

# ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

11(207)  
2013

ТЕОРЕТИЧЕСКИЙ И ПРИКЛАДНОЙ НАУЧНО-ТЕХНИЧЕСКИЙ ЖУРНАЛ

Издается с ноября 1995 г.

УЧРЕДИТЕЛЬ  
Издательство "Новые технологии"

## СОДЕРЖАНИЕ

### ОБЩИЕ ВОПРОСЫ

Майоров А. А., Цветков В. Я. Геоинформатика как важнейшее направление развития информатики . . . . . 2

### ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ

Филимонов А. Б., Кромин О. А. Механизмы коммуникаций семантических JADE-агентов. . . . . 8

### МОДЕЛИРОВАНИЕ И ОПТИМИЗАЦИЯ

Васильев А. В. Численно-аналитическое исследование математической модели дробного осциллятора методом S-преобразования. . . . . 15

Казakov П. В. Модель динамического изменения размера популяции для генетических алгоритмов многокритериальной оптимизации . . . . . 20

Федосов В. В., Федосова А. В. Стохастический поиск защитных нормативов объектов (зон), размещенных на общей территории с группой источников промышленных выбросов . . . . . 27

Максименко-Шейко К. В., Толлок А. В., Шейко Т. И. Сопряженная задача конвективного теплообмена в топливной кассете ТВЭЛов . . . . . 32

### БЕЗОПАСНОСТЬ ИНФОРМАЦИИ

Востриков А. А. О матрицах Адамара—Мерсенна и маскировании изображений . 37

Беляева Е. А. Сравнительная оценка аппаратно-программных модулей доверенной загрузки на основе классификации их функциональных возможностей . . 40

Еременко А. В., Сулавко А. Е. Исследование алгоритма генерации криптографических ключей из биометрической информации пользователей компьютерных систем . . . . . 47

### КОДИРОВАНИЕ И ОБРАБОТКА СИГНАЛОВ

Домунян А. А. Метод распознавания жестов руки на основе позиционного расстояния . . . . . 52

### Журнал в журнале НЕЙРОСЕТЕВЫЕ ТЕХНОЛОГИИ

Гриняк В. М., Девятисильный А. С. Нейронечеткая обучаемая система распознавания воздушных объектов . . . . . 58

Горбатков С. А., Белолипцев И. И., Солнцев А. М. О состоятельности процедур регуляризации нейросетевых моделей налогового и финансового контроля . . 64

Contents . . . . . 69

Приложение. Бородкин А. А., Некрасов И. В., Толчеев В. О. Методы повышения быстродействия непараметрических классификаторов в задачах обработки и анализа библиографических текстовых документов

Главный редактор:  
СТЕМПКОВСКИЙ А. Л.,  
акад. РАН, д. т. н., проф.

Зам. главного редактора:  
ДИМИТРИЕНКО Ю. И.,  
д. ф.-м. н., проф.  
ФИЛИМОНОВ Н. Б., д. т. н., с.н.с.

Редакционный совет:  
БЫЧКОВ И. В., акад. РАН, д. т. н.  
ЖУРАВЛЕВ Ю. И.,  
акад. РАН, д. ф.-м. н., проф.  
КУЛЕШОВ А. П.,  
акад. РАН, д. т. н., проф.  
ПОПКОВ Ю. С.,  
чл.-корр. РАН, д. т. н., проф.  
РУСАКОВ С. Г.,  
чл.-корр. РАН, д. т. н., проф.  
СОЙФЕР В. А.,  
чл.-корр. РАН, д. т. н., проф.  
СОКОЛОВ И. А.,  
акад. РАН, д. т. н., проф.  
СУЕТИН Н. В., д. ф.-м. н., проф.  
ЧАПЛЫГИН Ю. А.,  
чл.-корр. РАН, д. т. н., проф.  
ШАХНОВ В. А.,  
чл.-корр. РАН, д. т. н., проф.  
ШОКИН Ю. И.,  
акад. РАН, д. т. н., проф.  
ЮСУПОВ Р. М.,  
чл.-корр. РАН, д. т. н., проф.

Редакционная коллегия:  
АВДОШИН С. М., к. т. н., доц.  
АНТОНОВ Б. И.  
БАРСКИЙ А. Б., д. т. н., проф.  
ВАСЕНИН В. А., д. ф.-м. н., проф.  
ГАЛУШКИН А. И., д. т. н., проф.  
ДОМРАЧЕВ В. Г., д. т. н., проф.  
ЗАГИДУЛЛИН Р. Ш., к. т. н., доц.  
ЗАРУБИН В. С., д. т. н., проф.  
ИВАННИКОВ А. Д., д. т. н., проф.  
ИСАЕНКО Р. О., к. т. н., с.н.с.  
КАРПЕНКО А. П., д. ф.-м. н., проф.  
КОЛИН К. К., д. т. н., проф.  
КУЛАГИН В. П., д. т. н., проф.  
КУРЕЙЧИК В. М., д. т. н., проф.  
КУХАРЕНКО Б. Г., к. ф.-м. н., доц.  
ЛЬВОВИЧ Я. Е., д. т. н., проф.  
МАЛЬЦЕВ П. П., д. т. н., проф.  
МИХАЙЛОВ Б. М., д. т. н., проф.  
НЕЧАЕВ В. В., к. т. н., проф.  
ПАВЛОВ В. В., д. т. н., проф.  
РЯБОВ Г. Г., чл.-корр. РАН, д. т. н., проф.  
СОКОЛОВ Б. В., д. т. н., проф.  
УСКОВ В. Л., к. т. н. (США)  
ФОМИЧЕВ В. А., д. т. н., проф.  
ЧЕРМОШЕНЦЕВ С. Ф., д. т. н., проф.  
ШИЛОВ В. В., к. т. н., доц.

Редакция:  
БЕЗМЕНОВА М. Ю.  
ГРИГОРИН-РЯБОВА Е. В.  
ЛЫСЕНКО А. В.  
ЧГУНОВА А. В.

Информация о журнале доступна по сети Internet по адресу <http://novtex.ru/IT>.  
Журнал включен в систему Российского индекса научного цитирования.  
Журнал входит в Перечень научных журналов, в которых по рекомендации ВАК РФ должны быть опубликованы научные результаты диссертаций на соискание ученой степени доктора и кандидата наук.

УДК 528:004.528

**А. А. Майоров**, д-р техн. наук, ректор,  
**В. Я. Цветков**, д-р техн. наук, проф.,  
Московский государственный университет  
геодезии и картографии,  
e-mail: cvj2@mail.ru

## Геоинформатика как важнейшее направление развития информатики

*Рассмотрена эволюция информатики и геоинформатики. Показаны стадии эволюции информатики в виде "информатики 1", "информатики 2", "информатики 3". Рассмотрены истоки происхождения и особенности развития геоинформатики, а также современные проблемы этой науки. Приведены сходство и различие между информатикой и геоинформатикой. Показано, что информатика в предметных областях строится на специализации методов обработки предметной области и специализации данных на этой основе, а геоинформатика — на интеграции данных и применении их в разных областях. Рассмотрены проблемы информатики и геоинформатики. Отмечена необходимость интеграции этих направлений. Показано, что геоинформатика является интеграционной основой объединения "информатики 1", "информатики 2", "информатики 3".*

**Ключевые слова:** геоинформатика, информатика, геоаналитика, геоинформационный подход, геоинформационные технологии

### Введение

Современное развитие цивилизации тесно связано с освоением Земли и с науками о Земле. Одной из первых наук о Земле была геометрия. В 1878 г. в Париже была основана Federation Internationale des Geometres (FIG) [1]. Геометрия переводится как "измерение Земли", а "геодезия" — как "деление Земли". Однако с течением времени геометрия стала разделом математики и теоретической наукой, а практической наукой, связанной с измерениями и исследованиями земной поверхности, стала геодезия. Поэтому в настоящее время в иностранных и российских источниках FIG интерпретируется как Международная федерация геодезистов.

Дифференциация и специализация наук о Земле породили различные направления: картографию, фотограмметрию, геологию, геодезию, географию, геодинамику и др. Освоение космического про-

странства привело к появлению нового направления в исследовании Земли — дистанционного космического зондирования. Огромное число данных, получаемых в разных направлениях, особенно в аспекте глобального изучения процессов на земной поверхности и в околоземном пространстве, потребовало их интеграции, обобщения и совместного анализа. Ответом на это требование стало появление нового научного направления — геоинформатики (геоматики). Геоинформатика интегрирует основные направления наук о Земле на основе информационных технологий, систем автоматизированного проектирования и методов обработки информации.

Дополнительным аргументом в создании геоинформатики явились работы в области получения пространственного знания. Начиная с 60-х годов прошлого столетия, специалисты в области искусственного интеллекта занимаются проблемой получения и применения пространственного знания [2]. Появление геоинформатики сблизило абстрактное пространственное знание в области искусственного интеллекта с реальным пространством и реальными пространственными объектами геоинформатики.

### Эволюция информатики

Информатика возникла как наука о программировании (computer science) и обработке информации. Термин "информатика" возник в 1960-х годах во Франции для названия области, занимающейся автоматизированной переработкой информации, как слияние французских слов information и automatique (F. Dreyfus, 1962 г.).

В разных странах информатика (ср. нем. informatik, англ. Information technology, фр. informatique, англ. computer science — компьютерная наука — в США, англ. computing science — вычислительная наука — в Великобритании) интерпретировалась как наука о способах получения, накопления, хранения, преобразования, передачи, защиты и использования информации. При этом термин "информация" не связывался с какой-либо предметной областью, а был обобщением информации как *объекта обработки* на компьютере.

В СССР/России эволюция термина "информатика" имеет свою историю.

Довольно долго в СССР/России термин "информатика" использовался как синоним термина "программирование". В России курсы по изучению информатики длительное время включали изучение программирования, в первую очередь, и применение информационных систем — во вторую.

Объектом изучения и областью исследований информатики, или ее доминантой, являются методы обработки информации, компьютерные модели, алгоритмы анализа и вычислений — *безотносительно* к области использования этой информации. Следует подчеркнуть, что информатика занимается именно обработкой информации, а не теорией обработки.

