);
- 2) все экземпляры класса «APM» обладают двумя типами свойств: «Подверженность DDoS» и «Время отклика на сетевой запрос»;
- 3) у класса «APM» есть подкласс — «APM_1» (в рамках типа «APM» выделяется подтип «APM_1»);
- 4) имеется класс «Кибер_Противник» («КП»);
- 5) все экземпляры класса «КП» обладают свойством «Способность осуществлять DDoS атаку», которое порождает способность осуществлять определенный тип действий: «Осуществление DDoS атаки»;
- 6) тип «КП» разделяется на два подтипа: «КП_1» и «КП_2».

Данные, непосредственно введенные в базу знаний (БЗ) экспертом, приведены в таблице 4 в пунктах 1-9, и в таблице 5 в пунктах 1-7. Указанные данные возможно обработать, применяя правила, примеры которых приведены в таблицах 2 и 3, и породить новые знания, представленные в таблице 4 в пунктах 10-15 и в таблице 5 в пунктах 8-31. Следует отметить, что при порождении новых знаний осуществляется как построение новых ролей, так и порождение новых концептов, не вводимых экспертом.

Таблица 4. Перечень концептов, введенных экспертом (экспертами) и сформированных ИС

№	Тип концепта	Имя концепта
1	Объект	APM
2	Объект	APM_1
3	Объект	Кибер_Противник (КП)
4	Объект	КП_1
5	Объект	КП_2
6	Свойство	Подверженность DDoS
7	Свойство	Время отклика на сетевой запрос
8	Свойство	Способность осуществлять DDoS атаку
9	Действие	Осуществление DDoS атаки
10	Свойство	Время отклика на сетевой запрос[APM_1]
11	Свойство	Подверженность DDoS[APM_1]
12	Свойство	Способность осуществлять DDoS атаку[КП_1]
13	Свойство	Способность осуществлять DDoS атаку[КП_2]
14	Действие	Осуществление DDoS атаки [Способность осуществлять DDoS атаку[КП_1]]
15	Действие	Осуществление DDoS атаки [Способность осуществлять DDoS атаку[КП_2]]
16	Объект	APM_1[Подверженность DDoS[APM_1], Осуществление DDoS Атаки [Способность осуществлять DDoS атаку[КП_1]]]
17	Свойство	Подверженность DDoS [APM_1[Подверженность DDoS[APM_1], Осуществление DDoS Атаки [Способность осуществлять DDoS атаку[КП_1]]]]
18	Свойство	Время отклика на сетевой запрос[APM_1 [Подверженность DDoS[APM_1], Осуществление DDoS Атаки [Способность осуществлять DDoS атаку[КП_1]]]]
19	Объект	APM_1[Подверженность DDoS[APM_1], Осуществление DDoS Атаки [Способность осуществлять DDoS атаку[КП_2]]]
20	Свойство	Подверженность DDoS [APM_1[Подверженность DDoS[APM_1], Осуществление DDoS Атаки [Способность осуществлять DDoS атаку[КП_2]]]]
21	Свойство	Время отклика на сетевой запрос[APM_1 [Подверженность DDoS[APM_1], Осуществление DDoS Атаки [Способность осуществлять DDoS атаку[КП_2]]]]

В таблице 5 в поле «О» указывается объем понятия: «А» — все (каждый, любой), «S» — некоторые (есть такие, которые...). В поле «Т» указывается тип концепта из состава рассматриваемого семантического звена: «О» — объект, «Р» — свойство, «А» — действие.

Таблица 5. Семантические звенья, введенные экспертом (экспертами) и выведенные ИС

№	О	Т	Концепт №1	Тип ребра	Т	Концепт №2
1	А	О	АРМ	Обладает свойством	Р	Подверженность DDoS
2	А	О	АРМ	Обладает свойством	Р	Время отклика на сетевой запрос
3	А	О	АРМ_1	Аппр.-т	О	АРМ
4	А	О	Кибер Противник (КП)	Обладает свойством	Р	Способность осуществлять DDoS атаку
5	А	Р	Способность осуществлять DDoS атаку	Порождает способность к	А	Осуществление DDoS атаки
6	А	О	КП_1	Аппр.-т	О	Кибер Противник (КП)
7	А	О	КП_2	Аппр.-т	О	Кибер Противник (КП)
8	А	О	АРМ_1	Обладает свойством	Р	Время отклика на сетевой запрос[АРМ_1]
9	А	О	АРМ_1	Обладает свойством	Р	Подверженность DDoS[АРМ_1]
10	S	О	АРМ	Обладает свойством	Р	Время отклика на сетевой запрос[АРМ_1]
11	S	О	АРМ	Обладает свойством	Р	Подверженность DDoS[АРМ_1]
12	А	Р	Время отклика на сетевой запрос[АРМ_1]	Аппр.-т	Р	Время отклика на сетевой запрос
13	А	О	АРМ_1	Обладает свойством	Р	Время отклика на сетевой запрос
14	А	Р	Подверженность DDoS[АРМ_1]	Аппр.-т	Р	Подверженность DDoS
15	А	О	АРМ_1	Обладает свойством	Р	Подверженность DDoS
16	А	О	КП_1	Обладает свойством	Р	Способность осуществлять DDoS атаку
17	А	О	КП_2	Обладает свойством	Р	Способность осуществлять DDoS атаку
18	А	О	КП_1	Обладает свойством	Р	Способность осуществлять DDoS атаку[КП_1]

Продолжение таблицы 5

№	O	T	Концепт №1	Тип ребра	T	Концепт №2
19	A	O	КП_1	Обладает свойством	P	Способность осуществлять DDoS атаку[КП_2]
20	A	P	Способность осуществлять DDoS атаку[КП_1]	Аппр.-т	P	Способность осуществлять DDoS атаку
21	A	P	Способность осуществлять DDoS атаку[КП_2]	Аппр.-т	P	Способность осуществлять DDoS атаку
22	S	O	Кибер_Противник (КП)	Обладает свойством	P	Способность осуществлять DDoS атаку[КП_1]
23	S	O	Кибер_Противник (КП)	Обладает свойством	P	Способность осуществлять DDoS атаку[КП_1]
24	A	P	Способность осуществлять DDoS атаку[КП_1]	Порождает способность к	A	Осуществление DDoS атаки
25	A	P	Способность осуществлять DDoS атаку[КП_2]	Порождает способность к	A	Осуществление DDoS атаки
26	A	P	Способность осуществлять DDoS атаку[КП_1]	Порождает способность к	A	Осуществление DDoS атаки [Способность осуществлять DDoS атаку[КП_1]]
27	A	P	Способность осуществлять DDoS атаку[КП_2]	Порождает способность к	A	Осуществление DDoS атаки [Способность осуществлять DDoS атаку[КП_2]]
28	A	A	Осуществление DDoS атаки [Способность осуществлять DDoS атаку[КП_1]]	Аппр.-т	A	Осуществление DDoS атаки
29	A	A	Осуществление DDoS атаки [Способность осуществлять DDoS атаку[КП_2]]	Аппр.-т	A	Осуществление DDoS атаки
30	S	P	Способность осуществлять DDoS атаку	Порождает способность к	A	Осуществление DDoS атаки [Способность осуществлять DDoS атаку[КП_1]]
31	S	P	Способность осуществлять DDoS атаку	Порождает способность к	A	Осуществление DDoS атаки [Способность осуществлять DDoS атаку[КП_2]]

Продолжение таблицы 5

32	A	A	Осуществление DDoS атаки [Способность осуществлять DDoS атаку[КП_1]]	Пригодно для воздействия на	P	Подверженность DDoS[APM_1]
33	A	O	APM_1	Продуцирует	O	APM_1[Подверженность DDoS [APM_1], Осуществление DDoS Атаки [Способность осуществлять DDoS атаку[КП_1]]]
34	A	O	APM_1[Подверженность DDoS [APM_1], Осуществление DDoS Атаки [Способность осуществлять DDoS атаку[КП_1]]]	Аппр.-т	O	APM
35	A	O	APM_1[Подверженность DDoS [APM_1], Осуществление DDoS Атаки [Способность осуществлять DDoS атаку[КП_1]]]	Обладает свойством	P	Подверженность DDoS [APM_1 [Подверженность DDoS [APM_1], Осуществление DDoS Атаки [Способность осуществлять DDoS атаку[КП_1]]]]
36	S	O	APM	Обладает свойством	P	Подверженность DDoS [APM_1[Подверженность DDoS[APM_1], Осуществление DDoS Атаки [Способность осуществлять DDoS атаку[КП_1]]]]
37	A	P	Подверженность DDoS [APM_1 [Подверженность DDoS [APM_1], Осуществление DDoS Атаки [Способность осуществлять DDoS атаку[КП_1]]]]	Аппр.-т		Подверженность DDoS
38	A	O	APM_1[Подверженность DDoS [APM_1], Осуществление DDoS Атаки [Способность осуществлять DDoS атаку[КП_1]]]	Обладает свойством	P	Время отклика на сетевой запрос[APM_1 [Подверженность DDoS[APM_1], Осуществление DDoS Атаки [Способность осуществлять DDoS атаку[КП_1]]]]

Продолжение таблицы 5

39	S	O	APM	Обладает свойством	P	Время отклика на сетевой запрос[APM_1 [Подверженность DDoS[APM_1], Осуществление DDoS Атаки [Способность осуществлять DDoS атаку[КП_1]]]]
40	A	P	Время отклика на сетевой запрос[APM_1 [Подверженность DDoS[APM_1], Осуществление DDoS Атаки [Способность осуществлять DDoS атаку[КП_1]]]]	Аппр.-т	P	Время отклика на сетевой запрос
41	A	A	<i>Осуществление DDoS атаки [Способность осуществлять DDoS атаку[КП_2]]]</i>	<i>Пригодно для воздействия на</i>	P	<i>Подверженность DDoS[APM_1]</i>
42	A	O	APM_1	Продуцирует	O	APM_1[Подверженность DDoS[APM_1], Осуществление DDoS Атаки [Способность осуществлять DDoS атаку[КП_2]]]]
43	A	O	APM_1[Подверженность DDoS [APM_1], Осуществление DDoS Атаки [Способность осуществлять DDoS атаку[КП_2]]]]	Аппр.-т	O	APM
44	A	O	APM_1[Подверженность DDoS [APM_1], Осуществление DDoS Атаки [Способность осуществлять DDoS атаку[КП_2]]]]	Обладает свойством	P	Подверженность DDoS [APM_1[Подверженность DDoS[APM_1], Осуществление DDoS Атаки [Способность осуществлять DDoS атаку[КП_2]]]]
45	S	O	APM	Обладает свойством	P	Подверженность DDoS [APM_1[Подверженность DDoS[APM_1], Осуществление DDoS Атаки [Способность осуществлять DDoS атаку[КП_2]]]]

Продолжение таблицы 5

46	A	P	Подверженность DDoS [APM_1 [Подверженность DDoS [APM_1], Осуществление DDoS Атаки [Способность осуществлять DDoS атаку [КП_2]]]]	Аппр.-т	P	Подверженность DDoS
47	A	O	APM_1[Подверженность DDoS [APM_1], Осуществление DDoS Атаки [Способность осуществлять DDoS атаку[КП_2]]]	Обладает свойством	P	Время отклика на сетевой запрос [APM_1 [Подверженность DDoS [APM_1], Осуществление DDoS Атаки [Способность осуществлять DDoS атаку[КП_2]]]]
48	S	O	APM	Обладает свойством	P	Время отклика на сетевой запрос [APM_1 [Подверженность DDoS [APM_1], Осуществление DDoS Атаки [Способность осуществлять DDoS атаку[КП_2]]]]
49	A	P	Время отклика на сетевой запрос[APM_1 [[Подверженность DDoS [APM_1], Осуществление DDoS Атаки [Способность осуществлять DDoS атаку [КП_2]]]]	Аппр.-т	P	Время отклика на сетевой запрос

Следует пояснить порядок формирования имен концептов, порожденных в ходе осуществления интенционального расширения онтологии и представленных в таблицах 4 и 5. Так, например, «*Осуществление DDoS атаки [Способность осуществлять DDoS атаку[КП_1]]*» означает возможность осуществлять конкретный тип действий («*Осуществление DDoS атаки*») конкретным типом объектов («*КП_1*») ввиду того, что они (объекты) обладают определенным типом свойств («*Способность осуществлять DDoS атаку*»).

Далее, если эксперт укажет на то, что «*Осуществление DDoS атаки [Способность осуществлять DDoS атаку[КП_1]]*» пригодно для воздействия на свойство «*Подверженность DDoS[APM_1]*» (см. п. 32 таблицы 5), а также, что и «*Осуществление DDoS атаки [Способность осуществлять DDoS атаку[КП_2]]*» пригодно для воздействия на свойство «*Подверженность DDoS[APM_1]*» (см. п. 41 таблицы 5), то система знаний способна породить ряд концептов (см. п.п. 16-21 таблицы 4) и семантических звеньев (см. п.п. 33-40 и 42-49 таблицы 5). При этом имя концепта «*APM_1[Подверженность DDoS[APM_1], Осуществление DDoS Атаки [Способность осуществлять DDoS атаку[КП_2]]]*» хранит историю его формирования и означает «*APM_1*», но после того, как на него осуществит воздействие «*КП_2*» путем «*Осуществление DDoS Атаки*» через свойство «*Подверженность DDoS*».

Порожденный объект обладает также двумя свойствами («*Подверженность DDoS[APM_1][Подверженность DDoS[APM_1], Осуществление DDoS Атаки [Способность осуществлять DDoS атаку[КП_1]]]*» и «*Время отклика на сетевой запрос[APM_1 [Подверженность DDoS[APM_1], Осуществление DDoS Атаки [Способность осуществлять DDoS атаку[КП_1]]]*»), как и его «родитель», но эти свойства уже не вносятся экспертом, а порождаются ИС самостоятельно, что позволяет сократить время описания рассматриваемой предметной области и количество возможных ошибок, которые может совершить эксперт. К основным из которых можно отнести упущение некоторых фактов.

Приведенный пример демонстрирует принципиальную способность ИС пополнять собственную БЗ новыми концептами и/или семантическими звеньями, не заносимыми в БЗ непосредственно создателем (экспертом), либо самой интеллектуальной системой по результатам наблюдения за явлениями во «Внешнем Мире», что несомненно способствует повышению уровня интеллектуального содержания БЗ и снижению времени ее пополнения.

Кроме того, порождение концептов типа: «*Время отклика на сетевой запрос[APM_1 [Подверженность DDoS[APM_1], Осуществление DDoS Атаки [Способность осуществлять DDoS атаку[КП_1]]]*» позволяет говорить о возможности построения в ходе предвычислений своего рода протофреймов, которые потенциально могут быть в дальнейшем заполнены конкретными значениями. Означивание может быть осуществлено как экспертами рассматриваемой предметной области, так и в ходе измерения параметров явлений, наблюдаемых через систему сенсоров самой ИС.

6. Оценивание показателей пополнения онтологии системы знаний. При оценивании результатов работы ИС, заключающейся в осуществлении ею правдоподобных умозаключений, предлагается оценить прирост информационного содержания БЗ.

В рамках рассматриваемого выше примера экспертом было введено 9 концептов из 21 и 9 ролей из 49, остальные были порождены системой знаний (таблица 6). Из значений, приведенных в таблице 6, можно сделать вывод о том, что в результате работы системы знаний было изменено (пополнено) информационное содержание ее БЗ. Для измерения информационного содержания БЗ можно использовать тезаурусную меру информации.

Таблица 6. Количественные параметры рассматриваемого примера

Предметная область	Кол-во Концептов	Кол-во СЗ
Всего введено экспертами	9	9
Построено системой знаний	12	40
Всего в БЗ	21	49

Под тезаурусом будем понимать словарь, в котором слова, относящиеся к какой-либо области знания, расположены по тематическому принципу и показаны семантические отношения между лексическими единицами [9]. Исходя из приведенного определения, элементом тезауруса предлагается считать СЗ. В пользу данного предложения можно отнести положение, обоснованное в работе [10]: «Чем больше элементов в системе, тем большую долю содержащейся в ней информации составляет информация, содержащаяся во взаимосвязях ее элементов».

Учитывая, что согласно определению тезаурусной меры информации [11] количеством семантической информации, содержащейся в тексте T относительно тезауруса Q , называется степень изменения тезауруса Q под воздействием текста T ($I(T, Q) = (Q(T) - Q) / Q$), то изменение информационного содержания БЗ ИС по результатам осуществления ею интеллектуальной обработки поступающих данных можно определить как:

$$I_T(Z_M, X_{SM}) = \sum_{i=1}^k \frac{N(Z_M(Z_M, X_{SM_i})) - N(Z_M)}{N(Z_M)},$$

где X_{SM} — Семантическое Звено (элемент тезауруса);

$N(Z_M)$ — количество СЗ в БЗ.

Приведенное выражение позволяет оценить количественные результаты работы машины логического вывода, реализованной в ИС. Можно определить:

I_{T_1} — тезаурусную меру информации, если под тезаурусом понимать совокупность концептов и ролей: (т.е. структурных элементов онтологии): $I_{T_1} = ((21+49) - (9+9)) / (9+9) \approx 2,9$;

I_{T_2} — тезаурусную меру информации, если под тезаурусом понимать совокупность СЗ: $I_{T_2} = (49 - 9) / 9 \approx 4,4$.

На первый взгляд может показаться, что тезаурусная мера информации I_{T_1} более привлекательна, так как она непосредственно учитывает все структурные элементы из которых формируется онтология: все концепты и роли. Однако предлагается применять на практике меру I_{T_2} . Аргументы в пользу данного выбора лучше продемонстрировать на конкретном примере (рисунок 5 и таблица 3).

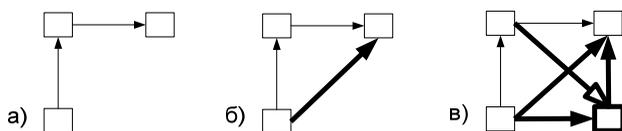


Рис. 5. Иллюстрация к примеру расчета тезаурусной меры информации

Пусть на вход ИС поступила простейшая семантическая конструкция, представленная на рисунке 5(а). Тогда в том случае, если система знаний не осуществляет построение интенционалов (как делает большинство машин логического вывода, применяемых для пополнения онтологий [12-16]), то в результате осуществления логических выводов может быть получена конструкция, приведенная на рисунке 5(б). Если же система знаний в состоянии порождать не только новые роли, но и концепты, то в результате ее работы может быть построен фрагмент онтологии, изображенный на рисунке 5(в).

Если рассчитать изменение информационного содержания БЗ (I_{T_1} и I_{T_2}) для рассмотренных построений (рисунок 5(а) и рисунок 5(б)), то можно получить следующие значения:

— для рис. 5(б) $I_{T_1} = 1/5$, $I_{T_2} = 1/2$;

— для рис. 5(в) $I_{T_1} = 1$, $I_{T_2} = 2$.

Прежде чем оценивать полученные значения, полезно отметить, что I_T указывает на то, насколько увеличилось информационное содержание БЗ относительно уже имеющейся информации: если в БЗ

имелось X информации, то после работы машины логического вывода ее количество становится равным $X + X \times I_T$.

Исходя из приведенного примера и указанного замечания можно отметить, что при оценивании количественных результатов работы машины логического вывода следует вычислять I_{T2} , так как данная оценка в большей степени соответствует эмпирическим знаниям.

При необходимости можно проанализировать графики, отображающие темпы изменения информационного содержания БЗ ИС в ходе занесения в онтологию новых знаний и их первичной интеллектуальной обработки, но ввиду ограничений на объем статьи данная динамика не приводится.

7. Заключение. Предложенный механизм обработки поступающих на вход интеллектуальной системы фактов направлен на их автоматическую интеграцию в имеющуюся у нее онтологию, в ходе которой порождаются новые знания. Порождение новых знаний осуществляется в ходе построения умозаключений на основе простого категорического силлогизма с учетом семантики используемых ролей, а также в ходе пополнения онтологии порожденными интенционалами. Оба указанных метода реализованы в системы знаний, примером которой является гиромат [1], и выполняются во взаимосвязи друг с другом, что позволяет ей в ходе осуществления предвычислений не только дополнять онтологию достроенными ролями, но и встраивать в упорядоченное множество концептов новые концепты, порожденные ей. Автоматически осуществляемое гироматом пополнение онтологии, значительно повышает ее информационное содержание и очевидно сокращает время ее пополнения.

Литература

1. *Бирюков Д.Н.* Когнитивно-функциональная спецификация памяти для моделирования целенаправленного поведения киберсистем // Труды СПИИРАН. 2015. Вып. 3(40). С. 55–76.
2. *Бирюков Д.Н., Ломако А.Г.* Подход к построению системы предотвращения киберугроз // Проблемы информационной безопасности. Компьютерные системы. С-Пб.: Издательство Политехнического университета. 2013. №2. С. 13–19.
3. *Бирюков Д.Н., Ломако А.Г., Ростовцев Ю.Г.* Облик антиципирующих систем предотвращения рисков реализации киберугроз // Труды СПИИРАН. 2015. Вып. 2(39). С. 5–25.
4. *Бирюков Д.Н., Ломако А.Г., Сабиров Т.Р.* Многоуровневое моделирование сценариев упреждающего поведения // Проблемы информационной безопасности. Компьютерные системы. С-Пб.: Издательство Политехнического университета. 2014. №4. С. 41–50.
5. *Бирюков Д.Н., Ломако А.Г.* Денотационная семантика контекстов знаний при онтологическом моделировании предметных областей конфликта // Труды СПИИРАН. 2015. Вып. 5(42). С. 155–179.

6. *Бочаров В.А., Маркин В.И.* Основы логики // М.: МГУ. 2008.
7. *Бирюков Д.Н., Ломако А.Г.* Формализация семантики для представления знаний о поведении конфликтующих сторон // Материалы 22-й научно-практической конференции «Методы и технические средства обеспечения безопасности информации». С-Пб.: Издательство Политехнического университета. 2013. С. 8–11.
8. *Кистлеров В.Л.* Принципы построения языка алгебраических вычислений FLAC // М.: Препринт. 1987. 41 с.
9. *Шрейдер Ю.А.* Тезаурусы в информатике и теоретической семантике // «Научно-техническая информация. Сер. 2». 1971. № 3. С. 7–12
10. *Луценко Е.В., Коржаков В.Е.* Количественные меры уровня системности и степени детерминированности в рамках СТИ // Вестник Адыгейского государственного университета. 2006. № 4. С. 169–177.
11. *Барановская Т.П., Лойко В.И., Семенов М.И., Трубилин А.И.* Информационные системы и технологии в экономике. М.: Финансы и статистика. 2005. 416 с.
12. What is RacerPro?: Racer Systems GmbH & Co.KG. 2009 // URL: <http://www.racer-systems.com/products/racerpro/> (дата обращения 14.12.2015).
13. *Shearer R., Motik B., Horrocks I.* Hermit: A Highly-Efficient OWL Reasoner // Proc. of the 5th Int. Workshop on OWL: Experiences and Directions (OWLED 2008 EU), Karlsruhe, Germany. 2008. 10 p.
14. *Tsarkov D., Horrocks I.* FaCT++ description logic reasoner: System description // In Proc. of the Int. Joint Conf. on Automated Reasoning (IJCAR 2006). 2006. 5 p.
15. *Sirin E., Parsia B., Gran B., Kalyanpur A., Katz Y.* Pellet: A Practical OWL-DL Reasoner // Journal Web Semantics: Science, Services and Agents on the World Wide Web archive. 2007. vol. 5. no 2, pp. 51–53.
16. *Haarslev V., Moller R.* RACER: A core inference engine for the Semantic Web // Proc. of the 2nd Intern. workshop on evaluation of ontology-based tools (EON-2003), Florida. USA. 2003. pp. 27–36.
17. *Calvanese D., Giacomo G., Lenzerini M., Rosati R.* View-based query answering in description logics: Semantics and complexity // J. of Computer and System Sciences. 2012. vol. 78(1). pp. 26–46.
18. *Eiter T., Ortiz M., Simkus M.* Conjunctive query answering in the description logic SH using knots. // J. of Computer and System Sciences 2012. vol. 78(1). pp. 47–85.
19. *Eiter T., Ortiz M., Simkus M., Tran T.K., Xiao G.* Query rewriting for Horn-SHIQ plus rules // Proc. of the 26th Nat. Conf. on Artificial Intelligence (AAAI 2010). AAAI Press. 2012. pp. 726–733.
20. *Kikot S., Zolin E.* Modal definability of first-order formulas with free variables and query answering // Journal of Applied Logic. 2013. vol. 11. no. 2. pp. 190–216.
21. *Kikot S., Tsarkov D., Zakharyashev M., Zolin E.* Query Answering via Modal Definability with FaCT++: First Blood // Informal Proceedings of DL 2013: 26th International Workshop on Description Logics. CEUR Workshop Proceedings, 2013. vol. 1014. pp. 328–340.
22. *Kikot S., Konchakov R., Podolskii V., Zakharyashev M.* Query Rewriting over Shallow Ontologies // Informal Proceedings of DL 2013: 26th International Workshop on Description Logics. CEUR Workshop Proceedings. 2013. vol. 1014. pp. 316–327
23. *Maier F., Ma Y., Hitzler P.* Paraconsistent OWL and related logics // Semantic Web. 2013. vol. 4(4). pp. 395–427.
24. *Zhang X., Wang K., Wang Z., Ma Y., Qi G.* A distance-based paraconsistent semantics for DL-Lite // International Conference on Knowledge Science, Engineering and Management. 2015. LNAI 9403. pp. 1–13.
25. *Goncalves R., Matentzoglou N., Parsia B., Sattler U.* The empirical robustness of description logic classification // Description Logics. 2013. vol. 1014. pp. 197–208.

26. *Goncalves R. et al.* OWL reasoner evaluation (ORE) workshop 2013 results: Short report // 2nd International Workshop on OWL Reasoner Evaluation (ORE-2013). 2013. vol. 1015. pp.1–18.
27. *Glimm B., Horrocks I., Motik B., Stoilos G., Wang Z.* Hermit: An owl 2 reasoner // *Autom. Reason.* 2014. vol. 53(3). pp. 245–269.
28. *Amir S., Ait-Kaci H.* Cedar: a fast taxonomic reasoned based on lattice operations // *International Semantic Web Conference (Posters & Demos)*. 2013. pp. 9–12.

References

1. Biryukov D.N. [The Cognitive and Functional Specification of Memory for Modeling of Purposeful Behavior of Cybersystems]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2015. vol. 3(40). pp. 55–76. (In Russ.).
2. Biryukov D.N., Lomako A.G. [Approach to creation of system of cyber-threats preventing]. *Problemy informatsionnoy bezopasnosti. Kompyuternie sistemy - Problems of information security. Computer systems*. SPB: St. Petersburg Polytechnical University. 2013. vol. 2. pp. 13–19. (In Russ.).
3. Biryukov D.N., Lomako A.G., Rostovtsev Y.G. [The Appearance Cyber Threats Risk Prevention Systems]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2015. vol. 2(39). pp. 5–25. (In Russ.).
4. Biryukov D.N., Lomako A.G., Sabirov T.R. [Multilevel modeling of scenarios of anticipatory behavior]. *Problemy informatsionnoy bezopasnosti. Kompyuternie sistemy – Problems of information security. Computer systems*. SPB: St. Petersburg Polytechnical University 2014. vol. 4. pp. 41–50. (In Russ.).
5. Biryukov D.N., Lomako A.G. [Denotational semantics of knowledge contexts in ontologic modeling of subject domains of the conflict]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2015. vol. 5(42). pp. 155–179. (In Russ.).
6. Bocharov V.A., Markin V.I. *Osnovy logiki* [Logic bases]. M.: MSU. 2008. (In Russ.).
7. Biryukov D.N., Lomako A.G. [Formalization of semantics for representation of knowledge of behavior of conflicting parties]. *Materialy 22 nauchno-prakticheskoy konferentsiyi "Metody I tehnichestkiye sredstva obespecheniya bezopasnosti informacii"* [Materials of the 22nd scientific and practical conference "Methods and Technical Means of Safety of Information"]. SPB: St. Petersburg Polytechnical University. 2013. pp. 8–11. (In Russ.).
8. Kistlerov V.L. [Design principles for the language of algebraic computations FLAC]. M.: Preprint. 1987. 41 p. (In Russ.).
9. Schreider Y.A. [Thesauri in Informatics and in theoretical semantics]. *Nauchno-tehnicheskaya informatsiya. Ser. 2* – *Scientific and technical information. Ser. 2*. 1971. vol. 3. pp. 7–12. (In Russ.).
10. Lutsenko E.V., Korzhakov V.E. [Quantitative measures of the level of consistency and a degree of determinancy within the framework of STI]. *Vestnik Adygejskogo gosudarstvennogo universiteta – the Bulletin of Adyge state University*. 2006. vol. 4. pp. 169–177. (In Russ.).
11. Baranovskaya T.P., Loyko V.I., Semenov M.I., Trubilin A.I. [Information systems and technologies in Economics]. M.: *Finansy i statistika – Finance and statistics*. 2005. 416 p. (In Russ.).
12. What is RacerPro?: Racer Systems GmbH & Co.KG. 2009. Available at: <http://www.racer-systems.com/products/racerpro/> (accessed: 14.12.2015).
13. Shearer R., Motik B., Horrocks I. Hermit: A Highly-Efficient OWL Reasoner. Proc. of the 5th Int. Workshop on OWL: Experiences and Directions (OWLED 2008 EU). Karlsruhe. Germany. 2008. 10 p.
14. Tsarkov D., Horrocks I. FaCT++ description logic reasoner: System description. In Proc. of the Int. Joint Conf. on Automated Reasoning (IJCAR 2006). 2006. 5 p.

15. Sirin E., Parsia B., Gran B., Kalyanpur A., Katz Y. Pellet: A Practical OWL-DL Reasoner. *Journal Web Semantics: Science, Services and Agents on the World Wide Web archive*, 2007. vol. 5. no. 2. pp. 51–53.
16. Haarslev V., Moller R. RACER: A core inference engine for the Semantic Web. Proc. of the 2nd Intern. workshop on evaluation of ontology-based tools (EON-2003). Florida. USA. 2003. pp. 27–36.
17. Calvanese D., Giacomo G., Lenzerini M., Rosati R. View-based query answering in description logics: Semantics and complexity. *J. of Computer and System Sciences*. 2012. vol. 78(1). pp. 26–46.
18. Eiter T., Ortiz M., Simkus M. Conjunctive query answering in the description logic SH using knots. *J. of Computer and System Sciences*. 2012. vol. 78(1). pp. 47–85.
19. Eiter T., Ortiz M., Simkus, M., Tran T.K., Xiao G. Query rewriting for Horn-SHIQ plus rules. Proc. of the 26th Nat. Conf. on Artificial Intelligence (AAAI 2010). AAAI Press. 2012. pp. 726–733.
20. Kikot S., Zolin E. Modal definability of first-order formulas with free variables and query answering. *Journal of Applied Logic*. 2013. vol. 11. no. 2. pp. 190–216.
21. Kikot S., Tsarkov D., Zakharyashev M., Zolin E. Query Answering via Modal Definability with FaCT⁺⁺: First Blood. Informal Proceedings of DL 2013: 26th International Workshop on Description Logics. CEUR Workshop Proceedings. 2013. vol. 1014. pp. 328–340.
22. Kikot S., Kontchakov R., Podolskii V., Zakharyashev M. Query Rewriting over Shallow Ontologies. *Informal Proceedings of DL 2013: 26th International Workshop on Description Logics. CEUR Workshop Proceedings*. 2013. vol. 1014. pp. 316–327.
23. Maier F., Ma Y., Hitzler P. Paraconsistent OWL and related logics. *Semantic Web*. 2013. vol. 4(4). pp. 395–427.
24. Zhang X., Wang K., Wang Z., Ma Y., Qi G. A distance-based paraconsistent semantics for DL-Lite. International Conference on Knowledge Science, Engineering and Management. 2015. LNAI 9403. pp. 1–13.
25. Goncalves R., Matentzoglou N., Parsia B., Sattler U. The empirical robustness of description logic classification. *Description Logics*. 2013. vol. 1014. pp. 197–208.
26. Goncalves R., Bail S., Jiménez-Ruiz E., Matentzoglou N., Parsia B., Glimm B., Kazakov Y. OWL reasoner evaluation (ORE) workshop 2013 results: Short report. 2nd International Workshop on OWL Reasoner Evaluation (ORE-2013). 2013. vol. 1015. pp. 1–18.
27. Glimm B., Horrocks I., Motik B., Stoilos G., Wang Z. Hermit: An owl 2 reasoner. *Autom. Reason*. 2014. vol. 53(3). pp. 245–269.
28. Amir S., Ait-Kaci H. Cedar: a fast taxonomic reasoner based on lattice operations. International Semantic Web Conference (Posters & Demos). 2013. pp. 9–12.

Бирюков Денис Николаевич — д-р техн. наук, профессор кафедры систем сбора и обработки информации, Военно-космическая академия имени А.Ф. Можайского (ВКА им. А.Ф. Можайского). Область научных интересов: системный анализ, защита информации, интеллектуальная поддержка принятия решений. Число научных публикаций — 70. Biryukov.D.N@yandex.ru; ул. Ждановская, 13, Санкт-Петербург, 197198; р.т.: +7(812) 237-19-60.

Biryukov Denis Nikolaevich — Ph.D., Dr. Sci., professor of systems for collecting and processing information department, Mozhaisky Military Space Academy. Research interests: system analyses, IT-Security, intelligent decision support. The number of publications — 70. Biryukov.D.N@yandex.ru; 13, Zhdanovskaya street, St.-Petersburg, 197198, Russia; office phone: +7(812) 237-19-60.

Ломako Александр Григорьевич — д-р техн. наук, профессор кафедры систем сбора и обработки информации, Военно-космическая академия имени А.Ф. Можайского (ВКА им. А.Ф. Можайского). Область научных интересов: информационная безопасность, теоретическое и системное программирование, синтез и верификация корректности моделей программ. Число научных публикаций — 290. lomako_ag@mail.ru; ул. Ждановская, 13, Санкт-Петербург, 197198; р.т.: +7(812) 237-19-60.

Lomako Aleksandr Grigor'evich — Ph.D., Dr. Sci., professor of system for collecting and processing information department, Mozhaisky Military Space Academy. Research interests: information security, theoretical and system programming, synthesis and verification of program models. The number of publications — 290. lomako_ag@mail.ru; 13, Zhdanovskaya street, St.-Petersburg, 197198, Russia; office phone: +7(812) 237-19-60.

Жолус Роман Борисович — к-т биол. наук, соискатель кафедры систем сбора и обработки информации, Военно-космическая академия имени А.Ф. Можайского (ВКА им. А.Ф. Можайского). Область научных интересов: информационная безопасность, моделирование социальных систем. Число научных публикаций — 10. p.glybovsky@yandex.ru; ул. Ждановская, 13, Санкт-Петербург, 197198; р.т.: +7(812) 237-19-60.

Zholus Roman Borisovich — Ph.D., applicant of system for collecting and processing information department, Mozhaisky Military Space Academy. Research interests: information security, modeling social systems. The number of publications — 10. p.glybovsky@yandex.ru; 13, Zhdanovskaya street, St.-Petersburg, 197198, Russia; office phone: +7(812) 237-19-60.

РЕФЕРАТ

Бирюков Д.Н., Ломако А.Г., Жолус Р.Б. **Пополнение онтологических систем знаний на основе моделирования умозаключений с учетом семантики ролей.**

Известно, что способность человека увязывать предъявляемые ему факты с имеющимися у него знаниями и ценностными приоритетами значительно обогащает и актуализирует их. В этой связи предлагается в интеллектуальной системе, призванной синтезировать сценарии целенаправленного поведения, реализовать механизм автоматической обработки поступающих на ее вход данных с целью их интеграции в имеющуюся у нее онтологию. Предложенный механизм интеграции поступающих в память интеллектуальной системы данных предполагает не только автоматическое построение ролей онтологии, но и автоматическое формирование концептов, что значительно повышает информационное содержание онтологии и сокращает время ее пополнения.

SUMMARY

Biryukov D.N., Lomako A.G., Zholus R.B. **Ontological Knowledge System Completion Based on Modeling Inferences Taking into Account Role Semantics.**

It is known that the human ability to link obtained facts with existing knowledge and value priorities enriches and updates the latter. In this connection, it is suggested that the intelligent system designed to synthesize purposeful behavior scenarios should implement the mechanism of automatic processing of input data with the aim of their integration into existing ontology. The proposed mechanism for integrating input data into the intelligent system's memory involves not only automatic construction of ontology roles but also automatic formation of concepts, which greatly increases ontology information content and decreases time of its completion.