Теория информации нужна философам для познания сущности мира, лингвистам и филологам — для понимания смысла, историкам — для отражения связи времен, анализа развития общества. Смысл и знание — вот в чем нуждались гуманитарии. Впоследствии эта область была обозначена как "информатика 1" [3]. В англоязычной научной литературе она имеет аналог — information sciences.

К концу 1940-х годов в связи с частым употреблением словосочетания "научная информация" ученые ВИНТИ предложили термин "теория научной информации". Эта позиция нашла отражение в работах А. И. Михайлова, В. А. Полушкина, А. И. Черного, Р. С. Гиляревского и др. (1962—1965 гг.) [3].

Работа К. Э. Шеннона [4] при ее написании никакого отношения не имела к теории информации, ее целевым назначением было создание общей теории связи [5].

Термин "информатика" с момента его появления в 1963 г. [6] приобрел в нашей стране несколько значений. Воспользуемся анализом, приведенным в работе [3].

В 1952 г. в Москве в системе Академии наук был создан Институт научной информации (ИНИ), в 1955 г. переименованный в ВИНТИ — Всесоюзный институт научной и технической информации. В Постановлении АН СССР от 25 июля 1952 г. № 458 указывалось, что Институт научной информации создается "в целях систематической информации научных и инженерно-технических работников промышленности, научно-исследовательских учреждений и высших учебных заведений о новых работах в области науки и техники, публикуемых в СССР и зарубежных странах".

В 1966 г. авторы монографии "Основы научной информации" [7] опубликовали статью "Информатика — новое название теории научной информации" [8], которая была с интересом встречена мировым научным сообществом и переведена на английский, японский и немецкий языки.

В 1968 г. вышла в свет новая книга А. И. Михайлова, А. И. Черного и Р. С. Гиляревского "Основы информатики" [9] — переработанное и расширенное издание монографии "Основы научной информации". Книга была полностью переведена и издана в ГДР (1970 г.), ФРГ (1970 г.), Италии (1973 г.) и на Кубе (1973 г.).

Так возникло первое направление информатики в России, которое интерпретируется как "Теория научно-информационной деятельности", или "информатика-1".

С современных позиций можно определить "информатику 1" как семантико-когнитивное направле-

ние исследования информационных процессов и информационных объектов, направленное, в первую очередь, на исследование семантики и содержательности информации и информационных сообщений.

Применение "информатики 1" в любой области основано на анализе содержательности, семантики и когнитивных процессов в этой области. В результате такого исследования осуществляется обобщение, накопление и междисциплинарный перенос знаний.

В 1976 г. ровно через 10 лет после выхода работы [8] был опубликован перевод книги профессоров Мюнхенского технического университета Ф. Л. Бауэра и Г. Гооза "Информатика. Вводный курс" [10] под редакцией член-корр. АН СССР А. П. Ершова.

Приведем фрагменты из предисловия редактора перевода.

1. "Можно было бы не тратить слов на объяснение перевода названия, если бы не "перехват" термина, осуществленный десять лет назад специалистами по документалистике и информационно-поисковым системам. Редактор имеет в виду книгу А. И. Михайлова и А. И. Черного "Основы информатики" (Наука, М., 1968) [9], в которой описываются методы существенно более узкой области применения ЭВМ". ...<sup>1</sup>

2. "Отличительной особенностью информатики является в целом удавшаяся попытка авторов преподавать основы программирования и обработки информации как курс, излагающий сложившиеся концепции и понятия, которые, с одной стороны, формализуются в некотором математическом аппарате и, с другой стороны, раскрываются в виде конструкций обобщенного машинного языка" ...<sup>2</sup>

3. "На первый взгляд, особенно при чтении начальных глав, изложение может показаться несколько калейдоскопическим".<sup>3</sup>

Подчеркнутое в пп. 1, 2, по мнению авторов статьи, не соответствует действительности и опровергается временем. Но в целом так формировалась еще одна информатика, которую называют "информатика-2" [3] — "наука о вычислительных машинах и их применении", которая имеет зарубежный аналог как computer science.

Как показали прошедшие десятилетия, кое-что из книги [10] не прижилось и не вошло в курсы информатики, читаемые в СССР/России. Это, прежде всего, психофизические аспекты интерпретации

<sup>1</sup> С точки зрения истории и логики ни о каком "перехвате" и речи не могло быть. Ученые этого направления последовательно и сознательно, самостоятельно пришли к этому термину, вложив в него определенный смысл.

<sup>2</sup> С точки зрения истории и логики о сложившихся концепциях было говорить рано. Тематика книги по охвату много шире, чем программирование. Она включает кодирование и распознавание образов, психофизические явления и психофизическую интерпретацию информационных процессов. Это было исключено в той информатике, о которой говорит А. П. Ершов. Ученые этого направления, взяв зарубежный термин, сузили объект его описания и вложили в него иной смысл, чем авторы монографии [10].

<sup>3</sup> Последняя фраза касается книги Ф. Л. Бауэра и Г. Гооза [10]. Уже тогда А. П. Ершов чувствовал несоответствие между программированием и содержанием книги.

информационных образов (это область "информатики 1") и собственно теория информации, а не фрагменты из теории связи в виде отдельных формул. Поэтому мнение А. П. Ершова о книге [10] следует считать преувеличенным.

С современных позиций можно определить "информатику 2" как научное направление исследования информационных процессов, систем и технологий, касающееся, в первую очередь, исследования сбора, обработки и представления информации и информационных сообщений.

С технологических позиций "информатика 2" — посредник между математикой и логикой, с одной стороны, и прикладными науками — с другой. Она возникла и развивается как наука об обработке информации с помощью вычислительных машин. На рис. 1 показано отношение информатики и других наук.

Применение "информатики 2" в любой области основано на *специализации и адаптации "информатики 2"* к методам, задачам и данным этой предметной области. В результате такой специализации появляется специфическая информатика в данной предметной области (см. рис. 1).

Эта специализированная информатика возникает на основе *дифференциации информатики* применительно к проблемной области.

Однако важно подчеркнуть, что при этом вектор такой специализированной информатики меняется. Главным для нее является не обработка информации как таковая (доминанта "информатики 2"), а *решение проблем и задач в своей области* с использованием информатики.

Следует отметить, что любая специализированная информатика это, в первую очередь, *решение задач предметной области* и только затем — обработка информации, или "информатика 2". По номенклатуре специальностей ВАК "информатика 2" относится к категории 05.00.00 — технические науки.

"Специализированные информатики" возникают на основе *дифференциации "информатики 2"* применительно к проблемной области.

Таким образом, развитие "информатики 2" происходит путем совершенствования методов обработки, с одной стороны, и ее дифференциации и адаптации к разным предметным областям — с другой. Данные, применяемые в "информатике 2", — это компьютерные модели, отражающие предметную область исследования.



Рис. 1. Информатика и ее взаимосвязь с другими науками

Одним из итогов развития "информатики 2" можно считать "информациологию" И. И. Юзвщина [11], написанную на основе субстанционального понимания информации.

В работе [12] автор признает неудачным название термина "информатика" и предлагает другой, более подходящий по его мнению, "компьютинг". Это еще раз дает основание считать ошибочным мнение А. П. Ершова "о перехвате термина".

Но в целом можно говорить об определенном застое в развитии информатики. В первую очередь это связано с тем, что теория информации оказалась вне "информатики 2" и в рамках "информатики 2" не развивалась. Следствием этого явилось новое направление, возглавляемое Лучиано Флориди [13, 14], по созданию семантической теории информации. Оно существенно ближе "информатике 1".

В ИНИ АН РАН с 2011 г. в рамках "информатики 1" проводится семинар "Методологические проблемы наук об информации". Начало было положено обсуждением работы Л. Флориди "Открытые проблемы в философии информации" [15], опубликованной в 2004 г., в которой сформулированы 18 проблем. Одна из них гласит: "Возможна ли большая объединенная теория информации?". Дискуссия продолжается.

#### Эволюция геоинформатики

Геоинформатика [16, 17] возникла вследствие необходимости обобщения методов в области наук о Земле (гео) и интеграции методов получения пространственного знания (частично "информатика 1", но более близки методы искусственного интеллекта) с науками о Земле на основе систем и методов автоматизированного проектирования (САПР) и методов обработки информации ("информатика 2"). Впоследствии появился ее двойник — геоматика.

Объективная потребность в интеграции наук о Земле: геодезии (25.00.32), фотограмметрии (25.00.34), картографии (25.00.33), дистанционного зондирования Земли (25.00.34), землеустройства, кадастра и мониторинга земель (25.00.26) — назрела давно. В скобках дана номенклатура специальностей ВАК. То есть по номенклатуре специальностей геоинформатика относится к иному классу 25.00.00 — Наука о Земле.

Преподавание перечисленных дисциплин отличалось большим взаимопроникновением. На практике часто геодезисты работали в фотограмметрии, а фотограмметристы работали на геодезических работах или занимались составлением карт и т. д.

На основе развития САПР и "информатики 2" появилась возможность интеграции наук о Земле в единую систему наук [18]. Эта *интегрированная система наук* о Земле называется **геоинформатика** (25.00.35) (рис. 2).

Следует отметить, что наряду с геоинформатикой существует дисциплина "информатика в геодезии" (до недавнего времени такие специалисты

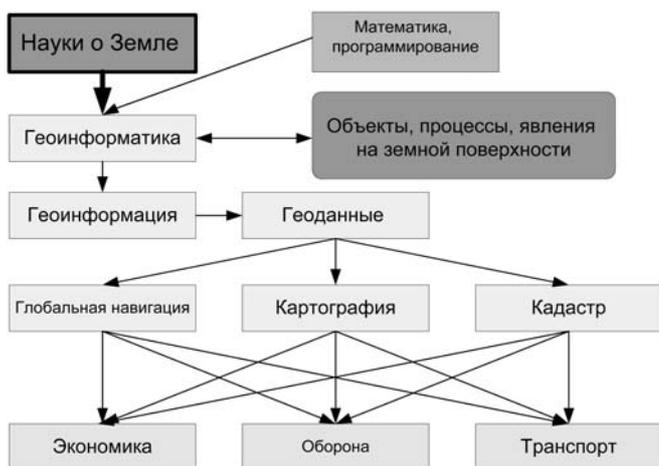


Рис. 2. Геоинформатика и ее взаимосвязь с другими науками

готовились в вузах России), "информатика в геологии", "информатика в фотограмметрии", "информатика в картографии". Но в отличие от геоинформатики это дифференцированные науки, которые создавались на основе дифференциации информатики. Геоинформатика создавалась на основе интеграции. Зарубежные синонимы геоинформатики имеют такие значения: "geo science", "geoinformatica" и "geo computer science".

Геоинформатика имеет свои специфические данные, эти данные называют геоданными [19, 20].

Для понимания геоданных необходимо обратиться к словам, имеющим приставку "гео". *Гео* (от греч. гея — Земля) — часть сложных слов, означающая: относящийся к Земле, к ее изучению. В название ряда наук "гео" входит формально и содержательно как составная часть (геометрия, геодезия, география, геология, геодинамика, геоинформатика, геоматика, геомаркетинг и др.), но в ряде наук "гео" в явном виде не входит, но входит содержательно (транспорт, архитектура, земледелие, землепользование, кадастр, управление недвижимостью, распределенные системы, логистика, космические исследования, фотограмметрия, картография, мировая экономика, социальные процессы и явления, развитие человеческого общества и др.). Таким образом, области, на которые распространяется содержательная часть "гео", приводит к понятию геоданных.

*Геоданные* — тематические, пространственные и временные данные, отражающие свойства объектов, процессов и явлений, происходящих на Земле, в околоземном пространстве, под земной поверхностью.

Следует отметить, что в науках о Земле (не только в геоинформатике, но и в геодезии, геологии, геодинамике и др.) иногда злоупотребляют приставкой "гео".

Говорят о геопространственных данных, которые по законам формирования понятий являются и подмножеством данных, и подмножеством геоданных. Однако часто этот термин используют как синоним или альтернативу термину "геоданные", т. е.

подмножество отождествляют с множеством. Термин "геопространственный" имеет право на существование, но по законам образования дефиниций этот термин имеет меньший объем понятия, чем "пространственный".

Известно, что пространство и время — это разные категории. Поскольку термин "пространственный" не описывает временные характеристики и не эквивалентен термину "временной", он не может служить описанием временных характеристик. В отличие от этого, геоданные включают три группы: "место", "время", "тема".

Поэтому использование термина "геопространственные данные" не допустимо как эквивалент терминов "геоданные" или "пространственные данные". Однако это имеет место в геоинформатике и в геодезии.

В геоинформатике и в геодезии говорят о геопорталах, но часто бывает, что это по существу обычные порталы, которые хранят пространственную информацию.

Некоторые специалисты говорят о геопространстве или о геоинформационном пространстве, но не дают четкого определения, что это такое и как оно соотносится с другими видами пространств. Говорят о геопространственной информации как о тождестве геоинформации.

Некоторые специалисты в области геодезии и геоинформатики уже начинают разрабатывать свой математический аппарат. Например, вводят термин "метрические графы", но не определяют, что это такое и чем "метрический граф" отличается от "графа в метрическом пространстве". В отдельных работах по геоинформатике дают свое толкование термину "базы данных" или термину "метаданные". Все это примеры научной некорректности, с которыми надо бороться.

Можно констатировать, что развитие науки "геоинформатика" породило много спекуляций на термине "гео" и попытках вместо создания новых теорий вводить новые термины как альтернативу существующим или новые термины, которые не укладываются в существующую в науке (геоинформатике и информатике) систему терминологических отношений [21]. Часто отсутствие научной новизны не только в геоинформатике, но даже в геодезии и в оптике прикрывают введением нового необоснованного термина, который имеет известный эквивалент, или термина, который не имеет объяснения.

Объектом изучения геоинформатики или ее доминантой являются пространственные отношения [22], процессы и явления, происходящие на Земле, в околоземном пространстве, под земной поверхностью. В геоинформатике геоданные изначально структурированы. Они содержат три группы:

$$GD = \Phi\{(C1, C2, \dots, Cn), \\ (Pt1, Pt2, \dots, Ptm), (A1, A2, \dots, Al)\}.$$

Здесь  $C_i$  — совокупность координатных (пространственных) параметров ( $i = 1, \dots, n$ );  $Pt_i$  — совокупность временных параметров ( $i = 1, \dots, m$ );

$A_i$  — совокупность тематических характеристик ( $i = 1, \dots, k$ ).

Такая структура данных делает их удобными для моделирования в пространстве и времени. Геоданные преобразуются в *цифровые карты*, *цифровые модели* [23] и *тематические модели предметной области*. Затем геоданные преобразуют в *геоинформационные модели* (модели описания) и в *интегрированную информационную основу* (основу обработки геоинформации), в которой содержатся данные *разных предметных областей*.

В результате комплексной обработки с использованием интегрированной информационной основы получают специализированные комплексные наборы данных как описание для каждой предметной области. Особенностью является интегрированная информационная основа [24], которая включает данные разных предметных областей. Но при решении практических задач используют те данные, которые необходимы.

Интересно сравнение применения топологии в геоинформатике и информатике. В информатике топология используется в алгоритмическом аспекте как инструмент анализа информационных потоков (потоков данных). В геоинформатике топология применяется в алгоритмическом аспекте как инструмент обработки геоинформации, а в пространственном аспекте — как инструмент пространственного описания геоданных. Это можно рассматривать как развитие информатики.

Сбор данных в "информатике 2" и геоинформатике осуществляется по-разному. В информатике данные — это *специализированные* данные одной предметной области, предназначенные для компьютерной обработки. Такие данные предназначены для решения специализированных задач только в этой предметной области.

Например, информатика в геодезии предназначена для решения задач в области геодезии с использованием информационных систем и информационных технологий и не решает задачи в области картографии, фотограмметрии, космической съемки, в области кадастра, мониторинга земель и т. д.

Сбор геоданных в геоинформатике осуществляется набором разных технологий: воздушная и наземная фотограмметрия, геодезия, картография, геостатистика, статистика и пр. В геоинформатике большое число разнообразных первичных данных подвергают предобработке, которая включает: унификацию, коррекцию и интеграцию данных в единую среду, введение топологии и ассоциативных связей.

Данные геоинформатики, применяемые при решении практических задач, — это *интегрированные* данные, предназначенные для использования в разных областях. Например, интегрированные данные в ГИС позволяют решать задачи в области картографии, фотограмметрии, космической съемки, в области кадастра, мониторинга земель и т. д. То есть налицо полная противоположность *информатике в геодезии* (см. выше).

Геоинформатика имеет сходство с "информатикой 2" в части хранения и обработки геоинформации [25], поэтому уместно сравнить геоинформационный и информационные подходы.

Информационный и геоинформационный подходы включают использование: информационных единиц, (гео)информационных моделей, (гео)информационного мониторинга [26], (гео)информационного моделирования, информационных потоков.

Основой геоинформационного подхода является выделение трех групп геоданных: "место", "время", "тема", которые позволяют проводить классификацию различных данных и последующую интеграцию их в единую информационную основу.

Интегрирующей функцией при геоинформационном подходе обладают цифровые модели и цифровые карты [23], что создает возможность объединять разнородные информационные ресурсы. Интегрирующая функция дает возможность создавать гипертекстовую структуру данных, входящих в информационную основу. Графические и картографические визуальные модели в геоинформатике отображают обширное информационное пространство. В это пространство входит множество отношений между объектами реального мира и их атрибутами.

Стратификация данных с включением отношений иерархии означает создание иерархической модели, визуально представляемой как совокупность слоев с общей тематикой и признаками. Принципиальным в таком описании является возможность использования "в чистом виде" логических и теоретико-множественных операций с геоданными для получения новых знаний об объектах и явлениях на поверхности Земли.

Выявление пространственных отношений дает возможность находить слабые и сильные, явные и неявные связи между объектами, находящимися в разных точках пространства. Визуальное моделирование является ключевым в представлении, интерпретации и обработке геоданных.

Проанализируем различия между информационным и геоинформационным подходами. Информационный подход направлен на обработку любой информации и решение любых задач, но требует адаптации к предметной области.

Геоинформационный подход направлен на обработку пространственной информации, геоданных и решение задач, связанных с положением, размещением и перемещением объектов на земной поверхности. Он направлен на решение задач, связанных с возникновением, протеканием и исчезновением различных процессов и явлений на поверхности Земли. Он не требует адаптации, поскольку использует интегрированные данные, пригодные для использования в различных областях без адаптации.

Информационный подход играет роль посредника в обработке исходных данных, собираемых пользователем, и решении задач, поставленных пользователем. Геоинформационный подход играет

роль прикладного инструмента при решении задач пользователя.

Информационный подход в большей степени ориентирован на обработку, безотносительно к приложениям. Это определяет его инструментальный характер и позволяет рассматривать как инструмент посредника (программиста).

Геоинформационный подход ориентирован не только на обработку, но и на обобщение и анализ информации с целевым выходом — получение знаний [27] или информации для поддержки принятия решений. Это определяет его интеграционный характер с приложениями и позволяет рассматривать как инструмент предметника пользователя — исследователя процессов и явлений на земной поверхности.

Информационный подход появился раньше и был ориентирован на обработку данных безотносительно к пространственным отношениям. Исторически геоинформационный подход появился позже, поэтому в нем добавлена специфика, позволяющая находить и использовать пространственные отношения для решения комплекса задач.

Фактор координатной среды в информационном подходе отсутствует. Фактор координатной среды присутствует и играет в геоинформационном подходе интегрирующую ключевую роль.

Информационный подход направлен на выявление и моделирование связей. Геоинформационный подход направлен на выявление и использование связей и отношений, среди которых ведущую роль играют пространственные [23].

В информационном подходе характерен посреднический аспект между методами обработки и приложениями. Это определяет в нем аспект дифференциации как основной. В геоинформационном подходе характерен интеграционный аспект.

Таким образом, геоинформационный подход имеет свою специфику и может быть рассмотрен как развитие информационного подхода.

### Заключение

В настоящее время предложено создать "информатику 3" или "интегральную информатику". В ее состав предполагается включить [12] теоретическую информатику, социальную информатику, техническую информатику и даже биологическую информатику и физическую информатику. Однако интеграционная основа для такого перечня не определена.