А.А. ЛИПАТОВ, В.Н. УШАКОВ, М.В. НИКИТИНА
**МЕТОДЫ ВЫЯВЛЕНИЯ ОДНОРОДНЫХ И НЕОДНОРОДНЫХ
ГРУПП ОБЪЕКТОВ НА ОСНОВЕ НЕОПРЕДЕЛЕННЫХ
КАЧЕСТВЕННЫХ ДАННЫХ**

Липатов А.А., Ушаков В.Н., Никитина М.В. Методы выявления однородных и неоднородных групп объектов на основе неопределенных качественных данных.

Аннотация. В статье предложены критерии выявления групп объектов на основе неопределенных оценок значений их качественных признаков.

Решаются задачи выявления однородных и неоднородных групп объектов. В однородных группах значения одноименных качественных признаков всех объектов совпадают. В неоднородных группах значения таких признаков могут не совпадать, однако должны удовлетворять априорно заданным допустимым сочетаниям.

Для выявления групп объектов применяется теоретико-графовый подход. При этом неопределенные оценки признаков объектов представляются в виде множеств их возможных значений. При принятии решения о принадлежности пары объектов к одной группе используется трехзначная логика, что позволяет выявлять возможные и достоверные группы.

Ключевые слова: группирование, качественные данные, неопределенные данные, недоопределенные вычисления, многозначная логика, теория графов.

Lipatov A. A., Ushakov V.N., Nikitina M.V. Methods of Detecting Homogeneous and Heterogeneous Groups of Objects based on Uncertain Qualitative Data.

Abstract. In this paper, the criteria of detection of groups of objects are suggested. These criteria are based on uncertain estimates of objects qualitative attributes. The tasks of the homogeneous and heterogeneous detection of groups of objects are solved. In the homogeneous groups, the values of cognominal qualitative attributes are equal. In the heterogeneous groups, the values of such attributes may differ, but they have to match the a priori set valid combinations. The groups detection is based on the graph-theoretical approach. The decision about the pair of objects belonging to the same group is made by the ternary logic. That allows detecting possible and reliable object groups.

Keywords: grouping, qualitative data, uncertain data, subdefinite computations, multiple-valued logic, graph theory.

1. Введение. Задача выявления групп (группирования) взаимодействующих объектов встречается во многих предметных областях, в частности, в радиолокации, информационной безопасности и др. В человеко-машинных системах, операторы которых должны воспринимать и анализировать информацию о большом количестве объектов за ограниченное время, применяется обобщение отображаемой информации на основе группирования наблюдаемых объектов.

Группирование объектов особенно актуально в условиях, когда их взаимодействие может вызвать «эффект роя» [1-3].

Группируемые объекты описываются набором своих координат состояния (КС), которые могут быть как количественными (например,

координаты в пространстве, скорость и курс подвижных объектов), так и качественными (класс, тип, принадлежность какому-либо субъекту и др.).

Признаком принадлежности объектов к одной группе может быть их близость в пространстве КС. Понятие близости объектов формализуется в виде некоторых показателей, на основе которых формулируются критерии принадлежности объектов к одной группе.

Так, в пространстве количественных КС показателями близости объектов могут служить евклидово расстояние, квадрат евклидова расстояния, взвешенное евклидово расстояние между объектами и др. В [4] в качестве показателя близости предложено отношение r_{cp}/R_{cp} , где r_{cp} — среднее расстояние между объектами в группе, а R_{cp} — среднеерасстояние между группами, в [5] показателями близости служат модули разностей значений одноименных КС объектов.

Критерии принадлежности объектов к одной группе могут быть как экстремальными [4], так и пороговыми [5]. При этом в ряде практически значимых приложений для группирования могут использоваться априорно заданные пороговые значения [5].

В пространстве качественных КС (признаков) будем рассматривать два критерия принадлежности объектов к одной группе. Первый из них требует совпадения значений одноименных признаков объектов. Группы объектов, выделенные по этому критерию, будем называть *однородными*.

Второй критерий требует, чтобы значения одноименных качественных признаков всех объектов группы образовывали допустимые сочетания, заданные априорно. Такие группы объектов будем называть *неоднородными*.

Следует учитывать, что оценки КС многих реальных объектов характеризуются неопределенностью, обусловленной как ошибками их оценивания, так и неполнотой информации об объектах. В связи с этим актуальной является разработка методов группирования объектов на основе неопределенных количественных и качественных данных.

Целью данной статьи является разработка методов выявления однородных и неоднородных групп объектов на основе неопределенных значений их качественных признаков.

2. Методы группирования объектов. В настоящее время известны методы группирования объектов, предназначенные для использования в различных предметных областях [4-18]. С помощью этих методов решаются задачи обнаружения и сопровождения множественных объектов [6], сегментации совместно движущихся объектов [7], распознавания их групповых действий [8].

Примеры методов группирования, ориентированных на использование точных оценок КС объектов, представлены в [4, 5, 9, 10].

Широкое применение для группирования объектов в условиях неопределенности нашли подходы нечеткой логики и нечеткой кластеризации [11-14]. Используются также генетические алгоритмы [15], нейронные сети [16], недоопределенные вычисления [17, 18] и другие подходы из области искусственного интеллекта.

При группировании объекты описываются набором их КС $X = \{x_1, x_2, \dots, x_k, x_{k+1}, \dots, x_n\}$, где $X_{quant} = \{x_1, x_2, \dots, x_k\}$ – количественные КС, а $X_{qual} = \{x_{k+1}, x_{k+2}, \dots, x_n\}$ — качественные признаки объекта.

Для каждой КС x_i , $i = \overline{1, n}$ известно множество A_i ее допустимых точных значений. Для числовых КС — это множества целых или действительных чисел, а для качественных признаков – множества символьных значений, соответствующих тем или иным возможным состояниям данного параметра объекта.

Если используются точные данные об объекте, то оценки КС x_i , $i = \overline{1, n}$ имеют значения в виде элементов множеств A_i .

Группирование объектов может быть основано на проверке пар объектов на принадлежность к одной группе [5]. Обозначим критерий принадлежности пары объектов p и q к одной группе $Cg(p, q)$. Для проверки его выполнения используется ряд частных критериев близости объектов в некоторых подпространствах пространства КС объектов.

Например, для каждой i -ой КС пары объектов $i = \overline{1, n}$ задается критерий $Cr_i(x_{pi}, x_{qi})$, где x_{pi} и x_{qi} — значения этой координаты объектов p и q соответственно. Пара объектов p и q считается принадлежащей к одной группе, если для любого $i = \overline{1, n}$ критерий $Cr_i(x_{pi}, x_{qi})$ выполняется.

Для количественных КС $Cr_i(x_{pi}, x_{qi})$ может иметь вид [5]:

$$\left| x_{pi} - x_{qi} \right| < \varepsilon_i, \quad i = \overline{1, k}, \quad (1)$$

где ε_i — априорно заданное максимально допустимое различие значений i -х КС.

Также для любого множества КС $X' \subseteq X_{quant}$ может быть задан критерий $Cr'(X'_p, X'_q)$ близости объектов p и q в пространстве ко-

ординат X' . Такой критерий может быть основан на евклидовом расстоянии между точками в этом пространстве, квадрате евклидова расстояния, взвешенном евклидовом расстоянии и др.

Для качественных признаков критерий принадлежности пары объектов p и q к *однородной* группе имеет вид:

$$x_{pi} = x_{qi}, \quad i = \overline{k+1, n}. \quad (2)$$

Критерий принадлежности пары объектов p и q к *неоднородной* группе формулируется следующим образом:

$$x_{pi} \in D_i, \quad x_{qi} \in D_i, \quad i = \overline{k+1, n}, \quad (3)$$

где $D_i \subset A_i$ – допустимое сочетание значений i -го признака объектов в группе.

С использованием указанных критериев вывод о принадлежности пары объектов p и q к одной группе можно сделать в следующем случае: для любого $l = \overline{1, k}$ выполняется критерий $Cr_l(x_{pl}, x_{ql})$ либо $Cr'(X'_p, X'_q)$, $x_{pl} \in X'_p$, $x_{ql} \in X'_q$, и для любого $l = \overline{k+1, n}$ выполняется критерий $Cr_l(x_{pl}, x_{ql})$.

Для выявления групп объектов на основе описанных выше критериев можно использовать методы теории графов [5]. Строится граф, вершины которого соответствуют объектам, а ребра обозначают принадлежность пары объектов к одной группе. Задача группирования объектов решается путем выделения связанных компонент графа с помощью алгоритма поиска в глубину [19].

3. Методы группирования объектов с неопределенными координатами состояния. Наиболее простым способом учета неопределенности оценок значений КС объектов является представление количественных КС в виде числовых интервалов, а качественных признаков в виде множеств их возможных значений, включающих в себя действительные, но неизвестные в данный момент значения этих КС.

В этом случае объект представляется набором своих КС $\{^*x_1, ^*x_2, \dots, ^*x_k, ^*x_{k+1}, \dots, ^*x_n\}$, где $^*x_i, i = \overline{1, k}$ — его количественные координаты, а $^*x_i, i = \overline{k+1, n}$ — качественные признаки.

Каждая количественная КС $^*x_i, i = \overline{1, k}$, может быть представлена интервалом с нижней границей $x_{iН} = \hat{x} - \Delta x_i$ и верхней границей

$x_{iB} = \hat{x} + \Delta x_i$, где $\hat{x}_i = x_i + \xi_i$ — текущая оценка i -й КС объекта (число), x_i — ее истинное значение, ξ_i — ошибка ее измерения, $\Delta x_i = 3\sigma_i$, σ_i — среднеквадратичное отклонение величины ξ_i .

Для качественных признаков x_i^* , $i = \overline{k+1, n}$ известны конечные множества их возможных значений A_i , а текущие оценки их значений представлены множествами $B_i \subseteq A_i$. При $|B_i| = 1$ значение x_i^* известно точно, а случай, когда $|B_i| = |A_i|$ соответствует полной неопределенности значения x_i^* .

Учитывая, что в условиях неопределенности исходных данных заключение о принадлежности объекта к той или иной группе также может быть неоднозначным, целесообразно ввести два типа групп: *возможные* и *достоверные*. Представлять такие группы можно с помощью недоопределенных множеств (Н-множеств) [20].

Н-множество S представляется тремя компонентами:

$$S = S^+, S^-, M,$$

где S^+ — множество элементов, которые точно принадлежат S , S^- — множество элементов, которые точно не принадлежат S , M — мощность Н-множества S . Мощность M Н-множества S представляет собой множество целых чисел (как правило, интервал), удовлетворяющих следующим условиям:

$$\forall m \in M \quad m \geq \text{card}(S^+), m \leq \text{card}(U) - \text{card}(S^-),$$

где $\text{card}(Y)$ — мощность множества Y , U — множество всех потенциальных элементов Н-множества S . Мощность множества M служит индикатором определенности задания соответствующего Н-множества. Чем она меньше, тем точнее определено Н-множество. При $\text{card}(M) = 1$ S представляет собой обычное множество, все элементы которого точно известны, и может отображать достоверную группу объектов. Если $\text{card}(M) > 1$, то S — это Н-множество, и с его помощью представляется возможная группа объектов.

Для обработки КС и групп объектов, представленных таким образом, можно использовать методы недоопределенных вычислений [21].

Выявление возможных и достоверных групп осуществляется на основе логических условий, выраженных с помощью недоопределенного расширения булевой алгебры, построенного по правилам, предложенным в [21]. Такое недоопределенное расширение имеет логические значения $\{0\}$, $\{1\}$, $\{0,1\}$, которые интерпретируются как Ложь, Истина и Неопределенность соответственно. С содержательной точки зрения Неопределенность представляет возможное выполнение проверяемого условия.

На основе этого подхода ранее были разработаны способы группирования объектов с учетом их неопределенных количественных КС [17, 18]. В данной работе предлагается способ, который позволяет проводить группирование с использованием как количественных, так и качественных КС.

Как и при использовании точных оценок КС объектов, для каждой КС *x_i , $i = \overline{1, n}$ может быть задано условие, необходимое, но не достаточное для отнесения пары объектов p и q к одной группе. Однако, учитывая неопределенность оценок КС *x_i , о таком условии можно сказать, что оно либо выполняется, либо не выполняется, либо возможно выполняется. Чтобы отличать такое условие от приведенных выше критериев вида $Cr_i(x_{pi}, x_{qi})$, будем обозначать его ${}^*Cr_i({}^*x_{pi}, {}^*x_{qi})$.

Тогда критерий принадлежности пары объектов p и q к одной группе ${}^*Cg(p, q)$ можно сформулировать следующим образом.

Если существует $i \in \{1, 2, \dots, n\}$, для которого ${}^*Cr_i({}^*x_{pi}, {}^*x_{qi})$ не выполняется, то объекты p и q *не принадлежат* к одной группе.

Если не существует $i \in \{1, 2, \dots, n\}$, для которого ${}^*Cr_i({}^*x_{pi}, {}^*x_{qi})$ не выполняется, но существует $i \in \{1, 2, \dots, n\}$, для которого ${}^*Cr_i({}^*x_{pi}, {}^*x_{qi})$ возможно выполняется, то объекты p и q *возможно принадлежат* к одной группе.

Если для любого $i \in \{1, 2, \dots, n\}$ условие ${}^*Cr_i({}^*x_{pi}, {}^*x_{qi})$ выполняется, то p и q *достоверно принадлежат* к одной группе.

Влияние частных критериев ${}^*Cr_i({}^*x_{pi}, {}^*x_{qi})$ на результат проверки обобщенного критерия ${}^*Cg(p, q)$ представлено в таблице 1.

Таблица 1. Правила проверки пары объектов на принадлежность к одной группе

Критерий	Выполнение		
	Не выполняется	Возможно выполняется	Выполняется
${}^*Cg(p, q)$	Не выполняется	Возможно выполняется	Выполняется
${}^*Cr_1(x_{p1}, x_{q1})$	Выполняется	Выполняется	Выполняется
${}^*Cr_2(x_{p2}, x_{q2})$	Возможно выполняется	Выполняется	Выполняется
...
${}^*Cr_i(x_{pi}, x_{qi})$	Не выполняется	Возможно выполняется	Выполняется
...
${}^*Cr_n(x_{pn}, x_{qn})$	Выполняется	Выполняется	Выполняется

Варианты критерия ${}^*Cr_i(x_{pi}, x_{qi})$ для неопределенных количественных КС предложены в [17, 18]. Один из них представляет собой недоопределенное расширение критерия (1) и сформулирован следующим образом [18].

Если $|\hat{x}_{pi} - \hat{x}_{qi}| > \varepsilon_i + \Delta x_{pi} + \Delta x_{qi}$, то ${}^*Cr_i(x_{pi}, x_{qi})$ не выполняется.

Если $|\hat{x}_{pi} - \hat{x}_{qi}| \leq \varepsilon_i + \Delta x_{pi} + \Delta x_{qi}$ и $|\hat{x}_{pi} - \hat{x}_{qi}| > \varepsilon_i - \Delta x_{pi} - \Delta x_{qi}$, то ${}^*Cr_i(x_{pi}, x_{qi})$ возможно выполняется.

Если $|\hat{x}_{pi} - \hat{x}_{qi}| \leq \varepsilon_i - \Delta x_{pi} - \Delta x_{qi}$, то ${}^*Cr_i(x_{pi}, x_{qi})$ достоверно выполняется.

Частный критерий принадлежности пары объектов p и q к однородной группе по качественному признаку $x_i, i = \overline{k+1, n}$ формулируется следующим образом.

Если $B_{pi} \cap B_{qi} = \emptyset$, то ${}^*Cr_i(x_{pi}, x_{qi})$ не выполняется.

Если $|B_{pi}| > 1$ и (или) $|B_{qi}| > 1$ и $B_{pi} \cap B_{qi} \neq \emptyset$, то ${}^*Cr_i(x_{pi}, x_{qi})$ возможно выполняется.

Если $|B_{pi}| = |B_{qi}| = 1$ и $B_{pi} = B_{qi}$, то ${}^*Cr_i(x_{pi}, x_{qi})$ достоверно выполняется.

Примеры проверки критерия принадлежности пары объектов к однородной группе ${}^*Cr_i(x_{pi}, x_{qi})$ для качественных признаков приведены на рисунке 1.

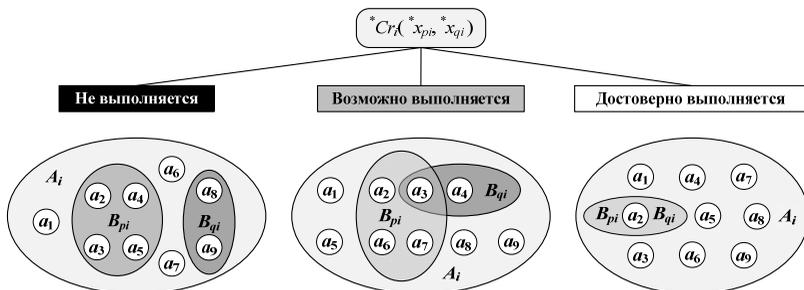


Рис. 1. Проверка критерия принадлежности объектов p и q к однородной группе по i -му качественному признаку

Частный критерий принадлежности пары объектов p и q к *неоднородной* группе по качественному признаку *x_i , $i = k+1, n$ при допустимом сочетании значений данного признака в группе $D_i \subset A_i$ формулируется следующим образом.

Если $B_{pi} \cap D_i = \emptyset$ и (или) $B_{qi} \cap D_i = \emptyset$, то ${}^*Cr_i(x_{pi}, x_{qi})$ *не выполняется*.

Если $B_{pi} \cap D_i \neq \emptyset$ и $B_{qi} \cap D_i \neq \emptyset$ и $B_{pi} \not\subset D_i$ и (или) $B_{qi} \not\subset D_i$, то ${}^*Cr_i(x_{pi}, x_{qi})$ *возможно выполняется*.

Если $B_{pi} \subseteq D_i$ и $B_{qi} \subseteq D_i$, то ${}^*Cr_i(x_{pi}, x_{qi})$ *достоверно выполняется*.

Пример проверки данного критерия для i -го качественного признака представлен на рисунке 2.

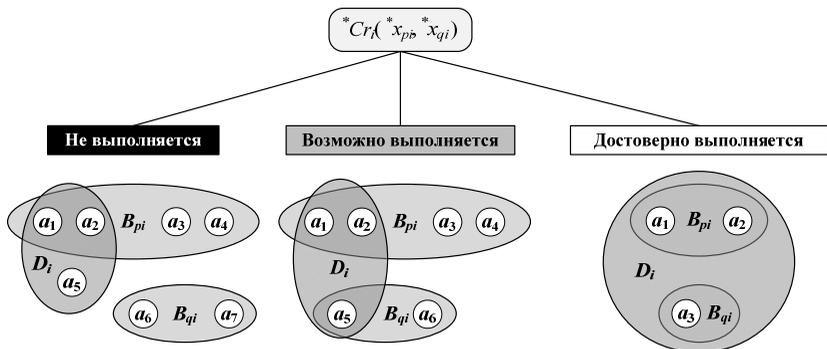


Рис. 2. Проверка критерия принадлежности объектов p и q к неоднородной группе по i -му качественному признаку

Для выявления групп объектов строится граф, вершины которого соответствуют объектам, а ребра, в отличие от графа, предложенного в [5], могут быть двух типов и обозначают достоверную или возможную принадлежность пары объектов к одной группе [17]. В таком графе возможной группе соответствует связная компонента, в которой есть не менее одной пары вершин, любой путь между которыми включает в себя хотя бы одно возможное ребро. Достоверной группе соответствует связная компонента графа, в которой между любой парой вершин существует не менее одного пути, состоящего только из достоверных ребер.

Выявление связных компонент графа выполняется в два этапа с помощью алгоритма, предложенного в [17]. На первом этапе с использованием известного алгоритма поиска в глубину [19] выделяются достоверные группы объектов. Затем путем попарного сравнения вершин из различных достоверных групп выделяются возможные группы объектов.

4. Заключение. В данной работе рассмотрены способы группирования объектов без учета и с учетом неопределенности оценок их количественных и качественных КС.

Предложены критерии определения принадлежности пары объектов к одной возможной или достоверной группе на основе неопределенных количественных и качественных данных об объектах. Эти критерии, с учетом возможных ограничений на сочетания значений одноименных качественных признаков объектов в группе, позволяют выявлять однородные и неоднородные группы объектов.

С использованием предложенных критериев выявление групп объектов осуществляется на основе их представления в виде графа, вершины которого соответствуют объектам, а ребра обозначают достоверную или возможную принадлежность пары объектов к одной группе. Группам объектов соответствуют связные компоненты данного графа, которые выделяются с помощью модифицированного алгоритма поиска в глубину.

Литература

1. *Канащенков А.И., Меркулов В.И., Самарин О.Ф.* Облик перспективных бортовых радиолокационных систем // Возможности и ограничения. М.: ИПРЖР. 2002.
2. *Li X., Yao X.* Cooperatively coevolving particle swarms for large scale optimization // IEEE Trans. Evol. Comput. 2012. vol. 16. no. 2. pp. 210–224.
3. *Верба В.С., Меркулов В.И.* Теоретические и прикладные проблемы разработки систем радиоуправления нового поколения // Радиотехника. 2013. № 1. С. 64–68.
4. *Горощенко Л.Б.* Методы координированного наведения и атаки несколькими истребителями группы самолетов противника // Полет. 2000. № 6. С. 32–36.
5. *Кирсанов А.П., Сорвенков О.С., Сузанский Д.Н.* Методы обработки радиолокационной информации при сопровождении компактных групп воздушных объектов // Радиотехника. 1996. № 10. С. 102–108.

6. Kryś S., Jankowski S. Extended hierarchical temporal memory for visual object tracking // *Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments 2011*. International Society for Optics and Photonics. 2011. vol. 8008. pp. 80081C–80081C-9.
7. Wu S., San Wong H. Joint segmentation of collectively moving objects using a bag-of-words model and level set evolution // *Pattern Recognition*. 2012. vol. 45. no. 9. pp. 3389–3401.
8. Lan T. et al. Discriminative latent models for recognizing contextual group activities // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2012. vol. 34. no. 8. pp. 1549–1562.
9. Лучков Н.В., Пчелин Н.А. Группирование объектов на основе радиолокационных наблюдений. // *Автоматизация процессов управления*. 2014. № 2(36). С. 16–19.
10. Вахненко В.А., Матвеев П.А., Цишук А.А. Группирование воздушных целей // *Успехи современной радиоэлектроники*. 2014. № 3. С. 11–13.
11. Pedrycz W., Bargiela A. An Optimization of Allocation of Information Granularity in the Interpretation of Data Structures: Toward Granular Fuzzy Clustering // *IEEE Trans. Syst. Man. And Cybern. B*. 2012. vol. 42. no. 3. pp. 582–590.
12. Mei J.P., Chen L. A fuzzy approach for multitype relational data clustering // *IEEE transactions on Fuzzy systems*. 2012. vol. 20. no. 2. pp. 358–371.
13. Chen D. et al. A novel algorithm for finding reducts with fuzzy rough sets // *IEEE Transactions on Fuzzy Systems*. 2012. vol. 20. no. 2. pp. 385–389.
14. Abdallah A.C.B., Frigui H., Gader P. Adaptive local fusion with fuzzy integrals // *IEEE Transactions on Fuzzy Systems*. 2012. vol. 20. no. 5. pp. 849–864.
15. Lewis R., Pullin E. Revisiting the Restricted Growth Function Genetic Algorithm for Grouping Problems // *Evol. Comput.* 2011. vol. 19. no. 4. pp. 693–704.
16. Karasulu B., Balli S. Image Segmentation Using Fuzzy Logic, Neural Networks and Genetic Algorithms: Survey and Trends // *Mach. Graph. and Vision*. 2010. vol. 19. no. 4. pp. 367–409.
17. Лунатов А.А. Метод и алгоритм формирования групп наблюдаемых воздушных объектов с неточными координатами состояния // *Радиотехника*. 2011. №8. С. 80–83.
18. Верба В.С., Лунатов А.А., Федисов А.Н. Выявление групп воздушных объектов с учетом неопределенности их координат состояния // *Успехи современной радиоэлектроники*. 2014. №1. С. 24–29.
19. Ахо А., Хопкрофт Дж., Ульман Дж. Построение и анализ вычислительных алгоритмов: пер. с англ. // М.: Мир. 1979. 536 с.
20. Нариньяни А.С. Недоопределенные множества – новый тип данных для представления знаний // Новосибирск. 1980. 28 с.
21. Нариньяни А.С. Недоопределенность в системе представления и обработки знаний // *Изв. АН СССР. Техн. кибернетика*. 1986. № 5. С. 3–28.

References

1. Kanaschenkov, A.I., Merkulov, V.I., Samarin, O.F. *Oblik perspektivnyh bortovyyh radiolokacionnyh sistem. Vozmozhnosti i ogranichenija*. [Image of advanced airborne radar systems. The possibilities and limitations]. Moscow: IPRZHR. 2002. (In Russ.).
2. Li X., Yao X. Cooperatively coevolving particle swarms for large scale optimization. *IEEE Trans. Evol. Comput.* 2012. vol. 16. no. 2. pp. 210–224.
3. Verba V.S., Merkulov V.I. [Theoretical and applied problems of developing a new generation of radio systems]. *Radiotekhnika – Radiotechnics*. 2013. vol. 1. pp. 64–68. (In Russ.).
4. Goroshchenko L.B. [Methods of coordinated attacks targeting several groups of enemy aircraft fighters]. *Polyet – Fly*. 2000. vol. 6. pp. 32–36. (In Russ.).

5. Kirsanov, A.P., Sorvenkov, O.S., Suzansky, D.N. [Methods of processing radar information when accompanied by compact groups of air objects]. *Radiotekhnika – Radiotechnics*. 1996. vol. 10. pp. 102–108. (In Russ.).
6. Kryś S., Jankowski S. Extended hierarchical temporal memory for visual object tracking. *Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments 2011*. International Society for Optics and Photonics. 2011. vol. 8008. pp. 80081C–80081C-9.
7. Wu S., San Wong H. Joint segmentation of collectively moving objects using a bag-of-words model and level set evolution. *Pattern Recognition*. 2012. vol. 45. no. 9. pp. 3389–3401.
8. Lan T. et al. Discriminative latent models for recognizing contextual group activities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2012. vol. 34. no. 8. pp. 1549–1562.
9. Luchkoff N.V., Pchelin N.A. [Grouping objects from radar observation]. *Avtomatika i processi upravleniya – Automatics and control processes*. 2014. vol. 2(36). pp. 16–19. (In Russ.).
10. Vakhnenko V.A., Matveev P.A., Tsishuk A.A. [Grouping of air targets]. *Uspehi sovremennoy radioelektroniki – The successes of modern radio electronics*. 2014. vol. 3. pp. 11–13. (In Russ.).
11. Pedrycz W., Bargiela A. An Optimization of Allocation of Information Granularity in the Interpretation of Data Structures: Toward Granular Fuzzy Clustering. *IEEE Trans. Syst. Man. And Cybern. B*. 2012. vol. 42. no. 3. pp. 582–590.
12. Mei J.P., Chen L. A fuzzy approach for multitype relational data clustering // *IEEE transactions on Fuzzy systems*. 2012. vol. 20. no. 2. pp. 358–371.
13. Chen D. et al. A novel algorithm for finding reducts with fuzzy rough sets // *IEEE Transactions on Fuzzy Systems*. 2012. vol. 20. no. 2. pp. 385–389.
14. Abdallah A.C.B., Frigui H., Gader P. Adaptive local fusion with fuzzy integrals // *IEEE Transactions on Fuzzy Systems*. 2012. vol. 20. no. 5. pp. 849–864.
15. Lewis R., Pullin E. Revisiting the Restricted Growth Function Genetic Algorithm for Grouping Problems. *Evol. Comput.* 2011. vol. 19. no. 4. pp. 693–704.
16. Karasulu B., Balli S. Image Segmentation Using Fuzzy Logic, Neural Networks and Genetic Algorithms: Survey and Trends // *Mach. Graph. and Vision*. 2010. vol. 19. no. 4. pp. 367–409.
17. Lipatov A.A. [The method and algorithm of forming groups of observed air objects with inaccurate coordinates state]. *Radiotekhnika – Radiotechnics*. 2011. vol. 8. pp. 80–83. (In Russ.).
18. Verba V.S., Lipatov A.A., Fedisov A.N. [Identify groups of air objects, taking into account the uncertainty of the state of origin]. *Uspehi sovremennoy radioelektroniki – The successes of modern radio electronics*. 2014. vol. 1. pp. 24–29. (In Russ.).
19. Aho A., Hopcroft J., Ullman J. *Postroenie i analiz vychislitel'nyh algoritmov: per. s angl.* [The Design and Analysis of Computer Algorithms: Translation from English]. Moscow: Mir. 1979. 536 p. (In Russ.).
20. Narin'yani A.S. *Nedopredelennye mnozhestva – novyj tip dannyh dlja predstavlenija znaniy* [Underdetermined sets - a new data type for representing knowledge]. Novosibirsk. 1980. 28 p. (In Russ.).
21. Narin'yani A.S. [Subdefiniteness in the knowledge representation and processing system]. *Izvestia AN SSSR. Tehnicheskya Kibernetika – Proceedings of the USSR Academy of Sciences. Technical Cybernetics*. 1986. vol. 5. pp. 3–28 (In Russ.).

Липатов Алексей Андреевич — к-т техн. наук, начальник лаборатории, АО "Концерн "Вега". Область научных интересов: искусственный интеллект, поддержка принятия решений, программирование в ограничениях. Число научных публикаций — 40. alexey.lipatov75@mail.ru; Кутузовский проспект, 34, Москва, 121170; р.т.: +7(499)753-40-04(92-42).

Lipatov Alexey Andreevich — Ph.D., head of laboratory, Joint-Stock Company «Radio Engineering Corporation «Vega». Research interests: artificial intelligence, decision support, constraint programming, cognitive computer graphic. The number of publications — 40. alexey.lipatov75@mail.ru; 34, Kutuzov avenue, Moscow, 121170, Russia; office phone: +7(499)753-40-04(92-42).

Ушаков Валерий Николаевич — к-т техн. наук, ведущий научный сотрудник АО "Концерн "Вега", референт отдела автоматизации и электроники Федеральное государственное бюджетное учреждение науки Всероссийский институт научной и технической информации Российской академии наук (ВИНИТИ РАН). Область научных интересов: автоматика и радиоэлектроника, радиолокационные системы, робототехнические комплексы авиационного базирования, алгоритмы и программные средства, методы управления и обработки информации. Число научных публикаций — 25. vanick@bk.ru, http://www.vega.su; Кутузовский проспект, 34, Москва, 121170; р.т.: +7(499)753-40-04.

Ushakov Valery Nikolaevich — Ph.D., leading researcher, Joint-Stock Company «Radio Engineering Corporation «Vega», referent, Russian Institute for Scientific and Technical Information of Russian academy of science (VINITI RAS). Research interests: automation and radionics, radar systems, air-based robotic systems, algorithms and software, control and information processing. The number of publications — 25. vanick@bk.ru, http://www.vega.su; 34, Kutuzov avenue, Moscow, 121170, Russia; office phone: +7(499)753-40-04.

Никитина Маргарита Валерьевна — инженер, АО "Концерн "Вега". Область научных интересов: биоинформатика, искусственный интеллект, поддержка принятия решений, программирование в ограничениях. margarita307@gmail.com; Кутузовский проспект, 34, Москва, 121170; р.т.: +7(499)753-40-04.

Nikitina Margarita Valerievna — engineer, Joint-Stock Company «Radio Engineering Corporation «Vega». Research interests: bioinformatics, artificial intelligence, decision support, constraint programming. margarita307@gmail.com; 34, Kutuzov avenue, Moscow, 121170, Russia; office phone: +7(499)753-40-04.

Поддержка исследований. Работа выполнена при финансовой поддержке РФФИ (проект № 15-08-04000-а).

Acknowledgements. This research is supported by RFBR (grant 15-08-04000-a).

РЕФЕРАТ

Липатов А.А., Ушаков В.Н., Никитина М.В. **Методы выявления однородных и неоднородных групп объектов на основе неопределенных качественных данных.**

В данной статье рассматривается задача группирования объектов, описываемых вектором их количественных и качественных координат состояния (КС). Оценки значений КС объектов могут быть неопределенными вследствие ошибок измерения (оценивания) или неполноты информации об объектах и представляют собой множества. Совокупность объектов, составляющих группу, должна определяться критерием, учитывающим как количественные, так и качественные КС.

Ранее были предложены способы выявления групп объектов на основе неопределенных значений их количественных КС.

На основе качественных КС могут выявляться однородные группы, в которых значения всех одноименных качественных КС всех объектов совпадают, и неоднородные группы, в которых значения одноименных качественных КС могут не совпадать, однако должны удовлетворять априорно заданным допустимым сочетаниям значений данных КС.

Целью данной статьи является разработка способа выявления однородных и неоднородных групп объектов на основе неопределенных значений их количественных и качественных КС.

Предлагаемый способ основан на проверке принадлежности пар объектов к одной группе в соответствии с критерием, который выражается через частные критерии близости значений одноименных КС объектов. Все указанные критерии могут, с учетом неопределенности оценок КС, не выполняться, возможно выполняться либо достоверно выполняться. Это позволяет выявлять возможные и достоверные группы объектов.

Ранее были предложены частные критерии близости значений одноименных количественных КС, а также качественных КС для однородных групп. В данной статье предлагается критерий близости значений качественных КС для неоднородных групп, отличающийся тем, что значения данных КС сравниваются не друг с другом, а с априорно заданными множествами-образцами, представляющими допустимые сочетания значений в группе.

Путем проверки пар объектов на принадлежность к группе строится граф их взаимодействия, в котором, с помощью модифицированного алгоритма поиска в глубину, выделяются связанные компоненты, представляющие возможные и достоверные группы объектов.

SUMMARY

Lipatov A.A., Ushakov V.N., Nikitina M.V. **Methods of Detecting Homogeneous and Heterogeneous Groups of Objects based on Uncertain Qualitative Data.**

In this paper, the problem of grouping objects described by the vector of their quantitative and qualitative state coordinates (SC) is considered. Estimates of objects' SC values can be uncertain due to the errors of measurements or incomplete information about the objects. SC estimates are represented as sets. The set of objects that make up the group should be determined according to the criterion that takes into account both quantitative and qualitative SC.

Methods enabling to detect groups of objects based on their uncertain quantitative SC were proposed previously.

Based on the qualitative SC homogeneous groups may be detected. In these groups the values of all cognominal qualitative SC of the objects are equal. In heterogeneous groups the values of cognominal qualitative SC may differ, but have to match the a priori set valid combinations of the corresponding SC.

The aim of this paper is to provide a method of detecting homogeneous and heterogeneous groups of objects based on the uncertain values of their quantitative and qualitative SC.

The proposed method is based on the verification of pairs of objects belonging to the same group according to the criterion expressed through individual criteria of proximity of the cognominal objects' SC values. These criteria may either be not fulfilled, or possibly fulfilled, or reliably fulfilled.

The criteria of object groups detection based on the uncertain quantitative SC and homogeneous groups detection based on the uncertain qualitative SC were suggested previously. In this paper, the criterion of detection of heterogeneous object groups is proposed. According to this criterion, the objects' SC uncertain values are compared with the a priori set sample sets representing the allowable combinations of the values in the group.

Verification of object pairs belonging to the same group enables one to construct a graph. In this graph the nodes correspond to the objects and the edges indicate the object pairs' group affiliation. Possible and reliable object groups are represented as connected components of this graph. These groups are detected by a modified depth-first search algorithm.

Б.Г. ИЛЬЯСОВ, И.Б. ГЕРАСИМОВА, А.Г. КАРАМЗИНА, Ю.Р. СУЯРГУЛОВА
**МОДЕЛЬ НАКОПЛЕНИЯ ЗНАНИЙ В НАУЧНОЙ ШКОЛЕ В
РЕЗУЛЬТАТЕ САМООРГАНИЗАЦИИ ПРОЦЕССА ОБМЕНА
ИНФОРМАЦИЕЙ**

Ильясов Б.Г., Герасимова И.Б., Карамзина А.Г., Суюргулова Ю.Р. Модель накопления знаний в научной школе в результате самоорганизации процесса обмена информацией.

Аннотация. В статье впервые предложена нелинейная многосвязная мультиагентная модель накопления знаний в научной школе в результате самоорганизации процесса обмена научной информацией. Выделены три группы агентов (научных сотрудников) как носителей знаний, которые нелинейно взаимодействуют между собой и контролируют свою деятельность по накоплению знаний. При этом рассматриваются два режима процесса обмена научной информацией агентами: свободный (в форме дискуссий) и деловой (в форме выполнения совместного проекта). Предложенные модели носят абстрактно-обобщающий характер, представляют собой систему нелинейных дифференциальных уравнений, описывающих процессы накопления знаний в результате обмена научной информацией с соблюдением сложившихся структурных отношений между агентами научной школы. Данная статья преследует цель раскрытия сущности сложного механизма процесса накопления знаний в научной школе за счет различных форм активного взаимодействия агентов, представляющих собой единство «человек-ЭВМ-база знаний».

Ключевые слова: научная школа, носители знаний, агенты, математическая модель, обмен и накопление знаний, системный подход, самоорганизация, многосвязные процессы.

Ilyasov B.G., Gerasimova I.B., Karamzina A.G., Suyargulova Yu.R. A Model of Knowledge Accumulation in the Scientific School as a Result of Self-Organization of Information Exchange Process.

Abstract. A nonlinear multivariable multi-agent model of knowledge accumulation in the scientific school as a result of self-organization of scientific information exchange process is first proposed in the paper. Three groups of agents (researchers) as knowledge carriers interacting nonlinearly with each other and monitoring their activity on the knowledge accumulation are distinguished. Two modes of agents' scientific information exchange are considered: free mode (in the form of discussion) and business mode (in the form of joint project implementation). The proposed models have the abstract-generalizing character. They represent a system of nonlinear differential equations describing processes of knowledge accumulation as a result of scientific information exchange in compliance with existing structural relations between the agents of the scientific school. The aim of this paper is to disclose the complex mechanism of knowledge accumulation in the scientific school through various forms of active interaction of the scientific school agents representing the unity of the "man-computer-knowledge base".

Keywords: scientific school; knowledge carriers; mathematical model; exchange and accumulation of knowledge; systematic approach; self-organization; multiply connected processes.

1. Введение. Анализ проблемы развития науки является объектом изучения многих исследователей. Особый интерес к этой проблеме возрос вследствие научно-технических революций во многих обла-

стях знаний (информатика, биология, биохимия, робототехника, телекоммуникационные и информационные технологии, системы безопасности и т.д.). При этом под знанием понимается продукт общественной, материальной и духовной деятельности людей, выраженный в какой-либо форме (знаковой, модельной, образной, письменной, словесной и т.п.), которая отражает объективные свойства материального мира, общества. Знания обладают свойством преобразования и перехода из одних форм представления в другие, более понятные и доступные исследователю. Знание обладает свойством сохранения его (база знаний), передачи, обмена, накопления и развития. Эти свойства знания являются базой и опорой для последующей деятельности человечества. В каждой области знаний имеются свои общепринятые условные единицы измерения знаний, его объема и качества. При этом знания, накопленные в течение длительного времени и реализованные в какой-либо форме, делаются на отдельные поколения (например, поколения летательных аппаратов в технике). Переход от одного поколения знаний к другому и определяет рост накопления знаний. Например, никто не сомневается в том, что темпы накопления знаний во второй половине XX века были значительно выше, чем в XIX веке.

Помогая человеку использовать в своей практической деятельности все новые достижения научного познания, наука вносит в жизнь общества элемент динамичности, революционизирует общественную практику. На современном уровне развития науки получение принципиально новых знаний дается все труднее и характеризуется возрастанием трудоемкости и стоимости познавательного процесса.

Анализируя работы отечественных исследователей в области науковедения, философии науки (теории познания) можно выделить три направления исследований.

Первое направление связано с проблемами развития науки и организации науки в России, построением различных моделей развития науки и выявления закономерностей [2, 3]. В работах [1, 4, 5, 6, 21, 22] рассматривается общая структура науки, структура научного знания, особенности современного этапа развития науки, динамика науки. В работе [7] в основном рассматривается математическая модель динамики фундаментальной науки в зависимости от политики финансирования процесса развития науки, оценивается скорость роста стоимости науки и стабилизация уровня мирового продукта, оспаривается идея о том, что наука не является единственным методом познания.

Во втором направлении изучаются малые научные группы, их структура. В работе проф. И.Д. Софронова [17] рассматриваются модели архитектуры научного коллектива и вопросы распределения при-

оритета и вознаграждений в моделях творческого коллектива в советский период развития науки. В работе [8] рассматривается структура и взаимодействие малых научных групп при выполнении проекта, рассматривается влияние на производительность научного коллектива индивидуальных характеристик агентов, таких как деловая активность, уровень интеллекта, уровень эмоциональности.

Третье направление носит исторический характер и связано с воспоминаниями о научных школах, научных руководителях и их выдающихся результатах. В работе [9], например, излагается история создания вычислительной техники и перспектив ее развития и т.д.

Однако среди этих трех направлений нет работ, посвященных изучению вопросов о внутреннем взаимодействии коллектива при получении знаний, не раскрываются механизмы формирования и накопления знаний. По нашему мнению, научные школы должны быть самостоятельным предметом изучения науковедения.

Моделирование процессов накопления и обмена знаниями в научных школах позволяет выявить и оценить проявляющийся синергетический эффект. Синергетика позволяет установить связь и взаимодействие между микро- и макропроцессами, накоплением и обменом знаниями на различных организационных уровнях [18-20]. Основная идея синергетики: сложные системы изменяются в результате изменений, происходящих на микроуровне, но они недоступны непосредственному наблюдению, а доступен наблюдению их совокупный результат, который описывается управляющими параметрами системы. Авторы статьи придерживаются такого же мнения, считая, что малые знания агентов в процессе общения друг с другом системно сливаются, интегрируются в знания более высокого уровня.

Заметим, что слишком большой объем накопленной фактической информации в различных научных направлениях делает этот процесс многомерным, а относительная сложность связи между отдельными направлениями делает затруднительным применение описательных и чисто содержательных подходов к анализу процессов развития конкретного научного направления.

В силу сказанного, представляет интерес изучение процесса развития научных школ как эволюции многомерных многосвязных нелинейных самоорганизующихся систем, осуществляемой на качественном абстрактном уровне. Справедливость этого подхода основана на том, что отсутствуют точные количественные характеристики, отражающие содержание научного знания, ценность научного результата и продуктивность научного труда. Практически невозможно установить связь между количеством научных публикаций и качеством

научной деятельности [4, 12, 13, 16]. Однако сам процесс накопления знаний подчиняется объективным системным законам. При этом, по-видимому, можно говорить и о темпах накопления некоторого объема знаний за определенный период времени.

Далее рассмотрим процесс развития научной школы, опираясь не столько на количественный анализ, сколько на качественную сторону интерпретации внутренних механизмов функционирования и развития научной школы за счет обмена и накопления знаний между различными агентами (группами) как носителями знаний в результате их обмена научными знаниями.

2. Структура научной школы. Научная школа (НШ) рассматривается как человеко-машинная система, состоящая из множества взаимосвязанных и взаимодействующих агентов, представляющих собой триаду «человек-ЭВМ-база знаний», образующих единое целое. Выделим три группы (прослойки) агентов, отличающихся качеством и уровнем профессиональных знаний (компетенциями). При этом НШ может образовывать из этих групп, например, иерархическую (трехуровневую) организационную структуру.

Научная школа может рассматриваться как множество взаимосвязанных процессов деятельности агентов, направленной либо на обмен знаниями (обучение, семинары, круглые столы, конференции, симпозиумы и т.п.), либо на достижение определенной цели (решение проблемы, реализацию проекта, разработку технологии или конструкции, написание монографий).

Первый режим работы НШ называется свободным (обучающим, коммуникационным), второй — научно-деловым (активным, трансляционным) [21].

Развитие НШ рассматривается как процесс накопления знаний, полученных в результате совместных действий агентов и как процесс использования этих знаний при достижении поставленных целей. Особенность НШ при университетах заключается в том, что агенты выполняют двойную роль: с одной стороны, они участвуют в образовательном процессе, а с другой стороны, они осуществляют научную деятельность.

Таким образом, выделены три группы агентов как носителей и накопителей знаний:

- N_1 — агенты, отвечающие за накопление знаний на уровне магистрантов, аспирантов, ассистентов и младших научных сотрудников;
- N_2 — агенты, отвечающие за накопление знаний на уровне доцентов, старших научных сотрудников и докторантов;

– N_3 — агенты, отвечающие за накопление знаний на уровне докторов наук, профессоров, ведущих и главных научных сотрудников.

Учитывая, что в результате научно-технического прогресса идет рост новых знаний, то процесс накопления знаний является для всех трех агентов вполне естественным и закономерным.

Агенты как носители специализированных знаний взаимодействуют между собой, самостоятельно осмысливают полученные новые знания, принимают решения и контролируют свою деятельность (процесс самоорганизации). Графически взаимодействие трех типов агентов представлено на рисунке 1.

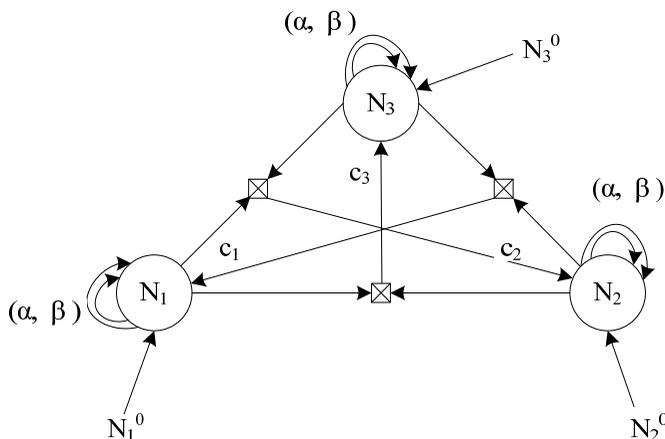


Рис. 1. Нелинейная модель взаимодействия различных агентов N_i как носителей знаний

Накопление знаний условно можно рассматривать как результативность научной деятельности агентов, выраженную в различной форме (например, 12 публикаций в год, получение 3 патентов в год, написание 1 монографии в год, 4 доклада на международной конференции, подготовка 2 магистров и 1 кандидата наук в год и т.д.), т.е. как накопление определенного типа и объема знаний за определенный период времени. Более объективным будет оценка накопления знаний в научной школе в процессе выполнения средне- и крупномасштабных научных проектов за некоторый отрезок времени. Здесь проект рассматривается в пространственно-временном плане как множество задач, требующих использование не только старых, но и синтез и применение новых знаний, например, в виде инновационных технологий.

Процесс обмена знаниями в научной школе происходит в определенные промежутки времени (например, на интервале рабочего дня), поэтому его можно считать непрерывным дискретно-временным процессом.

Таким образом, целесообразно изучать закономерности процесса накопления знаний в научной школе в результате обмена знаниями между агентами как тремя типами носителей знаний.

3. Нелинейная многосвязная мультиагентная модель научной школы в режиме свободного обмена информацией. В статье идет речь об абстрактных моделях накопления знаний в результате обмена научной информацией между агентами. В этом случае нельзя говорить о конкретной научной школе и конкретном типе агента. Эта модель есть результат многолетнего обобщения процессов получения знаний в малых научных группах как основе научной школы. Адекватность (вариабельность) предлагаемых моделей не должна вызывать сомнений, т.к., с одной стороны, все участники процесса являются представителями любой научной школы и имеют единую структуру и их отношения основаны на традиционной структуре научной школы (учитель–ученик), а с другой стороны, эксперименты их взаимодействия проведены многократно мировым сообществом ученых в виде научно-технических конференций, семинаров и т.д. Адекватность процессов научного общения реальным процессам основана на социологических методах наблюдения (простом и включенном), анализе документаций (статей, диссертаций и т.п.), многолетнем опыте руководства и проведения предзащит и защит диссертаций, организации, проведения и участия на многочисленных научно-технических конференциях различного уровня, круглых столах, научных дискуссиях, а также на методе интервью и экспертных оценок, данных специалистами в той или иной области знаний [10, 11].

В научной литературе в области науковедения, философии, социологии не описана специфика взаимодействия агентов в научной школе. В данной статье процесс научного общения представляется в виде трехсвязного процесса с участием трех типов агентов. Рассмотрим нелинейную модель НШ в режиме свободного взаимодействия агентов. При этом сделаны следующие допущения:

– каждая группа носителей знаний накапливает знания самостоятельно в соответствии со своим нелинейным механизмом самоорганизации. При этом нелинейные механизмы самоорганизации представляют собой суммарное действие положительной (в линейной форме) и отрицательной обратной связи (в квадратичной или кубической форме). Такая форма самоорганизации (самоуправления) часто встречается в синергетике [14, 15];

– отношения двух типов носителей знаний проявляются к третьему типу знаний в нелинейной форме (например, в виде произведения), т.е. два типа групп, объединяясь в коалицию, помогают повысить интенсивность освоения знаний третьей группе. Такая форма отношений встречается, например, в дискуссиях и также является формой самоорганизации.

В связи с вышесказанным, нелинейная модель НШ в режиме свободного взаимодействия (обмена знаниями) имеет вид:

$$\begin{cases} \dot{N}_1 = \alpha_1 N_1 - \beta_1 N_1^2 + c_1 N_2 N_3 + N_1^0, \\ \dot{N}_2 = \alpha_2 N_2 - \beta_2 N_2^2 + c_2 N_1 N_3 + N_2^0, \\ \dot{N}_3 = \alpha_3 N_3 - \beta_3 N_3^2 + c_3 N_1 N_2 + N_3^0, \end{cases} \quad (1)$$

где коэффициент α_i ($i=1,2,3$) отражает мотивацию i -ой группы увеличить интенсивность накопления знаний, β_i — отражает силу стабилизации процесса накопления знаний, при этом условие статической устойчивости в автономном режиме:

$$N_i > \frac{\alpha_i}{\beta_i}.$$

\dot{N}_i — скорость действия агентов по обмену и накоплению знаний;

N_i^0 — агент с начальным уровнем накопленных знаний по конкретному вопросу в конкретной области знаний при его работе в автономном режиме (условно приравнивается к 1).

Значения коэффициентов α_i и β_i для каждого агента определяется им самим (методом проб и ошибок). Здесь коэффициент c_i отражает долю нелинейного влияния остальных двух групп в виде положительной обратной связи ($c_i > 0$), способствующей росту знаний i -го агента. Отрицательная обратная связь ($c_i < 0$) тормозит процесс накопления знаний и здесь не рассматривается. В силу природы выделенных групп: $c_1 > c_2 > c_3$. Эта цепь неравенств отражает реальную закономерность взаимодействия между агентами с учетом принятой иерархии качества их знаний: «аспирант-доцент-профессор». Конкретные значения для каждой научной школы определяются экспериментально либо экспертным путем. Набор этих значений отражает мотивацию агентов оказывать помощь друг другу в ускорении процесса накопле-

ния знаний в научной школе. Варьируя параметры (α_i, β_i, c_i) агентов можно получить достаточно высокую интенсивность накопления знаний.

При отсутствии нелинейного влияния ($c_i=0$) между агентами и наличии высокой степени самоорганизации ($\alpha_i = 1, \beta_i = 2$) система (1) является устойчивой и не превышает начальный уровень $N_i^0 = 1$ накопленных знаний каждым агентом в автономном режиме. Накопление знаний во всех трех группах отсутствует.

Уменьшение коэффициентов β_i повышает уровень накопленных знаний в автономном режиме (при $c_i=0$). Например, при условии $\alpha_i = 1, \beta_i = 0,5$ в системе уровень самостоятельно накопленных знаний каждым агентом в 2,7 раз превышает начальный. Это свидетельствует о том, что при высоком уровне самоорганизации, агенты могут индивидуально накапливать значительный объем знаний, опираясь на дополнительные источники знаний (библиотека, Internet).

Увеличение положительного нелинейного взаимного влияния агентов приводит к росту уровня накопления знаний. Система, например, устойчива и достигает относительно высокой интенсивности накопления различных типов знаний при условии, что $\alpha_i = 1, \beta_i = 2$ и $c_1 = 2, c_2 = 1, c_3 = 0,5$. В этом случае объем накопления знаний первой группой агентов превышает начальный ($N_i^0 = 1$) в 1,9 раза, второй группой — в 1,6 раз и третьей — в 1,4 раз (рисунок 2). Следовательно, процесс свободного обмена знаниями на семинарах, свободных дискуссиях создает эффект эмерджентности, т.е. интегрального (системного) накопления знаний в каждой группе агентов, а следовательно, и в научной школе в целом.

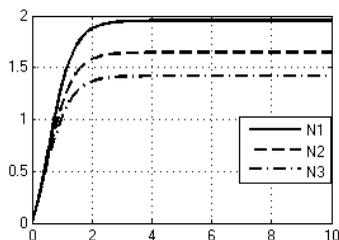


Рис. 2. Интегральное (системное) накопление знаний

Изменение значения коэффициента положительных нелинейных отношений c_i между агентами приводит к аналогичному изменению:

уменьшение положительных значений коэффициентов приводит к уменьшению накопления знаний, а их увеличение — к повышению. При одновременном уменьшении коэффициентов β_i и c_i следующим образом: $\alpha_i = 1$, $\beta_i = 1$, $c_1 = 0,5$, $c_2 = 0,25$, $c_3 = 0,1$ накопление знаний возрастает: первая группа агентов увеличивает знания, превышая начальный уровень в 2,2 раза, вторая группа агентов — в 2 раза и третья — в 1,8 раз.

Причиной потери устойчивости и возникновения колебательно-го движения может служить наличие чистого запаздывания при обмене знаниями. Согласно [14] чистые временные запаздывания в передаче информации можно классифицировать по двум типам: так называемое запаздывание в нелинейных связях и собственное запаздывание в формировании сигнала стабилизации. Рассмотрим влияние чистого запаздывания в нелинейном канале обратной связи накопления знаний. Собственное чистое запаздывание обуславливается тем, что восприятие и усвоение информации по стабилизации каждым агентом идет не мгновенно, а с некоторым запаздыванием во времени. Учитывая вышесказанное, нелинейную модель взаимодействия агентов в режиме свободного взаимодействия можно представить в виде:

$$\dot{N}_i(t) = \alpha_i N_i(t) - \beta_i N_i^2(t - \tau_i) + c_i N_j(t) N_k(t) + N_i^0, \quad (2)$$

где $i, j, k = 1, 2, 3$; $i \neq j \neq k$,

τ_i — собственное чистое запаздывание в действии i -го агента по нелинейной стабилизации темпов накопления знаний.

При значении коэффициентов $\alpha_i = 1$, $\beta_i = 2$, $c_1 = 2$, $c_2 = 1$, $c_3 = 0,5$ и величине чистого запаздывания $0 \leq \tau_i \leq 0,13$ с система является устойчивой. В таком режиме процесс накопления знаний агентами превышает начальный уровень и характеризуется перерегулированием. Объем накопления знаний первой группой агентов в 1,4 раза превосходит уровень автономного режима, второй группой — в 1,2 раза, третьей группой — в 1,1 раз. Процессы обмена знаниями в виде затухающих колебаний в системе наблюдаются при значении чистого запаздывания $0,14с \leq \tau_i \leq 0,17с$. С увеличением в указанных пределах значений чистого запаздывания повышается частота колебаний.

При чистом запаздывании $0,18с \leq \tau_{икр} \leq 0,19с$ в системе (2) наблюдается образование устойчивых предельных циклов. Это свидетельствует о том, что коллектив не пришел к единому мнению, в связи с этим проблема остается нерешенной до конца. Процесс накопления

знаний на уровне агентов N_1 в виде переходного процесса и фазового портрета (N_1, \dot{N}_1) представлены на рисунке 3.

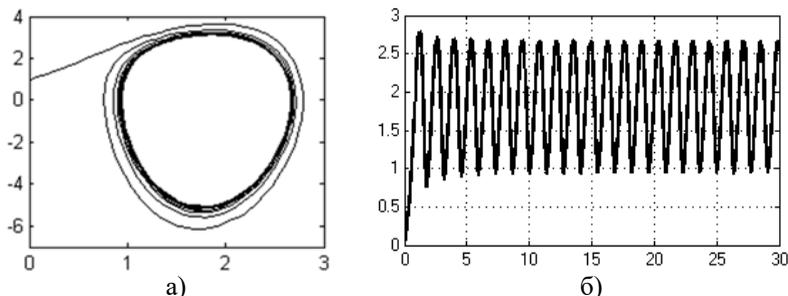


Рис. 3. Автоколебательный процесс в системе в режиме свободного взаимодействия с запаздыванием в формировании сигнала стабилизации при $\tau_{икр} = 0,18$ с: (а) фазовый портрет N_1 ; (б) график переходного процесса

Аналогичный процесс наблюдается и в других группах агентов как носителей знаний. Дальнейшее увеличение чистого запаздывания приводит систему в статически неустойчивое состояние (происходит «срыв» системы).

Уменьшение значения коэффициента положительных нелинейных отношений c_i между агентами увеличивает до некоторого значения запас устойчивости процессов накопления знаний. При одновременном изменении коэффициентов β_i и c_i ($\alpha_i = 1, \beta_i = 1, c_1 = 0,5, c_2 = 0,25, c_3 = 0,1$) увеличивается критическое значение чистого запаздывания: затухание колебаний наблюдаются при $\tau_{икр} \leq 0,3$ с, предельные циклы образуются при $0,31$ с $\leq \tau_{икр} \leq 0,33$ с. Хаотическое движение процессов накопления знаний наблюдается при $\tau_{икр} = 0,34$ с (рисунок 4).

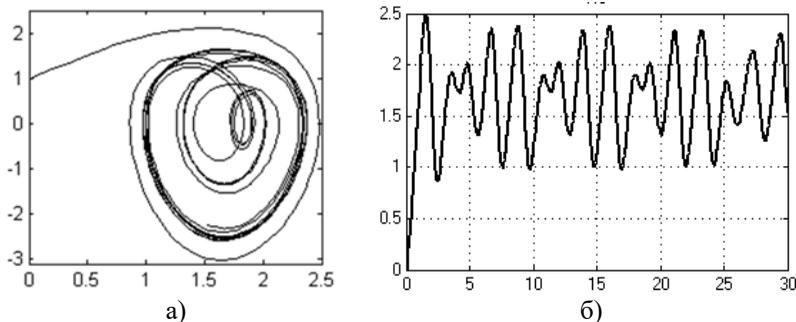


Рис. 4. Хаотическое движение системы в режиме свободного взаимодействия при наличии запаздывания в формировании сигнала стабилизации, равном $\tau_{кр}=0,34c$: (а) фазовый портрет N_2 ; (б) график переходного процесса

Процессы накопления знаний становятся статически неустойчивыми при дальнейшем увеличении значения чистого запаздывания. Таким образом, в режиме свободного обмена знаниями при правильной организации взаимодействия агентов и их высокой самоорганизации можно достичь высокого уровня накопления знаний как каждой группой агентов, так и научной школой в целом.

4. Нелинейная многосвязная мультиагентная модель накопления знаний в научной школе в режиме делового взаимодействия. В деловом (активном) режиме взаимодействия агенты как носители знаний N_i используют свои знания, накапливаемые в нелинейном процессе обмена информацией для решения научной проблемы, выполнения научного проекта. Мультипликативная форма подчеркивает равноценность используемых знаний, накопленных всеми тремя группами агентов. Без знания любого агента проект не может быть выполнен.

Рассмотрим нелинейную систему самоорганизации в виде введения положительных линейных и отрицательных квадратичных обратных связей для всех агентов и для агента-накопителя знаний.

$$\begin{cases} \dot{N}_1 = \alpha_1 N_1 - \beta_1 N_1^2 + c_1 N_2 N_3 + \lambda_1 Z + N_1^0, \\ \dot{N}_2 = \alpha_2 N_2 - \beta_2 N_2^2 + c_2 N_1 N_3 + \lambda_2 Z + N_2^0, \\ \dot{N}_3 = \alpha_3 N_3 - \beta_3 N_3^2 + c_3 N_1 N_2 + \lambda_3 Z + N_3^0, \\ \dot{Z} = \alpha_4 Z - \beta_4 Z^2 + N_1 N_2 N_3, \end{cases} \quad (3)$$

где \dot{Z} — скорость действия агента по накоплению интегральных знаний для выполнения проекта.

Z — агент, отвечающий за объем коллективно накопленных интегральных системных знаний, необходимых для выполнения проекта.

λ_i — коэффициенты положительной обратной связи, оказывающие влияние на процесс накопления знаний i -ым агентом. Величина $\lambda_i Z$ — положительный эффект, способствующий увеличению уровня накопления знаний i -ым агентом при успешном выполнении проекта, при этом $\lambda_1 > \lambda_2 > \lambda_3$ в силу организации структуры взаимоотношений сформированных групп агентов.

Рассмотрим влияние эффекта самоорганизации и взаимных контактов на процессы накопления знаний в системе.

Отсутствие накопления знаний у i -го агента ($N_i = N_i^0 = 1$) наблюдается при отсутствии влияния нелинейных отношений ($c_i = 0$) между агентами, положительного вклада $\lambda_i Z$ ($\lambda_i = 0$) и при наличии самоорганизации $\alpha_i = 1$, $\beta_i = 2$. Объем коллективно накопленных знаний Z также выходит на уровень, равный 1. Примем $N_i = 1$ и $Z = 1$ за базовый уровень. На основе результатов моделирования можно сделать вывод о том, что если в системе каждый агент работает самостоятельно вне коллектива, то накопление знаний в коллективе будет определяться только объемом знаний его членов и не использоваться при их совместной (коллективной) работе.

Одновременное увеличение c_i и λ_i ($\alpha_i = 1, \beta_i = 2, c_1 = 2, c_2 = 1, c_3 = 0,5, \lambda_1 = 2, \lambda_2 = 1, \lambda_3 = 0,5$) приводит к росту накопления знаний агентами в несколько раз по сравнению с базовым: N_1 в 3,4 раза превосходит начальный уровень, N_2 — в 2,7 раз, N_3 — в 2,2 раза, а объем коллективно накопленных знаний Z — в 3,45 раза, получаемый в результате решения уравнения (4):

$$2Z^2 - Z - 20,2 = 0. \quad (4)$$

Уменьшение значения коэффициентов нелинейных отношений c_i между агентами уменьшает объем накапливаемых знаний. Уменьшение значения коэффициентов обратной связи λ_i также снижает уровень накопления знаний. Одновременное уменьшение значений коэффициентов c_i и λ_i приводит к более интенсивному снижению уровней накопления знаний. При уменьшении коэффициента $\beta_i = 1$ и неизменных значениях остальных коэффициентов ($\alpha_i = 1, c_1 = 2, c_2 = 1, c_3 = 0,5, \lambda_1 = 2, \lambda_2 = 1, \lambda_3 = 0,5$) наблюдается неустойчивость системы. Изменение распределения параметров следующим образом: $\alpha_i = 1, \beta_i = 1, c_1 = 0,5, c_2 = 0,25, c_3 = 0,1, \lambda_1 = 0,5, \lambda_2 = 0,25, \lambda_3 = 0,1$ приводит к устойчивости системы и изменениям в объемах накопления знаний. Накопление знаний на уровне первого типа агентов N_1 превосходит в 2,9 раза базовый, N_2 — в 2,5 раз, N_3 — в 2 раза, а объем коллективно накопленных знаний Z — в 4,4 раза, получаемый в результате решения уравнения (5):

$$Z^2 - Z - 14,5 = 0. \quad (5)$$

Отрицательное влияние результатов выполнения проекта в виде отрицательной обратной связи ($\lambda_i < 0$) уменьшает значения накопления знаний у N_i и Z , т.е. затормаживает процессы накопления знаний у всех представленных агентов.

Нестабильность процессов накопления наблюдается при высокой интенсивности взаимных контактов агентов (c_i) и сильной положительной обратной связи по выполнению проекта (λ_i).

Пусть в режиме делового взаимодействия дополнительно накладывается ограничение в виде запаздывания на процесс интегрального накопления знаний в научной школе при выполнении проекта:

$$\dot{Z}(t) = \alpha_4 Z(t) - \beta_4 Z^2(t - \tau_4) + \prod_{i=1}^3 N_i(t). \quad (6)$$

Здесь τ_4 — собственное чистое запаздывание в стабилизации процесса интегрального накопления знаний при выполнении проекта.

Небольшое запаздывание $\tau_{кр} = 0,1с$ в обратных связях стабилизации при условии, что коэффициенты принимают следующие значения $\alpha_i = 1$, $\beta_i = 2$, $c_1 = 2$, $c_2 = 1$, $c_3 = 0,5$, $\lambda_1 = 2$, $\lambda_2 = 1$, $\lambda_3 = 0,5$, приводит к колебаниям переходных процессов с образованием предельных циклов (рисунок 5). В этом случае не представляется возможным успешно выполнить проект, а также эффективно накапливать знания отдельными агентами.

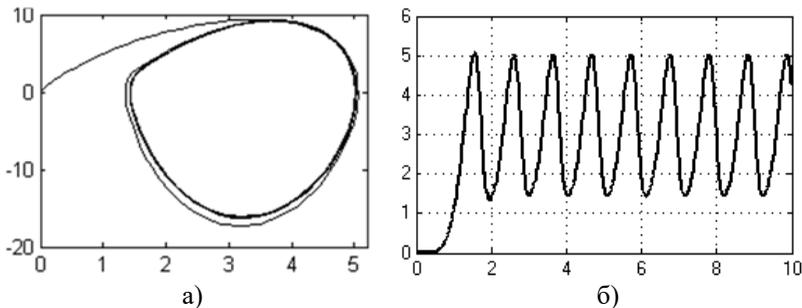


Рис. 5. Автоколебательный процесс в системе при $\tau_{кр}=0,1с$: (а) фазовый портрет Z ; (б) график переходного процесса

Увеличение чистого запаздывания приводит к неустойчивости системы.

Устойчивыми процессы накопления знаний становятся при условии: $\alpha_i = 1$, $\beta_i = 1$, $c_1 = 0,5$, $c_2 = 0,25$, $c_3 = 0,1$, $\lambda_1 = 0,5$,

$\lambda_2 = 0,25$, $\lambda_3 = 0,1$; перерегулирование процессов наблюдается при $0,12c \leq \tau_{икр} \leq 0,15c$, затухание колебаний — при $0,16c \leq \tau_{икр} \leq 0,2c$ (с увеличением чистого запаздывания повышаются частота и амплитуда колебаний). Дальнейшее повышение чистого запаздывания в накопленные знания $\tau_{икр} > 0,2c$ приводит систему в статически неустойчивое состояние («срыв»).

Таким образом, в режиме делового взаимодействия между агентами можно достичь устойчивого накопления интегральных знаний как в целом, так и для каждого носителя знаний (агента) в частности.

5. Заключение. В представленной статье впервые рассмотрена проблема изучения в абстрактной форме многомерных взаимосвязанных нелинейных процессов накопления знаний в результате самоорганизации и обмена информацией между агентами научной школы, в которой агенты представлены в виде системной триады: «человек-ЭВМ-база знаний», т.е. как элементы человеко-машинной системы. По результатам исследований можно отметить следующее:

1) Предложена нелинейная динамическая многоагентная мультиагентная модель накопления знаний в научной школе в результате самоорганизации процессов научного обмена информацией между агентами как носителями различных типов знаний.

2) Абстрактная модель взаимодействия агентов, представленная в виде системы нелинейных дифференциальных уравнений, позволяет избежать системных ошибок (в виде влияния отрицательных результатов выполнения проекта на продуктивность агентов) и повысить эффективность процессов накопления знаний как в научной школе, так и у агентов за счет системной организации нелинейного группового взаимодействия агентов друг с другом, т.е. как организации положительной коллективной обратной связи, так и за счет правильного сочетания отрицательной и положительной обратной связи при самоорганизации каждого агента.

3) Сильное коллективное давление агентов друг на друга (коэффициенты (c_i)), а также стремление каждого агента (коэффициенты $\alpha_i > 0$) получить быстрее и больше новые знания могут привести к неустойчивости («срыву») процесса накопления знаний или к заикливанию процесса (переход к длительным безрезультатным дискуссиям)). Эти процессы наблюдаются и в реальной жизни.

4) Выявлено влияние чистого запаздывания в действии i -го агента по нелинейной стабилизации темпов накопления знаний. Показано, что при различных значениях величины чистого запаздывания в системе могут возникнуть как устойчивые, так и неустойчивые режи-

мы, а также автоколебания. Это еще раз подчеркивает сложность многосвязных динамических процессов накопления знаний агентами научной школы при их нелинейном взаимодействии и слабой самоорганизации этого процесса.

Литература

1. Кун Т. Структура научных революций // М.: АСТ Москва. 2009. 320 с.
2. Служба в двух академиях. Юсупов Рафаэль Мидхатович: к 80-летию со дня рождения и 55-летию науч.-пед. деятельности: сост. Д. В. Бакурадзе, Р.И. Белова // СПб.: Анатолия. 2014. 272 с.
3. Юсупов Р.М. Наука и национальная безопасность. 2-е издание переработанное и дополненное. СПб.: Наука, 2011. 369 с.
4. Лебедев С.А. Российская наука и образование: проблемы и перспективы // Высшее образование в России. 2012. № 11. С. 82–89.
5. Лебедев С.А. Философия науки: 5-е изд., перераб. и доп // М.: Академический проект. Альфа Матер. 2007. 731 с.
6. Лешкевич Т.Г. Философия науки: Учеб. Пособие // М.: ИНФРА-М, 2010. 272 с.
7. Панов А.Д. Макроэволюция и наука // Научноисследовательские исследования. 2013. С. 225–256.
8. Ильясов Б.Г., Герасимова И.Б., Мухамедрахимова Л.Н. Особенности моделирования социальной творческой личности // Вестник УГАТУ. 2014. Вып. 18. № 1(62). С. 156–160.
9. Васильев В.И., Котенко П.С. История и перспективы развития вычислительной техники: 2-ое издание, переработанное и дополненное // М.: Машиностроение, 2013. 496 с.
10. Системный анализ и принятие решений: Словарь-справочник / Под ред. В.Н. Волковой, В.Н. Козлова // М.: Высш. шк. 2004. 616 с.
11. Граждан В.Д. Социология управления: учебник: 2-е изд., перераб. // М.: КНОРУС. 2009. 512 с.
12. Шкляр М. Ф. Основы научных исследований: учебное пособие: 5-е изд. Москва: Дашков и К. 2014. 244 с.
13. Бажанов В.А. Рефлексия в современном науковедении // Рефлексивные процессы и управление. 2002. Вып. 2. № 2. С. 73–89.
14. Трубецков Д.И., Мchedлова Е.С., Красичков Л.В. Введение в теорию самоорганизации открытых систем: 2-е изд. перераб. и доп. // М.: Издательство Физико-математической литературы. 2005. 212 с.
15. Яблонский А.И. Модели и методы исследования науки // М.: Эдуториал УРСС. 2001. 400 с.
16. Шарабчиев Ю.Т. Аттестация научных кадров: количество и качество научных публикаций и результатов научной деятельности // Медицинские новости. 2015. № 1. С. 21–32.
17. Софронов И.Д. Избранные труды. Математическое моделирование и вычислительные системы // Саров: ФГУП «РФЯЦ-ВНИИЭФ». 2005. 564 с.
18. Безручко Б.П. и др. Путь в синергетику. Экскурс в десяти лекциях // М.: Ком-Книга. 2010. 304 с.
19. Малинецкий Г.Г. Математические основы синергетики. Хаос, структуры, вычислительный эксперимент // М.: Либроком. 2015. 310 с.
20. Трубецков Д.И. Введение в синергетику. Хаос и структуры // М.: Либроком. 2014. 210 с.
21. Кохановский В.П. Основы философии науки. Учебное пособие для аспирантов: изд. 3-е. // Ростов-на-Дону: Феникс. 2006. 608 с.
22. Основы науковедения // Москва: «Наука». 1985. 431 с.