Геоинформатика включает элементы "информатики 1", использует "информатику 2" и обладает интеграционными возможностями, необходимыми "для информатики 3". Разумеется, нельзя отождествлять все эти направления. Но, как направление развития информатики геоинформатика может быть использована при условии более тесного сотрудничества представителей всех информатик с геоинформатикой и совместной борьбой с терминологическим мусором в геоинформатике.

### Список литературы

1. **Цветков В. Я.** Международная конференция "Образование в области геодезии, кадастра и землеустройства: тенденции глобализации и конвергенции" // Инженерные изыскания. 2012. № 11. С. 12—14.
2. **Kuipers B.** Modeling Spatial Knowledge // Cognitive Science. 1978. N 2. P. 129—153.
3. **Черный Ю. Ю.** Полисемия в науке: когда она вредна? (на примере информатики) Ч. 1. // Открытое образование. 2010. № 6. С. 97—106.
4. **Shannon C. E. A.** Mathematical Theory of Communication // Bell System Technical Journal. 1948. Vol. 27. P. 379—423 & 623—656.
5. **Cover T. M., Thomas J. A.** Elements of Information Theory. New York: J. Wiley, Sons, 1991.
6. **Темников Ф. Е.** Информатика // Изв. высших учебных заведений. Электротехника. 1963. № 11. С. 1277.
7. **Михайлов А. И., Черный А. И., Гиляревский Р. С.** Основы научной информации. М.: Наука, 1965. 655 с.
8. **Михайлов А. И., Черный А. И., Гиляревский Р. С.** Информатика — новое название теории научной информации // Научно-техническая информация. 1966. № 12. С. 35—39.
9. **Михайлов А. И., Черный А. И., Гиляревский Р. С.** Основы информатики. М.: Наука, 1968. 756 с.
10. **Бауэр Ф. Л., Гооз Г.** Информатика. Вводный курс. М.: Мир, 1976. 484 с.
11. **Юзвишин И. И.** Информациология или закономерности информационных процессов. М.: Высшая Школа. 2001.
12. **Коллин К. К.** Становление информатики как фундаментальной науки и комплексной научной проблемы // Системы и средства информатики. Спец. выпуск. Научно-методологические проблемы информатики / Под ред. К. К. Колина. М.: ИПИ РАН, 2006. С. 7—58.
13. **Floridi L.** Semantic Conceptions of Information First published Wed Oct 5. 2005; substantive revision Fri Jan 28, 2011. URL: <http://plato.stanford.edu/entries/information=semantic>.
14. **Tsvetkov V. Ya.** Semantic Information Units as L. Floridi's Ideas Development // European Researcher. 2012. Vol. 25, N 7. P. 1036—1041.
15. **Хлебников Г. В.** Философия Лучано Флориди // Доклад на заседании семинара "Методологические проблемы наук об информации". М.: ИНИОН. 2011.
16. **Цветков В. Я.** Геоинформационные системы и технологии. М.: Финансы и статистика, 1998. 288 с.
17. **Геодезия, картография, геоинформатика, кадастр:** Энциклопедия. В 2-х т. / Под ред. А. В. Бородко, В. П. Савиных. М.: ООО "Геодезкартгиздат", 2008. Т. I, 496 с.
18. **Савиных В. П., Цветков В. Я.** Геоинформатика как система наук // Геодезия и картография. 2013. № 4. С. 52—57.
19. **Цветков В. Я., Домницкая Э. В.** Геоданные как основа цифрового моделирования // Современные наукоемкие технологии. 2008. № 4. С. 100—101.
20. **Майоров А. А.** Современное состояние геоинформатики // Инженерные изыскания. 2012. № 7. С. 12—15.
21. **Тихонов А. Н., Иванников А. Д., Цветков В. Я.** Терминологические отношения // Фундаментальные исследования. 2009. № 5. С. 146—148.
22. **Цветков В. Я.** Пространственные отношения в геоинформатике // Международный научно-техн. и производ. журнал "Науки о Земле". 2012. Вып. 01. С. 59—61.
23. **Цветков В. Я.** Цифровые карты и цифровые модели // Геодезия и аэрофотосъемка. 2000. № 2. С. 147—155.
24. **Цветков В. Я.** Создание интегрированной информационной основы ГИС // Геодезия и аэрофотосъемка. 2000. № 4. С. 150—154.
25. **Кулагин В. П., Цветков В. Я.** Геоинформационные и информационные технологии // Геодезия и картография. 2002. № 3. С. 41—43.
26. **Иванников А. Д., Кулагин В. П., Мордвинов В. А.** и др. Получение знаний для формирования информационных образовательных ресурсов. М.: ФГУ ГНИИ ИТТ "Информика", 2008. 440 с.
27. **Hill Linda L.** Georeferencing: The Geographic Associations of Information // MIT Press Cambridge, Massachusetts, London, 2009. 272 p.

УДК 004.42:004.896

**А. Б. Филимонов**, д-р техн. наук, проф.,  
e-mail: filimon\_ab@mail.ru

**О. А. Кромин**, аспирант,  
Московский государственный университет  
приборостроения и информатики

## Механизмы коммуникаций семантических JADE-агентов

*Описываются механизмы коммуникаций агентов на программной платформе JADE. Рассматривается структура семантического языка коммуникаций агентов FIPA-SL. Показываются особенности программирования семантических агентов в среде фреймворка JSA.*

**Ключевые слова:** мультиагентные системы, формат сообщений, семантика коммуникационных актов, фреймворк программирования семантических агентов JSA

### Введение

В последнее время агентные технологии получают все большее признание в сфере интеллектуализации автоматизированных информационно-управляющих систем [1–6]. Агентный подход позволяет естественным образом решать трудно формализуемые задачи, причем в агентной архитектуре заложен потенциал дальнейшей модернизации системы и расширения ее функциональных возможностей.

Важнейшим аспектом разработки мультиагентных систем (МАС) являются вопросы организации взаимодействия агентов. Наиболее гибкий механизм взаимодействия удастся реализовать посредством BDI-архитектуры агентов, которая базируется на расширениях модальных логик и позволяет выражать *ментальные* свойства агентов: убеждения, намерения, обязательства [7]. Такую возможность предоставляет семантическая платформа JADE Semantic Agent Framework. В ней воплощена BDI-подобная абстракция агентов, причем семантическая интерпретация сообщений основывается на логическом языке FIPA-SL, предложения которого выражаются логикой ментальных отношений и действий. Целью статьи является обсуждение архитектуры JSA-агентов и особенностей механизма семантической обработки сообщений, реализуемого фреймворком JSA.

### Агенты

Под *агентом* понимается любая сущность, которая может воспринимать среду обитания и воздействовать на нее. При этом агент способен автономно решать возложенные на него задачи, адаптироваться к изменениям во внешнем окружении, а также общаться с другими агентами для достижения заданных глобальных целей.

В рамках мультиагентного подхода информационно-управляющая система строится как совокупность программных агентов. В МАС агенты могут взаимодействовать между собой на основе обмена текущей информацией. При этом комплекс решаемых задач распределяется между агентами. Мультиагентная архитектура придает системам гибкость, которую не могут обеспечить классические клиент-серверные и сервис-ориентированные архитектуры.

Агенты можно классифицировать по степени развития внутреннего представления внешнего мира и способу поведения. Различают два основных класса архитектур агентов [2]:

- *реактивные* (reactive architecture), основанные на поведенческих моделях типа "стимул—реакция";
- *делиберативные* (deliberative architecture), базирующиеся на принципах и методах работы со знаниями.

*Реактивные* агенты имеют довольно бедную (или не имеют вовсе) внутреннюю модель внешней среды. Они действуют в режиме реального времени, основываясь на небольшом количестве информации от своих сенсоров и простых правилах "ситуация—действие" и практически не способны планировать свои действия. В простейшем варианте поведение реактивных агентов имеет конечно-автоматную структуру.

Делиберативные агенты, которые также называют разумными, должны обладать *собственной символьной моделью внешнего мира* и принимать решения на основе формально-логических рассуждений. Им присущи следующие "ментальные свойства":

- *знания* (knowledge) — постоянные, неизменяемые в процессе функционирования знания агента о себе, среде и других агентах;
- *убеждения* (beliefs) — знания агента о среде (в том числе о других агентах), которые могут с течением времени изменяться и даже становиться неверными;

- *желания* (desires) — желаемые состояния внешней среды, причем они могут быть противоречивыми и недостижимыми;
- *цели* (goals) — множество совместимых и достижимых желаний;
- *намерения* (intentions) — множество избранных целей, которые агент собирается достигнуть вследствие своих желаний или в силу взятых на себя обязательств;
- *обязательства* (commitments) — задачи, которые агент берет на себя по просьбе (поручению) других агентов в рамках кооперативных целей или целей отдельных агентов в рамках сотрудничества.

Belief-Desire-Intention (BDI)-архитектура — частный тип делиберативных архитектур, базирующийся на расширениях модальных логик и позволяющий выражать *ментальные* свойства агента: желания, намерения, обязательства.

### Агентная платформа JADE

Для обеспечения *интероперабельности* агентов и агентных сервисов необходима стандартизация агентных систем и технологий. Этой цели служит международная организация FIPA (Foundation of Physical Intelligent Agents, сайт <http://www.fipa.org>), которая занимается стандартизацией механизмов коммуникации агентов, транспортировки сообщений, управления агентами, а также определяет абстрактную архитектуру MAS. Необходимо отметить, что стандарты FIPA не затрагивают архитектуру и способ реализации отдельного агента, что оставляет разработчикам широкий выбор.

Спецификации FIPA определяют понятие *агентной платформы* как промежуточного исполнительного уровня (middleware) между программными агентами и операционной системой. MAS работает "поверх" агентной платформы и использует ее сервисы. Для многих приложений наиболее предпочтительной является агентная платформа JADE (Java Agent Development Framework, сайт <http://jade.tilab.com/>), предназначенная для создания распределенных мультиагентных приложений на основе транспортной архитектуры "точка—точка". JADE полностью реализована на языке Java и поддерживает FIPA-стандарты создания MAS.

JADE включает среду исполнения для программных агентов, программные библиотеки, необходимые для разработки прикладных агентов, а также набор графических инструментов для администрирования и мониторинга запущенных агентов.