References

1. Kun T. *Struktura nauchnyh revolucij* [Structure of Scientific Revolutions]. M.: AST Moskva. 2009. 320 p. (In Russ.).
2. *Sluzhba v dvuh akademijah. Jusupov Rafajel' Midhatovich: k 80-letiju so dnja rozhdenija i 55-letiju nauch.-ped. dejatel'nosti* [Service in the two academies. Rafael Jusupov Midhatovich: the 80th anniversary and 55th anniversary of scientific-pedagogical activity] SPb.: Anatolija, 2014. 272 p. (In Russ.).
3. Jusupov R.M. *Nauka i nacional'naja bezopasnost'* [Science and National Security]. SPb.: Nauka. 2011. 369 p. (In Russ.).
4. Lebedev S.A. [Russian science and education: problems and prospects]. *Vyshee obrazovanie v Rossii – Higher education in Russia*. 2012. vol. 11. pp. 82–89. (In Russ.).
5. Lebedev S.A. *Filosofija nauki* [Philosophy of science]. M.: Akademicheskij proekt, Al'ma Mater, 2007. 731 p. (In Russ.).
6. Leshkevich T.G. *Filosofija nauki: Ucheb. Posobie* [Philosophy of science]. M.: INFRA-M. 2010. 272 p. (In Russ.).
7. Panov A.D. [Macroevolution and Science]. *Naukovedcheskie issledovanija – Scientific research*. 2013. pp. 225–256. (In Russ.).
8. Il'jasov B.G., Gerasimova I.B., Muhamedrahimova L.N. [Features of simulation social creative personality]. *Vestnik UGATU – Vestnik USATU*. 2014. Issue 18. vol. 1(62). pp. 156–160. (In Russ.).
9. Vasil'ev V.I., Kotenko P.S. *Istorija i perspektivy razvitija vychislitel'noj tehniki* [History and perspectives of computer technology development]. M. Mashinostroenie 2013. 496 p. (In Russ.).
10. *Sistemnyj analiz i prinjatje reshenij: Slovar'-spravochnik. Pod red. V.N. Volkovoj, V.N. Kozlova* [System Analysis and Decision Making: Reference Dictionary. Edited by V.N. Volkova, V. N. Kozlov]. M.: Vyssh. shk. 2004. 616 p. (In Russ.).
11. Grazhdan V.D. *Sociologija upravlenija: uchebnik* [Sociology of Management: tutorial]. M.: KNORUS. 2009. 512 p. (In Russ.).
12. Shkljar M.F. *Osnovy nauchnyh issledovanij: uchebnoe posobie* [Basics of research]. Moskva: Dashkov i K. 2014. 244 p. (In Russ.).
13. Bazhanov V.A. [Reflexion in the modern science of science]. *Refleksivnye processy i upravlenie – Reflective processes and management*. 2002. vol. 2. no. 2. pp. 73–89. (In Russ.).
14. Trubeckov D.I., Mchedlova E.S., Krasichkov L.V. *Vvedenie v teoriju samoorganizacii otkrytyh sistem* [Introduction to the theory of self-organizing open systems]. M.: Izdatel'stvo Fiziko-matematicheskoj literatury. 2005. 212 p. (In Russ.).
15. Jablonskij A.I. *Modeli i metody issledovanija nauki* [Models and methods research of science]. M.: Jeditorial URSS. 2001. 400 p. (In Russ.).
16. Sharabchiev Ju.T. [Certification of scientific personnel: the quantity and quality of scientific publications and results of scientific activity]. *Medicinskie novosti – Medical news*. 2015. vol. 1. pp. 21–32. (In Russ.).
17. Sofronov I.D. *Izbrannye trudy. Matematicheskoe modelirovanie i vychislitel'nye sistemy* [Selected works. Mathematical modeling and computational systems]. Sarov: FGUP «RFJaC-VNIJJeF». 2005. 564 p. (In Russ.).
18. Bezruchko B.P., et al. *Put' v sinergetiku. Jekskurs v desjati lekcijah* [Path in synergetics. Excursus into ten lectures]. Moscow. KomKniga. 2010. 304 p. (In Russ.).
19. Malineckij G.G. *Matematicheskie osnovy sinergetiki. Haos, struktury, vychislitel'nyj jeksperiment* [Mathematical foundations of synergetics. Chaos, structure, computing experiment]. M.: Librokom. 2015. 310 p. (In Russ.).
20. Trubeckov D.I. *Vvedenie v sinergetiku. Haos i struktury* [Introduction to the synergetics. Chaos and structures]. M.: Librokom, 2014. 210 p. (In Russ.).

21. Kohanovskij V.P., et al. *Osnovy filosofii nauki. Uchebnoe posobie dlja aspirantov* [Fundamentals of the philosophy of science. Study guide for graduate students]. Rostov-na-Donu: Feniks. 2006. 608 p. (In Russ.).
22. *Osnovy naukovedenija* [Basics of science of science]. Edited by Stefanov N. et al. «Nauka». 1985. 431 p. (In Russ.).

Ильясов Барый Галеевич — д-р техн. наук, профессор, академик АН РБ, профессор кафедры технической кибернетики факультета информатики и робототехники, ФГБОУВПО Уфимский государственный авиационный технический университет (УГАТУ). Область научных интересов: системный анализ, управление в технических и социально-экономических системах, моделирование, теория систем. Число научных публикаций — 450. ilyasov@tc.ugatu.ac.ru; ул. К. Маркса, 12, Уфа, 450000; р.т.: +7(347)2737835, Факс: +7(347)273-78-35..

Ильясов Барый Галеевич — Ph.D., Dr. Sci., professor, academician of Academy of Sciences of Bashkortostan, professor of technical cybernetics department of computer science and robotics faculty, Federal state budgetary educational institution of higher professional education Ufa State Aviation Technical University (USATU). Research interests: system analysis, control of complex technical, social-economic systems. The number of publications — 450. ilyasov@tc.ugatu.ac.ru; 12, K. Marksastreet, Ufa, 450000; office phone: +7(347)2737835, Fax: +7(347)273-78-35..

Герасимова Ильмира Барыевна — д-р техн. наук, доцент, профессор кафедры автоматизированных систем управления факультета информатики и робототехники, ФГБОУВПО Уфимский государственный авиационный технический университет (УГАТУ). Область научных интересов: анализ и управление научно-образовательными системами. Число научных публикаций — 80. tarot_gera@mail.ru; ул. К. Маркса, 12, Уфа, 450000; р.т.: +7(347) 273-78-23, Факс: +7(347) 273-78-23.

Gerasimova Ilmira Baryevna — Ph.D., Dr. Sci., associate professor, professor of automated control systems department of computer science and robotics faculty, Federal state budgetary educational institution of higher professional education Ufa State Aviation Technical University (USATU). Research interests: analysis and control of the scientific-educational systems. The number of publications — 80. tarot_gera@mail.ru; 12, K. Marksastreet, Ufa, 450000; office phone: +7(347) 273-78-23, Fax: +7(347) 273-78-23.

Карамзина Анастасия Геннадьевна — к-т техн. наук, доцент кафедры технической кибернетики факультета информатики и робототехники, ФГБОУВПО Уфимский государственный авиационный технический университет (УГАТУ). Область научных интересов: системный анализ и управление сложными организационными системами. Число научных публикаций — 76. karamzina@tc.ugatu.ac.ru; ул. К. Маркса, 12, Уфа, 450000; р.т.: +7(347)273-78-35, Факс: +7(347)273-78-35.

Karamzina Anastasiya Gennadievna — Ph.D., associate professor of technical cybernetics department of computer science and robotics faculty, Federal state budgetary educational institution of higher professional education Ufa State Aviation Technical University (USATU). Research interests: system analysis and control of complex organization systems, modeling, systems theory. The number of publications — 76. karamzina@tc.ugatu.ac.ru; 12, K. Marksastreet, Ufa, 450000; office phone: +7(347)273-78-35, Fax: +7(347)273-78-35.

Суяргулова Юлия Рустамовна — аспирант кафедры технической кибернетики факультета информатики и робототехники, ФГБОУВПО Уфимский государственный авиационный технический университет (УГАТУ), инженер кафедры технической кибернетики

ки факультета информатики и робототехники, ФГБОУВПО Уфимский государственный авиационный технический университет (УГАТУ). Область научных интересов: управление в социально-экономических системах. Число научных публикаций — 23. yulya_fazletdino@mail.ru; ул. К. Маркса, 12, Уфа, 450000; р.т.: +7(347)273-78-35, Факс: +7(347)273-78-35.

Suyargulova Iuliia Rustamovna — Ph.D. student of technical cybernetics department of computer science and robotics faculty, Federal state budgetary educational institution of higher professional education Ufa State Aviation Technical University (USATU), engineer of technical cybernetics department of computer science and robotics faculty, Federal state budgetary educational institution of higher professional education Ufa State Aviation Technical University (USATU). Research interests: control in and socio-economic systems. The number of publications — 23. yulya_fazletdino@mail.ru; 12, K. Marksastreet, Ufa, 450000; office phone: +7(347)273-78-35, Fax: +7(347)273-78-35.

Поддержка исследований. Работа выполнена при финансовой поддержке РФФИ (проект № 15-08-01334-а)

Acknowledgements. This research is supported by RFBR (grant 15-08-01334-a)

РЕФЕРАТ

Ильясов Б.Г., Герасимова И.Б., Карамзина А.Г., Суяргулова Ю.Р. **Модель накопления знаний в научной школе в результате самоорганизации процесса обмена информации.**

В статье рассматривается вопрос функционирования научных школ как элементов, способствующих движению науки в целом. Приводятся результаты изучения процесса взаимодействия различных агентов в научных школах как многомерных многосвязных системах. Научная школа рассматривается как система, состоящая из взаимосвязанных и взаимодействующих агентов как носителей знаний, которые самостоятельно осмысливают полученные знания, принимают решения и контролируют свою деятельность. В работе выделены процессы накопления знаний для трех групп агентов. Определены два режима работы научной школы: свободный (обучающий, коммуникационный) и деловой (активный, трансляционный). Для первого режима характерен обмен знаниями в процессе проведения семинаров, конференций и т.п., для второго характерно достижение определенной цели (выполнение проекта, написание монографии и т.п.).

Авторами предложены нелинейные модели научной школы в этих двух режимах. Выполнено компьютерное моделирование процессов обмена и накопления знаний. Исследовано влияние эффекта самоорганизации и взаимных контактов, процесса накопления знаний для выполнения проекта на процессы накопления знаний различных агентов, чистого запаздывания при обмене и накоплении знаний.

SUMMARY

Ilyasov B.G., Gerasimova I.B., Karamzina A.G., Suyargulova Iu.R. **A Model of Knowledge Accumulation in the Scientific School as a Result of Self-Organization of Information Exchange Process.**

The functioning of scientific schools as elements that contribute to the movement sciences in general is discussed in the article. Results of studying of process of interaction of various groups in scientific schools as multidimensional multiply connected systems are resulted. Scientific school is considered as a system consisting of interrelated and interacting groups of knowledge holders which independently understand purchased knowledge, make decisions and control their activities. The processes of knowledge accumulation for three groups of knowledge holders are identified in the article. Two modes of operation of the scientific school: free (training, communicative) and businesses (active, translational) were defined. The first mode is characterized by sharing of knowledge in the process of conducting seminars, conferences, etc., the second is characterized by the achievement of a specific purpose (complete the project, writing a monograph, etc.).

The authors proposed non-linear models of the scientific school in these two modes. Computer modeling of the processes of exchange and accumulation of knowledge was performed. The effect of self-organization and mutual contacts, influence of the process of knowledge accumulation to perform the project on processes of knowledge accumulation by different groups of holders, the pure delay in the exchange and accumulation of knowledge are investigated in the article.

А.Я. ФРИДМАН, В.Г. КУРБАНОВ
**ИНФОРМАЦИОННАЯ ТЕХНОЛОГИЯ ИНТЕГРАЦИИ
ПРОСТРАНСТВЕННЫХ ДАННЫХ В СИТУАЦИОННУЮ
СИСТЕМУ МОДЕЛИРОВАНИЯ**

Фридман А.Я., Курбанов В.Г. Информационная технология интеграции пространственных данных в ситуационную систему моделирования.

Аннотация. В работе представлены основные принципы интеграции пространственных данных в ситуационную систему моделирования с целью их равноправной обработки с другими типами информации. В результате пространственный блок становится полнофункциональной управляемой подсистемой системы моделирования, то есть не только обеспечивает визуализацию, оперативное измерение графических характеристик элементов модели и выдачу их в расчетные модули, но и модификацию по результатам расчетов, а также участвует в постановке задачи моделирования и контроле корректности модели.

Ключевые слова: динамический пространственный объект, ситуационное управление, настраиваемый интерфейс, концептуальная модель.

Fridman A.Ja., Kurbanov V.G. Information Technology to Integrate Spatial Data into the Situational Modelling System.

Abstract. This paper introduces the basics of integrating spatial data into the situational modelling system (SMS) in order to provide their equitable processing together with other kinds of information. As a result, the spatial unit becomes a full-scale controlled subsystem of SMS. Besides ensuring visualization, operative measurement of graphic parameters of model elements and transferring them to calculating modules, this unit modifies these parameters according to results of calculations as well as takes part in setting modelling tasks and controlling correctness of the whole model.

Keywords: dynamical spatial object; situational control; customizable interface; conceptual model.

1. Введение. Настоящая публикация продолжает изложенную в предшествующих работах авторов тематику ситуационного моделирования динамических пространственных объектов, в частности промышленно-природных комплексов (ППК).

Ввиду сложности объектов класса ППК структура их математической модели не может быть задана априорно и, следовательно, систему моделирования нельзя строить как законченный пакет прикладных программ. Исходная модель, отражающая представления специалиста о предметной области, изменяется и уточняется в итеративном режиме по результатам проверки на реальных данных. Программное обеспечение таких систем моделирования базируется на информационной модели, отображающей семантическую связь понятий предметной области на потоки обрабатываемых данных. При построении информационных моделей сложных объектов эффективно использование различных приемов декомпозиции, с помощью которых задача моделирования сводится к совокупности более простых для формализации

и численного решения подзадач, а объект исследования представляется в виде некоторой иерархической системы. Анализ способов управления такими системами на языке теории множеств в рамках традиционного оптимизационного подхода был начат фундаментальными работами М. Месаровича и Я. Такахары и привел к появлению общей теории систем [1], где изучаемый объект моделируется некоторой формальной системой. Однако отмеченная выше необходимость многостороннего описания объектов изучаемого класса приводит к тому, что на практике их функционирование оценивается набором критериев, определенных на переменных состояния объекта. Альтернативный подход основывается на работах Р. Шеннона, а в нашей стране — Н.П. Бусленко и Н.Н. Моисеева и их последователей; этот подход был назван имитационным моделированием.

Еще одно направление отхода от классической теории управления связано с работами школы Д.А. Поспелова по ситуационному управлению, основанному на замене описания объекта формальной системой его описанием в виде набора (в общем случае — сети) таких формальных систем, причем переход от одной формальной системы к другой производится на основе анализа ситуации на объекте и применения адекватных ей правил модификации компонентов формальной системы — логико-трансформационных правил [2]. При этом, как и в прикладных системах искусственного интеллекта, целью управления становится поиск не оптимальных, а приемлемых решений. Однако методы ситуационного управления не поддерживают иерархическое описание объекта.

Другой метод построения иерархий был развит на основе работ по концептуальному проектированию баз данных (Дж. Мартин, Дж. Ульман). Он заключается в образовании новых объектов предметной области из имеющихся путем обобщения (введения объекта, для которого каждый из подчиненных ему представляет некоторый частный случай) и агрегации (конструирования объекта из компонентов). Обобщение и агрегация позволяют при многократном их применении построить иерархию концептуальных объектов. В дальнейшем, с появлением работ Э. Йодана (E. Yourdon) это направление оформилось в структурный анализ [3, 4], где модель получается декомпозицией объекта исследования по одному или нескольким параметрам и/или функциям. Школа В.В. Фильчакова дополнила методы структурного анализа методами синтеза среды исполнения и представлениями результатов исследования модели, что создало возможность комплексной автоматизации всех этапов моделирования [5]. К этой же парадигме относятся работы Е.И. Перовской и ее коллег по имитационному моделированию гибких дискретных систем (ГДС) на базе общей теории систем,

где сделаны первые шаги по интеграции геоинформационной системы (ГИС) в систему концептуального моделирования [6]. Однако в работах по ГДС не рассматриваются технологические вопросы реализации концептуальной модели.

Разрабатываемый авторами подход к моделированию ППК детализирует идеи ситуационного управления [2] для новой иерархической концептуальной модели предметной области, которая соответствует принципам общей теории иерархических многоуровневых систем [7] и открыта для оперативной модификации пользователем. В отличие от многочисленных исследований, рассматривающих методологические вопросы структурирования процессов создания и моделирования широкого класса сложных систем с точки зрения их семантики (начиная от UML и до, например, техники ситуационного метода (Situational Method Engineering) [8, 9], распознавания ситуаций (situation awareness) [10–12], общих вопросов интеграции данных, в частности моделей DF&RM и JDL [13–16] и др.), описываемая система ограничена задачами программно-алгоритмической организации имитационного моделирования сложных пространственных объектов и ориентирована на автоматизацию всех этапов моделирования, широкое применение экспертных знаний, использование ГИС-технологии для графического представления составных частей объекта, постановки задачи и представления результатов моделирования, а также для выполнения пространственно-зависимых расчетов. В рамках этого исследования разработана ситуационная система моделирования (ССМ) [17, 18]. Терминология ситуационного моделирования более конкретизирована по сравнению с перечисленными технологиями моделирования (вкратце она пояснена ниже, детали можно найти в публикациях авторов). В частности, термин «интеграция данных» тоже сужен: под ним в ССМ понимается возможность однотипно и равноправно обрабатывать информацию как от расчетных модулей ССМ, так и от встроженных геоинформационной и специализированной экспертной систем. С технологической точки зрения, ССМ должна заполнить имеющийся пробел в средствах моделирования организационных систем в рамках структурного подхода к построению открытой модели предметной области. Доказано, что ССМ позволяет моделировать все основные виды иерархий, рассматриваемых в [7] (стратифицированные, многослойные и многоэшелонные иерархии).

2. Терминология ситуационного моделирования. Чтобы использовать ССМ, ППК необходимо представить в виде иерархически упорядоченного множества *объектов* (составных частей). Эта иерархия отражает организационные взаимоотношения объектов. Каждому объекту может приписываться набор *процессов*, имитирующих преобразо-

вание некоторого набора входных *ресурсов* (или данных) в выходные. Одной из основных характеристик процесса является его *исполнитель*, который определяет его динамические свойства и способ реализации в компьютере. Исполнителя можно задать либо непосредственно (в виде функции), либо ссылкой на имя реализующего этот процесс программного модуля. Ресурсы атрибутированы списками допустимых значений, и также могут иметь исполнителей, например, порождающий их процесс или базу данных, где хранятся временные ряды значений. Списки используются и для числовых, и для ранжированных переменных, чтобы избежать вычислительных проблем, связанных с малыми изменениями данных, и обеспечить их совместную расчетно-логическую обработку. Логическая обработка данных реализуется посредством встроенной экспертной системы, которая может быть назначена исполнителем любого ресурса или процесса. Тогда этот ресурс или все выходные ресурсы этого процесса должны присутствовать в правых частях некоторого набора правил экспертной системы ССМ.

Подлежащие сопоставлению альтернативы реализации ППК вносятся в модель на этапе ее конструирования либо путем декомпозиции некоторого объекта на подобъекты по типу «или», либо заданием альтернативных наборов ресурсов на входе некоторого процесса.

После формирования иерархии объектов необходимо задать ресурсы, которыми обмениваются объекты (сформировать списки входных и выходных ресурсов объектов). Множество ресурсов включает настроечные ресурсы, передаваемые вдоль дуг графа объектов и обеспечивающие корректность иерархии объектов, а также «материальные» ресурсы, описывающие другие типы взаимодействий между объектами. Каждый входной ресурс должен либо порождаться другим объектом, либо помечаться как внешний, что означает его реализацию из некоторой базы данных или на выходе некоторой функции, заданной пользователем.

Таким образом, путем формирования иерархии объектов, описания связывающих их ресурсов и процессов, которые реализуют внутри объектов преобразование входных ресурсов в выходные, строится декларативная *ситуационная концептуальная модель* ППК. В ходе ее построения с помощью отношений иерархии любой составной объект модели однозначно сопоставляется с некоторым подмножеством ГИС-элементов (точек, дуг и полигонов), формирующих его графическое представление. Это позволяет автоматически измерять графические характеристики объектов для использования в расчетах и отображать результаты моделирования на карте.

На такой модели можно исследовать различные варианты реализации ППК и сопоставлять их в статике и динамике.

Теперь об организации моделирования ППК. Базовой формой представления информации в ССМ является *факт*, содержащий сведения о значении некоторого ресурса. *Исходной ситуацией* называется конечный список фактов о состоянии объекта исследования, вводимый пользователем. На основе анализа исходной ситуации и концептуальной модели экспертная система (задавая при необходимости дополнительные вопросы пользователю) доопределяет исходную ситуацию до *полной ситуации*, которой соответствует связный фрагмент концептуальной модели, возможно, включающий некоторые альтернативы реализации структуры объекта моделирования. *Достаточная ситуация* получается из соответствующей ей полной ситуации путем выбора альтернатив, предпочтительных по заданным критериям качества (в ходе решения задачи классификации ситуаций). Достаточная ситуация является управляющей, если требует изменения структуры объекта. При невозможности выбора единственной предпочтительной ситуации имеющиеся альтернативы могут исследоваться в имитационном режиме. *Сценарий* представляет собой последовательность достаточных ситуаций, генерируемых одним и тем же фрагментом модели, и определяет конкретный вариант моделирования.

Таким образом, ССМ решает задачу интеграции различных форм представления знаний с целью применения современных методов исследования и прогнозирования поведения сложных нестационарных природно-промышленных комплексов как многоуровневых многокомпонентных пространственных объектов.

К особенностям ССМ можно отнести следующее:

- автоматизация всех этапов моделирования;
- широкое применение экспертных знаний;
- использование семантически значимых понятий предметной области для построения ее формального описания;
- использование ГИС-технологии не только для графического представления составных частей объекта и результатов моделирования, но также для постановки задачи и выполнения пространственно-зависимых расчетов;
- поддержка современных сценарных подходов к моделированию;
- наличие средств детального формального анализа модели;
- развитый аппарат сопоставительного анализа ситуаций в статике и динамике;
- автоматический синтез исполнительной среды моделирования;
- использование единой инструментальной среды и методологии для моделирования как нормальных, так и критических режимов функционирования ППК;

– возможность интеграции знаний экспертов по различным аспектам функционирования ППК для обоснования принятия решений.

Представленная ситуационная система моделирования позволяет использовать различные формы представления знаний и методы исследования плохо формализованных комплексов, принадлежащих к классу нестационарных пространственных объектов. Она обеспечивает интеграцию уже проверенных надежных решений и анализ возможных сценариев поведения ППК.

ССМ предоставляет ЛПР алгоритмическую поддержку для обоснования его решений об изменении или сохранении структуры подчиненного ему объекта и в этом смысле является альтернативой экспертному совету.

Теперь приведем краткое описание свойств ситуационной концептуальной модели.

3. Основы метода ситуационного моделирования. Географическая информационная система является наиболее «консервативной» составной частью ССМ в плане сложности ее встраивания в систему моделирования, поскольку в традиционных применениях ГИС используются в основном для разработки покрытий и обработки графических характеристик элементов этих покрытий с помощью небольшого набора дорогостоящих (по крайней мере для российского рынка) средств пространственного анализа данных.

КМПО включает: множество объектов модели; отношения, задаваемые над этим множеством; множества атрибутов объектов и отношений; множества функций (функциональных отношений) между информационными объектами, процессами и их атрибутами, а также ресурсов (количественных характеристик) объектов и процессов.

Одно из отличий ССМ от прототипов заключается в том, что наряду с типами элементов модели, формируемыми пользователем и участвующими в контроле корректности модели, для обеспечения более детального формального анализа КМПО в описание ее элементов введены категории объектов, процессов и данных, которые не назначаются пользователем, а присваиваются автоматически. Все объекты модели имеют непосредственную или косвенную (путем задания списка их подобъектов) географическую привязку, которая обеспечивает их согласованное представление в ГИС и КМПО. Привязка объектов к картографической основе производится в ходе конструирования КМПО на этапе контроля и присвоения категорий объектам. Прежде чем рассмотреть эти процедуры, кратко опишем использующиеся в них допустимые категории элементов модели.

В ССМ допускается два вида данных. Если значения некоторого данного есть строковые константы, то такое данное называется параметром (категория PAR), а имеющее числовые значения именуется переменной. В дальнейшем, если сказанное относится к данным любого разрешенного в ССМ типа (и к переменным, и к параметрам), употребляется термин «данное». Данные являются ресурсами (количественными характеристиками) объектов или процессов (категория RES), переменные могут также использоваться как настроечные параметры функций (критериев) качества функционирования элементов КМПО (категория ADJ). При использовании в КМПО каждому данному присваивается имя и список допустимых значений. Отдельную категорию — (категию GIS) — составляют графические характеристики объектов КМПО, непосредственно вычисляемые средствами ГИС. Все они относятся к переменным и могут использоваться только как входные ресурсы объектов или процессов.

Объекты КМПО в ССМ имеют три основные характеристики: имя, функциональный тип, который определяет структуру и функции объекта и используется в процессе анализа корректности КМПО, и имя суперобъекта, который доминирует данный объект в КМПО (отсутствует для объекта верхнего уровня). Далее для простоты рассмотрена двумерная ГИС, на которой реализована исследовательская версия ССМ, хотя аналогичная организация интерфейса возможна и для 2.5 или 3-мерной ГИС, лишь бы объект моделирования имел иерархическую организационную структуру.

По положению в дереве объектов и на карте выделяются три категории объектов КМПО: примитивы (категория LEAF), структурно неделимые с точки зрения глобальной цели моделирования; элементарные объекты (категория GISC), географически связанные с одним ГИС-элементом (полигоном, дугой или точкой какого-либо покрытия); и составные объекты (категория COMP), состоящие из элементарных и/или составных объектов.

Кроме перечисленных основных характеристик объектов, используются следующие дополнительные характеристики, относящиеся к теме настоящей работы:

– положение, которое можно задавать либо непосредственно (координатами или ссылкой на имя ГИС-элемента), либо опосредованно (списком составных частей объекта — подобъектов). Как уже говорилось, примитивы КМПО также могут состоять из нескольких географически различных объектов, но их подобъектами должны быть только ГИС-элементы. Для удобства анализа тип элементарных объектов и ГИС-элементов должен начинаться тремя символами, соответствующи-

ми географическому начертанию объекта (для полигонов — *pol*, для дуг — *arc*, для точечных объектов — *dot*). Всем подобъектам некоторого элементарного объекта автоматически присваивается его положение;

- список входных данных *list_in*;
- список выходных данных *list_out*.

Для составных и элементарных объектов также задается список подобъектов, из которых они состоят.

Каждому объекту сопоставляется в КМПО фрейм следующего вида (часть слотов, не относящихся к теме настоящей работы, опущена):

$$o_i ::= \langle \alpha _ \beta _ \text{имя} \rangle, f_o, [t_h, \langle \text{list_comp} \rangle / \langle \text{area} \rangle], \langle \text{list_in} \rangle, \langle \text{list_out} \rangle, \langle \text{cat}_o \rangle, \rangle, \quad (1)$$

где: α — номер уровня дерева объектов, на котором расположен данный объект, β — его порядковый номер на этом уровне;

f_o — определяет функциональный тип объекта;

$\langle \text{cat}_o \rangle ::= \text{LEAF} \mid \text{COMP} \mid \text{GISC}$ — категория объекта, вносимая в его фрейм в ходе анализа КМПО;

$\langle \text{list_comp} \rangle$ — список имен подобъектов нижележащего уровня декомпозиции, из которых состоит данный объект. Если этот список не задается, то объект рассматривается как примитив КМПО (категория LEAF). То же происходит в случае, когда в список подобъектов входят только ГИС-элементы. Возможен и обратный случай, когда составной объект имеет тип ГИС-элемента (категория GISC). Тогда всем подчиненным ему объектам автоматически присваивается положение его самого. В общем случае, если некоторый объект введен в КМПО с помощью списка его подобъектов, то их графические атрибуты автоматически связываются с графическими ресурсами этого объекта (см. ниже). Альтернативный способ определения объекта КМПО состоит в задании опции $\langle \text{area} \rangle$, которая связывает объект с именем некоторого служебного ГИС-покрытия, определяющего территорию данного объекта. При наличии этой опции в состав объекта автоматически включаются все элементы КМПО, находящиеся на его территории, соответственно, не допускается пересечение территорий различных объектов, имеющих опцию $\langle \text{area} \rangle$;

t_h — тип отношения иерархии, связывающий объект с его подобъектами и обязательный при наличии $\langle \text{list_comp} \rangle$.

Таким образом, декомпозиция объектов в КМПО отражает их взаимное пространственное расположение и организационные связи.

Процессы в КМПО отображают преобразования данных и реализуются в КМПО различными способами в зависимости от присвоенной процессу одной из трех следующих категорий: внутренние про-

цессы (категория INNER), все их входные и выходные данные относятся к одному объекту; внутриуровневые процессы (категория INTRA), связывающие объекты одного уровня КМПО; межуровневые процессы (категория INTER), описывающие передачу данных между объектом и подобъектами или между объектом и суперобъектом. Введенное категорирование процессов несколько усложняет процесс создания КМПО (в некоторых случаях может потребоваться создавать фиктивные процессы, обеспечивающие такую типизацию), но позволяет сделать процедуры формального контроля КМПО значительно более полными и детальными.

Основными характеристиками процессов являются: уникальное имя, характеристика исполнителя процесса и функциональный тип процесса, который определяет тип преобразований, им осуществляемых, и используется в процессе анализа корректности КМПО; дополнительно используются список входных данных и их допустимых граничных значений; список выходных данных и их допустимых граничных значений.

Исполнитель процесса определяет его динамические свойства и способ реализации в компьютере. Исполнитель можно задать либо непосредственно (в виде разностного уравнения), либо косвенной ссылкой на имя реализующего этот процесс программного модуля.

Этап спецификации связей объектов КМПО считается завершенным, когда для каждого входного ресурса (кроме внешних данных) любого объекта модели $o_i \in O$ назначены порождающие этот ресурс объекты:

$$[res_p \in list_in(o_i)] \leftrightarrow [res_p \in list_out(o_j)], j \neq i$$

и каждому уже описанному в модели ресурсу res_p назначены исполнители.

Значения ресурсов могут поступать в ССМ извне либо вычисляться в ходе моделирования. В первом случае это могут быть только входные ресурсы объектов или процессов КМПО, им назначается исполнитель типа DB, если значения ресурса хранятся в базе исходных данных (БИД) или GEN, если они генерируются некоторой функцией или программой. Кроме того, входные ресурсы элементов КМПО могут иметь тип исполнителя GIS, если относятся к графическим характеристикам объектов модели и вычисляются средствами ГИС, и ES, если вычисляются экспертной системой.

При разработке ССМ, как уже отмечалось, значительные усилия были направлены на детализацию автоматического контроля корректности модели, чтобы избежать потери ее обзорности при росте и/или

модификации. Ниже описано применение категорий элементов модели для контроля ее корректности.

4. Оперативный контроль корректности модели. После присвоения категорий ресурсам автоматически запускается приведенный далее алгоритм, присваивающий и контролирующей категории объектов. Его основной задачей является не первичное присвоение категорий объектам, которые ее еще не имеют, что достаточно легко сделать, а выявление ошибок присвоения категории, могущих возникнуть при редактировании КМПО. Поэтому он, как и другие алгоритмы автоматического присвоения атрибутов в ССМ, может показаться несколько избыточным. По той же причине все конфликты, возникающие при автоматическом категоризировании, не разрешаются автоматически, а выдаются для разрешения пользователю. Для упрощения анализа алгоритм нумерует однотипные элементы одного уровня декомпозиции непосредственно друг за другом.

Алгоритм категоризирования и присвоения уровней объектам КМПО.

Шаг 1. Найти все объекты, имеющие в качестве входных ресурсов только внешние данные (ресурсы с типами исполнителя DB, GEN или GIS). Если какой-либо из этих объектов имеет подобъект типа, отличающегося от типов ГИС-элементов, то стоп — конфликт категоризирования объектов. Иначе присвоить всем этим объектам категорию LEAF.

Шаг 2. Найти все еще не имеющие категории объекты, имеющие тип ГИС-элементов. Присвоить категорию GISC им и всем объектам, входящим в их множества подчиненности в дереве объектов и еще не имеющим категории.

Шаг 3. Найти объект $o_0 \in O$, ни один выходной ресурс которого, кроме настроечных параметров, являющихся входными ресурсами его непосредственных подобъектов, не является входным ресурсом ни одного объекта из множества O . Если таких объектов нет или их несколько, то стоп — конфликт категоризирования объектов.

Шаг 4. Присвоить объекту o_0 уровень декомпозиции 0, изменив соответственно его имя в его во фрейме (1). Если список его подобъектов пуст, то стоп — конфликт категоризирования объектов. Если этот список не пуст, то внести все входящие в него объекты во множество M , $\alpha := 1$, $\beta := 1$ и присвоить объекту o_0 категорию COMP.

Шаг 5. Выбрать объект o_i из множества M , внести α и β в его фрейм, зафиксировать его тип $f_o := f_o(o_i)$, $\beta := \beta + 1$.

Шаг 6. Исключить объект o_i из множества M . Если o_i имеет категорию LEAF, то перейти к шагу 10. Если o_i имеет категорию GISC,

то перейти к шагу 8. Если o_i имеет категорию COMP, то стоп — конфликт категоризирования объектов.

Шаг 7. Если o_i не имеет подобъектов без категории, то стоп — конфликт категоризирования объектов. Если имеет, то присвоить ему категорию COMP.

Шаг 8. Если o_i имеет подобъекты без категории или с категорией COMP, то стоп — конфликт категоризирования объектов.

Шаг 9. Включить все подобъекты объекта o_i во множество Q.

Шаг 10. $\beta := \beta + 1$. Если во множестве M есть еще объекты типа f_0 , то выбрать один из них в качестве o_i и перейти к шагу 6.

Шаг 11. Если во множестве M есть еще объекты, то перейти к шагу 5.

Шаг 12. Если множество Q пусто, то перейти к шагу 13. Если нет, то $M ::= Q$, $Q ::= \emptyset$, $\beta := 1$, $\alpha := \alpha + 1$, и перейти к шагу 5.

Шаг 13. Если во множестве O не осталось объектов, не имеющих категории, то алгоритм успешно завершен, иначе конфликт: КМПО не является связанной по иерархии объектов.

Теорема. Алгоритм категоризирования объектов и предшествующий ему контроль КМПО ССМ выявляют все возможные ошибки категоризирования объектов модели.

Доказательство. Для доказательства достаточно показать ситуации, в которых будут выявлены любые ошибочно присвоенные категории объектов КМПО.

Если некоторому объекту присвоена категория LEAF вместо категорий GISC или COMP, то будет выявлен конфликт при проверке типов исполнителей его входных ресурсов, которые не все будут относиться к внешним данным.

Если некоторому объекту присвоена категория GISC вместо категории COMP, то будет выявлен конфликт на шаге 8 алгоритма категоризирования объектов.

Если некоторому объекту присвоена категория GISC или COMP вместо категории LEAF, то будет выявлен конфликт при проверке типов его подобъектов, не все из которых будут ГИС-элементами.

Если некоторому объекту присвоена категория COMP вместо категории GISC, то будет выявлен конфликт на шаге 7 алгоритма категоризирования объектов. Конец доказательства.

В ходе категоризирования объектов КМПО формируется их ГИС-представление. Данные об объектах КМПО в ГИС заключаются в файл базы данных (БД) следующей структуры. Первая запись этого файла содержит данные об объекте самого высокого уровня проблемной области. В поле «Name» занесено имя объекта (аргумент $\langle \alpha_ \beta_ \text{имя} \rangle$ фрейма

объектов), в поле «Type» — категория объекта (аргумент <cat,> фрейма объектов). Числовое поле «Num» содержит количество его подобъектов, и мето-поле «Objects» включает список имен подобъектов нижележащего уровня декомпозиции, из которых состоит данный объект (аргумент <list_comp> фрейма объектов). Имеются также два поля « α » и « β », которые отражают вспомогательные данные: номер уровня декомпозиции, на котором находится данный объект (α), и номер местоположения объекта внутри этого уровня (β).

Следующие n записей файла содержат вышеперечисленные данные о каждом из подобъектов вершины (n равно количеству подобъектов). Следующие m записей представляют данные об m объектах нижележащего уровня и так далее. Таким образом описаны все уровни проблемной области КМПО. Объектами самого нижнего уровня, представленными в файле, должны быть элементарные объекты (категория GIS). Для объекта категории GIS в поле «Objects» хранится значение идентификатора GIS, что находится в соответствии со структурой фрейма объектов.

5. Привязка концептуальной модели к карте. Взаимодействие GIS с КМПО начинается на стадии конструирования последней: при задании одному из объектов модели типа, определенного для GIS-элементов, пользователю предлагается выбрать этот объект на карте. Результат выбора сохраняется в справочнике «GIS-элементы КМПО». Одно из условий адекватности модели объекта исследования состоит в наличии GIS-привязки всех его составных частей — объектов КМПО, то есть в возможности однозначного сопоставления любого составного объекта модели с некоторым подмножеством GIS-элементов. Процесс привязки имеет дополнительную ветвь при задании для некоторого объекта модели опции «agea». В таком случае пользователь создает и сообщает системе имя специального GIS-покрытия, содержащего изображение границы объекта. Это имя заносится в БД «Связи объектов», и GIS осуществляет выборку всех GIS-элементов, полностью или в большей части находящихся в пределах указанной границы. Для отобранных элементов с помощью разработанного алгоритма выявляются все «старшие» объекты, не имеющие суперобъектов среди отобранных. Множество «старших» объектов предьявляется пользователю для утверждения в качестве подобъектов объекта, имеющего опцию «agea».