Каждая запущенная динамическая среда JADE называется *контейнером*. Набор активных контейнеров образует *платформу*, архитектуру которой поясняет рис. 1. Один из контейнеров является главным (main) и должен быть всегда активным на платформе. Все остальные контейнеры регистрируются и связываются с ним в момент запуска.

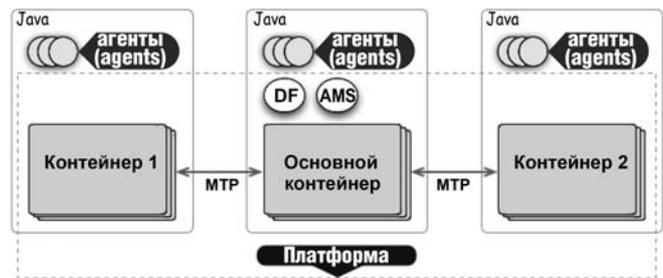


Рис. 1. Общий вид платформы JADE

Контейнеры могут располагаться на разных физических или виртуальных серверах и объединяются между собой с помощью сервиса MTP (Message Transport Protocol). В JADE доступны различные типы протоколов MTP. По умолчанию используется протокол HTTP для передачи данных между контейнерами с адресацией <http://<ip-адрес>:7778/acc>.

Главный контейнер наряду с прикладными агентами содержит также два системных агента — AMS и DF, которые автоматически активируются при запуске агентной платформы.

Агент AMS (Agent Management System) создает среду функционирования агентов, предоставляет агентам разнообразные сервисы и обеспечивает различные операции платформы: создание, управление жизненным циклом и ликвидацией агентов, удаление контейнеров, завершение работы платформы. Каждому агенту необходимо зарегистрироваться в AMS для получения персонального идентификатора AID.

Агент DF (Directory Facilitator) обеспечивает регистрацию сервисов и поиск агента по сервису "желтых страниц": агенты платформы могут подписываться у DF-агента на получение информации о регистрации необходимого сервиса.

### Жизненный цикл агента

Прикладной агент — обычный экземпляр определяемого разработчиком Java-класса, который расширяет общий базовый класс *Agent* для JADE-агентов и наследует все его виртуальные методы. Вычислительная модель агента является многозадачной, причем отдельные задачи могут выполняться одновременно.

Алгоритм работы агента представляет рис. 2. Каждая его фактическая работа выполняется в рамках *поведения* (behaviour) — отдельного задания, которое агент может выполнить.

Метод *setup()* вызывается JADE-средой при запуске агента, инициализирует все необходимые агенту дополнительные сервисы и добавляет поведения, которые будут реализовывать логику агента.

Функционирование агентов JADE задается с помощью наследников класса *Behavior*, объекты которых агент декларирует в качестве своего поведения. Класс *Agent* предоставляет методы *addBehaviour()*

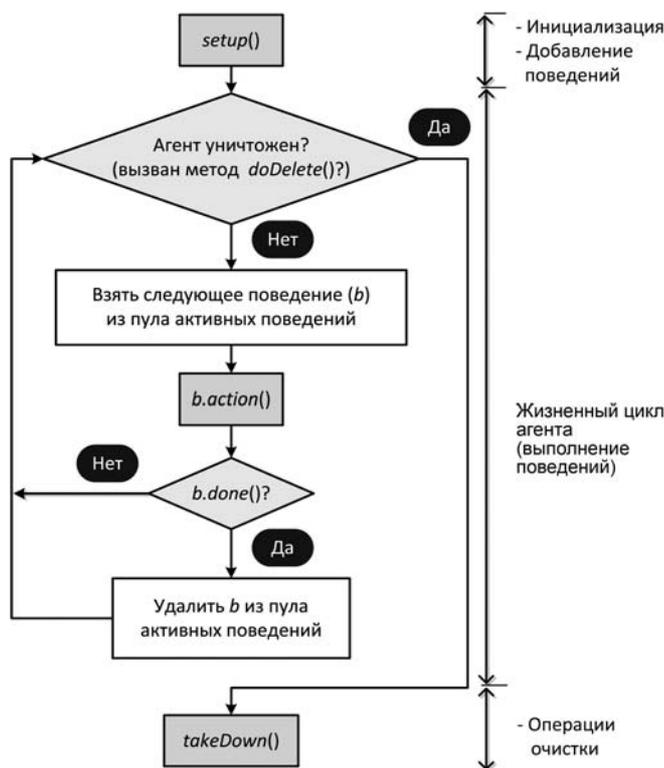


Рис. 2. Алгоритм работы агента

и *removeBehaviour()*, позволяющие управлять готовой очередью задач данного агента — добавлять поведение и удалять поведение агента.

В жизненном цикле агента JADE берет поведение из пула активных поведений и вызывает метод *action()*, который определяет операции для данного поведения. В случае успешного завершения поведения метод *done()* возвращает булево значение "истина", и поведение удаляется из пула активных поведений.

Существуют различные предопределенные модели поведения, которые выполняются однократно, циклически или по расписанию. Из них возможно создавать сложносоставные модели поведения.

Пока агент не уничтожен, он выполняет последовательно все поведения, которые были созданы во время инициализации. Во время выполнения агент может отправлять или принимать сообщения и менять свое поведение.

Для прекращения работы агента должен быть вызван метод *doDelete()*, после выполнения которого платформой вызывается метод *takeDown()*, включающий очистные операции агента.

### Коммуникации между агентами

Агенты могут общаться посредством сообщений. Механизм обмена сообщениями схож с обменом электронной почтой. Процесс обработки сообщений в JADE асинхронный. Сообщения обрабатываются агентом по принципу FIFO (First Input First



Рис. 3. Асинхронная обработка сообщений в JADE

Output). Каждый агент имеет свою очередь сообщений, которую он обрабатывает по мере необходимости (рис. 3).

Коммуникации между агентами могут быть основаны на теории речевых актов Дж. Остина [8, 9], согласно которой языковой (речевой) акт рассматривается как минимальная и базисная единица коммуникации.

Спецификации FIPA определяют язык коммуникаций агентов ACL (Agent Communication Language). FIPA ACL основан на теории речевых актов: сообщения предназначены для выполнения некоторых действий на основании содержимого сообщения. Сообщения, которыми обмениваются между собой агенты JADE, имеют формат языка ACL. Этот формат включает следующие поля [7]:

```
(
<performative> // тип коммуникативного акта;
:sender // отправитель сообщения,
:receiver // получатели сообщения,
:content // содержание сообщения,
:reply-with // метка исходящего сообщения,
:reply-to // адреса агентов, которым отправляется ответ,
:in-reply-to // ссылка на входящее сообщение,
:reply-by // лимит времени на ответ,
:language // язык сообщения,
:encoding // кодировка сообщения,
:ontology // онтология,
:protocol // используемый протокол общения,
:conversation-id // идентификатор разговора
)
```

Поле *performative* выражает коммуникативные намерения отправителя сообщения и может принимать следующие значения:

```
inform — информирование,
request — запрос выполнения действия,
query-ref — запрос на получение информации,
query-if — запрос на получение информации по условию,
cfr (call for proposal) — звонок по предложению,
propose — предложение,
accept-proposal — принятие предложения,
reject-proposal — отклонение предложения и др.
```

Языки контента ACL-сообщений необходимы агентам для обмена знаниями, намерениями, опытом, интерпретации сообщений и формализации рассуждений. Таковыми являются языки KIF, Prolog, Clips, SQL, FIPA-SL, FIPA-CCL, FIPA-KIF, DAML.

Опишем назначение и формат двух основных перформативов, используемых при разработке МАС: *inform* и *query-ref*.

Речевой акт *inform* означает, что агент-отправитель информирует агента-получателя, что предикат поля *content* является истинным.

**Пример.** Агент *i* на языке Prolog информирует агента *j*, что сейчас дверь открыта:

```
(inform
:sender (agent-identifier :name i)
:receiver (agent-identifier :name j)
:content "door (now, open)"
:language Prolog)
```

Речевой акт *query-ref* означает, что один агент спрашивает у другого список ссылок на все объекты, удовлетворяющие предикату в поле *content*.

Среди составных термов языка FIPA-ACL отметим формулы

<quant><выражение> ,

где <quant> — один из квантификаторов any/all/itoa/some. Данные формулы позволяют представлять множества объектов, удовлетворяющих некоторым ограничениям, и относятся к запросам вида *query-if* и *query-ref*.

**Пример.** Агент *i* спрашивает агента *j* обо всех его доступных услугах:

```
(query-ref
:sender (agent-identifier :name i)
:receiver (set (agent-identifier :name j))
:content
((all ?x (available-service j ?x)))
)
```

### Структура языка FIPA-SL

FIPA ACL обеспечивает семантическую детализацию смысла возможных коммуникативных актов посредством семантического языка FIPA-SL, который, в сущности, является модальной логикой, оперирующей с убеждениями (beliefs), намерениями (intentions) и действиями (action).

FIPA-SL — расширенный язык первого порядка с синтаксисом, подобным LISP. SL-предложения выражаются логикой ментальных отношений и действий.

Конструкции языка:

```
FORMULA ::= ATOMIC_FORMULA
| MODAL_LOGIC_FORMULA
| ACTION_FORMULA
| QUANTIFIED_FORMULA
| UNARY_LOGICAL_FORMULA
| BINARY_LOGICAL_FORMULA;
```

```
ATOMIC_FORMULA ::= proposition_symbol
| result
| predicate
| true
| false
| equals;
```

```
MODAL_LOGIC_FORMULA ::= belief
| uncertainty
| intention
| persistent_goal;
```

```
ACTION_FORMULA ::= feasible | done (характеристики action — "осуществимо", "готово");
```

```
QUANTIFIED_FORMULA ::= forall | exists (кванторы);
```

```
UNARY_LOGICAL_FORMULA ::= not (логический оператор);
```

```
BINARY_LOGICAL_FORMULA ::= and | or | implies | equiv (логические операции).
```

Ментальная модель агента основана на представлении трех примитивов: убеждении, неопределенности, выборе.

FIPA-SL определяет следующие модальности [10]:

- (B *i* φ): агент *i* убежден (believe) в φ;
- (U *i* φ): агент *i* не уверен (uncertain) в φ;
- (I *i* φ): агент *i* намерен (intend) сделать φ;
- (done *a* φ): действие *a* только что произошло, а φ уже выполнено;
- (feasible *a* φ): действие *a* может произойти, φ будет иметь место.

**Пример.** В следующем FIPA-ACL-сообщении используется коммуникативный акт *inform*, где в качестве отправителя выступает *agent1*, получателя — *agent2*, а контент выражает модальность убеждения:

```
(inform
:sender agent1
:receiver agent2
:content
(B agent1 (sensorValue 100))
:language fipa-sl
)
```

Из информационных сообщений наиболее употребительны сообщения об убеждении (belief) и о намерении (intention).

**Пример.** FIPA-SL-формула

```
(I agentController (B sensorTerm (temperature 10)))
```

имеет следующее толкование: агент *agentController* намерен принять убеждение от агента *sensorTerm*, что значение параметра *temperature* равно 10. Схематически формулу представляет граф на рис. 4.

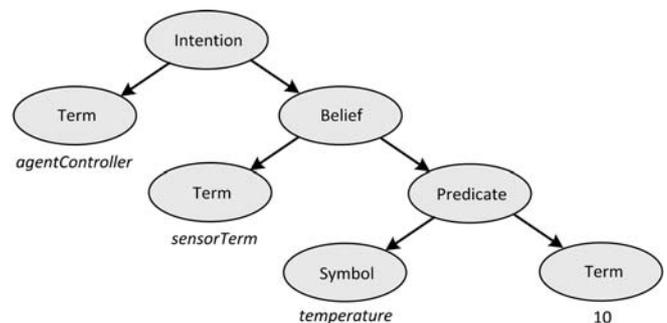


Рис. 4. Схематичное представление формулы

## Фреймворк JSA

Фреймворк JADE Semantic Add-on/JADE Semantic Agent Framework (JSA) разработан компанией France Telecom Research&Development [7, 11–13] как надстройка над фреймворком JADE и дополнен набором инструментов, необходимых для семантической обработки сообщений FIPA-ACL. JSA в полной мере реализует возможности языка FIPA-SL. Процесс обработки сообщений в JSA отражает рис. 5.

Основу функционирования JSA-агентов составляет механизм *семантической интерпретации* (Semantic Interpreter Behaviour) входящих FIPA-ACL-сообщений [14]. В нем ключевую роль играют *семантические репрезентации* (SR — Semantic Representation) — FIPA-SL-выражения, которые представляют смысл сообщений. Основная деятельность агента (activity) состоит в дедукции воспринимаемых событий и соответствующей модификации убеждений и шаблонов его поведения на основе обработки SRs, которая осуществляется посредством применения *правил семантической интерпретации* (SIP — Semantic Interpretation Principle).

SIPs реализуются *SIP-адаптерами* — обработчиками информационных сообщений в формате FIPA-SL. При создании SIP-адаптера задается вид сообщения, которое данный адаптер будет обрабатывать.

Различают SIPs двух типов: *предопределенные*, которые определены FIPA-ACL-семантикой, и *пользовательские*, которые могут быть использованы для настройки поведения агентов в конкретной предметной области, осуществления действий или планирования. В JSA существует 27 предопределенных SIP-адаптеров [7], причем каждый из них имеет свою область применения и обрабатывает определенный тип сообщений.

Фреймворк JSA состоит из следующих основных компонентов (рис. 5):

- *база убеждений* (Belief base) — хранит убеждения агента. Убеждения являются предикатными формулами;

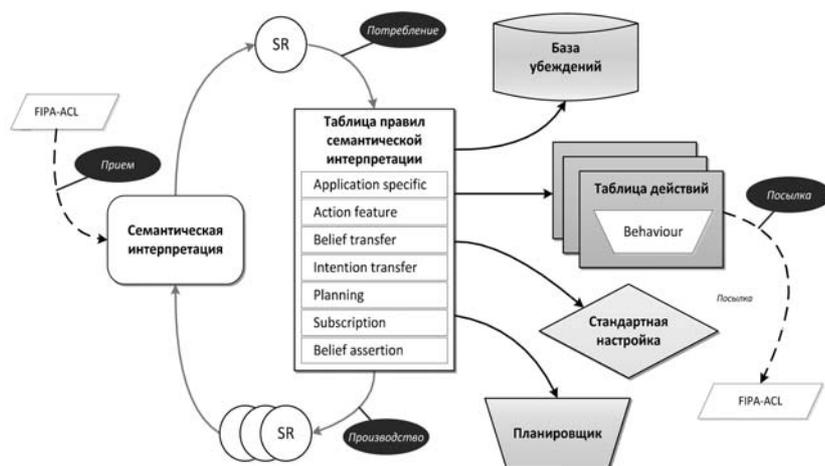


Рис. 5. Схема обработки сообщений в JSA

- *таблица действий* (Action table) — хранит описания (пред- и постусловия) и коды всех действий, которые может выполнять агент;
- *стандартная настройка* (Standard customization) — средство настройки механизма семантической интерпретации;
- *планировщик* (Planner) — осуществляет поиск плана достижения заданной цели;
- *таблица правил семантической интерпретации* (Semantic Interpretation Principles table) — содержит базовые SIPs. Каждое такое правило предназначено для управления частным видом SR. Соответствующий SIP-адаптер может порождать другие SRs, добавлять или удалять образцы поведения, обновлять базу убеждений.

В цикле интерпретации посредством SIPs реализуются функции *производства* (production) и *потребления* (consumption) SRs. Функция производства читает входящие сообщения и производит SRs согласно типу и содержанию сообщения. Функция потребления читает эти SRs и действует на них, например, декларируя убеждения и/или производя другие SRs. Применение данных функций приводит к обновлению базы убеждений и таблицы действий агента.

SIP таблица разбита на классы, основными из которых являются:

- *Application specific*. Содержит SIPs, специфичные для данного агента или приложения.
- *Action features*. Эти SIPs производят несколько SRs из полученного сообщения, которые представляют собой семантические особенности соответствующего коммуникативного действия согласно FIPA-ACL-спецификациям. Сгенерированные SRs далее интерпретируются другими SIPs.
- *Belief transfer*. Эти SIPs управляют принятием убеждений, предложенными внешними агентами. Они применяются к SRs, утверждающим (stating), что некий внешний агент намеревается убедить данного агента в конкретном факте. SIP-адаптер проверяет, что семантический агент убежден в данном факте, и производит SRs с соответствующим назначением. Такие SRs полезны при интерпретации входящих информационных сообщений (Inform messages).
- *Intention transfer*. Эти SIPs применяются к SRs, утверждающим, что намерения некоего внешнего агента включают конкретную цель. Они проверяют, что семантический агент принял (adopt) эту цель и производит SRs соответствующего назначения. Такие SRs являются средством формирования коопераций и полезны при интерпретации всех форм запросов (Request messages).

- *Planning*. Эти Sips применяются к SRs, утверждающим, что семантический агент намеревается достигнуть конкретной цели. Они добавляют соответствующий шаблон поведения. Среди предопределенных SIPs — Action Performance SIP и Rationality Principle SIP, которые выполняют примитивное действие, выбранное из таблицы действий (action table). Разработчики комплектуют этот класс SIPs, чтобы управлять (handle) более тонко возможностями планирования.
- *Subscription*. Эти SIPs применяются к SRs, утверждающим, что некий внешний агент намеревается быть уведомленным (notified) о некотором событии. SIP-адаптеры устанавливают поведение семантического агента, выполняющее мониторинг событий, и посылают ожидаемые уведомления. В частности, такие предопределенные SRs интерпретируют акты подписки (Subscribe) и запросы типа "когда" (request-when) и "всякий раз" (request-whenever).
- *Belief assertion*. Эти SIPs применяются к SRs, утверждающим, что семантический агент уверен в конкретном факте и поэтому добавляют его в базу убеждений.

Поясним схему обработки сообщений в JSA-агенте (рис. 5). Входящее FIPA-ACL-сообщение сначала проверяется на соответствие FIPA-SL-спецификациям. Предопределенные SIPs выполняют его семантический разбор. Выделенные из сообщения SRs передаются соответствующим SIP-адаптерам. Последние генерируют новые SRs, которые передаются в основной цикл интерпретации JSA-агента. Они пополняют базу убеждений либо обрабатываются пользовательскими SIP-адаптерами и фильтрами, которые создают исходящие FIPA-ACL-сообщения и далее передают его внешнему агенту.

Первыми всегда выполняются SIPs класса "Action feature". Они применяются к исходной SR входящего FIPA-ACL-сообщения — проверяется возможность выполнения данной SR и создаются четыре новых SR:

- предварительное условие (Feasibility precondition) (не используется);
- постоянное условие осуществимости (Persistent precondition);
- рациональное действие (Rational effect);
- постусловие (Postcondition).

SIP "Action feature" в случае входящих сообщений-запросов (request messages) создает SR для SIP "Intention transfer".

При взаимодействии агентов чаще всего используются сообщения об убеждениях и о намерениях. Соответствующие SRs имеют вид (В ??myself ...) или (I ??myself ...), где мета-ссылка ??myself подставляет имя агента, отправившего сообщение.

Сообщения типа *inform*, поступившие в цикл семантической интерпретации, подвергаются проверке

SIP-адаптерами класса "Belief transfer" с целью отклонить или принять данное сообщение. Во втором случае JSA сохраняет сообщение в базе убеждений.

При создании SIP-адаптера задается паттерн, который данный адаптер будет обрабатывать. Паттерн (шаблон) — это общий вид SR, где вместо конкретных значений используются маски, например:

(В ??myself (sensorValue ?x)).

Здесь ??myself — маска, которая будет заменена именем агента, формирующим SR, а ?x — значение "sensorValue".

Особо отметим SIP-адаптеры двух типов:

- SIP-адаптеры типа "Belief transfer" обрабатывают SRs, которые содержат факты и имеют вид (??fact ??value). Генерируемые такими SIP-адаптерами SRs попадают в базу убеждений.
- SIP-адаптеры типа "Intention transfer" обрабатывают высказывания о намерениях вида

(I <agentname> (<выражение>)),

давая возможность принять этот перформатив к исполнению или отказать, сгенерировав сообщение о таком намерении.