На этапе определения ресурсов КМПО происходит формирование множества данных, используемых GIS [18], для чего предварительно, с учетом особенностей языка GIS, формируются функции разрешенных типов графических характеристик для различных типов

объектов, рассматриваемых как аргументы этих функций. Далее без потери общности приводится описание интерфейса для широко распространенных языков ГИС Avenue и SML.

Получить графические характеристики или координаты объекта (пары объектов) можно, посылая ему (им) на языке Avenue запросы.

Из-за ограниченных возможностей макроязыков ГИС в ССМ не предусмотрена автоматическая модификация графических характеристик ГИС-элементов (для такой модификации необходимо применение ручных операций). Однако в языках ГИС имеются некоторые средства автоматической генерации новых элементов на основе имеющихся ГИС-элементов. Получающиеся при этом элементы (назовем их трансформированными ГИС-элементами — ТГЭ) являются результатами применения к одному или нескольким (в исследовательской версии ССМ — двум) элементам множества графических объектов выбранной пользователем ГИС-операции из набора допустимых операций, которые рассмотрены ниже. Общий формат команды применения операции имеет вид:

```
<операция> ::= <код_операции> <код_ГИС-элемента> [<код_ГИС-элемента>] <параметры_операции> [<легенда>] [<комментарий>],
```

где: <код_операции> — значение кода из справочника доступных ГИС-операций;

<код_ГИС-элемента> — коды одного или двух ГИС-объектов, к которым применяется операция;

<параметры_операции> — список кодов ресурсов — параметров операции в соответствии с полями $N_{\text{пар}}$ и $\text{list}_{\text{пар}}$ справочника ГИС-операций;

<легенда> — строка, используемая для надписи на ГИС-изображении результата операции;

<комментарий> — пользовательский текст, используемый в качестве комментария для результата операции.

Основные ГИС-операции для исследовательской версии ССМ приведены в таблице 1.

Применение ГИС-операций задается пользователем на этапе конструирования связей элементов модели путем выбора ГИС в качестве потребителя того или иного ресурса. В таком случае пользователю предлагается выбрать на карте ГИС-элемент(ы) — участник(и) операции, по их ГИС типам формируется список допустимых операций, после выбора желаемой операции задаются остальные ее параметры.

Таблица 1. ГИС-операции для исследовательской версии ССМ

Служебный тип ресурса	Назначение ресурса	Допустимые типы объектов
are	площадь	полигон
arem	площадь объединения	любая пара, состав.объект
areo	площадь непересекающихся частей	2 полигона
areo1	площадь непересекающейся части первого	2 полигона
ares	площадь пересечения	2 полигона
cen	координаты геометрического центра	линия, полигон
cene	координаты геом. центра экстента	любая пара, состав.объект
cenm	коорд. геометрического центра объединения	любая пара
cens	коорд. геометрического центра пересечения	линия-полигон, 2полигона
cox	координата X	точка
coy	координата Y	точка
dis	минимальное расстояние	любая пара
disc	расстояние между центрами	любая пара
disb	расстояние между началами	точка-линия, 2 линии
dise	расстояние между концами	точка-линия, 2 линии
disf	миним. расст. между границами экстентов	2 составных объекта
disfc	расстояние между центрами экстентов	2 составных объекта
end	координаты конца	линия
ext	описывающий прямоугольник (экстент)	линия, полигон
extm	описывающий прямоугольник объединения	любая пара, состав.объект
len	длина линии	линия, граница полигона
lene	длина границы экстента	составной объект
lenm	длина объединения	2 линии, 2 полигона
leno	длина непересекающейся части линии	2 линии, линия-полигон
leno1	длина непересекающейся части первого	2 полигона
lens	длина линии пересечения	2 линии, линия-полигон
per	положение точки на линии в %	точка-линия
sec	координаты точек пересечения	2 линии, линия-полигон, 2 полигона
sta	координаты начала	линия

Результатом ГИС-операции в общем случае является некоторый ТГЭ, его реализация зависит от категории ресурсов — параметров операции его создания. Если они имеют категорию ТМР, то ТГЭ создается только в оперативной памяти ГИС, если задана категория СМР, то создается специальное ГИС-покрытие, отображающее результат операции и сохраняемое в базе соответствующего сценария для использования при сопоставлении результатов моделирования различных сценариев. При назначении операций в соответствии с принципами построения организационных иерархий предполагается, что некоторый объект КМПО имеет право применять ГИС-операции только к объектам из его множества подчиненности.

Для представления результатов моделирования на картографической основе, используется описанный выше набор ГИС-операций над ГИС-элементами КМПО, который зависит от языка используемой ГИС. Для применения этих операций в БД ГИС имеется справочник «ГИС-операции», связывающий служебный код операции ($код_{оп}$) с ее именем на языке ГИС ($имя_{оп}$), количеством участвующих в операции объектов $N_{оп}$ (в исследовательской версии ССМ это число равно 1 или 2) и списком их ГИС-типов ($типы_{о}$), а также количеством параметров операции ($N_{пар}$) и списком их ГИС-имен ($list_{пар}$).

После того как описание проблемной области КМПО полностью завершено, файл БД сохраняется на внешнем запоминающем устройстве для последующего чтения и интерпретации ГИС-приложением. На основе этого файла ГИС-приложение создает объект виртуальной таблицы, к которому могут быть направлены запросы в ходе моделирования.

Предварительный этап интерпретации данных состоит в упорядочивании таблицы. На первом шаге она индексируется по полю « α », чтобы установить численные значения уровней декомпозиции в порядке возрастания. На втором шаге — по полю « β », чтобы упорядочить внутриуровневые номера объектов. Затем при помощи поля «Туре» осуществляется выборка объектов, имеющих категорию GISC. Формируется список имен этих объектов, полученных из поля «Name». Также составляется словарь, где ключами являются имена ГИС-объектов, а ассоциированными с ними значениями — их идентификаторы. Далее выборка инвертируется, вследствие чего выбранными становятся объекты других категорий КМПО — COMP и LEAF. В новой выборке процесс считывания информации осуществляется сверху вниз по записям таблицы. Имена объектов также объединяются в список. Параллельно идет процесс составления словаря-справочника,

где ключом является имя, а ассоциированным значением — список подобъектов из поля «Objects».

Записи, относящиеся к составным объектам (COMP) или примитивам КМПО (LEAF), выполняют в ГИС только вспомогательную функцию. Имена этих объектов используются на заключительном этапе ГИС-сеанса в качестве названий для пунктов меню, таблиц, полей, подписей к графикам, диаграммам. Такой справочный материал дает возможность правильно сгруппировать результаты и обеспечить удобство их представления на экране. Списки имен подобъектов поля «Objects» послужат для формирования различных выборок пространственных данных на основе элементарных объектов, что существенно в многоступенчатых задачах.

Всю основную для себя информацию ГИС-приложение считывает из множества записей элементарных объектов. Как уже было упомянуто, объекты категории GIS являются ГИС-элементами и имеют конкретную географическую привязку. Эти объекты распознаются средой ГИС при помощи их служебных идентификаторов. Каждый идентификатор уникален и состоит из имени покрытия и внутреннего номера, который приписан каждому объекту на этом покрытии. На первом этапе процедуры распознавания определяется количество и поименный список всех покрытий ГИС-приложения, участвующих в решении конкретной задачи. Каждому выбранному покрытию назначается статус видимого и активного. Это означает, что только эти покрытия (помимо базового) будут отображены на карте.

На втором этапе распознавания производятся выборки объектов из каждого покрытия по их внутренним номерам. Выборки осуществляются при помощи ряда запросов и могут систематизироваться тематически (по отдельным покрытиям) либо отбираются элементарные объекты ГИС, относящиеся к тому или иному составному объекту или примитиву. Как правило, это несколько элементов ГИС, представленных на разных покрытиях, но расположенных близко друг к другу в географическом отношении. Прежде всего, покрытиям назначается статус активных. Затем перебираются атрибутивные таблицы этих покрытий с выделенными объектами, в результате создается новый графический объект *gr*. Список всех созданных графических объектов заносится в выборку *grList*. На основе списка графики *grList* и словаря-справочника составных объектов и примитивов формируется новый словарь для объектов категорий COMP и LEAF, ключами в котором будут служить также имена этих объектов, а ассоциированными значениями — тождественные их подобъектам графические объекты из списка *grList*. Этот графический справочник необходим на завершающем этапе представления

результатов. Он дает возможность быстрого доступа к готовым наборам ГИС-объектов, которые по-разному группируются на соответствующих картах в зависимости от их принадлежности тем или иным объектам более высокого уровня декомпозиции.

Так же составляются списки и формируются выборки объектов, для которых требуется определить графические или другие характеристики, если эти атрибуты будут использоваться внешними программами или расчетными модулями. Графические характеристики ГИС-объектов включают в себя:

- X и Y — координаты точки;
 - Dimension — размерность, точке соответствует размерность 0, линии — 1, полигону — 2;
 - Length — длина, периметр или окружность графического объекта (только с размерностью 1 или 2 для двумерной ГИС);
 - Area — площадь двумерных объектов (полигонов);
 - Center — точка, представляющая собой математический центр объекта, который не обязательно находится внутри объекта;
 - Extent — наименьший прямоугольник, полностью содержащий в себе графический объект.
- Получить графические характеристики или координаты объекта можно, посылая ему специальные запросы на языке Avenue: GetX, GetY, GetDimension, ReturnArea, ReturnCenter, ReturnExtent, ReturnLength.

К графическим характеристикам также могут быть отнесены и графические результаты, получаемые при преобразовании или взаимодействии элементарных объектов:

- запросы LineIntersection, PointIntersection и ReturnIntersection возвращают графический объект, являющийся результатом пересечения двух объектов;
- запросы ReturnMerged и ReturnUnion — возвращают объединение двух объектов;
- запрос ReturnClipped(Rect) — возвращает графический объект, вырезанный прямоугольником Rect;
- при помощи запроса ReturnBuffered можно построить вокруг объекта буферную зону;
- при помощи запроса Distance — определить минимальное расстояние между двумя объектами.

Для линейных объектов к атрибутам можно причислить и графические решения двух обратных задач: возврат положения точки на линии в процентном отношении от ее начала по запросу PointPosition и получение точечного объекта, заданного на линии в процентах P_r по

запросу `Along(Pr)`, а также расстояние между линией и точкой по запросу `QueryPointDistance`, который возвращает отрицательное числовое значение, если точечный объект находится слева от линейного.

Выборки элементарных объектов и их графические атрибуты записываются в специальные текстовые или табличные файлы, которые хранятся вне ГИС-приложения и доступны любому программному средству. В случае необходимости выборки экспортируются в новые шейп-файлы и впоследствии могут добавляться на карту в виде самостоятельных покрытий, содержащих только объекты выборки.

При определении географической зоны, связанной с решаемой задачей, в ГИС-приложении широко используется графическая характеристика `Extent` (экстент). Она имеет существенное значение, так как непосредственно применяется для определения и коррекции границ проблемной области при подготовке картографического материала и визуализации его на экране монитора.

Как правило, ситуационная модель для конкретного варианта моделирования строится на основе многоуровневого фрагмента КМПО. Поэтому для полного отображения ситуации в ГИС-приложении требуется предоставление нескольких экранов картографической информации с разными наборами подобъектов проблемной области. Эта задача решается при помощи ранее составленных списков и словарей-справочников для разных категорий объектов, а также списка и справочника графических шейпов.

Для каждого составного объекта или примитива уже сформирована выборка подобъектов категории `GISC` и соответствующая ей выборка графики, для которой может быть определен `Extent`. Для начала необходимо получить в графическом справочнике ассоциированный с именем объекта набор графических шейпов и установить при помощи запроса `SetSelected` каждому элементу этого набора статус выбранного. Затем выполняется ряд запросов для сформированного ранее списка графики `grList`, которые увеличивают экстент (`ext`) выбранной графики (`grSel`) до размера картографического окна ГИС-приложения. В результате на экране монитора отображаются в оптимальном масштабе все элементарные объекты, входящие в составной объект или примитив, вместе с частью прилегающей географической зоны.

Аналогичная процедура проделывается для всех объектов категорий `COMP` и `LEAF`. Соответственно, общее количество подготовленных картографических окон будет соответствовать числу этих объектов. Окна систематизируются в логические группы по принадлежности отображаемых объектов к одному и тому же уровню декомпозиции и графически передают все ступени иерархии ситуационной модели.

Для того чтобы визуализировать картографические окна, средствами ГИС подготавливается система временных пунктов меню, которая также имеет свою иерархическую структуру.

На этом шаге процесс интерпретации и подготовки данных заканчивается. Завершающим этапом становится просмотр полученных результатов и расчетов по анализируемой ситуации. Помимо описанных выше картографических данных, представление результатов может заключаться в таблицах БД, диаграммах, графиках, схемах и т.д., содержащих возможно полную информацию о ситуационной модели. Вся информация поступает в ГИС-приложение из внешней базы данных и расчетных программ в виде текстовых файлов и файлов БД. Она распознается и систематизируется по принадлежности к ГИС-объектам.

Выше описан случай, при котором для постановки задачи использовалась КМПО. Другой, визуальный, вариант постановки задачи осуществляется в ГИС-приложении при непосредственном использовании карты. На ней «мышью» выбирается интересующая географическая зона и назначаются покрытия, которые будут принимать участие в моделировании ситуации. Элементарные объекты, заключенные внутри указанной зоны, идентифицируются ГИС-приложением. К их именам и идентификаторам, организованным в текстовый файл, получает доступ КМПО, которая интерпретирует и систематизирует данные своими внутренними средствами. В результате вновь имеется фрагмент КМПО, описывающий проблемную область, и дальнейшее решение задачи сводится к тому, что уже было описано ранее.

6. Пример применения ГИС-интерфейса. Рассмотрим алгоритм геокодирования, использованный для операций с протяженными (линейными) объектами в программной системе, которая описана в [19]. Задача этого алгоритма — обеспечить поиск и отображение (при необходимости — в увеличенном масштабе) окрестности определенных участков линейных объектов, имеющих на одном из покрытий ГИС-приложения. В описанной там первой версии упомянутой системы моделирования линейные объекты представлены линиями электропередач (ЛЭП), целью поиска является поиск аварийных участков ЛЭП и отображение подъездных путей к ним для удобства работы ремонтных служб.

Информация о неисправностях ЛЭП поступает с дистанционных датчиков и может быть двух типов: неисправность возникла на таком-то километре некоторой ЛЭП, либо одновременно появились неисправности двух ЛЭП, обычно это связано с замыканием между ними. Поэтому реализованы два варианта поиска: по заданному рас-

стоянию от крайнего узла выбранного объекта и поиск окрестности точки пересечения двух линейных объектов по их именам.

Такие задачи не решались внутренними средствами модуля отображения покрытий и макроязыка SML используемой ГИС. Потребовалось разработать процедуру адресации внутренних участков линейных объектов, для этого был применен дополнительный пакет ГИС, который поддерживает функцию геокодирования.

Геокодирование — механизм для создания непосредственной связи между объектами покрытия и их адресами на уровне базы данных. Для того чтобы воспользоваться его возможностями, требуется некоторая подготовительная работа.

В первую очередь необходимо создать адресный файл [info_file]. Для этого ЛЭП на всем своем протяжении условно рассматривается как цепь отрезков фиксированной длины. Всем узлам, полученным при разбиении, дается собственное наименование. После этого можно создавать файл БД и заполнять таблицу атрибутов адресными компонентами. Имена полей БД таковы:

- name — наименование узла разбиения;
- number — номер этого узла;
- LEP_name — наименование ЛЭП;
- LEP_type — характеристика типа ЛЭП.

Количество записей в базе данных равно количеству участков, на которые условно разделена сеть линий электропередач.

Когда адресный файл подготовлен, выполняется команда:

```
APARSE [info_file] [out_item] [number_house_numbers],
```

которая форматирует адресные составляющие в единую запись.

Аргумент [number_house_numbers] определяет структуру адресной записи. В геокодировании существует четыре вида адресов, которые различаются по количеству числовых составляющих. Для каждого вида поддерживается определенный набор и порядок расположения адресных элементов. В созданном файле базы данных присутствует только один числовой компонент для отдельно взятой записи. Такая структура соответствует аргументу [number_house_numbers], равному 1.

После выполнения команды APARSE в адресном файле появляется новое поле [out_item]. В нем содержатся в корректной форме адреса всех существующих узлов разбиения ЛЭП. В дальнейшем, при использовании команды AMATCH, результирующее поле [out_item] используется как адресное поле [address_item].

На втором этапе строится адресное покрытие. Прежде всего, нужно добавить в таблицу атрибутов объектов несколько новых полей, которые содержат следующие адресные компоненты:

– `from_num` — номер первого узла при разбиении ЛЭП на условные равновеликие отрезки (все значения этого поля равны единице);

– `to_num` — номер последнего узла при разбиении ЛЭП на отрезки;

– `LEP_name` — наименование ЛЭП;

– `LEP_type` — характеристика типа ЛЭП.

В данном случае в ряду адресных компонентов для каждого линейного объекта (ЛЭП) используются две числовые составляющие — номера начального и конечного узлов разбиения. Такой форме записи соответствует аргумент `[number_house_numbers]`, равный 2.

Новые поля необходимо заполнить таким образом, чтобы в каждой строке таблицы наименованию данной ЛЭП соответствовал номер ее пользовательского идентификатора, который находится в поле `cover_ID`. После этого можно приступать к построению адресного покрытия при помощи команды:

```
ACREATE [cover] [LINE] [number_house_numbers].
```

После ее выполнения налаживается взаимодействие между адресными атрибутами и соответствующими элементами покрытия `[cover]`. В базе данных `[cover]` создается специальный файл `Cover.ADD`, который сохраняет атрибуты адресного покрытия.

Для того чтобы перестроить существующие взаимоотношения покрытий с учетом вновь созданного, служит команда:

```
ABUILD [cover] [LINE] {SOUNDEX}.
```

После ее применения адресное покрытие может полноценно использоваться в ГИС.

Аргумент `SOUNDEX` указывает на то, что этой командой будет создано определенным образом поле эквивалентных значений для `LEP_name`. Оно позволит ускорить процесс сравнения адресов, который рассматривается ниже.

Наличие отдельного покрытия дает ряд преимуществ при работе с адресами. В файле `Cover.ADD` можно сохранять для каждого объекта любое количество записей, а также все связи между ними. Это делается при помощи внутреннего индекса `Cover_` и пользовательского `Cover_ID`, которые, соответственно, равны во всех записях, сделанных для одного объекта. Назначение уже существующему адресу псевдо-

нима (или альтернативного адреса) позволяет значительно облегчить обращение к нему в адресном покрытии и, как следствие, идентификацию объекта, расположенного по этому адресу.

Последний шаг подготовительной работы состоит в сопоставлении записей, сделанных в поле [address_item] адресного файла [info_file], и аналогичных атрибутов адресного покрытия [match_cov]. Делается это при помощи команды:

```
AMATCH [info_file] [address_item] [match_cov] [out_cov] ##  
{NOREJECTS/REJECTS}.
```

Процесс сравнения осуществляется поэтапно. Разработан его алгоритм и специальная тестовая таблица, в которой подсчитываются очки, набранные каждым кандидатом. Сначала сопоставление происходит для поля SOUNDEX и числовых значений адресных компонентов. Всем выявленным при этом возможным кандидатам присваивается минимальный счет — 100 очков. Далее сравнению подвергаются LEP_name и LEP_type. При каждом совпадении счет остается прежним, а при несовпадении — уменьшается на обусловленное счетной таблицей количество очков. Для того чтобы сопоставление прошло успешно, возможному кандидату нельзя терять ни одного очка, т.е. в результате тестирования сохранить все 100.

Если сравнение удалось, и испытуемые адреса полностью совпали, то устанавливается взаимодействие между адресным файлом и адресным покрытием. По итогам сравнения строится покрытие [out_cov]. Положительные результаты заносятся в таблицу атрибутов файла Out_cover. PAT, а отрицательные — в Out_cover. REJ.

В случае, когда стопроцентное совпадение не достигнуто, можно использовать опцию REJECTS. Указание этого аргумента в команде AMATCH вызывает к действию интерактивную программу Reject processing (Обработка отказов). Эта программа выводит на экран поочередно все несовпавшие адреса, сопровождая их диагностическими сообщениями, что позволяет оценить и попытаться решить проблему.

Применение программы Reject processing позволяет, в большинстве случаев, устранить все неудачи сопоставления и добиться успешного выполнения команды AMATCH.

Поиск по полученному покрытию производится с помощью команды:

```
WHEREIS [LIST / symbol] [address /street /intersection].
```

Эта команда идентифицирует адресные записи и обозначает на электронной карте местоположение точечного адреса [address], линей-

ного объекта [street] или точки пересечения двух протяженных объектов [intersection]. Выделение указанных адресов происходит при помощи цветowych меток и символов [symbol], назначаемых пользователем.

Опция LIST выдает в верхней строке экрана монитора адресную запись (или несколько записей) из таблицы атрибутов объектов покрытия, которая соответствует заданному второму аргументу команды WHEREIS.

Отображение на экране и выделение искомых объектов происходит вместе с частью окружающей местности сразу же на самой подробной крупномасштабной карте, имеющейся в наличии. Для реализации процесса автоматического увеличения используется ряд команд языка SML.

7. Заключение. С помощью описанных в настоящей работе команд ГИС создаются картографические слои, содержащие ГИС-объекты КМПО ССМ, и строится набор идентификаторов, которые позволяют взаимно однозначно связывать фрагменты концептуальной модели с наборами графических атрибутов, входящих в эти фрагменты объектов, и использовать указанные атрибуты при моделировании.

На примере ГИС PC ARC/INFO, использованной при реализации исследовательской версии ССМ, ее макроязыка SML и более гибкого языка Avenue, предназначенного для разработки графических интерфейсов пользователя, показана организация запросов к данным ГИС, а также особенности применения указанных языков для решения типичных задач, возникающих при моделировании нестационарных пространственных объектов: выборки объектов вручную или по логическому условию; просмотра графических и других атрибутов объектов; определения относительного расположения и длин участков линейных объектов с применением механизма геокодирования.

ГИС является эффективным инструментом для автоматической обработки пространственных характеристик промышленно-природных систем и выборки необходимых параметров для имитационного моделирования. Она также обеспечивает возможность постановки задачи анализа сценариев и представления результатов моделирования в наглядной форме непосредственно на географической карте или масштабном плане объекта управления.

В целом имеющийся опыт применения ситуационного подхода в моделировании различных аспектов функционирования ППК позволяет утверждать, что эта разработка может представлять интерес в научно-практической деятельности по исследованию природно-промышленных комплексов и поддержке принятия решений по управлению объектами такого класса. ССМ способна служить основой для

разработки многоуровневых систем регионального управления, работающих в единой среде и не предъявляющих жестких требований к компьютерной квалификации пользователей.

Литература

1. Месарович М., Такахара Я. Общая теория систем: математические основы // М.: Мир. 1978. 312 с.
2. Поспелов Д.А. Ситуационное управление: теория и практика // М.: Наука. 1986. 288 с.
3. Йодан Э. Структурное проектирование и конструирование программ // М.: Мир. 1979. 360 с.
4. Yourdon E. Modern Structured Analysis // Prentice-Hall. 1989. 688 p.
5. Бржезовский А.В., Фильчаков В.В. Концептуальный анализ вычислительных систем // СПб: ЛИАП. 1991. 78 с.
6. Перовская Е.И. Основы гибкой автоматизации // Л.: ЛИАП. 1986. 32 с.
7. Месарович М., Мако Д., Такахара И. Теория иерархических многоуровневых систем // М.: Мир. 1973. 344 с.
8. Henderson-Sellers B., Ralyté J. Situational Method Engineering: State-of-the-Art Review // J. of Universal Computer Science. 2010. vol. 16. no. 3. pp. 424–478.
9. Gericke A., Fill H.-G., Karagiannis D., Winter R. Situational Method Engineering for Governance, Risk and Compliance Information Systems // Proceedings of the 4th Int. Conf. on Design Science Research in Information Systems and Technology: ACM Press. 2009. pp. 24.
10. Попович В.В., Прокаев А.Н., Сорокин Р.П., Смирнова О.В. О распознавании ситуации на основе технологии искусственного интеллекта // Труды СПИИРАН. 2008. № 7. С. 93–104.
11. Endsley M.R. Theoretical underpinnings of Situation Awareness: a critical review // Situation awareness analysis and measurement. 2000. pp 3–32.
12. Endsley M.R. Final Reflections: Situation Awareness Models and Measures // J. of Cognitive Engineering and Decision Making. 2015. vol. 9. no. 1. pp. 101–111.
13. Steinberg A.N., Bowman C.L. Revisions to the JDL data fusion model // Handbook of Multisensor Data. 2009. pp. 45–68.
14. Steinberg A.N. Foundations of Situation and Threat Assessment // Handbook of Multisensor Data. 2009. pp. 437–502.
15. Steinberg A.N., Bowman C.L. Rethinking the JDL data fusion model // Proceedings of the MSS National Symposium on Sensor and Data Fusion. 2004. vol. 1. pp. 79–97.
16. Bowman C.L. The Dual Node Network (DNN) Data Fusion & Resource Management (DF&RM) Architecture // Proceedings of AIAA Intelligent Systems Conference. 2004. vol. 50. pp. 134.
17. Фридман А.Я., Курбанов В.Г. Формальная концептуальная модель промышленно-природного комплекса как средство управления вычислительным экспериментом // Труды СПИИРАН. 2014. № 6(37). С. 424–453.
18. Фридман А.Я., Курбанов В.Г. Ситуационное моделирование надежности и безопасности промышленно-природных систем // Информационно-управляющие системы. 2014. № 4(71). С. 1–10.
19. Сагидова М.Л., Фридман А.Я. Реализация интерфейса ГИС и концептуальной модели в ситуационной системе вычислительного эксперимента // Модели социальных, технологических и образовательных процессов. 2001. С. 93–98.

References

1. Mesarovic M., Takahara Ya. *Obshchaia teoriia sistem: matematicheskie osnovy* [General systems theory: mathematical foundations]. Moscow: Pub. Mir. 1978. 312 p. (In Russ.).
2. Pospelov D.A. *Situatsionnoe upravlenie: teoriia i praktika* [Situational control: theory and practice]. Moscow: Pub. Nauka. 1986. 288 p. (In Russ.).
3. Yourdon E. *Strukturnoe proektirovanie i konstruirovaniye programm* [Structural design and programs construction]. Moscow: Pub. Mir. 1979. 360 p. (In Russ.).
4. Yourdon E. *Modern Structured Analysis*. Prentice-Hall. 1989. 688 p.
5. Brzhezovskii A.V., Fil'chakov V.V. *Kontseptual'nyi analiz vychislitel'nykh sistem* [Conceptual analysis of computer systems]. Saint-Petersburg: Pub. LIAP. 1991. 78 p. (In Russ.).
6. Perovskaia E.I. *Osnovy gibkoi avtomatizatsii* [Basics of flexible automation.]. Leningrad: Pub. LIAP. 1986. 32 p. (In Russ.).
7. Mesarovich M., Mako D., Takahara Ya. *Teoriia ierarhicheskikh mnogourovnevnykh sistem* [The theory of hierarchical multilevel systems]. Moscow: Pub. Mir. 1973. 344 p. (In Russ.).
8. Henderson-Sellers B., Ralyté J. Situational Method Engineering: State-of-the-Art Review. *J. of Universal Computer Science*. 2010. vol. 16. no. 3. pp. 424–478.
9. Gericke A., Fill H.-G., Karagiannis D., Winter R. Situational Method Engineering for Governance, Risk and Compliance Information Systems. Proceedings of the 4th Int. Conf. on Design Science Research in Information Systems and Technology: ACM Press. 2009. pp. 24.
10. Popovich V.V., Prokaev A.N., Sorokin R.P., Smirnova O.V. [On recognition of the situation, based on artificial intelligence technology]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2008. vol. 7. pp. 93–104. (In Russ.).
11. Endsley M.R. Theoretical underpinnings of Situation Awareness: a critical review. *Situation Awareness Analysis and Measurement*. 2000. pp. 3–32.
12. Endsley M.R. Final Reflections: Situation Awareness Models and Measures. *Cognitive Engineering and Decision Making*. 2015. vol. 9. no. 1. pp. 101–111.
13. Steinberg A.N., Bowman C.L. Revisions to the JDL data fusion model. *Handbook of Multisensor Data*. 2009. pp. 45–68.
14. Steinberg A.N. Foundations of Situation and Threat Assessment. *Handbook of Multisensor Data*. 2009. pp. 437–502.
15. Steinberg A.N., Bowman C.L.. Rethinking the JDL data fusion model. Proceedings of the MSS National Symposium on Sensor and Data Fusion. 2004. vol. 1. pp. 79–97.
16. Bowman C.L. The Dual Node Network (DNN) Data Fusion & Resource Management (DF&RM) Architecture. Proceedings of AIAA Intelligent Systems Conference. 2004. vol. 50. p. 134.
17. Fridman A.Ya., Kurbanov V.G. [Formal Conceptual Model of Industry-Natural Complexes as a Mean to Organize Computing Experiments]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2014. vol. 6(37). pp. 424–453. (In Russ.).
18. Fridman A.Ya., Kurbanov V.G. [Situational Modeling of Reliability and Safety for Industry-Natural Systems]. *Informatsionno-upravliaushchie sistemy — Information and Control Systems*. 2014. vol. 4(71). pp. 68–77. (In Russ.).
19. Sagidova M.L., Fridman A.Ya. [The implementation of a GIS interface and the conceptual model of situational system of computational experiment]. *Modeli sotsial'nykh, tekhnologicheskikh i obrazovatel'nykh protsessov – Models of social, technological and educational processes*. 2001. pp. 93–98. (In Russ.).

Фридман Александр Яковлевич — д-р техн. наук, профессор, ведущий научный сотрудник лаборатории информационных технологий управления промышленно-природными системами, Институт информатики и математического моделирования

технологических процессов Кольского научного центра Российской академии наук. Область научных интересов: моделирование комплексных технологий и их воздействия на окружающую среду, прикладные интеллектуализированные системы. Число научных публикаций — 260. fridman@iimm.kolasc.net.ru; 184209, Мурманская обл., Апатиты, ул.Ферсмана, 24а; р.т.: +7(81555) 79782.

Fridman Alexander Jakovlevich — Ph.D., Dr. Sci., professor, leading researcher of information technologies for nature-industrial systems control laboratory, Institute for Informatics and Mathematical Modelling of Technological Processes of the Kola Science Center of Russian Academy of Sciences. Research interests: modeling of complex technologies and their impact on the environment, application intellectualized system. The number of publications — 260. fridman@iimm.kolasc.net.ru; 24A, Fersman st., 184209, Apatity, Murmansk Region, Russian Federation; office phone: +7(81555) 79782.

Курбанов Вугар Гариб оглы — доцент, старший научный сотрудник, старший научный сотрудник, Институт проблем машиноведения Российской Академии Наук (ИПМАШ РАН). Область научных интересов: общие вопросы математики, математическая кибернетика, математическое моделирование, оптимальные системы. Число научных публикаций — 70. vugar_borchali@yahoo.com; Большой пр. В.О., 61, Санкт - Петербург, 196135; р.т.: +7(812)3219007.

Kurbanov Vugar Garib ogli — Ph.D., associate professor, senior researcher, Institute of Problems in Mechanical Engineering of Russian Academy of Sciences (IPME RAS). Research interests: general questions of mathematical cybernetics, mathematical modeling, optimum systems. The number of publications — 70. vugar_borchali@yahoo.com; 61, Bol'shoi av. V.O., Saint-Petersburg, 199178, Russian Federation; office phone: +7(812)3219007.

Поддержка исследований. Работа выполнена при поддержке РФФИ (проекты №№ 14-07-00256-а, 14-07-00257-а, 14-07-00205-а, 13-07-00318-а)

Acknowledgements. This research is supported by RFBR (grants 15-07-04760, 15-07-02757, 16-29-04424, 16-29-12901)

РЕФЕРАТ

Фридман А.Я., Курбанов В.Г. **Информационная технология интеграции пространственных данных в ситуационную систему моделирования.**

Настоящая публикация продолжает тематику ситуационного моделирования динамических пространственных объектов, в частности промышленно-природных комплексов (ППК), изложенную в предшествующих работах авторов. Ядром системы моделирования является иерархическая концептуальная модель (ИКМ) ППК, которая применяется не только для фиксации структуры объекта и статического контроля взаимосвязей его компонентов, но и как основное средство обработки ситуаций на всех этапах моделирования. На этапе постановки задачи моделирования ИКМ управляет выбором релевантных исходных данных, участвует в процедурах пополнения и доопределения ситуаций. При реализации имитационного режима ИКМ диспетчирует вызов программных модулей и управление потоками данных. По завершении расчетов эта модель организует распределение результатов по модулям хранения и представление результатов моделирования. Такой подход обеспечивает автоматизацию всех этапов вычислительного эксперимента с возможностью равноправной обработки информации от расчетных модулей, имитирующих составные части ППК, и интегрированных с ней ГИС и экспертной системы. Особенности предложенного подхода состоят в широком применении экспертных знаний, поддержке современных сценарных подходов к моделированию. В данной статье представлены основные принципы организации взаимодействия геоинформационной системы с другими составными частями системы ситуационного моделирования. При этом ГИС используется не только для визуализации результатов моделирования, но является полнофункциональной управляемой подсистемой, то есть обеспечивает оперативное измерение графических характеристик элементов модели, выдачу их в расчетные модули и модификацию по результатам расчетов. Кроме того, ГИС участвует в постановке задачи моделирования и контроле корректности модели.

SUMMARY

Fridman A.Ja., Kurbanov V.G. **Information Technology to Integrate Spatial Data into the Situational Modelling System.**

This paper continues to describe the situational approach to modelling of dynamic spatial objects, in particular industry-nature complexes (INC), introduced in earlier publications of the authors. A hierarchical conceptual model (ICM) of INC is at the core of our modelling system. This model not only formalizes the INC structure and implements static control of interconnections among the INC elements, but also provides situations processing during all stages of modelling. At the stage of modelling task setting, ICM controls the choice of relevant source data, takes part in completion of situations. During simulation, ICM dispatches invoking of program modules and controls data flows. When calculations are over, ICM manages distribution of the calculated results among storages and provides for their visualization. The described approach to modelling provides for automation of every modelling stage and for equitable processing of the information from calculating modules which simulate some INC components, embedded geographic information (GIS) and expert systems. Distinctive features of this approach include the wide use of expert knowledge and support of modern scenario-based implementation of modelling.

The given paper introduces the main principles of GIS interaction with other components of the situational modelling system. Besides ensuring visualization of modelling results, the GIS also functions as a full-scale controlled subsystem. It provides for the operative measurement of graphical parameters of model elements, transfers them to calculating modules and modifies them according to results of calculations. Above this, the GIS takes part in the setting of modelling tasks and in the correctness control of the whole model.

В.В. БУРАКОВ, В.Ф. ВОЛКОВ, С.А. ПОТРЯСАЕВ, В.И. САЛУХОВ,
Н.А. ШЕДЬКО
**ФОРМАЛИЗОВАННЫЕ ПОДХОДЫ К ИССЛЕДОВАНИЮ
ТОЧНОСТИ МОДЕЛЕЙ ЭКОНОМИЧЕСКИХ СИСТЕМ**

Бураков В.В., Волков В.Ф., Потрысаев С.А., Салухов В.И., Шедько Н.А. **Формализованные подходы к исследованию точности моделей экономических систем.**

Аннотация. В статье на примерах наиболее распространенных на микро- и макроуровнях экономико-математических моделей предложен подход к построению методики «оптимальной сложности», обеспечивающей минимальную величину суммарной ошибки при заданной продолжительности решения информационно-расчетных задач в рамках модельных исследований организационно-экономических систем. Кроме того, данный подход позволяет обосновать требования к точности входной информации. Показано, что для обеспечения рационального уровня точности моделирования орган управления (заказчик модели) должен учитывать складывающиеся в реальности соотношения точности исходной информации, структурной точности модели, функциональной точности модели и точности вычислительных алгоритмов.

Ключевые слова: точность исходной информации, оптимальная сложность модели, метод линеаризации, ошибки расчета, методические ошибки, результативность операции.

Burakov V.V., Volkov V.F., Potrysaev S.A., Salukhov V.I., Shedko N.A. **Formalized Approaches to the Study of the Accuracy of Economic Systems Models.**

Abstract. In the article, we propose an approach to the creation of a method of optimal complexity through examples of economic and mathematical models that are most common at micro - and macro-levels. This method minimizes the total error for given duration of solving information and calculation tasks in the model studies of organizational-economic systems. In addition, this approach allows us to substantiate the requirements for the accuracy of input information. It is shown that to ensure the sustainable level of simulation accuracy, the authority (a model customer) needs to consider real relations of the accuracy of initial information, the structural accuracy of the model, the functional model accuracy and the accuracy of numerical algorithms.

Keywords: initial information accuracy; optimal complexity; linearization method; computational error; procedural error; operation effectiveness.

1. Введение. Анализ литературы показывает, что известные постановки задач обоснования экономических управленческих решений базируются на допущении о наличии всех необходимых исходных данных. Однако, как на этапе проектирования, так и на этапах планирования и, тем более, в повседневной управленческой деятельности большая часть исходных данных, необходимых для вычисления соответствующих показателей и описания сложившейся рыночной ситуации, известна либо приближенно, с определенными ошибками, либо указывается в некотором диапазоне. Сами показатели (рейтинги, индексы) на практике также определяются с некоторыми ошибками, обусловленными погрешностями численных

алгоритмов и вычислений, и, кроме того, они во многом зависят от «неколичественной» информации, например, корректности формулировок вопросов при маркетинговых исследованиях.

Следовательно, возникают вопросы: во-первых, всегда ли целесообразно реализовывать «абсолютно точные» экономико-математические методы и модели, ведь наличие ошибок в исходных данных предопределяет ошибки и в расчете показателей эффективности, которые могут «свести на нет» выигрыш от оптимизации; во-вторых, как выявить априори, не решая в строгой постановке задачу оптимизации (по виду матриц с исходной информацией), ситуацию, в которой целесообразнее реализовать упрощенные оценочные алгоритмы, т. к. они не связаны со своеобразной платой за оптимальность решения – большими вычислительными затратами и расходами на сбор информации и дают ту же точность расчетов показателей выходного эффекта. Судя по известной литературе, ранее эти вопросы различными исследователями в области экономико-математического моделирования (ЭММ) не ставились. Вместе с тем необходимо отметить работу А.Е. Кобринского [2], являющуюся одной из первых в области точности ЭММ, монографию Р.М. Юсупова, Е.Н. Розенвассера [1], в которой рассмотрены методы оценивания чувствительности решений оптимизационных задач и исследования В.П. Заболотского [3], связанные с оцениванием точности прогнозов затрат в сфере ВПК.

Отдельным вопросам названной проблемы посвящено большое количество научных работ. В частности, в трудах Г.П. Фомина, Г.Б. Петухова, В.В. Глухова [5, 6, 8] рассматриваются особенности применения математического инструментария для исследования организационных и экономических систем. В публикациях А.С. Елисеева, М.Б. Гитмана, И.А. Максимцева, Б.И. Герасимова [12, 13, 16] анализируются вопросы обеспечения стабильности деятельности экономических систем, их целевых процессов и оценивания погрешностей расчетов контролируемых показателей. Известны также труды А.Д. Припадчева, А.Э. Александрова, Н.В. Якушева, И.А. Бородиновой, Л.А. Сараева, В.Н. Буре, В.В. Мазалова, М.М. Ерихова, Е.В. Карасевой, М.А. Татарникова [7, 9–11, 15], посвященные разработке подходов к повышению точности экономико-математических моделей применительно к крупным транспортным организациям. Однако изложенные в указанных и во многих других работах методы обеспечения требуемой точности моделей носят достаточно частный характер и применимы в

относительно узких предметных областях. Общие подходы к повышению точности модельных исследований организационно-экономических систем в настоящее время изучены недостаточно, отсутствуют универсальные методы и системы показателей, применимые для различных аспектов экономики и управления. Как следствие, актуальным является разработка соответствующего экономико-математического инструментария, применимого для оценивания и обеспечения необходимых точностных характеристик экономических моделей.

В общем случае можно указать три основные причины, вызывающие ошибки при моделировании экономических систем (ЭС):

- неточность входной информации;
- методические ошибки, связанные с упрощением модели и неучетом ею тех или иных факторов и характеризующиеся величиной σ_M ;
- ошибки расчетов (например, при методе статистических испытаний ввиду малости выборки, при численном интегрировании за счет величины шага, вообще за счет округлений и т.п.), характеризующиеся величиной σ_p .

Ошибка в определении некоторого интегрального экономического показателя K , обусловленная всеми независимыми ошибками этих трех групп, может оцениваться с помощью формулы:

$$\sigma_K = \sqrt{\sum_{i=1}^n \left(\frac{\partial K}{\partial A_i}\right)^2 \sigma_{A_i}^2 + \sigma_M^2 + \sigma_p^2}, \quad (1)$$

где K — показатель, определяемый при расчетах параметров модели (например, вероятность какого-то события, математическое ожидание какой-либо величины и др.); A_i — входные данные, влияющие на величину показателя; σ_{A_i} — величины среднеквадратичных ошибок в их определении.

Изменение величин составляющих суммарной ошибки в тех случаях, когда они заметно меньше остальных, не приводит к существенному изменению итоговой ошибки. Поэтому, если модель является грубой (велики σ_M и σ_p) или часть информации, вводимой в

модель, определена с большими ошибками (часть членов $\frac{\partial K}{\partial A_k} \sigma_{A_i}$ велика), ошибки при неизвестной информации могут быть определены весьма приближенно. Во всех случаях необходимо проанализировать,

в какой степени неизвестная информация сказывается на общей ошибке в определении показателя. При применении метода имитационного моделирования всегда следует проанализировать возможность упрощения экономической модели. Это увеличивает методическую ошибку, но сокращает затраты времени на получение одной реализации и тем самым при том же времени моделирования позволяет увеличить число реализаций и, следовательно, уменьшить ошибку расчета. Таким образом, можно найти «оптимальную сложность» экономической модели, обеспечивающую минимальную величину суммарной ошибки при фиксированном времени модельных исследований. Во всех случаях построения моделей следует выбирать оптимальное сочетание сложности модели (определяющей методическую ошибку), метода расчетов (определяющего ошибку расчета) и точности входной информации.

2. Оценка точности результатов экономического моделирования методами теории массового обслуживания. При оценке качества моделей экономических систем, описанных с помощью теории массового обслуживания целесообразно использовать метод линеаризации функций случайных аргументов. Он дает особенно хорошие результаты, если отклонения значений исходных данных от их истинных значений сравнительно невелики и зависимость критерия от исходных параметров на исследуемом участке близка к линейной.

В общем виде зависимости, по которым производится определение критериев качества функционирования ЭС как системы массового обслуживания (СМО), могут быть записаны следующим образом:

$$K = f(x_1, x_2, \dots, x_i, \dots, x_n),$$

где K — определяемый показатель; x_i — исходный параметр.

Формулу для определения суммарной ошибки (дисперсии) исследуемого показателя K в зависимости от дисперсий исходных величин D_{x_i} можно записать так:

$$D_k = \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i}\right)_m^2 D_{x_i} + 2 \sum_{i < j} \left(\frac{\partial f}{\partial x_i}\right)_m \left(\frac{\partial f}{\partial x_j}\right)_m k_{ij}, \quad (2)$$

где D_{x_i} — дисперсия исходной величины x_i ; k_{ij} — корреляционная функция;

— $\left(\frac{\partial f}{\partial x_i} \right)$ величина первой производной от функции $f(x_1, x_2, \dots, x_i, \dots, x_n)$ по переменной x_i .

Очень часто величины x_i и x_j не коррелированы, и тогда формула (2) значительно упрощается:

$$D_k = \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i} \right)_m^2 D_{x_i}. \quad (3)$$

Рассмотрим подходы к определению ошибок, связанных с неточностью получения исходных данных, на примерах типовых СМО, традиционно используемых при построении моделей экономических систем (транспортных, строительных, торговых, телекоммуникационных).

Проанализируем простейшую систему массового обслуживания с отказами, состоящую из n однотипных устройств, в которую поступает пуассоновский поток заявок, а время обслуживания заявок устройствами распределено по показательному закону. Вероятность P_n состояния, при котором все устройства заняты, определяется по формуле:

$$P_n = \frac{a^n}{n!} = P_{омк}, \quad (4)$$

$$\sum_{k=1}^n \frac{a^k}{k!}$$

где $a = \lambda \bar{t}_{обс}$; λ — плотность поступающих заявок; $\bar{t}_{обс}$ — среднее время, необходимое для обслуживания одной заявки.

Формула (4) справедлива для установившегося процесса. Вероятность того, что каждая заявка будет обслужена, равна:

$$P_{обс} = 1 - P_{омк}. \quad (5)$$

Основной исходной характеристикой, необходимой для расчетов, является $a = \lambda \bar{t}_{обс}$. Входящие в нее величины $\bar{t}_{обс}$ и λ на практике определяются с ошибками, причем и сами величины в течение исследуемого интервала времени могут принимать различные значения в пределах некоторой области. На величину параметров λ и $\bar{t}_{обс}$ влияет большое количество различных факторов, анализ работы отдельных систем массового обслуживания показывает, что часто эти

параметры подчиняются нормальному закону распределения. Рассмотрим, как влияют ошибки в определении λ и $\bar{t}_{обс}$ на точность получения значения вероятности $P_{отк}$. Если отклонения величин λ и $\bar{t}_{обс}$ сравнительно малы, то можно воспользоваться методом линеаризации функции случайных аргументов. Тогда:

$$D_{P_n} = \left(\frac{\partial P_n}{\partial \lambda}\right)^2 D_\lambda + \left(\frac{\partial P_n}{\partial \bar{t}_{обс}}\right)^2 D_{\bar{t}_{обс}}, \quad (6)$$

где D_λ , $D_{\bar{t}_{обс}}$ — дисперсии величин λ и $\bar{t}_{обс}$, $\left(\frac{\partial P_n}{\partial \lambda}\right)$, $\left(\frac{\partial P_n}{\partial \bar{t}_{обс}}\right)$ — частные производные от функции $P_n = f(\lambda, \bar{t}_{обс})$ по аргументам λ и $\bar{t}_{обс}$.

Функция $P_n = f(\lambda, \bar{t}_{обс})$ непрерывна и дифференцируема. Вначале целесообразно определить производную от этой функции по аргументу a :

$$\frac{\partial P_n}{\partial a} = \frac{n-a}{a} P_n + P_n^2. \quad (7)$$

Отсюда дисперсия D_{P_n} равна:

$$D_{P_n} = \left(\frac{n-a}{a} P_n + P_n^2\right)^2 \bar{t}_{обс}^2 \sigma_\lambda^2 + \left(\frac{n-a}{a} P_n + P_n^2\right)^2 \lambda^2 \sigma_{\bar{t}_{обс}}^2. \quad (8)$$

Обозначим $\frac{n-a}{a} P_n + P_n^2 = A$; с увеличением коэффициента A увеличивается и дисперсия D_{P_n} . Поэтому представляет интерес найти его экстремальное значение. Необходимое условие экстремума определяется выражением:

$$\frac{\partial A}{\partial a} = -\frac{n}{a^2} P_n + \frac{n-a}{a} \left(\frac{n-a}{a} P_n + P_n^2\right) + 2P_n \left(\frac{n-a}{a} P_n + P_n^2\right) = 0. \quad (9)$$

После преобразований получаем значения a^* , для которых функция $A=A(a)$ достигает экстремума:

$$a_{1,2}^* = \frac{-n(3P_n - 2) \pm \sqrt{n^2(3P_n - 2)^2 - 4(2P_n^2 - 3P_n + 1)n(n-1)}}{2(2P_n^2 - 3P_n + 1)}. \quad (10)$$

Как видно из последнего равенства, экстремум функции $A = A(a)$ определяется выражением, заданным в неявной форме. Можно показать, что величина относительной ошибки равна:

$$\frac{\Delta P_{\text{обсмакс}}}{P_{\text{обс}}} = 3 \left(\frac{n-a}{a} P_n + P_n^2 \right) \frac{Ra}{1-P_n}. \quad (11)$$

Если известны σ_λ и $\sigma_{\bar{r}_{\text{обс}}}$, то не представляет большого труда оценивать ошибки в определении показателя $P_{\text{обс}}$ по формуле (8). По аналогичной схеме могут быть выведены выражения для расчета D_{P_n} для других типов СМО экономической направленности (системы с ожиданием; системы с ограниченным временем ожидания и т.п.).

3. Точность оценивания эффективности сложных организационно-технических систем. В основу исследования «экономической эффективности» создаваемой организационно-технической системы (ОТС) обычно закладывается анализ соотношения «эффективность — стоимость» [8]. Однако чаще всего при этом происходит подмена понятий. В термин «эффективность» разные авторы вкладывают различный смысл. У одних — это «полезность», у других — «прибыль», у третьих «наносимый конкуренту или предотвращенный собственный ущерб». В работе [5] показано, что перечисленные показатели описывают одно и тоже свойство операций — результативность или способность давать целевой (полезный) эффект. Поэтому авторы, выбирающие перечисленные показатели эффективности, на самом деле анализируют проблему не «эффективность-стоимость», а проблему «эффект-стоимость». Для решения этой проблемы они же обычно используют отношения количественных мер упомянутых показателей целевых эффектов к стоимостям соответствующих проектов. Это решение не корректно по той причине, что даже найденный таким способом лучший вариант выполнения проекта не гарантируют достижения цели по экономическим показателям, так как в критерии отсутствует значение допустимых затрат.

В работе [5] предложен подход, согласно которому наиболее объективной характеристикой проекта является вероятность достижения цели $P_{\text{дл}}$. Этот показатель определяется соотношением фактических и требуемых значений таких частных показателей, как результативность,

ресурсоемкость, оперативность. Тогда проблема «эффект-стоимость» решается «автоматически», за счет построения зависимости $P_{\text{дц}}$ от компонентов векторов возможных $Y_{\langle 3 \rangle} = \langle \hat{g}, \hat{c}, \hat{t} \rangle$ и требуемых $Z_{\langle 3 \rangle} = \langle \hat{g}T, \hat{c}d, \hat{t}d \rangle$ характеристик проекта (расходы ресурсов всегда имеют стоимостной эквивалент \hat{c} и \hat{c}_d). Следовательно, проблема «эффективность-стоимость» в рамках создания или оценивания (ОТС) возникает при обосновании требуемого значения показателя эффективности ($P_{\text{дц}}^{\text{ТР}}$) и необходимого числа ее элементов $N_{\text{ТС}}$.

В свою очередь, проблема обоснования требуемого уровня эффективности функционирования экономической системы должна решаться в зависимости от типов проектов, в которых она участвует. Так, если проект имеет уникальный и бескомпромиссный характер, то в качестве требуемого уровня эффективности должен выступать максимально возможный уровень, обеспечивший принцип практической уверенности (оцениваемой, например, через вероятность практически достоверного события) и средства, выделяемые на подготовку такого проекта, должны соответствовать этому значению эффективности. В том случае, когда проекты носят «массовый характер», возможен компромисс между требуемой эффективностью отдельного из них и требуемой эффективностью экономической системы в целом, когда снижение требуемой эффективности отдельного проекта компенсируется большим количеством других выполняемых проектов. В этом случае обоснование требуемого уровня эффективности деятельности системы заключается в поиске такого его значения, увеличение которого уже не целесообразно с экономических позиций (выгоднее просто увеличить количество проектов и структурных подразделений ЭС, участвующих в их выполнении). Например, пусть стоимость обеспечения требуемой эффективности функционирования структурного элемента экономической системы определяется зависимостью:

$$C = EXP(A / (1 - P_{\text{дц}}^{\text{ТР}})) + B), \quad (12)$$

где A и B — статистические коэффициенты, $P_{\text{дц}}^{\text{ТР}}$ — требуемое значение показателя эффективности [3-5].

При повышении требований к уровню эффективности $P_{\text{дц}}^{\text{ТР}}$ на величину ΔP , с одной стороны, экономическая система понесет расходы в размере:

$$\Delta C_p = EXP\left(\frac{A}{1 - P_{ДЦ1}^{TP} - \Delta P} + B\right) - EXP\left(\frac{A}{1 - P_{ДЦ1}^{TP}} + B\right), \quad (13)$$

а, с другой стороны, она получит выгоду от возможности сократить количество участвующих в проекте таких элементов в объеме:

$$\Delta C_N = \bar{C}_1 \left[\frac{\ln(1 - P_{(ДЦ)}^{TP})}{\ln(1 - P_{(ДЦ1)}^{TP})} - \frac{\ln(1 - P_{(ДЦ)}^{TP})}{\ln(1 - P_{(ДЦ1)}^{TP} - \Delta P)} \right], \quad (14)$$

где \bar{C}_1 — средняя стоимость использования одного структурного элемента ЭС.

Тогда (рисунок 1), построив графики функций $\Delta C_p(\Delta P)$ и $\Delta C_N(\Delta P)$ и, найдя точку ΔP , в которой $\Delta C_p(\Delta P) = \Delta C_N(\Delta P)$, мы можем определить ΔP^{opt} , откуда $P_{ДЦ1}^{TP} = (P_{ДЦ1}^{TP})^0 + \Delta P^{opt}$.

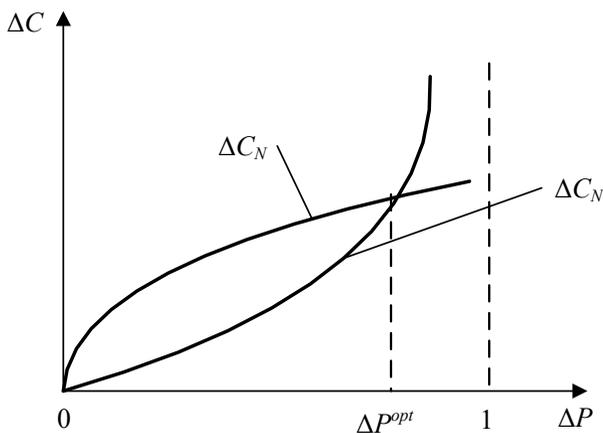


Рис. 1. Зависимости выигрыша (потерь) системы от величины ΔP

4. Точность исходной информации в иерархических моделях экономических систем. Как известно, по месту на иерархической лестнице модели экономических систем можно разделить на три класса [3-17]:

— модели, в которых рассматривается взаимодействие

структурных элементов разных видов, имеющих коренное различие в функционировании, например, производственные комплексы разных направлений деятельности или даже различные секторы экономики (первый класс моделей);

— модели, в которых рассматривается взаимодействие структурных элементов одного вида (т. е. не имеющих коренного различия в функционировании), но разных типов, например, производственные комплексы одного направления деятельности, но применяющие разные технологии (второй класс);

— модели, в которых рассматривается взаимодействие структурных элементов одного вида и одного типа (третий класс).

Особенности экономических моделей первого класса:

— невозможность полной замены одного вида структурных элементов другими и зачастую невозможность функционирования одного их вида без других, например, выполнение крупных национальных проектов, где взаимодействуют разные виды производственных комплексов, причем недостаток одних из них может быть только частично компенсирован другими при обязательном согласовании деятельности; вследствие этого существенной частью таких моделей является наличие большого числа логических связей;

— большой размер таких моделей не позволяет, как правило, рассматривать все единицы всех типов структурных элементов каждого вида, вследствие чего в рассмотрение вводится обобщенный элемент (или множество однотипных элементов) каждого вида, выбранное оптимальным образом; поэтому при исследовании таких моделей необходимо пользоваться информацией, получаемой из моделей классов, в которых должны быть выбраны оптимальные сочетания оптимальных элементов каждого типа;

— модели этого класса, как правило, предназначаются для решения распределительных задач; в них выбираются оптимальные сочетания структурных элементов разных видов, производится распределение задач, решаемых элементами разных видов.

Особенности экономических моделей второго класса:

— действие структурных элементов одного вида, но разных типов, которые выполняют не все задачи крупного проекта, а только часть их; поэтому для исследования моделей этого класса необходима информация об объеме задач, решаемых элементами данного вида, получаемая из моделей первого класса;

— при исследовании моделей этого класса предполагается, что структурные элементы каждого типа обладают оптимальными

характеристиками, поэтому для их решения необходима информация, получаемая из моделей третьего класса;

— в задачах исследования моделей этого класса, как правило, определяются основные характеристики структурных элементов рассматриваемого вида и оптимальное количество типов элементов данного вида, поэтому такие задачи иногда называют задачами выбора оптимальных шкал значений параметров; в этих задачах определяются и объемы работ проекта, выполняемых структурными элементами каждого типа.

Особенности экономических моделей третьего класса:

— деятельность структурных элементов одного типа и одного вида, которые выполняют часть задач экономической системы; вследствие этого для их исследования необходима информация об объеме задач, получаемая из моделей второго класса;

— в моделях этого класса определяются оптимальные параметры структурных элементов каждого типа при заданных основных характеристиках их (собственно, и определяющих тип); задачи исследования таких моделей, по существу, являются обобщенными задачами оптимального проектирования.

Из изложенного выше видна связь между моделями всех классов и необходимость одновременного исследования всей «лестницы» моделей, что практически осуществить весьма сложно. Ранее в работах [3-17] предлагался следующий подход для задач минимизации затрат при фиксированном критерии эффективности.

1. Принимается допущение о том, что функции затрат и критерия эффективности дважды непрерывно дифференцируемы по всем аргументам.

2. Предполагается, что относительное изменение показателя эффективности приближенно пропорционально соответствующим относительным изменениям затрат.

3. Оценивается относительное изменение общих затрат $\frac{\Delta C}{C}$,

вызванное неточностью:

$$\frac{\Delta C}{C} = \frac{D^{n+1}}{2^{n+1}}, \quad (15)$$

где D — характеристика точности (дисперсия) приближенного решения задачи оптимизации; n — модуль разности между номерами классов той задачи, в которой допущены неточности, и рассматриваемой. Так, например, если оптимальные значения

характеристик отдельных структурных элементов экономической системы определены в среднем с ошибкой в 45%, то в задаче этого класса относительное изменение общих затрат составит 10%, в задачах следующего класса относительное изменение общих затрат составит 0,5%, в задачах через один класс относительное изменение затрат составит 0,1%. Этот вывод позволяет расчленить решение полной задачи оптимизации на ряд отдельных. Рассмотрим этот же вопрос с других позиций. В тех случаях, когда функционирование рассматриваемой ОТС входит в более крупную систему целенаправленных действий (например, национальный проект), состоящих из множества альтернатив, образующих ветвящийся процесс достижения более глобальной цели, то каждая альтернативная ветвь такого процесса состоит из нескольких последовательных этапов и глобальный показатель эффективности (ПЭ) W в этом случае может быть представлен в следующем виде:

$$W = \sum_{\{m\}} P_{\text{ПР}m} \cdot P_{\text{ДЦ}m}, P_{\text{ДЦ}m} = P_{\text{ДЦ}m1} \cdot P_{\text{ДЦ}m2/1} \cdot \dots \cdot P_{\text{ДЦ}mi/j} \cdot \dots \cdot P_{\text{ДЦ}mn/n-1}, \quad (16)$$

где $P_{\text{ДЦ}m1}$ — вероятность достижения цели 1-го этапа m -й альтернативы целенаправленных действий; $P_{\text{ДЦ}mi/j}$ — условная вероятность достижения цели i -го этапа при условии достижения цели на j -м этапе m -й альтернативы целенаправленных действий; $P_{\text{ПР}m}$ — вероятность принятия решения о выборе m -й альтернативы целенаправленных действий; $\{m\}, \{n\}$ — множество альтернатив целенаправленных действий для достижения глобальной цели и множество этапов в каждой альтернативе, соответственно.

В более общей постановке математически модель показателя может быть обобщена в следующем виде:

$$P_{\text{ДЦ}} = F_p(A_{\langle s \rangle}), \quad (17)$$

где $A_{\langle s \rangle}$ — вектор параметров математической модели ПЭ; F_p — некоторая функция, описывающая зависимость ПЭ от вектора $A_{\langle s \rangle}$.

Используя известное выражение из теории погрешностей, можно записать:

$$\Delta F_p(a_1, a_2, \dots, a_s) = \sum_{i=1}^s \Delta a_i \left| \frac{dF_p(a_1, a_2, \dots, a_s)}{da_i} \right|, \quad (18)$$

где a_i — значение i -го параметра, которое является исходными

данными для вычисления ПЭ целенаправленной деятельности; Δa_i — погрешность в задании значения i -го параметра.

В рассматриваемом случае функция F_P имеет вид (18). Если допустить, что требования к относительной погрешности исходных данных являются одинаковыми для всех альтернатив, этапов и параметров экономической модели, то эти требования могут быть определены последовательно по следующим зависимостям:

$$P_{ДЦ_m} = \frac{\Delta W}{m}, \Delta P_1 = \frac{\Delta P_{ДЦ_m}}{(n-1) \prod_{i=2}^n P_{ДЦ_{i|j}}}, \quad (19)$$

$$\Delta P_{ДЦ_{i|j}} = \frac{\Delta P_{ДЦ_m}}{(n-1) P_1 \prod_{\substack{k=2, l=2 \\ i \neq k, j \neq l}}^n P_{ДЦ_{k|l}}}, \Delta a_i = \frac{\Delta P_{ДЦ_m}}{\sum_{i=1}^S |F'_P(a_1, a_2, \dots, a_S)|}$$

Анализ выражений (18) и (19) показывает, что точность исходных данных, обеспечивающую необходимую точность вычисления ПЭ, нужно согласовывать с количеством этапов реализации альтернативной последовательности действий экономической системы, которые собираются учитывать исследователь. С другой стороны, выражения (18) и (19) позволяют обосновать количество этапов, которые можно включить в модель показателя эффективности деятельности ЭС, при заранее известной точности исходных данных. Покажем это на отдельных примерах.

Пусть необходимо предъявить требования к точности исходных данных для оценивания эффективности функционирования

экономической системы с точностью $\Delta P_{ДЦ} \approx 0.1$ для модели ПЭ, имеющей вид $P_{ДЦ} = P_1(a_1) \cdot P_2(a_2) \cdot P_3(a_3)$. Из выше изложенного следует, что:

$$\Delta P_1 = \Delta P_2 = \Delta P_3 = 0.033, \\ \Delta a_1 = P_1^{-1}(0.033), \Delta a_2 = P_2^{-1}(0.033), \Delta a_3 = P_3^{-1}(0.033) (P_i^{-1}$$

— известные обратные функции).

Пусть необходимо решить, сколько последовательных этапов проекта можно учесть в модели ЭС, если известно, что точности параметров Δa_i , определяющих вероятность выполнения каждого из этапов, равны 0.02, а показатель эффективности требуется оценить с

точностью не ниже 0.1. Из вышеизложенного следует, что $P_1(0.02)P_2(0.02)\dots P_n(0.02)=0.1$.

Если предположить, что

$$P_1(0.02) = P_2(0.02) = \dots = P_n(0.02), \quad n = \frac{0.1}{P^{-1}(0.02)} \text{ например,}$$

если $P^{-1}(0.02) = 0.02$, то $n=5$, т.е. в модели ЭС целесообразно учесть 5 этапов.

5. Заключение. В статье на примерах наиболее распространенных на микро- и макро уровнях ЭММ обоснован подход к построению методики «оптимальной сложности», обеспечивающей минимальную величину суммарной ошибки при заданной продолжительности решения информационно-расчетных задач в рамках модельных исследований экономических систем. Кроме того, данный подход позволяет обосновать требования к точности входной информации.

Если методическая ошибка возникает вследствие невозможности описать влияние каких-либо факторов аналитическими зависимостями, то для контроля модели и определения или хотя бы оценивания методической ошибки неизбежно использование каких-либо других методов (например, оценивания при крайних допущениях). Вместе с тем следует отметить, что если определяющей ошибкой является погрешность из-за неточности определения входной информации, то стремиться к уменьшению остальных составляющих нет смысла.

Таким образом, для обеспечения рационального уровня точности моделирования ЛППР должен учитывать складывающееся в реальности соотношение точности исходной информации, структурной точности модели, функциональной точности модели и точности вычислительных алгоритмов.

Литература

1. Юсулов Р.М., Розенвассер Е.Н. Чувствительность систем управления // М.: Наука. 1981. 464 с.
2. Кобринский А.Е. Точность экономико-математических моделей // М.: Финансы и статистика. 1981. 255 с.
3. Заболотский В.П., Оводенко А.В. Математические модели в управлении // СПб: ГУАП. 2003. 196 с.
4. Фомин Г.П. Экономико-математические методы и модели в коммерческой деятельности // М.: Финансы и статистика. 2014. 464 с.
5. Петухов Г.Б., Якунин В.И. Методологические основы внешнего проектирования целенаправленных процессов и целеустремленных систем // М.: АСТ. 2006. 502 с.
6. Припадчев А.Д. Математическое моделирование структуры парка воздушных судов на основе симплекс-метода // Научно-технические ведомости СПбГПУ. 2010. №2(97). С. 117–120.

7. Глухов В.В., Медников М.Д., Коробко С.Б. Математические методы и модели для менеджмента // СПб: Лань. 2007. 528 с.
8. Александров А.Э., Якушев Н.В. Стохастическая постановка динамической транспортной задачи с задержками с учетом разброса времени доставки // Управление большими системами. 2006. №12–13. С. 5–14.
9. Бородинова И.А., Сараев Л.А. Стохастические транспортные задачи // Вестник Самарского ГУ. 2010. №7(81). С.1–20.
10. Буре В.Н., Мазалов В.В. Вычисление пассажиропотоков в транспортных системах // Управление большими системами. 2014. №47. С.77–91.
11. Елисеев А.С., Гитман М.Б. Оценка устойчивости производственного плана с учетом стохастичности ресурсных ограничений // Управление большими системами. 2013. №42. С.252–272.
12. Максимцева И.А. Основы наукоемкой экономики (Знания, Креативность, Инновации) // М.: Лань. 2011. 456 с.
13. Ерихов М.М., Карасева Е.В., Татарников М.А. Оптимизация маршрутных расписаний городского пассажирского транспорта // Научно-технические ведомости СПбГПУ. 2008. №3(56). С.121–124.
14. Герасимов Б.И. Модели погрешностей оценки качества: монография // Тамбов: ТГТУ. 2009. 180 с.
15. Максимова Т.Г., Шаныгин С.И. Учет устойчивости при моделировании организационных систем // Социально-экономические проблемы развития современного общества: сборник научных трудов. СПб.: СПбТЭИ 2005. С. 51–52.
16. Калинин В.В. Модель Альтмана на примере условной компании «Бизнес» // Консультант. 2006. №19. С.19–24.
17. Крюков А.Ф. Анализ методик прогнозирования критических ситуаций коммерческих организаций с использованием финансовых индикаторов // Менеджмент в России и за рубежом. 2001. №2. С. 25–34.

References

1. Yusupov R.M., Rozenvasser E.N. *Chuvstvitel'nost' sistem upravleniya* [Sensivity of management systems]. Moscow: Nauka. 1981. 464 p. (In Russ.).
2. Kobrinsky A.E. *Tochnost' jekonomiko-matematicheskikh modelej* [Accuracy of economical and mathematical models]. Moscow: Finansi i statistika. 1981. 255 p. (In Russ.).
3. Zabolotsky V.P., Ovodenko A.V. *Matematicheskie modeli v upravlenii* [Mathematical models in management]. St.Petersburg: GUAP. 2003. 196 p. (In Russ.).
4. Fomin G.P. *Jekonomiko-matematicheskie metody i modeli v kommercheskoj dejatel'nosti* [Economic and mathematical methods and models in commercial activity]. Moscow: Finansi i Statistika. 2014. 464 p. (In Russ.).
5. Petuhov G.B., Jakunin V.I. *Metodologicheskie osnovy vneshnego proektirovaniya celenapravlennyh processov i celeustremlynyh sistem* [Metodological basics of purposeful processes and systems external projecting]. Moscow. AST Publ. 2006. 502 p. (In Russ.).
6. Pripadchev A.D. [Mathematical modeling of aircraft park based on simplex-method]. *Nauchno-technicheskie vedomosti SPbGPU – St. Petersburg State Polytechnical University Journal*. 2010. vol. 2(97). pp. 117–120. (In Russ.).
7. Gluhov V.V., Mednikov M.D., Korobko S.B. *Matematicheskie metody i modeli dlja menedzhmenta* [Mathematical methods and models for management]. St.Petersburg: Lan'. 2007. 528 p. (In Russ.).
8. Aleksandrov A.E., Yakushev N.V. [Stochastic estimation of dynamic transport task with delays taking into account the spread of delivery time]. *Upravlenie bolshimi*

- systemami – Large-scale Systems Control*. Moscow. IPURAN. 2006. vol. 12–13. pp. 5–14. (In Russ.).
9. Borodinova I.A., Saraev L.A. [Stochastic transport tasks]. *Vestnik Samarskogo GU – Vestnik of Samara State University*. 2010. vol. 7(81). pp. 1–20. (In Russ.).
 10. Bure V.N., Mazalov V.V. [Calculation of passenger traffic in transport systems]. *Upravlenie bolshimi systemami – Large-scale Systems Control*. Moscow. IPU RAN. 2014. vol. 47. pp. 77–91. (In Russ.).
 11. Eliseev A.S., Gitman M.B. [Evaluation of the production plan sustainability taking into account stochastics of resource constraints] *Upravlenie bolshimi systemami – Large-scale Systems Control*. Moscow. IPU RAN. 2013. vol. 42. pp. 252–272. (In Russ.).
 12. Maksimtsev I.A. *Osnovy naukoemkoj jekonomiki (Znaniya, Kreativnost', Innovacii)* [Basics of knowledge-based economy (Knowledge, Creativity, Innovation)]. Moscow: Lan'. 2011. 456 p. (In Russ.).
 13. Erihov M.M., Karaseva E.V., Tatarnikov M.A. [Route optimization schedules of urban transport] *Nauchno-technicheskie vedomosti SPbGPU – St. Petersburg State Polytechnical University Journal*. 2008. vol. 3(56). pp. 121–124. (In Russ.).
 14. Gerasimov B.I. *Modeli pogreshnostej ocenki kachestva: monografija* [Error model quality evaluation: monography]. Tambov: TTGU. 2009. 180 p. (In Russ.).
 15. Maksimova T.G., Shanigin S.I. [Keeping stability in the modeling of organizational systems] *Sotsialno-ekonomicheskie problem sovremennogo obschestva. Sb. nauch. tr. – Social and economic problems of modern society: collected papers*. St. Petersburg: SPbTEI. 2005. pp.51–52. (In Russ.).
 16. Kalinin V.V [Altman model example of conditional company “Biznes”] *Konsultant – Consultant*. 2006. vol. 19. pp. 19–24. (In Russ.).
 17. Krukov A.F. [Analysis of critical situations forecasting methods in commercial organizations with using of financial indicators] *Management v Rossii i za rubezom – Management in Russia and Abroad*. 2001. vol. 2. pp. 25–34. (In Russ.).

Бураков Вадим Витальевич — д-р техн. наук, доцент, ведущий научный сотрудник лаборатории информационных технологий в системном анализе и моделировании, Федеральное государственное бюджетное учреждение науки Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН). Область научных интересов: методология методики оценивания качества программного обеспечения, рефакторинг программного обеспечения. Число научных публикаций — 160. Burakov@eureca.ru, <http://www.litsam.ru>; 14-я линия В.О., д. 39, Санкт-Петербург, 199178; р.т.: +7(812)328-0103, Факс: +7(812) 328-4450.

Burakov Vadim Vital'evich — Ph.D., Dr. Sci., associate professor, leading researcher of information technologies in system analysis and modeling laboratory, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS). Research interests: technology of quality management software tools, software engineering, evaluation, and quality improvement programs. The number of publications — 160. Burakov@eureca.ru, <http://www.litsam.ru>; 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone: +7(812)328-0103, Fax: +7(812) 328-4450.

Волков Валерий Федорович — д-р воен. наук, профессор, профессор кафедры системного анализа и управления, Военно-космическая академия имени А.Ф. Можайского. Область научных интересов: теория эффективности, исследование операций, системный анализ. Число научных публикаций — 250. valfedvolkov@gmail.com; ул. Ждановская 13, Санкт-Петербург, 197198; р.т.: +7 951 652 32 63.

Volkov Valery Fedorovich — Ph.D., Dr. Sci., professor, professor of systems analysis and control department, Mozhaisky Military Space Academy. Research interests: theory of effectiveness, operations research, system analysis. The number of publications — 250. vafedvolkov@gmail.com; 13, Zhdanovskaya street, St.-Petersburg, 197198, Russia; office phone: +7 951 652 32 63.

Потрясаев Семен Алексеевич — к-т техн. наук, старший научный сотрудник лаборатории информационных технологий в системном анализе и моделировании, Федеральное государственное бюджетное учреждение науки Санкт-Петербургского института информатики и автоматизации Российской академии наук (СПИИРАН). Область научных интересов: разработка научных основ теории управления структурной динамикой сложных организационно-технических систем. Число научных публикаций — 40. spotryasaev@gmail.com; 14-я линия В.О., 39, Санкт-Петербург, 199178; р.т.: (812)328-0103, Факс: (812)328-4450.