Этих двух типов SIP-адаптеров достаточно для реализации агентов — они обеспечивают необходимую функциональность для пополнения базы убеждений и обработки информационных сообщений.

Наряду с SIP-адаптерами обработка сообщений в JSA осуществляется посредством фильтров.

*Фильтры* — один из механизмов формирования ответа на запрос, использующих шаблоны убеждений. Основная функция фильтров — перехватывать запросы к базе убеждений в целях более оперативной реакции агента (рис. 6): агент сначала пытается передать управление соответствующему фильтру и сформировать ответ с его помощью, и только в случае неудачи осуществляется поиск в базе убеждений.

JSA в зависимости от запроса сформирует шаблоны убеждений для поиска результата. Например, из запроса

(any ?x (sensor ?x))

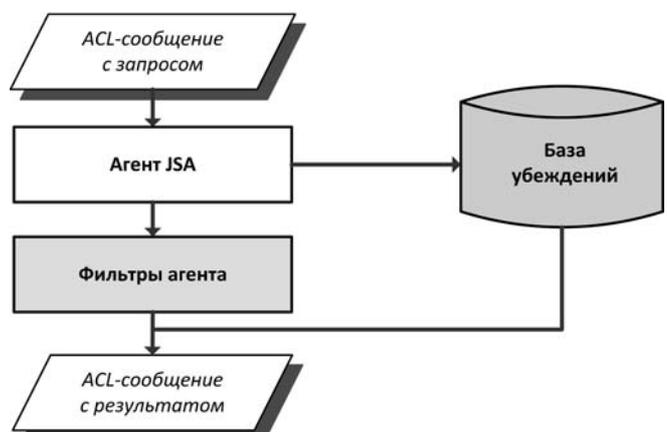


Рис. 6. Схема работы агента с фильтрами

будет сформирован шаблон поиска

```
(В ??myself (sensor ??x))
```

Если будет найден соответствующий фильтр, то значение ??x будет подставлено данным фильтром. Если же подходящий фильтр отсутствует, то запрос будет направлен в базу убеждений. Агент, отправивший запрос, получит сообщение в формате убеждения, например [10]

```
(В agentName (sensor 200))
```

В случае, если требуется особое поведение агента, существует возможность дополнения SIP-таблицы SIP-адаптера для специфических SRs.

### Заключение

Фреймворк JSA берет на себя функцию семантического разбора сообщений, устраняя ее из функциональности JSA-агентов. Кроме этого он автоматически осуществляет маршрутизацию сообщений между агентами после начала их взаимодействия, снимая с них необходимость выполнения коммуникационных актов, что позволяет ограничивать деятельность агента семантической обработкой SL-формул. Это позволяет абстрагировать архитектуру агента от специфики MAC, в которой он функционирует.

Таким образом, агент может отвечать за узкую область знаний, не имея представлений о всей системе в целом, действуя автономно, а коммуникативную функцию берет на себя JSA, не ограничивая сложность запросов агентов. Такой подход открывает возможности для унификации архитектуры агентов, существенно упрощая их разработку и понимание функционирования.

Основным недостатком JSA является избыточность получаемого программного кода. В связи с этим представляет интерес дальнейшее развитие JSA-технологии по пути создания надстройки над

классами агентов JSA, позволяющей скрывать неиспользуемые возможности фреймворка и сводить к минимуму число обработчиков, необходимое для организации взаимодействия агентов.

### Список литературы

1. **Jennings N. R., Wooldridge M. J.** Agent Technology. Berlin, Heidelberg, New-York; Springer-Verlag, 1998.
2. **Городецкий В. И., Грушинский М. С., Хабалов А. В.** Многоагентные системы (обзор) // Новости искусственного интеллекта. 1998. № 2. С. 64–116.
3. **Люгер Дж. Ф.** Искусственный интеллект: стратегия и методы решения сложных проблем. М.: Изд. дом "Вильямс", 2005.
4. **Тарасов В. Б.** От многоагентных систем к интеллектуальным организациям: философия, психология, информатика. М.: Эдиториал УРСС, 2002.
5. **Теряев Е. Д., Петрин К. В., Филимонов А. Б., Филимонов Н. Б.** Агентные технологии в автоматизированных информационно-управляющих системах. Ч. I. Основы агентного подхода // Мехатроника, автоматизация, управление, 2010. № 7. С. 11–27.
6. **Швецов А. Н.** Агентно-ориентированные системы: от формальных моделей к промышленным приложениям // Всероссийский конкурсный отбор обзорно-аналитических статей по приоритетному направлению "Информационно-телекоммуникационные системы", 2008. 101 с.
7. **Bellifemine F., Caire G., Greenwood D.** Developing Multi-Agent Systems with Jade. West Sussex: John Wiley & Sons, 2007.
8. **Harnish R. M.** Commitments and speech acts // Philosophica, 2005. V. 75. P. 11–41.
9. **Winograd T., Flores F.** Understanding Computers and Cognition: a New Foundation for Design. Norwood: Ablex, 1986.
10. **Lespérance Y., Shapiro S.** On Agent Programming Language Support for Rational Communication / N. Maudet, P.-Y. Schobbens, and M. Guyomard (Eds.), Modèles formels de l'interaction (MFI-9). University of Rennes 1 Press, Rennes, France, 2009. P. 223–232.
11. **Louis V., Martinez T.** The JADE Semantic Agent: towards an agent communication oriented middleware // AgentLink News. 2005. V. 18. P. 16–18.
12. **Pautret V.** Jade Semantics Add-on Programmer's guide. [Электронный ресурс]. URL: <http://jade.tilab.com/doc/tutorials/SemanticsProgrammerGuide.pdf>.
13. **Programming** institutional semantic agents. Tracking number 118 // Proc. of 8th Int. Conf on Autonomous Agents and Multiagent Systems (AAMAS 2009), Decker, Sichman, Sierra and Castelfranchi (eds.), May, 10–15, 2009, Budapest, Hungary.
14. **Louis V., Martinez T.** An operational model for the FIPA-ACL semantics // Agent Communication II, 2006.

УДК 621.372.061.9:517/9.001.57(045)

**А. В. Васильев**, канд. техн. наук, ст. науч. сотр.,  
Отделение гибридных моделирующих  
и управляющих систем в энергетике ИПМЭ  
им. Г. Е. Пухова НАН Украины,  
e-mail: oleksii.vasyliiev@gmail.com

## Численно-аналитическое исследование математической модели дробного осциллятора методом S-преобразования

*Метод S-преобразования применен к численно-аналитическому исследованию дифференциального уравнения дробного порядка — математической модели дробного осциллятора. Рассмотрен ряд иллюстративных примеров, показывающих порядок получения решений уравнения при различных порядках дробного дифференциального оператора, начальных условиях и внешних воздействиях. В качестве базисных систем использованы версии смещенных полиномов Лежандра. Вычислительные эксперименты выполнены в программной среде системы Mathematica®.*

**Ключевые слова:** дробное исчисление, дробный осциллятор, дробная производная Капуто, метод S-преобразования, полиномы Лежандра

### Введение

Как известно [2], моделью математического маятника (гармонического осциллятора) является обыкновенное дифференциальное уравнение второго порядка вида

$$\frac{d^2 x(t)}{dt^2} + \omega^2 x(t) = f(t), \quad x(0) = x_{00}, \quad \dot{x}(0) = x_{10},$$

где  $\omega$  — круговая частота.

Исследование колебательных процессов во фрактальных средах (пористые среды, вязко-упругие жидкости и др.) приводит к интегродифференциальным уравнениям с дробными интегралами и производными [7, 8], которые более адекватно описывают реальные процессы по сравнению с классическими методами, оперирующими с производными и интегралами целых порядков. Методы решения уравнений с дробными производными в настоящее время находятся в состоянии интенсивного развития [8]. В работе [6] рассмотрен декомпозиционный метод Адомиана применительно к дифференциальному уравнению дробного порядка — матема-

тической модели дробного осциллятора. В данной работе предлагается применить аппроксимационно-операционный метод S-преобразования [3], который, по нашему мнению, предоставляет широкие возможности для исследования и простоту реализации программными средствами компьютерной алгебры.

Работа построена следующим образом. В первом разделе приведены в ретроспективном плане основные сведения по аппроксимационно-операционному методу S-преобразования и его применению к дифференциальным уравнениям дробного порядка с дифференциальным оператором Капуто [4]. Второй раздел посвящен рассмотрению уравнения математической модели дробного осциллятора и приведению его к виду, удобному для применения S-преобразования. Иллюстративные примеры решения уравнения дробного осциллятора методом S-преобразования с базисной системой функций на основе полиномов Лежандра выполнены в программной среде системы Mathematica® [9] и рассмотрены в третьем разделе.

### 1. Операционный метод S-преобразования: применение к дифференциальным уравнениям дробного порядка

Прямое и обратное S-преобразование, как известно [3], определяется следующим образом. Для сигнала  $x(t)$  и системы линейно независимых базисных функций  $\mathbf{S}(t) = \{s_1(t), s_2(t), \dots, s_m(t)\}^T$ , определенных на одном и том же интервале изменения аргумента  $t \in [0, T)$ , прямое S-преобразование (1) сопоставляет сигналу его изображение  $x(t) \Rightarrow \mathbf{X}$  в виде вектора  $\mathbf{X}$  коэффициентов полинома наилучшего приближения:

$$\mathbf{X} = \left( \int_0^T \mathbf{S}(t) \cdot \mathbf{S}^T(t) dt \right)^{-1} \left( \int_0^T \mathbf{S}(t) \cdot x(t) dt \right), \quad (1)$$

где  $\mathbf{X} = \{X_1, X_2, \dots, X_m\}^T$ , а интегрирование векторно-матричных выражений здесь и далее в работе выполняется поэлементно. Обратное S-преобразование (2) реконструирует сигнал в виде его аппроксимации:

$$x_a(t) = \sum_{i=1}^m X_i s_i(t) = \mathbf{X}^T \cdot \mathbf{S}(t). \quad (2)$$

Операционный характер преобразованию придает выражение, позволяющее получить изображе-

ние интеграла сигнала по изображению подынтегральной функции без выполнения интегрирования, т. е. алгебраическим путем:

$$y(t) = \frac{1}{\Gamma(\beta)_0} \int_0^t (t - \tau)^{\beta-1} x(\tau) d\tau \Leftrightarrow \mathbf{Y} = \mathbf{P}^\beta \cdot \mathbf{X}. \quad (3)$$

Левая часть выражения (3) представляет собой интеграл Римана—Лиувилля порядка  $\beta$  [4, 7, 8]. При целых значениях порядка она представляет собой обычный повторный интеграл. В правой части выражения  $\mathbf{P}^\beta$  — операционная матрица интегрирования, зависящая от системы базисных функций, дробного порядка интегрального оператора и диапазона изменения аргумента  $T$ :

$$\mathbf{P}^\beta = \left( \int_0^T \mathbf{S}(t) \cdot \mathbf{S}^T(t) dt \right)^{-1} \times \left( \int_0^T \mathbf{S}(t) \cdot \left( \frac{1}{\Gamma(\beta)_0} \int_0^t (t - \tau)^{\beta-1} \cdot \mathbf{S}^T(\tau) d\tau \right) dt \right). \quad (4)$$

Выражение (4) позволяет определять операционные матрицы интегрирования как целого, так и дробного порядков.