Potryasaev Semen Alekseevich — Ph.D., senior researcher of laboratory for information technologies in systems analysis and modeling, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS). Research interests: analyses of systems, reliability theory, mathematical models and methods of decision-making support in complex technical-organizational systems with the use of aerospace data. The number of publications — 40. spotryasaev@gmail.com; 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone: (812)328-0103, Fax: (812)328-4450.

Салухов Владимир Иванович — к-т техн. наук, доцент, руководитель исследовательской группы информационных технологий в образовании, Федеральное государственное бюджетное учреждение науки Санкт-Петербургского института информатики и автоматизации Российской академии наук (СПИИРАН). Область научных интересов: информационные технологии в образовании, управление жизненным циклом инфотелекоммуникационных систем, анализ и разработка систем поддержки принятия решений на базе современных информационных технологий. Число научных публикаций — 50. vsaluhov@bk.ru; 14-я линия В.О., 39, Санкт-Петербург, 199178; р.т.: +7(812)3280382.

Salukhov Vladimir Ivanovich — Ph.D., associate professor, head of research group of information technologies in education, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS). Research interests: research and information technologies in education, lifecycle management infocommunication systems, analysis and development of support systems and decision making on the basis of modern information technologies. The number of publications — 50. vsaluhov@bk.ru; 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone: +7(812)3280382.

Шедько Наталья Адамовна — ассистент кафедры информатики и информационной безопасности, Петербургский государственный университет путей сообщения Императора Александра I (ПГУПС). Область научных интересов: управление проектами, исследование операций, системный анализ. Число научных публикаций — 3. shedkonatalia@yandex.ru; Московский пр. 9, Санкт-Петербург, 190031; р.т.: 89214026331.

Shedko Natalia Adamovna — assistant professor of computer science and information security department, Emperor Alexander I St. Petersburg State Transport University. Research interests: project management, operations research, system analysis. The number of

publications — 3. shedkonatalia@yandex.ru; 9, Moskovsky pr., Saint Petersburg, 190031, Russia; office phone: 89214026331.

Поддержка исследований. Работа выполнена при финансовой поддержке ведущих университетов Российской Федерации: СПбГПУ (мероприятие 6.1.1), ИТМО (субсидия 074–U01), гранта РФ № 16-19-00199, грантов РФФИ (№ 16-57-00172-Бел а, 16-07-00779, 16-08-01068, 16-07-01277, 15-07-08391, 15-08-08459, 16-08-00510, 16-07-00925) Программы фундаментальных научных исследований ОНИТ РАН (проект № 2.11)

Acknowledgements. This research is supported by the leading universities of the Russian Federation: STU (activity 6.1.1), the ITMO University (grant 074–U01), grant RSF № 16-19-00199, RFBR grants (No. 16-57-00172-Bela, 16-07-00779, 16-08-01068, 16-07-01277, 15-07-08391, 15-08-08459, 16-08-00510, 16-07-00925), Program of fundamental scientific research ONIT Russian Academies of Sciences (grant 2.11)

РЕФЕРАТ

Бураков В.В., Волков В.Ф., Потрясаев С.А., Салухов В.И., Шедько Н.А.
Формализованные подходы к исследованию точности моделей экономических систем.

Известные постановки задач обоснования экономических управленческих решений базируются на допущении о наличии всех необходимых исходных данных. Однако, как на этапе проектирования, так и на этапах планирования и тем более в «реальном масштабе времени» большая часть исходных данных, необходимых для вычисления показателей рыночной деятельности, будет либо известна приближенно, с определенными ошибками, либо указываться в некотором диапазоне. Сами показатели (рейтинги, индексы) на практике также определяются с некоторыми ошибками, обусловленными погрешностями численных алгоритмов или сбоями ЭВМ, и, кроме того, они во многом зависят от «неколичественной» информации, например, от корректности вопросов, задаваемых при маркетинговых исследованиях. В статье на примерах наиболее распространенных на микро- и макроуровнях экономико-математических моделей предложен подход к построению методики «оптимальной сложности», обеспечивающей минимальную величину суммарной ошибки при заданной продолжительности решения информационно-расчетных задач по оцениванию организационно-экономических систем (ОЭС). Кроме того, данный подход позволяет обосновать требования к точности входной информации.

Показано, что если проект ОЭС имеет уникальный и бескомпромиссный характер, то в качестве требуемого уровня эффективности должен выступать максимально возможный уровень, основанный на принципе практической уверенности (оцениваемой, например, через вероятность практически достоверного события), а степень точности модели должна быть максимальной. В том случае, когда проекты носят «массовый характер», возможен компромисс, и для обеспечения рационального уровня точности моделирования орган управления (заказчик модели) должен учитывать складывающиеся в реальности соотношения точности исходной информации, структурной точности модели, функциональной точности модели и точности вычислительных алгоритмов.

SUMMARY

Burakov V.V., Volkov V.F., Potryasaev S.A., Salukhov V.I., Shedko N.A.

Formalized Approaches to the Study of the Accuracy of Economic Systems Models.

Known formulations of the economic management decision problems are based on the assumption that there are all required initial data. However, during the stages of projecting, planning, and especially “in real time”, a larger half of initial data needed for calculating market activity indicators will be either known approximately (with certain mistakes) or specified roughly in a certain range. Indicators themselves (rat-ings, indices) are determined in practice with certain mistakes caused by errors of numerical algorithms or computer failures. Moreover, they largely depend on non-numerical information, e.g. on the correctness of questions asked at marketing re-search. In the article, we propose an approach to the creation of the optimal complexity method through examples of economic and mathematical models that are most common at micro - and macro-levels. This method minimizes the total error for given duration of solving information-computing tasks within the evaluation of or-ganizational and economic systems (OES). In addition, this approach allows us to substantiate requirements to the accuracy of input information.

It is shown that if the OES project is unique and uncompromising, then as the required efficiency level there must be the highest possible level. It must be based on the principle of practical certainty (estimated by the probability of a practically certain event). The degree of the model accuracy should be maximized. In case where projects are “massive”, a compromise is possible. To provide a sustainable level of the simulation accuracy, the authority (a model customer) must take into account real relations of the accuracy of initial information, the structural model accuracy, the functional model accuracy and the accuracy of numerical algorithms.

Д.С. ШИШИГИН
**К ВЫБОРУ ТЕХНОЛОГИИ ИНТЕГРАЦИИ ПРИКЛАДНОГО
ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ С САПР**

Шишигин Д.С. К выбору технологии интеграции прикладного программного обеспечения с САПР.

Аннотация. Компьютерные программы для научных исследований со сложными геометрическими моделями целесообразно интегрировать с САПР. Рассматриваются три подхода на основе: обмена данными в формате DXF (drawing exchange format), COM-технологии и прикладной программный интерфейс (API) ObjectARX. DXF-обмен данными с САПР — простой и универсальный способ, доступный для большинства исследователей, но исключающий интерактивное управление САПР. COM технологии предоставляют простые, надежные, но не самые быстродействующие механизмы интерактивного управления САПР из внешней программы. Комбинация с DXF повышает быстродействие COM. Для задач со сложными геометрическими моделями система AutoCAD предоставляет ObjectARX API и .NET API — низкоуровневые технологии, обеспечивающие наивысшую по сравнению с другими технологиями функциональность и производительность, но имеющие ограничения. Приведены листинги программ, упрощающие понимание рассматриваемых технологий, проведен анализ их быстродействия, даны рекомендации по их применению с позиции исследователя.

Ключевые слова: геометрическая модель, САПР, AutoCAD, DXF, COM, ObjectARX, автоматизация, научная графика.

Shishigin D.S. On Choosing the Technology of Application Software Integration with a CAD-System.

Abstract. Computer programs for scientific research with complex geometric models should be integrated with a CAD-system. In the paper, we consider three approaches based on data exchange in the DXF (drawing exchange format); COM-technologies; and application programming interface (API) ObjectARX. DXF data exchange with CAD is a simple and universal way available to most researchers, but it excludes interactive CAD control. COM technologies provide simple, reliable mechanisms of interactive CAD control from an external program, although they do not have the highest level of performance. Combination with DXF increases their performance. For problems with complex geometric models, Auto-CAD-system provides ObjectARX API and .NET API – low-level technologies that ensure the maximum possible functionality and performance, compared to other technologies, but have some limitations. We present program listings that simplify the understanding of considered technologies. Their performance analysis has been conducted, and recommendations for their use are given from the researcher’s perspective.

Keywords: geometric model, CAD, AutoCAD, DXF, COM, ObjectARX, automation, scientific graphics.

1. Введение. Компьютерные программы для научных исследований, проектирования и других задач со сложными геометрическими моделями и графическими выходными данными целесообразно разрабатывать на основе интеграции с САПР. Исследователь должен писать программный код только для проведения расчетов, а ввод геометриче-

ских данных и визуализацию результатов, включая анимацию динамических процессов, можно «поручить» САПР.

По степени интеграции с САПР различают автономные программы, которые обычно только считывают данные из графических моделей, и приложения, которые работают в среде САПР. В последнем случае обычно используется AutoCAD и ведется разработка AutoCAD-приложений.

Программные способы управления AutoCAD из приложения пользователя в наиболее общем виде рассмотрены в работе [1]. Для изучения СОМ-технологий полезна работа [2]. Технологии ObjectARX посвящены работы [3, 4]. Текущую информацию по способам управления AutoCAD содержит справочная система AutoCAD.

Эти данные рассчитаны на программистов, но их явно недостаточно, чтобы сделать САПР таким же привычным инструментом для исследователей, как системы компьютерной математики (СКМ). В то же время именно интеграция прикладных программ и СКМ с САПР — один из путей автоматизации научных исследований в задачах с геометрическим моделированием.

Среди успешных примеров интеграции прикладных программ с САПР задача генерации и визуализации нестандартных 3D сеток для расчетов электромагнитных полей методом конечных разностей во временной области [5]. Расчету и проектированию линий электропередач и распределительных сетей посвящены работы [6, 7]. К сожалению, число подобных научных публикаций очень незначительно и все они посвящены только технологии ObjectARX, поэтому потенциальные возможности интеграции прикладной программы с САПР для исследователя раскрыты далеко не полностью.

При разработке программы ЗУМ [8], предназначенной для расчета молниезащиты и заземления объектов электроэнергетики, мы прошли все ступени интеграции с AutoCAD и хотели бы обобщить результаты.

Показать и сопоставить возможные технологии интеграции прикладной программы или СКМ с САПР (на примере AutoCAD) с позиции исследователя, решающего задачи со сложными геометрическими моделями, — цель настоящей работы.

2. DXF (drawing exchange format). Передача геометрических данных в программу пользователя и обратно может быть выполнена в текстовых файлах, где геометрические данные записаны в формате dxf. Структура dxf-файла документирована в справочной системе AutoCAD. Данные представлены парами «код-значение» (таблица 1), которые записываются в двух строках (без комментариев).

Таблица 1. Фрагмент dxf-файла с линией

Код	Значение	Комментарий
...	...	
0	SECTION	Начало секции
2	ENTITIES	Секция графических данных
0	LINE	Линия
8	LayerName	Название слоя с линией
...	...	
10	0	X-координата начала
20	0	Y-координата начала
30	-0.7	Z-координата начала
11	100	X-координата конца
21	100	Y-координата конца
31	-0.7	Z-координата конца
...	...	
0	ENDSEC	Конец секции
...	...	

Пусть требуется прочитать геометрическую модель, состоящую из начерченных в AutoCAD линий. Сохраним файл в AutoCAD с расширением dxf, а затем последовательно (построчно) прочитаем его в собственной программе. Точно также можно записать линии в dxf-файл (листинг 1) и построить их в AutoCAD.

```
F:=TStringList.Create; //буфер
F.Add('0'); F.Add('SECTION');
F.Add('2'); F.Add('ENTITIES');
for i:=0 to high(x1) do
begin
  F.Add('0'); F.Add('Line');
  F.Add('8'); F.Add('SetOfLines'); //зададим имя слоя
  F.Add('10'); F.Add(FloatToStr( x1[i] )); //координаты начальной точки
  F.Add('20'); F.Add(FloatToStr( y1[i] ));
  F.Add('30'); F.Add(FloatToStr( z1[i] ));
  F.Add('11'); F.Add(FloatToStr( x2[i] )); //координаты конечной точки
  F.Add('21'); F.Add(FloatToStr( y2[i] ));
  F.Add('31'); F.Add(FloatToStr( z2[i] ));
end;
F.Add('0'); F.Add('ENDSEC');
F.Add('0'); F.Add('EOF');
F.SaveToFile('F.dxf');
ADoc.ActiveDocument.Import('F.dxf',Point,1); //загрузка в AutoCAD
```

Листинг 1. Создание набора линий в формате dxf и загрузка в AutoCAD

Обмен данными через dxf-файлы имеет ряд преимуществ. Он стал стандартом обмена файлами чертежей для всех САПР, потому универсален. Обработка dxf-файлов может осуществляться с помощью любого языка программирования и с высоким быстродействием. Размерность dxf-файла может превышать миллион строк.

Этот простой способ доступа к графическим данным применяется во многих профессиональных программах, например, продукции фирмы SES (sestech.com) — мирового лидера в области расчетов заземления и молниезащиты. Он годится для любого исследователя, позволяя расширить класс решаемых в СКМ задач.

Главным недостатком обмена через dxf-файлы является отсутствие возможности интерактивного взаимодействия пользователя с чертежом. Устранить его способами более сложные способы управления САПР.

3. СОМ-технологии. AutoCAD, как и другие профессиональные САПР, разработаны с учетом концепции СОМ (СОМ — Component Object Model, модель компонентных объектов, объектная модель) фирмы Microsoft, что дает доступ к объектам AutoCAD, их свойствам, методам, событиям из других систем и их приложений и наоборот [9].

Все геометрические примитивы (линии, полилинии, точки) текущего документа (назовем его *ADoc*) находятся в коллекции *ADoc.ModelSpace*. Число примитивов дает свойство *ADoc.ModelSpace.Count*. Для чтения всех геометрических объектов чертежа из собственной программы достаточно организовать цикл (листинг 2а).

```
var StartPoint, EndPoint: OleVariant; Line: IAcadLine;
begin
  for i:=0 to Acad.ActiveDocument.ModelSpace.Count-1 do
    if Acad.ActiveDocument.ModelSpace.Item(i).ObjectName = 'AcDbLine' then
      begin
        Line:=Acad.ActiveDocument.ModelSpace.Item(i) as IAcadLine;
        StartPoint:=Line.StartPoint;
        EndPoint:=Line.EndPoint;
        ...
      end;
    end;
end;
```

Листинг 2а. Чтение координат всех линий чертежа с помощью СОМ

Метод *AddLine* позволяет нарисовать линию в AutoCAD из собственной программы (листинг 2б). Циклически применяя рассмотренный метод можно нарисовать геометрический объект, состоящих из произ-

вольного числа линий. Получаемый объект новой линии (*Line*) обеспечивает доступ к ее свойствам, что позволяет изменить цвет, слой и т.д.

```
var StartPoint, EndPoint: OleVariant; Line: IAcadLine;
begin
  StartPoint:=VarArrayCreate([0,2], varDouble);
  EndPoint:=VarArrayCreate([0,2], varDouble);
  StartPoint[0]:=0; StartPoint[1]:=0; StartPoint[2]:=0;
  EndPoint[0]:=100; EndPoint[1]:=100; EndPoint[2]:=100;
  Line:=Acad.ActiveDocument.ModelSpace.AddLine(StartPoint, EndPoint);
end;
```

Листинг 26. Создание примитива линия с помощью COM

Данный способ является полноценным средством управления AutoCAD и обеспечивает необходимый уровень интерактивности (в отличие от способа с dxf-файлами). COM поддерживает широкий спектр языков программирования, включая самые популярные: C++, Delphi, Visual Basic, Java. В большинстве сред программирования реализованы модули, упрощающие использование объектной модели, например, каркас ActiveQt для среды Qt.

Стандарт COM спроектирован для межпроцессного взаимодействия, поэтому он позволяет разрабатывать внепроцессные (out-of-process) приложения, иначе говоря, программа пользователя не является частью процесса AutoCAD, что имеет достоинства и недостатки.

Главное достоинство — это высокая универсальность. Используя позднее связывание (COM-вызовы с помощью Invoke интерфейса IDispatch), мы можем добиться того, что связывание имен свойств и методов объекта с их кодом происходит не на этапе компиляции, а на этапе выполнения программы, что позволяет использовать одно приложение для всех AutoCAD-совместимых САПР.

Однако использование механизмов межпроцессного взаимодействия и позднего связывания приводит к значительному снижению производительности (в десятки и сотни раз по сравнению с внутрипроцессными взаимодействиями и ранним связываем). Построение десятков тысяч линий может занимать минуты, что неприемлемо.

Таким образом, COM-технологии являются достаточно простым и надежным инструментом для интерактивного обмена графической информацией программы пользователя с любым САПР, пригодным для большинства научных задач. COM-технологии пригодны как для разработки приложений к САПР, так и использования в обычных программах в качестве альтернативы DXF.

Но в задачах с большими графическими данными (тысячи геометрических примитивов и более), которые требуется быстро пере-

страивать, например, при анимации динамических процессов, подход на основе COM неэффективен из-за низкой производительности.

Для разработки профессиональной программы, управляющей САПР, с полным доступом к свойствам 3D объектов рекомендуются технологии ObjectARX и .NET.

4. ObjectARX. Это система разработки программных модулей для среды AutoCAD, создающая самый низкоуровневый код по сравнению с другими средствами автоматизации AutoCAD. Код стороннего приложения выполняется в том же адресном пространстве, что и код AutoCAD, поэтому взаимодействие между ними значительно эффективнее, поскольку данные могут передаваться через указатели, а не через механизм межпроцессного взаимодействия как в COM-приложениях. Модули, разработанные в ObjectARX, работают с основными структурами данных и кодом AutoCAD и имеют доступ непосредственно к графической системе и ядру геометрических построений. В результате достигается высокая функциональность и быстродействие, но при значительно более сложной структуре программы.

Разработка пользовательских приложений (ARX-приложений) производится с использованием неуправляемого кода на языке C++, т.е. на том же языке, что использован для разработки ядра AutoCAD.

Набор разработчика ObjectARX предоставляет специальные библиотеки, а также заголовочные файлы, с помощью которых приложение пользователя получает информацию о типах, классах и объектах AutoCAD. Подключая эти файлы в свой код, программа получает возможность взаимодействовать с AutoCAD, а также создавать свои классы и объекты. Например, примитив линия описан классом *AcDbLine*. Создав объект линии, можно добавить его в пространство модели чертежа, которое представляет собой блок объектов типа *AcDbBlockTableRecord*. Данные записи, в свою очередь, расположены в таблице блоков *AcDbBlockTable*, которая, как и все остальные таблицы, принадлежит базе данных чертежа *AcDbDatabase*.

Покажем алгоритм чтения координат всех линий чертежа (листинг 3а).

Первая часть кода служит для получения блока пространства модели (*ACDB_MODEL_SPACE*), в котором хранятся примитивы чертежа. Затем с помощью итератора *AcDbBlockTableRecordIterator* происходит получение всех примитивов в цикле. При работе с объектами базы данных их требуется открывать с нужными атрибутами (*AcDb::kForRead*, *AcDb::kForWrite*, т.д.) и обязательно закрывать. В случае, если полученный объект имеет тип *AcDbLine*, он преобразовывается, и получаются его координаты.

```

AcApDocument* pActiveDocument = acDocManager->mdiActiveDocument();
AcDbDatabase* pActiveDatabase = pActiveDocument->database();
AcDbBlockTable* pBlockTable;
pActiveDatabase->getSymbolTable(pBlockTable, AcDb::kForRead);
AcDbBlockTableRecord* pBlockTableRecord;
pBlockTable->getAt(ACDB_MODEL_SPACE, pBlockTableRecord,
AcDb::kForRead);
pBlockTable->close();
AcDbBlockTableRecordIterator* pIter;
pBlockTableRecord->newIterator(pIter);
for (; !pIter->done(); pIter->step()) {
    AcDbEntity* pEnt;
    pIter->getEntity(pEnt, AcDb::kForRead);
    if (pEnt->isA() == AcDbLine::desc()) {
        AcDbLine* pLine = static_cast<AcDbLine*>(pEnt);
        AcGePoint3d startPoint = pLine->startPoint();
        AcGePoint3d endPoint = pLine->endPoint();
        ...
    }
    pEnt->close();
}
delete pIter;
pBlockTableRecord->close();

```

Листинг 3а. Чтение координат всех линий чертежа с помощью ObjectARX

Покажем способ создания линии (листинг 3б).

```

AcApDocument* pActiveDocument = acDocManager->mdiActiveDocument();
AcDbDatabase* pActiveDatabase = pActiveDocument->database();
acDocManager->lockDocument(pActiveDocument);
AcDbBlockTable* pBlockTable;
pActiveDatabase->getSymbolTable(pBlockTable, AcDb::kForRead);
AcDbBlockTableRecord* pBlockTableRecord;
pBlockTable->getAt(ACDB_MODEL_SPACE, pBlockTableRe-
cord, AcDb::kForWrite);
pBlockTable->close();
AcGePoint3d startPoint(0.0, 0.0, 0.0);
AcGePoint3d endPoint(100.0, 100.0, 100.0);
AcDbLine* pLine = new AcDbLine(startPoint, endPoint);
pBlockTableRecord->appendAcDbEntity(pLine);
pLine->close();
pBlockTableRecord->close();
acDocManager->unlockDocument(pActiveDocument);

```

Листинг 3б. Создание примитива линия с помощью ObjectARX

Сначала, как и в листинге 3а, получается блок пространства модели, в который требуется поместить примитив. Затем создается новый объект *AcDbLine* (линия) со свойствами, принятыми по умолчанию (слой, тип линии, толщина, т.д.). Эти свойства можно изменить. Линия добавляется в блок с помощью метода *appendAcDbEntity*. Перед открытием объекта для записи необходимо заблокировать рабочий документ с помощью метода *lockDocument*, а в конце разблокировать с помощью *unlockDocument*.

Циклически добавляя примитивы в пространство модели AutoCAD, можно построить геометрические модели, сетки, научную графику или чертежи любой сложности.

Высокое быстродействие операций с графическими примитивами и функциональность, обеспечивающая доступ ко всем свойствам, методам, событиям с примитивами, — главное достоинство внутри-процессных API.

Главный недостаток ObjectARX — это требование к бинарной совместимости: пользовательский модуль должен быть создан той же версией компилятора, что и AutoCAD. Компания Autodesk нарушает бинарную совместимость AutoCAD (обычно каждые 3 выпуска), поэтому требуется создавать и поддерживать несколько версий пользовательских модулей (для разных версий AutoCAD), что усложняет разработку приложений и неудобно для пользователя. Напомним, что технология COM позволяет создавать программы, пригодные для любой версии AutoCAD, и применима для любого САПР.

Технология COM допускает выбор языка программирования, что может быть важно для исследователя, в то время как ObjectARX устанавливает язык C++. ObjectARX C++ «заточен» под библиотеку MFC (Microsoft Foundation Classes), использование которой не всегда приемлемо (пакет MFC не входит в бесплатную редакцию Visual Studio Express).

Усложнение кода ObjectARX (листинг 3) по сравнению с COM (листинг 2) также относится к недостаткам с позиции исследователя. Самостоятельная разработка ARX-приложений может оказаться для него сложной задачей.

Таким образом, технология ObjectARX позволяет эффективно решать задачи в AutoCAD с геометрическими моделями и научной графикой любой сложности, но ее реализация ориентирована на квалифицированных программистов.

5. Технология .NET. На основе платформы .NET Framework — еще одно средство разработки низкоуровневых AutoCAD-приложений, имеющее некоторые преимущества перед ObjectARX. В .NET расширен выбор языка программирования (C++, C#, VB .NET и др.). Созда-

ются dll-библиотеки с управляемым кодом (управляемые сборки), что открывает программисту механизмы управления памятью, а также обеспечивает простоту разработки пользовательских интерфейсов (с помощью библиотек .NET Windows Forms или .NET Windows Presentation Foundation, входящих в состав .NET Framework).

Однако, технология .NET не в состоянии устранить главный недостаток ObjectARX — требование к бинарной совместимости: пользовательский модуль должен быть создан на той же версии целевой платформы .NET Framework, что и AutoCAD. Таким образом, переход от ObjectARX к .NET не носит принципиального характера.

6. Анализ быстродействия. Скорость построения геометрических примитивов в AutoCAD из программы пользователя — один из основных критериев, определяющих область применения рассмотренных технологий. Будем строить линии в циклах переменной размерности N с помощью программных кодов, приведенных ранее, и отмечать время выполнения (рисунок 1).

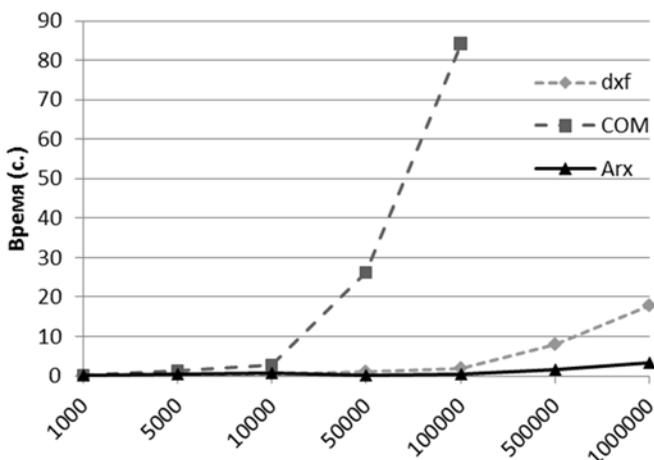


Рис. 1. Время построения линий (N – число линий) с использованием DXF, COM и ObjectARX (Компьютер: Windows 7 64-bit, Intel Core i5-4430 3.0 ГГц, ОЗУ 8 Гб)

В результате количественно подтверждены сделанные ранее выводы, а именно: межпроцессное взаимодействие через COM применимо только для задач ограниченной размерности (до 10000 геометрических примитивов). Низкоуровневые ARX-приложения обладают высокой производительностью, позволяющей проводить моделирование с большими геометрическими данными (1 млн. примитивов и более) или анимировать динамические процессы в сложных 3D системах.

Таким образом, имеем класс задач, эффективное решение которых возможно только в AutoCAD. Область применения других САПР на основе СОМ технологий сужена.

7. Комбинация СОМ с DXF. Скорость построения линий с помощью dxf сопоставима с ARX-технологией вплоть до 100000 элементов (рисунок 1), что вполне объяснимо — dxf-файл создается быстро, а его чтение — внутренние операции AutoCAD. Отсюда идея использования dxf-файла в качестве буфера обмена графической информацией с САПР на основе технологии СОМ [8]. Вместо прямого рисования линий в САПР (с использованием медленных механизмов межпроцессного взаимодействия) запишем их в dxf-файле и загрузим его в AutoCAD (листинг 1), что значительно быстрее. Этот способ позволяет использовать САПР для визуализации результатов научных исследований [8]. Он дает возможность строить графики из большего числа элементов, чем в СКМ, а послонная структура данных в САПР позволяет легко создавать выборки (включением/выключением слоев). Таким образом, использование dxf-файлов в качестве буфера обмена данными с САПР расширяет область применения СОМ-технологий.

8. Заключение. Компьютерные программы для научных исследований со сложными геометрическими моделями и задач проектирования целесообразно интегрировать с САПР. В зависимости от типа задачи возможны три основных пути: обмен данными в формате DXF, СОМ-технологии, прикладной программный интерфейс ObjectARX. DXF-обмен данными с САПР — простой и универсальный способ, доступный для большинства исследователей, но исключающий интерактивное управление САПР. Этот способ наиболее подходит для чтения геометрических данных при решении задач в СКМ. Его часто применяют в профессиональных программах с собственными геометрическими процессорами в качестве дополнительной опции.

СОМ технологии предоставляют простые, надежные механизмы интерактивного управления САПР из внешней программы, пригодные для большинства научных задач и задач проектирования. Разработка САПР-приложения на основе СОМ достаточно проста и может быть рекомендована исследователю. Низкая производительность в задачах с большими геометрическими данными — главный недостаток СОМ, однако производительность можно значительно повысить, если использовать dxf-файл в качестве буфера в операциях записи/чтения примитивов.

Для задач со сложными геометрическими моделями, анимацией динамических процессов рекомендуется система AutoCAD, которая предоставляет низкоуровневые ObjectARX API и .NET API технологии. Они обеспечивают наивысшую по сравнению с другими техноло-

гиями функциональность и производительность, хотя не лишены недостатков. Главным из них является требование бинарной совместимости программы пользователя и версии AutoCAD, другими словами, ARX-приложение работает только на определенных версиях AutoCAD, в отличие от универсальных программ, созданных по технологии COM (с использованием позднего связывания).

Литература

1. *Полещук Н.Н.* AutoCAD: разработка приложений, настройка и адаптация // СПб.: БХВ-Петербург. 2006. 992 с.
2. *Роджерсон Д.* Основы COM // М.: Изд.-торг. Дом «Русская редакция». 2000. 400 с.
3. *Kramer B.* ObjectARX Primer // Autodesk Press. 2000. 166 p.
4. *McAuley C.* Programming AutoCAD 2000 Using ObjectARX // Autodesk Press. 2000. 678 p.
5. *Jun Y., Dong-Lin S., Xiao-Ying Z.* The study and realization of automatic mesh generation based on electromagnetic simulation of FDTD // 3rd IEEE International Symposium on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications. 2009. pp. 1242–1245.
6. *Hou X., Lu X.* Automatic design of transmission lines using ObjectARX technology. International Conference on Control Engineering and Communication Technology (ICCECT). 2012. pp. 581–585.
7. *Huang Y. et al.* Design and implementation of reliability evaluation module for medium voltage distribution networks by objectarx technology // 8th International Conference on Advances in Power System Control, Operation and Management (APSCOM 2009). 2009. pp. 1–6.
8. *Шнишгин Д.С.* AutoCAD приложение для расчета молниезащиты и заземления объектов электроэнергетики // Автоматизация в промышленности. 2014. № 9. pp. 28–32.
9. *Полещук Н.Н.*, Программирование для AutoCAD 2013–2015 // М.: ДМК Пресс. 2015. 462 с.

References

1. *Poleshuk N.N.* AutoCAD: *razrabotka prilozhenij, nastrojka i adaptacija* [AutoCAD: application development, configuration and adaptation]. SPb.: BHV-Peterburg. 2006. 992 p. (In Russ.).
2. *Rodzherson D.* *Osnovy COM* [COM basics]. M.: Izd.-torg. Dom «Russkaja redakcija». 2000. 400 p. (In Russ.).
3. *Kramer B.* ObjectARX Primer. Autodesk Press. 2000. 166 p.
4. *McAuley C.* Programming AutoCAD 2000 Using ObjectARX. Autodesk Press. 2000. 678 p.
5. *Jun Y., Dong-Lin S., Xiao-Ying Z.* The study and realization of automatic mesh generation based on electromagnetic simulation of FDTD. 3rd IEEE International Symposium on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications. 2009. pp. 1242–1245.
6. *Hou X., Lu X.* Automatic design of transmission lines using ObjectARX technology. International Conference on Control Engineering and Communication Technology (ICCECT). 2012. pp. 581–585.
7. *Huang Y. et al.* Design and implementation of reliability evaluation module for medium voltage distribution networks by objectarx technology // 8th International Confer-

ence on Advances in Power System Control, Operation and Management (APSCOM 2009). 2009. pp. 1–6.

8. Shishigin D.S. [AUTOCAD application for calculation of lightning protection and grounding of electric power industry objects]. *Avtomatizacija v promyshlennosti – Automation in industry*. 2014. vol. 9. pp. 28–32. (In Russ.).
9. Poleshchuk N.N. *Programmirovanije dlja AutoCAD 2013–2015* [Programming for AutoCAD 2013-2015]. M.: DMK Press, 2015. 462 p. (In Russ.).

Шишигин Дмитрий Сергеевич — аспирант кафедры автоматизации и вычислительной техники Вологодского государственного университета. Область научных интересов: разработка программного обеспечения в области молниезащиты и заземления объектов электроэнергетики. Число научных публикаций — 25. shishigind@yandex.ru; ВоГУ, ул. Ленина, д. 15, г. Вологда, 160000, РФ; т. +79210646334.

Shishigin Dmitry Sergeevich — postgraduate student of the Automation and Computer Engineering department of the Vologda State University. Research interests: software development in the area of lightning protection and grounding of electric power industry objects. The number of publications — 25. shishigind@yandex.ru; VSU, 15, Lenina str., Vologda, 160000, Russia; p.n. +79210646334.

РЕФЕРАТ

Шиншигин Д.С. **К выбору технологии интеграции прикладного программного обеспечения с САПР.**

Компьютерные программы для научных исследований со сложными геометрическими моделями и задач проектирования целесообразно интегрировать с САПР. В зависимости от типа задачи возможны три основных пути: обмен данными в формате DXF, СОМ-технологии, прикладной программный интерфейс (API) ObjectARX. DXF-обмен данными с САПР — простой и универсальный способ, доступный для большинства исследователей, но исключающий интерактивное управление САПР. Этот способ наиболее подходит для чтения геометрических данных при решении задач в СКМ. Его часто применяют в профессиональных программах с собственными геометрическими процессорами в качестве дополнительной опции.

СОМ технологии предоставляют простые, надежные механизмы интерактивного управления САПР из внешней программы, пригодные для большинства научных задач и задач проектирования. Разработка САПР-приложения на основе СОМ достаточно проста и может быть рекомендована исследователю. Низкая производительность в задачах с большими геометрическими данными — главный недостаток СОМ, однако производительность можно значительно повысить, если использовать dxf-файл в качестве буфера в операциях записи/чтения примитивов.

Для задач со сложными геометрическими моделями, анимацией динамических процессов рекомендуется система AutoCAD, которая предоставляет низкоуровневые ObjectARX API и .NET API технологии. Они обеспечивают наивысшую по сравнению с другими технологиями функциональность и производительность, хотя не лишены недостатков. Главным из них является требование бинарной совместимости программы пользователя и версии AutoCAD, другими словами, ARX-приложение работает только на определенных версиях AutoCAD, в отличие от универсальных программ, созданных по технологии СОМ (с использованием позднего связывания).

SUMMARY

Shishigin D.S. **On Choosing the Technology of Application Software Integration with a CAD-System.**

Computer programs for scientific research with complex geometric models and for design problems should be integrated with CAD-systems. Depending on the problem, three main ways are possible: data exchange in the DXF format, COM-technologies, application programming interface (API) ObjectARX. DXF data exchange with CAD is a simple and universal way available to most researchers, but it excludes interactive CAD control. This approach is more appropriate for geometric data reading when solving problems in computer mathematics systems. It is often used in professional programs with their own geometric processors as an extra option.

COM technologies provide simple, reliable mechanisms of interactive CAD control from an external program, suitable for most research and design problems. CAD-application development based on COM is simple enough and can be recommended to the researcher. Low performance in problems with large geometric data is the main drawback of COM, however, the performance can be significantly increased if you use DXF file as a read/write buffer for geometric entities.

For problems with complex geometric models, dynamic processes animation AutoCAD is recommended which provides low-level ObjectARX API and .NET API technologies. They ensure the maximum possible functionality and performance, compared to other technologies, but have some drawbacks. The main among them is the binary compatibility requirement of user application and AutoCAD version. In other words, ARX-application works only with the certain AutoCAD versions, unlike universal programs based on the COM technology (late binding).

Л.В. Уткин, Ю.А. Жук
**ПОЛНОГЕНОМНЫЙ ПОИСК АССОЦИАЦИЙ С
ИСПОЛЬЗОВАНИЕМ МАТРИЦ ПАРНЫХ СРАВНЕНИЙ**

Уткин Л.В., Жук Ю.А. Полногеномный поиск ассоциаций с использованием матриц парных сравнений.

Аннотация. Предлагается простой метод определения значимости объектов популяции при установлении ассоциации между однонуклеотидными полиморфизмами и количественными признаками в полногеномном поиске ассоциаций. На первом этапе сравниваются пары объектов популяции с точки зрения расстояния между ними по фенотипу и генотипу. На втором этапе строятся матрицы парных сравнений объектов и вычисляются веса объектов в соответствии с аддитивной и мультипликативной шкалами. Показывается, как можно модифицировать метод Лассо с использованием весов. Числовые эксперименты с реальными данными иллюстрируют предлагаемый метод.

Ключевые слова: биоинформатика, регрессия, полногеномный поиск ассоциаций, парные сравнения, Лассо, аддитивные и мультипликативные шкалы.

Utkin L.V., Zhuk Y.A. A Genome-Wide Association Study using Pairwise Comparison Matrices.

Abstract. In this paper, we propose a simple method for assigning importance weights to individuals of a population to determine the association between single nucleotide polymorphisms and quantitative traits in the genome-wide association study. At the first step, pairs of individuals in population are compared in terms of distances between phenotypes and genotypes. At the second step, pairwise comparison matrices of individuals are constructed, and the weights of individuals are computed with respect to additive and multiplicative scales. It is shown how to modify the Lasso method using the weights. Numerical experiments with real data illustrate the proposed method.

Keywords: bioinformatics, regression, genome-wide association study, pairwise comparisons, Lasso, additive and multiplicative scales.

1. Введение. Одним из элементов биоинформатики как совокупности методов и подходов, включающих в себя математические методы компьютерного анализа, алгоритмов и программ в геномике, является полногеномный поиск ассоциаций (a genome-wide association study — GWAS). С точки зрения биологии основная цель GWAS заключается в поиске зависимости между значениями фенотипа объектов некоторой популяции и их генотипом. Другими словами, GWAS определяет, какие элементы ДНК или молекулярные маркеры, называемые однонуклеотидными полиморфизмами (single nucleotide polymorphisms — SNP), оказывают наибольшее влияние на значения фенотипа. С математической точки зрения цель GWAS — это построение зависимости между SNP, которые рассматриваются как независимые переменные, и значениями фенотипа как зависимой переменной. С точки зрения теории машинного обучения, GWAS — это задача отбора признаков, в которой признаками являются SNP, а обучающая

выборка — объекты анализируемой популяции, каждый из которых характеризуется определенным значением фенотипа и множеством значений генотипа (значениями зависимых переменных).

Вследствие важности решения задачи определения значимых маркеров, к настоящему времени разработано огромное количество статистических моделей и алгоритмов. Детальные обзоры существующих подходов можно найти в работах [6, 7, 11, 13, 15], согласно которым большинство моделей и алгоритмов GWAS используют:

1. методы фильтрации [3, 15], которые основаны на статистических свойствах SNP-маркеров для отбрасывания наименее информативных маркеров (критерий Фишера, дисперсионный анализ и т.д.);
2. методы упаковки [9], например, известный метод рекурсивного удаления признаков [10];
3. вложенные методы [14, 16], которые основаны на построении таких регрессионных моделей, как метод Лассо, гребневая регрессия, эластичная сеть, устанавливающих зависимость между SNP (независимые переменные) и количественным признаком или фенотипом (зависимая переменная).

Многими авторами отмечается, что методы, основанные на построении регрессионных моделей, являются наиболее перспективными и интересными, так как построение этих моделей охватывает одновременно все SNP, устанавливая в результате решения определенные веса SNP, которые и являются показателями их значимости в ассоциации.

Однако, несмотря на большое количество моделей и алгоритмов GWAS, в частности регрессионных моделей, большинство из них сталкиваются с рядом сложностей, главной из которых является то, что число SNP-маркеров обычно в десятки или сотни раз превышает число объектов в популяции или в обучающей выборке. Другими словами, размерность задачи значительно превышает количество точек в пространстве признаков. Это существенно снижает доверие к полученным результатам, особенно наиболее интересным регрессионным моделям. Следует также отметить, что значения фенотипа являются случайными величинами, которые определяются не только генотипом, но и условиями внешней среды, например, количеством осадков, количеством солнечных дней, средней температурой, влияние которых в явном виде обычно не известно. Кроме того, во многих случаях исходные данные (матрицы генотипов) могут содержать пропущенные значения, что также снижает качество моделей.

Одним из подходов повышения качества моделей является предварительная обработка имеющихся данных с целью выявления каких-либо закономерностей в данных или их свойств, которые могли

бы использоваться в качестве дополнительной информации при построении стандартных моделей или алгоритмов. В работе [1] предложена модификация метода Лассо, используемого для полногеномного поиска ассоциаций, на примере анализа удвоенных гаплоидных линий ячменя для учета дополнительной информации о целевых значениях фенотипа, определяемого некоторым свойством растений. В рамках модификации формализуется дополнительная информация о свойствах растений в виде пересечения двух множеств весов, приписываемых элементам обучающей выборки. Первое множество образовано при помощи интервальной модели засорения. Второе множество весов образуется последовательностью парных сравнений значений фенотипа. Это позволяет повысить качество модели. Однако данная модификация не учитывает внутренней структуры информации, что делает ее применение не столь эффективным.

Поэтому в представленной работе предлагается простой метод определения значимости объектов популяции или обучающей выборки с точки зрения представительности того или иного объекта в установлении ассоциации между SNP и количественными признаками. Значимость представляется в виде весов объектов и может в дальнейшем использоваться при «взвешивании» элементов обучающей выборки в любых моделях GWAS. Метод основан на двух идеях. Первая была предложена в другом методе GWAS [2], где сравниваются пары объектов популяции с точки зрения расстояния между ними по фенотипу и генотипу. Вторая идея заключается в построении матрицы парных сравнений объектов специальным образом и вычислении весов объектов, например, по аналогии с известным методом анализа иерархий [12]. Предлагаемый метод может применяться не только к объектам биоинформатики, но также использоваться и в других областях, где решается задача поиска значимых элементов, например, в задачах надежности больших систем, где матрица состояний элементов являются аналогом матрицы генотипов, а состояния системы — значения фенотипа.

2. Анализ пар объектов популяции и вычисление их весов.

Рассмотрим общее формальное определение задачи поиска ассоциаций с математической точки зрения. Пусть $\mathbf{X} = [X_1, \dots, X_p]$ — матрица генотипов для n объектов и p SNP. Со статистической точки зрения SNP можно рассматривать как независимые переменные, т.е. $X_j = (x_{1j}, \dots, x_{nj})^T$ — переменная, представляющая j -ый SNP, $j = 1, \dots, p$. Каждая переменная x_{ij} — аллель i -го объекта на локусе j -го SNP. Переменная может принимать значения $\{0, 1\}$, где 0 и 1 определяются исходя из частоты аллелей. Генотип может быть представлен также

одним из чисел $\{0, 1, 2\}$. В частности, для представления гомозиготных аллелей используются обозначения ($AA = 0$) и ($aa = 2$), для гетерозиготных — ($Aa / aA = 1$). Необходимо отметить, что для значений аллелей x_{ij} не установлен порядок, и соответствующие признаки являются номинальными. Поэтому, с одной стороны, вместо чисел 0, 1 или 2 могут использоваться произвольные обозначения. С другой стороны, для построения регрессионной модели необходимо использовать числа 0, 1 и 2. Вектор аллелей, соответствующих i -му объекту, является столбцом матрицы генотипов \mathbf{X} и будет обозначаться $\mathbf{x}_i^T = (x_{i1}, \dots, x_{ip})$, $i = 1, \dots, n$. Также обозначим вектор значений фенотипа всех объектов $\mathbf{Y} = (y_1, \dots, y_n)$, где $y_i \in \mathbf{R}$.

Таким образом, цель GWAS заключается в поиске значимых SNP в \mathbf{X} , от которых сильно зависит \mathbf{Y} . Цель данной работы — назначить веса $\mathbf{W} = (w_1, \dots, w_n)$ объектам популяции в соответствии с их значимостью в определении ассоциации между SNP и \mathbf{Y} .

Первая основная идея, лежащая в основе определения весов предлагаемого метода, основана на сравнении генотипов пар объектов и сравнение соответствующих значений фенотипов этих объектов в паре. При этом мы используем следующее интуитивное предположение. Если генотипы двух объектов близки друг к другу, и соответствующие значения фенотипов этих двух объектов различны, то SNP-маркеры, которые соответствуют различным элементам рассматриваемых двух генотипов, могут быть значимыми. Действительно, если два объекта различаются малым числом элементов генотипов, то естественно ожидать, что их фенотипы близки или равны. Однако, если соответствующие фенотипы различны, то можно предположить, что это малое число различных элементов генотипов и определяет различие фенотипов. Так как значение фенотипа определяется не только генотипом, но и влиянием внешней среды, то нельзя делать такое заключение по одной паре, а необходимо анализировать все пары объектов. Таким образом, с учетом расстояний между генотипами и фенотипами для каждой пары объектов популяции можно определить некоторый совместный показатель, характеризующий влияние данной пары на определение ассоциации между SNP и количественным признаком.

Вторая идея, положенная в основу предлагаемого метода, заключается в рассмотрении совместного показателя влияния пары как степени предпочтения одного объекта по сравнению с другим с точки зрения увеличения значения фенотипа. Такое представление позволяет построить матрицу попарных сравнений объектов и далее определить веса объектов.

Для формализации приведенных выше рассуждений необходимо определить понятия близости фенотипов и генотипов для каждой пары объектов, характеризующихся векторами аллелей \mathbf{x}_i , \mathbf{x}_j и расстояниями y_i , y_j . Обозначим расстояние между двумя векторами \mathbf{x}_i и \mathbf{x}_j в соответствии с некоторой метрикой как $\rho_{ij} = \rho(\mathbf{x}_i, \mathbf{x}_j)$. Учитывая тот факт, что элементы векторов аллелей являются номинальными, наиболее простой метрикой с вычислительной точки зрения является расстояние Хэмминга. Расстояние Хэмминга определяется как число позиций, в которых соответствующие символы двух векторов одинаковой длины различны. Расстояние между значениями фенотипа объектов обозначим d_{ij} , и оно определяется как $d_{ij} = y_i - y_j$.

Одним из вариантов совместного показателя, характеризующего пары объектов, является отношение d_{ij} к ρ_{ij} , обозначаемое r_{ij} , т.е. $r_{ij} = (d_{ij})^q / \rho_{ij}$. Здесь разность d_{ij} для общности берется в степени q . Чем больше d_{ij} и меньше ρ_{ij} , тем больше показатель r_{ij} , что соответствует большей степени предпочтения i -го объекта перед j -ым объектом.

Если теперь построить матрицу $R = \{r_{ij}\}_{n \times n}$ размерности $n \times n$, то полученная матрица имеет свойства, позволяющие говорить о ее согласованности как матрицы парных сравнений. Заметим, что каждый элемент матрицы определяется через разность $y_i - y_j$, поэтому выполняется соотношение $r_{ij} = -r_{ji}$, которое определяет аддитивную шкалу матрицы парных сравнений. Будем говорить, что i -ый объект предпочтительнее j -го объекта, если $r_{ij} > 0$.

Матрица R является слабо-транзитивной, если из условий $r_{ij} > 0$ и $r_{jk} > 0$ следует $r_{ik} > 0$. Докажем, что матрица R , построенная при помощи предлагаемого алгоритма, является слабо-транзитивной. Без потери общности будем полагать, что $q = 1$. Если объекты упорядочены в порядке убывания, т.е. $y_1 \geq \dots \geq y_n$, то можно записать следующее условие для разности фенотипов: $d_{ik} = d_{ij} + d_{jk}$, которое следует из $(y_i - y_j) + (y_j - y_k) = y_i - y_k$. Так как расстояние ρ_{ij} удовлетворяет неравенству треугольника $\rho_{ik} \leq \rho_{ij} + \rho_{jk}$, то:

$$r_{ik} = \frac{d_{ik}}{\rho_{ik}} \geq \frac{d_{ij} + d_{jk}}{\rho_{ij} + \rho_{jk}}.$$

Из условий $r_{ij} > 0$ и $r_{jk} > 0$ следует, что $d_{ij} > 0$ и $d_{jk} > 0$. Отсюда получаем:

$$r_{ik} \geq \frac{d_{ij} + d_{jk}}{\rho_{ij} + \rho_{jk}} > 0,$$

что и следовало доказать.

Условие слабой транзитивности является минимальным требованием согласованности парных сравнений. Оно позволяет нам говорить о вычислении весов объектов w_i , $i = 1, \dots, n$, которые можно рассматривать, как меры вклада каждого вектора \mathbf{X}_i в ассоциацию между соответствующими SNP и значениями фенотипа. Согласно работе [4], веса объектов при условии аддитивной шкалы определяются как:

$$w_i = \frac{1}{n} \sum_{j=1}^n r_{ij}.$$

Большой вес i -го объекта означает, что большинство пар объектов с индексами (i, j) при фиксированном i и при $j = 1, \dots, n$ являются определяющими с точки зрения изменений фенотипа.

Другой вариант совместного показателя, характеризующего пары объектов, является:

$$r_{ij} = \exp\left(\left(d_{ij}\right)^q / \rho_{ij}\right).$$

Если $d_{ij} < 0$, то $r_{ij} \in (0, 1]$, иначе $r_{ij} > 1$. Кроме того, выполняется соотношение $r_{ij} = 1 / r_{ji}$. Также заметим, что $r_{ii} = 1$. Тогда можно говорить о мультипликативной шкале предпочтений. Докажем, что условие слабой транзитивности выполняется, т.е. из условий $r_{ij} > 1$ и $r_{jk} > 1$ следует $r_{ik} > 1$. Без потери общности полагаем, что $q = 1$. Используя снова неравенство треугольника для расстояний, запишем:

$$r_{ik} = \exp\left(\frac{d_{ik}}{\rho_{ik}}\right) \geq \exp\left(\frac{d_{ij} + d_{jk}}{\rho_{ij} + \rho_{jk}}\right).$$

Из условий $r_{ij} > 1$ и $r_{jk} > 1$ следует, что $d_{ij} > 0$ и $d_{jk} > 0$. Отсюда получаем:

$$r_{ik} \geq \exp\left(\frac{d_{ij} + d_{jk}}{\rho_{ij} + \rho_{jk}}\right) > \exp(0) > 1,$$

что и следовало доказать.

Для вычисления весов объектов w_i , $i = 1, \dots, n$ используем результаты работы [4], где веса объектов при условии мультипликативной шкалы определяются как:

$$w_i = \left(\prod_{j=1}^n r_{ij}\right)^{1/n}.$$

После нормировки веса объектов могут использоваться для решения задачи GWAS при применении большинства известных методов, рассмотренных кратко во вводной части статьи. В частности, одним из наиболее успешных является метод Лассо. Поэтому рассмотрим, как изменится этот метод с учетом полученных весов.

3. Метод Лассо с весами. В соответствии с методом Лассо решается задача построения линейной регрессионной модели:

$$y = \mathbf{X}\mathbf{b}^T + b_0 + \varepsilon.$$

Здесь ε — ошибка, распределение которой подчиняется нормальному закону с нулевым математическим ожиданием; \mathbf{X} — вектор значений SNP-маркеров, для которого необходимо определить значение фенотипа y ; \mathbf{b} — вектор неизвестных параметров, каждый элемент которого определяет, как соответствующий SNP влияет на значения фенотипа. Определение вектора \mathbf{b} является основной задачей GWAS. Параметр b_0 — свободный член. Если все переменные центрированы, то можно исключить свободный член из рассмотрения.

Следует отметить, что в общем случае задачу построения регрессионной и классификационной модели можно записать в виде следующей задачи оптимизации:

$$\mathbf{b}^0 = \arg \min_{\mathbf{b}} \{l(Y, \mathbf{X}, \mathbf{b}) + \lambda \cdot Q_{\lambda}(\mathbf{b})\},$$

где $l(Y, \mathbf{X}, \mathbf{b})$ — функция потерь, $Q_{\lambda}(\mathbf{b})$ — штрафное слагаемое,

ограничивающее множество возможных решений, \mathbf{b}^0 — оптимальные значения параметров \mathbf{b} , λ — неотрицательный параметр регуляризации или сглаживания, характеризующий степень влияния второго слагаемого в регрессионной модели.

В методе Лассо используется стандартная функция потерь $\|Y - \mathbf{Xb}\|^2$ и штрафное слагаемое вида:

$$Q(\mathbf{b}) = \|\mathbf{b}\|^1 = \sum_{j=1}^m |b_j|.$$

Если обозначить вектор весов объектов как $\mathbf{w} = (w_1, w_2, \dots, w_n)$, то метод Лассо с учетом вектора весов можно записать в виде следующей задачи оптимизации:

$$\mathbf{b}^0 = \arg \min_{\mathbf{b}} \left\{ \|Y - \mathbf{Xb}\|^2 \cdot \mathbf{w}^T + \lambda \cdot \|\mathbf{b}\|^1 \right\}.$$

4. Числовые эксперименты. Числовые эксперименты осуществлялись для двух популяций удвоенных гаплоидных (УГ) линий ячменя. Первое множество данных состоит из 93 УГ линий ячменя, описанных в работе [5]. Данные по фенотипам и генотипам можно найти на сайте (<http://wheat.pw.usda.gov/ggpages/maps/OWB/>). В экспериментах анализировались линии в соответствии с семью количественными признаками: длина колоса (SL) в см.; количество зерен (GN); количество цветков (FN); вес зерен в 100-граммовых порциях (HGW); высота растений (PH) в см.; число колосков (SN); время колошения (HD) в днях. Карта сцепления состоит из 1328 SNP-маркеров.

Второе множество данных состоит из 92 УГ линий ячменя, полученных от скрещивания Dicktoo x Morex и описанных в работе [8]. Данные по фенотипам и генотипам можно найти на сайте <http://wheat.pw.usda.gov/ggpages/DxM/>. В экспериментах анализировались линии в соответствии с двумя количественными признаками: время колошения с яровизацией и без яровизации с режимом светового периода 8-ч света/16-ч темноты. Карта сцепления состоит из 117 SNP-маркеров.

Пропущенные данные оценивались посредством эвристической процедуры, предложенной в работе [1].

Из каждого множества данных случайным образом выбираются два подмножества: множество обучающих n примеров и множество тестирующих данных в количестве n_{test} примеров для оценки качества

алгоритмов, которое оценивается при помощи средней квадратической ошибки регрессии (СКОР), которая определяется как:

$$E = \frac{1}{n_{test}} \sum_{i=1}^{n_{test}} (y_i - \hat{f}(x_i))^2,$$

где \hat{f} — функция, оцененная предложенным методом, $\hat{f}(x_i)$ — прогнозируемое значение фенотипа y_i для каждого $i \in \{1, \dots, n_{test}\}$. Показатель ошибки СКОР вычисляется при помощи усреднения результатов многократно повторяющегося случайного выбора тестирующих данных. Чем меньше значение показателя, тем лучше соответствующий метод. Мы используем метод кросс-валидации с одним тестирующим элементом, т.е. $n_{test} = 1$. Это связано с тем, что число линий не велико по сравнению с числом SNP и каждое наблюдение важно.

Значения СКОР для первого множества данных при условии $q = 1$ и выборе аддитивной шкалы построения матрицы парных сравнений приведены в таблице 1. Первый столбец таблицы соответствует семи анализируемым признакам. Столбцы 2 и 3 содержат значения СКОР, полученные с использованием 20 наиболее значимых SNP. Столбцы 4 и 5 содержат значения СКОР, полученные с использованием всех SNP. Сокращения СМ и ПМ означают стандартный метод и предлагаемый метод с весами объектов соответственно. Можно заметить из таблицы 1, что предлагаемый метод обеспечивает лучшие результаты почти для всех количественных признаков. Кроме того, наибольший эффект достигается, когда используются только значимые SNP.

Таблица 1. СКОР при выборе аддитивной шкалы и $q = 1$

Признаки	Значимые SNP		Все SNP	
	СМ	ПМ	СМ	ПМ
SL	0.994	0.942	2.675	2.603
GN	85.914	85.634	140.237	140.242
FN	44.645	36.553	83.954	84.026
HGW	0.124	0.118	0.201	0.191
PH	82.422	72.054	159.735	155.271
SN	12.815	11.728	16.828	16.243
HD	55.537	48.419	90.236	91.116

Для исследования влияния значения параметра q на точность моделирования рассмотрен также случай $q = 2$, для которого результаты анализа приведены в таблице 2. Из таблицы видно, что точность

определения значимых SNP, в этом случае повышается. Отсюда можно сделать вывод, что «правильный» выбор параметра q может привести к существенно лучшим результатам.

Таблица 2. СКОР при выборе аддитивной шкалы и $q = 2$

Признаки	Значимые SNP		Все SNP	
	СМ	ПМ	СМ	ПМ
SL	0.994	0.933	2.675	2.586
GN	85.914	85.001	140.237	138.954
FN	44.645	34.221	83.954	85.930
HGW	0.124	0.116	0.201	0.189
PH	82.422	70.936	159.735	152.579
SN	12.815	12.249	16.828	16.802
HD	55.537	47.632	90.236	91.130

Для исследования того, как влияет шкала построения матрицы парных сравнений, были проведены эксперименты с использованием мультипликативной шкалы, результаты которых представлены в таблице 3. Из сравнения всех приведенных таблиц нельзя однозначно сделать вывод о преимуществе той или иной шкалы. Для одних количественных признаков, например, для HD, выбор мультипликативной шкалы дает существенное снижение СКОР по сравнению с аддитивной шкалой. Однако для других признаков такого существенного улучшения не наблюдается. Выбор шкалы может рассматриваться в качестве еще одного «параметра» настройки модели на конкретные данные.

Таблица 3. СКОР при выборе мультипликативной шкалы и $q = 1$

Признаки	Значимые SNP		Все SNP	
	СМ	ПМ	СМ	ПМ
SL	0.994	1.008	2.675	2.661
GN	85.914	84.001	140.237	139.971
FN	44.645	35.696	83.954	84.220
HGW	0.112	0.107	0.201	0.201
PH	82.422	78.350	159.735	157.549
SN	12.815	12.023	16.828	16.793
HD	55.537	43.442	90.236	90.008

Рассмотрим второе множество данных. Таблица 4 содержит показатели ошибки для множества данных Dicktoo x Morex при анализе двух количественных признаков: время колошения с яровизацией и без яровизации. Из таблицы 4 следует, что предлагаемый метод показывает значительно лучшие результаты при выборе аддитивной шкалы и для первого признака. Однако это улучшение не столь очевидно для второго признака. В то же время выбор мультипликативной шкалы дает лучшие резуль-

таты, которые представлены в таблице 5, где для первого и второго признаков использовались параметры $q = 1$ и $q = 1/2$ соответственно.

Таблица 4. SKOP при выборе аддитивной шкалы и $q = 2$

Признаки	Значимые SNP		Все SNP	
	СМ	ПМ	СМ	ПМ
без яровизации	29.849	25.261	46.169	43.753
с яровизацией	25.996	23.136	38.241	39.480

Таблица 5. SKOP при выборе мультипликативной шкалы и $q = 1$ и $q = 1/2$

Признаки	Значимые SNP		Все SNP	
	СМ	ПМ	СМ	ПМ
без яровизации	29.849	22.974	46.169	40.331
с яровизацией	25.996	25.295	38.241	37.580

5. Заключение. В работе был предложен простой с вычислительной точки зрения метод назначения весов объектам обучающей выборки, который позволяет учесть целый ряд факторов. Казалось бы, почему не приписать веса объектам в соответствии со значениями фенотипов или в соответствии с некоторой областью вокруг значений фенотипов, как это было сделано в работе [1]. Действительно, такой подход может быть использован. Однако он не учитывает два важных фактора. Во-первых, значения фенотипа являются случайными благодаря внешним условиям. Во-вторых, каждый объект характеризуется генотипом, который может определять изменения фенотипа, а может и не определять. Следует отметить, что второй фактор является намного более важным. Для учета одновременно обоих факторов было предложено сравнивать пары объектов с учетом расстояний между фенотипами и генотипами. Оказывается, что при определенном сравнении пары образуют матрицы парных сравнений, свойства согласованности которых позволяют определить веса объектов. Так как вес получается в результате усреднения по подмножеству пар, то случайные факторы «сглаживаются», что позволяет учесть первый фактор.

Метод был применен к известному алгоритму Лассо, который повсеместно используется в GWAS. Однако это не значит, что предлагаемый метод не может работать с другими инструментами GWAS. Известные методы фильтрации, основанные на проверке статистических гипотез, методы упаковки, например, рекурсивное удаление признаков, с таким же успехом могут быть модифицированы с учетом полученных весов объектов. Эффективность применения метода в сочетании с другими подходами в GWAS является одним из направлений дальнейших исследований.

Заметим, что при получении весов объектов не было использовано никакой дополнительной «внешней» информации, т.е. веса полностью определялись с использованием только матрицы генотипов и вектора фенотипов. Возникает еще один интересный вопрос, почему, например, метод Лассо не учитывает особенностей предлагаемого метода. Дело в том, что метод Лассо, а также многие другие методы предполагают независимость SNP, т.е. независимость строк матрицы генотипов, что в большинстве случаев является слишком оптимистичным предположением. Определение весов некоторым образом корректирует это предположение, так как в их вычислении используются столбцы матрицы генотипов.

Результаты числовых экспериментов с реальными данными продемонстрировали, что метод позволяет получить более точные характеристики по сравнению со стандартным методом Лассо. Однако эти результаты также показали, что требуется определенная настройка параметров для получения наибольшего эффекта.

Литература

1. Уткин Л.В., Жук Ю.А., Колен Ф. Робастная модификация метода Лассо для полногеномного поиска ассоциаций с учетом целевых значений фенотипа // Научно-технический вестник информационных технологий, механики и оптики. 2016. Вып. 16. № 1. С. 150–160.
2. Уткин Л.В., Уткина И.Л. Быстрый алгоритм полногеномного поиска ассоциаций по схеме случай–контроль // Известия Санкт-Петербургской лесотехнической академии. 2015. Вып. 213. С. 263–273.
3. Altidor W., Khoshgoftaar T.M., Van Hulse J., Napolitano A. Ensemble feature ranking methods for data intensive computing applications // Handbook of Data Intensive Computing. Springer. New York. 2011. pp. 349–376.
4. Barzilai J., Golany B. Deriving weights from pairwise comparison matrices: The additive case // Operations Research Letters. 1990. vol. 9. pp. 407–410.
5. Cistue L. et. al. Comparative mapping of the Oregon Wolfe barley using doubled haploid lines derived from female and male gametes // Theoretical and applied genetics. 2011. vol. 122(7). pp. 1399–1410.
6. Goddard M.E., Wray N.R., Verbyla K., Visscher P.M. Estimating effects and making predictions from genome-wide marker data // Statistical Science. 2009. vol. 24(4). pp. 517–529.
7. Hayes B. Overview of statistical methods for genome-wide association studies (GWAS) // Methods in Molecular Biology. 2013. vol. 1019. pp. 149–169.
8. Hayes P. et. al. The Dicktoo x Morex population // Plant Cold Hardiness. Springer US. 1997. pp. 77–87.
9. Kohavi R., John G.H. Wrappers for feature subset selection // Artificial Intelligence. 1997. vol. 97(1–2). pp. 273–324.
10. Guyon I., Weston J., Barnhill S., Vapnik V. Gene selection for cancer classification using support vector machines // Machine Learning. 2002. vol. 46. pp. 389–422.
11. Moore J.H., Asselbergs F.W., Williams S.M. Bioinformatics challenges for genome-wide association studies // Bioinformatics. 2010. vol. 26(4). pp. 445–455.

12. Saaty T.L. Multicriteria Decision Making: The Analytic Hierarchy Process. New York: McGraw Hill. 1980.
13. Szymczak S. et al. Machine learning in genome-wide association studies // *Genetic Epidemiology*. 2009. vol. 33. pp. 51–57.
14. Tibshirani R. Regression shrinkage and selection via the Lasso // *Journal of the Royal Statistical Society. Series B (Methodological)*. 1996. vol. 58(1). pp. 267–288.
15. Zhang X., Huang S., Zhang Z., Wang W. Chapter 10: Mining Genome-Wide Genetic Markers // *PLoS Computational Biology*. 2012. vol. 8(12). pp. e1002828.
16. Zou H., Hastie T. Regularization and variable selection via the elastic net // *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2005. vol. 67(2). pp. 301–320.

References

1. Utkin L.V., Zhuk Y.A., Coolen F. [Robust modification of the Lasso method for genome-wide association study in view of target phenotype values]. *Nauchno-tehnicheskij vestnik informacionnyh tehnologij, mehaniki i optiki – Scientific and Technical Journal of Information Technologies, Mechanics and Optics*. 2016. vol. 16. no. 1. pp. 150–160. (In Russ.).
2. Utkin L.V., Utkina I.L. [A fast algorithm of a case-control genome-wide association study]. *Izvestija Sankt-Peterburgskoj lesotehnicheskij akademii – Izvestia Sankt-Peterburgskoj Lesotehnicheskij Akademii*. 2015. vol. 213. pp. 263–273. (In Russ.).
3. Altidor W., Khoshgoftaar T.M., Van Hulse J., Napolitano A. Ensemble feature ranking methods for data intensive computing applications. *Handbook of Data Intensive Computing*, Springer. New York, 2011. pp. 349–376.
4. Barzilai J., Golany B. Deriving weights from pairwise comparison matrices: The additive case. *Operations Research Letters*. 1990. vol. 9. pp. 407–410.
5. Cistue L. et al. Comparative mapping of the Oregon Wolfe barley using doubled haploid lines derived from female and male gametes. *Theoretical and applied genetics*. 2011. vol. 122(7). pp. 1399–1410.
6. Goddard M.E., Wray N.R., Verbyla K., Visscher P.M. Estimating effects and making predictions from genome-wide marker data. *Statistical Science*. 2009. vol. 24(4), pp. 517–529.
7. Hayes B. Overview of statistical methods for genome-wide association studies (GWAS). *Methods in Molecular Biology*. 2013. vol. 1019. pp. 149–169.
8. Hayes P. et al. The Dicktoo x Morex population. *Plant Cold Hardiness*. Springer US. 1997. pp. 77–87.
9. Kohavi R., John G.H. Wrappers for feature subset selection. *Artificial Intelligence*. 1997. vol. 97(1–2). pp. 273–324.
10. Guyon I., Weston J., Barnhill S., Vapnik V. Gene selection for cancer classification using support vector machines. *Machine Learning*. 2002. vol. 46. pp. 389–422.
11. Moore J.H., Asselbergs F.W., Williams S.M. Bioinformatics challenges for genome-wide association studies. *Bioinformatics*. 2010. vol. 26(4). pp. 445–455.
12. Saaty T.L. Multicriteria Decision Making: The Analytic Hierarchy Process. New York: McGraw Hill. 1980.
13. Szymczak S. et al. Machine learning in genome-wide association studies. *Genetic Epidemiology*. 2009. vol. 33. pp. 51–57.
14. Tibshirani R. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1996. vol. 58(1). pp. 267–288.
15. Zhang X., Huang S., Zhang Z., Wang W. Chapter 10: Mining Genome-Wide Genetic Markers. *PLoS Computational Biology*. 2012. vol. 8(12). pp. e1002828.
16. Zou H., Hastie T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2005. vol. 67(2). pp. 301–320.

Уткин Лев Владимирович — д-р техн. наук, профессор, профессор кафедры телематики (при ЦНИИ РТК), Санкт-Петербургский политехнический университет Петра Великого (СПбПУ). Область научных интересов: интеллектуальный анализ данных, представление неопределенности, принятие решений при неполной информации, теория надежности, биоинформатика. Число научных публикаций — 328. lev.utkin@gmail.com, <http://levvu.narod.ru>; Политехническая, 29, Санкт-Петербург, 195251; р.т.: +7(812)5526521.

Utkin Lev Vladimirovich — Ph.D., Dr. Sci., professor, professor of telematics (under Russian state scientific center for robotics and technical cybernetics (RTC)) of Institute of applied mathematics and mechanics, Peter the Great Saint-Petersburg Polytechnic University. Research interests: machine learning, uncertainty representation, decision making under incomplete information, reliability theory, bioinformatics. The number of publications — 328. lev.utkin@gmail.com, <http://levvu.narod.ru>; 29, Polytechnicheskaya, St. Petersburg, 195251; office phone: +7(812)552-6521.

Жук Юлия Александровна — к-т пед. наук, доцент кафедры компьютерных образовательных технологий, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики (Университет ИТМО). Область научных интересов: когнитивные образовательные технологии, интеллектуальный анализ данных, мультимедийные технологии. Число научных публикаций — 73. zhuk_yua@mail.ru, <http://dce.ifmo.ru>; Кронверкский пр., 49, Санкт-Петербург, 197101; р.т.: +7(812)233-1961.

Zhuk Yulia Alexandrovna — Ph.D., associate professor of computer educational technologies department, ITMO University (Saint Petersburg National Research University of Information Technologies, Mechanics and Optics). Research interests: cognitive educational technologies, data mining, multimedia technologies. The number of publications — 73. zhuk_yua@mail.ru, <http://dce.ifmo.ru>; 49, Kronverksky pr., St. Petersburg, 197101; office phone: +7(812)233-1961.

РЕФЕРАТ

Уткин Л.В., Жук Ю.А. **Полногеномный поиск ассоциаций с использованием матрицы парных сравнений.**

Основная цель полногеномного поиска ассоциаций заключается в определении, какие ДНК-маркеры, называемые однонуклеотидными полиморфизмами, оказывают наибольшее влияние на значения фенотипа. С точки зрения теории машинного обучения имеем задачу отбора признаков.

Одним из подходов повышения качества моделей является предварительная обработка имеющихся данных с целью выявления каких-либо закономерностей в данных, которые могли бы использоваться в качестве дополнительной информации при построении стандартных моделей или алгоритмов. Предлагается простой метод определения значимости объектов популяции или обучающей выборки с точки зрения представительности того или иного объекта в установлении ассоциации между ДНК-маркерами и количественными признаками. Значимость представляется в виде весов объектов и может в дальнейшем использоваться при «взвешивании» элементов обучающей выборки в любых моделях полногеномного поиска ассоциаций. Метод основан на двух идеях. Первая идея — сравнение пар объектов популяции с точки зрения расстояния между ними по фенотипу и генотипу. Вторая идея заключается в построении матрицы парных сравнений объектов специальным образом и вычислении весов объектов, например, по аналогии с известным методом анализа иерархий.

Показано, как найденные веса могут быть использованы в методе Лассо. Результаты числовых экспериментов с реальными данными (удвоенные гаплоидные линии ячменя) продемонстрировали, что метод позволяет получить более точные характеристики по сравнению со стандартным методом Лассо.

SUMMARY

Utkin L.V., Zhuk Y.A. **A Genome-Wide Association Study using Pairwise Comparison Matrices.**

The main aim of a genome-wide association study is to determine which DNA-markers, called the single nucleotide polymorphisms, have the greatest influence on the phenotype values. This is a feature selection problem in the machine learning framework.

One of the approaches to enhance the model quality is preprocessing of available data in order to reveal some dependencies between data, which could be used as additional information for constructing the standard models and algorithms. We propose a simple method for determining the importance of individuals of a population or the training set elements according to the individual representativeness in looking for the association between DNA-markers and quantitative traits. The importance is represented in the form of individual weights and can be used for weighing elements of the training set in arbitrary genome-wide association study models. The method is based on two ideas. The first one is to compare pairs of the individuals of population in terms of distances between phenotypes and between genotypes. The second idea is to construct a pairwise comparison matrix for individuals in a special way and to compute the weights of individuals, for example, similarly to the well-known analytic hierarchy process.

It is shown how the computed weights can be applied to the Lasso method. The results of numerical experiments with real datasets (double haploid lines of barley) illustrated that the proposed method allows us to obtain more accurate measurements in comparison with the standard Lasso method.

РУКОВОДСТВО ДЛЯ АВТОРОВ

Взаимодействие автора с редакцией осуществляется через личный кабинет на сайте журнала «Труды СПИИРАН» <http://www.proceedings.spiiras.nw.ru>. При регистрации авторам рекомендуется заполнить все предложенные поля данных.

Подготовка статьи ведется с помощью текстовых редакторов MS Word 2007 и выше. Объем основного текста – от 15 до 25 страниц включительно. Формат страницы документа – А5 (148 мм ширина, 210 мм высота); ориентация – портретная; все поля – 20 мм. Верхний и нижний колонтитулы страницы – пустые. Основной шрифт документа – Times New Roman, основной кегль (размер) шрифта – 10 pt. Переносы разрешены. Абзацный отступ устанавливается размером в 10 мм. Межстрочный интервал – одинарный. Номера страниц не проставляются.

В основную часть допускается помещать рисунки, таблицы, листинги и формулы. Правила их оформления подробно рассмотрены на нашем сайте в разделе «Руководство для авторов».

AUTHOR GUIDELINES

Interaction between each potential author and the Editorial board is realized through the personal account on the website of the journal "Proceedings of SPIIRAS" <http://www.proceedings.spiiras.nw.ru>. At the registration the authors are requested to fill out all data fields in the proposed form.

The submissions should be prepared using MS Word 2007 text editor or higher versions, at that, only manuscripts in *.docx format will be considered. The text of the paper in the main part of it should be from 15 – 25 pages of A5 size that is 210 X 148 mm; orientation – portrait; all margins – 20 mm. The font of the main paper text is Times New Roman of 10 pt font size. The pages' headers and footers should be empty; indentation – 10 mm; line spacing – single; pages are not numbered; hyphenations are allowed.

Certain figures, tables, listings and formulas are allowed in the main section, and their typography is considered by the paper template in more detail in journal web.

ISSN 2078-9181



9 772078 918785 >



Вып. 4(47)

ТРУДЫ СПИИРАН
proceedings.spiiras.nw.ru

С-Пб
2016