## 2. Математическая модель дробного осциллятора

Математическая модель дробного осциллятора, приведенная в работе [6], имеет вид

$$\frac{d^\alpha x(t)}{dt^\alpha} + \omega^\alpha x(t) = f(t), \quad x(0) = x_{00}, \quad \dot{x}(0) = x_{10}, \quad 1 < \alpha \leq 2, \quad (5)$$

где  $\omega$  — частота,  $f(t)$  — функция внешней возбуждающей силы. Преобразуем выражение (5) к виду, удобному для применения метода S-преобразования. Введя новую функцию, обозначающую вторую производную решения уравнения, получим:

$$\frac{d^2 x(t)}{dt^2} = u(t); \quad (6)$$

$$\frac{dx(t)}{dt} = \dot{x}(0) + \int_0^t u(\tau) d\tau; \quad (7)$$

$$x(t) = x(0) + \dot{x}(0)t + \int_0^t \int_0^{\tau} u(\tau_1) d\tau_1 d\tau; \quad (8)$$

$$J^{2-\alpha} u(t) + \omega^\alpha \int_0^t \int_0^{\tau} u(\tau_1) d\tau_1 d\tau = f(t) - (x(0) + \dot{x}(0)t). \quad (9)$$

Выражение (9), в котором  $J^{2-\alpha}$  является оператором Римана—Лиувилля порядка  $1 - 2\alpha$ , эквивалентно исходному дифференциальному уравнению с заданными начальными условиями (5). Для последующего построения фазового портрета реше-

ния уравнения дробного осциллятора необходимо ввести функцию

$$r(t) = \frac{d^{\alpha/2} x(t)}{dt^{\alpha/2}} = J^{1-\alpha/2} \frac{dx(t)}{dt}. \quad (10)$$

Нетрудно видеть, что при  $\alpha = 2$  выражение (10) превращается в первую производную решения уравнения классического маятника.

Перейдем в операционное пространство S-преобразования:

$$\mathbf{P}^{2-\alpha} \cdot \mathbf{U} + \omega^\alpha \cdot \mathbf{P}^2 \cdot \mathbf{U} = \mathbf{F} - x(0) \cdot \mathbf{1} - \dot{x}(0) \cdot \mathbf{t}; \quad (11)$$

$$\mathbf{U} = (\mathbf{P}^{2-\alpha} + \omega^\alpha \cdot \mathbf{P}^2)^{-1} \cdot (\mathbf{F} - x(0) \cdot \mathbf{1} - \dot{x}(0) \cdot \mathbf{t}); \quad (12)$$

$$\dot{\mathbf{X}} = \dot{x}(0) \cdot \mathbf{1} + \mathbf{P}^1 \cdot \mathbf{U}; \quad (13)$$

$$\mathbf{X} = x(0) \cdot \mathbf{1} + \dot{x}(0) \cdot \mathbf{t} + \mathbf{P}^2 \cdot \mathbf{U}; \quad (14)$$

$$\mathbf{R} = \mathbf{P}^{1-\alpha/2} \cdot \dot{\mathbf{X}}; \quad (15)$$

$$x_a(t) = \mathbf{X}^T \cdot \mathbf{S}(t); \quad (16)$$

$$r_a(t) = \mathbf{R}^T \cdot \mathbf{S}(t). \quad (17)$$

Выражения (11), (13)—(15) являются операционными аналогами следующих зависимостей в пространстве оригиналов (временного аргумента): (11)  $\rightarrow$  (9), (13)  $\rightarrow$  (7), (14)  $\rightarrow$  (8), (15)  $\rightarrow$  (10). Выражение (12) является решением матричного уравнения (11), а (16) и (17) являются обратными S-преобразованиями (реконструкциями) функций  $x(t)$  и  $r(t)$ .

## 3. Численное исследование решений уравнений дробного осциллятора

Рассмотрим численные решения уравнения дробного осциллятора при значениях параметров, указанных в работе [6]:  $\alpha_1 = 1,9$ ;  $\alpha_2 = 1,75$ ;  $\alpha_3 = 1,5$ ;  $\omega = 1,5$ ;  $f_1 = 0$ ;  $f_2 = 1$ ;  $f_3 = \sin[\omega \cdot t]$ ;  $x_{01} = 1$ ;  $x_{d1} = 0$ ;  $x_{02} = 0$ ;  $x_{03} = 0$ .

Выберем в качестве базисной систему блочно-импульсных функций [5] (локальных полиномов Лежандра нулевого порядка) с числом локальных интервалов  $m = 200$  и значением шага по аргументу  $h = 0,05$ .

### Программа

- Задание числа локальных интервалов и значения шага по аргументу:  
 $m := 200; h := 0,05;$
- Определения изображений константы  $1$  и независимой переменной  $t$ :

$$\text{One} := \text{Table} \left[ \frac{1}{h} * \int_{(i-1)*h}^{i*h} 1 dt, \{i, m\} \right];$$

$$\text{Tim} := \text{Table} \left[ \frac{1}{h} * \int_{(i-1)*h}^{i*h} t dt, \{i, m\} \right];$$

- *Определение базисной системы функций:*

$s[t_, h_, i_] := \text{If}[(i - 1) \cdot h \leq t < i \cdot h, 1, 0];$   
 $s = \text{Table}[s[t, h, i], \{i, m\}];$

- *Определение операционной матрицы интегрирования порядка  $\beta$ :*

$$H[\beta_, h_, m_] := \frac{h^\beta}{\Gamma[\beta + 2]} *$$

$\text{Table}[\text{Which}[i < j, 0, i = j, 1, i > j,$   
 $(i - j + 1)^{\beta + 1} - 2(i - j)^{\beta + 1} +$   
 $+(i - j - 1)^{\beta + 1}], \{i, m\}, \{j, m\}];$

- *Определение изображения функции правой части уравнения:*

$$F := \text{Table}\left[\frac{1}{h} * \int_{(i-1)h}^{i \cdot h} t dt, \{i, m\}\right];$$

- *Задание числовых значений параметров, правой части и начальных условий:*

$\alpha_1 = 1,9; \alpha_2 = 1,75; \alpha_3 = 1,5; \omega = 1,5;$   
 $f_1 = 0; f_2 = 1; f_3 = \text{Sin}[\omega * t];$   
 $x_{o1} = 1; x_{o2} = 0; x_{o3} = 0;$

- *Нахождение операционных матриц интегрирования необходимых порядков:*

$P_1 := H[1, h, m]; P_2 := H[2, h, m];$   
 $P\alpha_1 = H[2 - \alpha_1, h, m]; P\alpha_2 = H[2 - \alpha_2, h, m];$   
 $P\alpha_3 = H[2 - \alpha_3, h, m];$

- *Нахождение изображений правых частей уравнения:*

$$F_1 := \text{Table}\left[\frac{1}{h} * \int_{(i-1)h}^{i \cdot h} f_1 dt, \{i, m\}\right];$$

$$F_2 := \text{Table}\left[\frac{1}{h} * \int_{(i-1)h}^{i \cdot h} f_2 dt, \{i, m\}\right];$$

$$F_3 := \text{Table}\left[\frac{1}{h} * \int_{(i-1)h}^{i \cdot h} f_3 dt, \{i, m\}\right].$$

**Пример 1.**  $\alpha_1 = 1,9; \alpha_2 = 1,75; \alpha_3 = 1,5;$   
 $\omega = 1,5; f_1 = 0; x_{o1} = 1; x_{o2} = 0; x_{o3} = 0;$

- *Нахождение решений уравнения в области  $S$ -преобразования:*

$U_1 = \text{Inverse}[P\alpha_1 + \omega^{\alpha_1} * P_2].(F_1 - \omega^{\alpha_1} * x_{o1} * \text{One});$

$U_2 = \text{Inverse}[P\alpha_2 + \omega^{\alpha_2} * P_2].(F_2 - \omega^{\alpha_2} * x_{o2} * \text{One});$

$U_3 = \text{Inverse}[P\alpha_3 + \omega^{\alpha_3} * P_2].(F_3 - \omega^{\alpha_3} * x_{o3} * \text{One});$

$X_1 = P_2.U_1 + x_{o1} * \text{One};$

$X_2 = P_2.U_2 + x_{o2} * \text{One};$

$X_3 = P_2.U_3 + x_{o3} * \text{One};$

$Xp_1 = H[1 - \alpha_1/2, h, m].P_1.U_1;$

$Xp_2 = H[1 - \alpha_2/2, h, m].P_1.U_2;$

$Xp_3 = H[1 - \alpha_3/2, h, m].P_1.U_3;$

- *Аппроксимация решений:*

$xa_1 = X_1.S;$

$xa_2 = X_2.S;$

$xa_3 = X_3.S;$

- *Визуализация решений и фазового портрета уравнения при значениях параметров примера 1 (рис. 1—3):*

$\text{ListPlot}[\{X_1, X_2, X_3\}, \text{AxesLabel} \rightarrow \{n, "X_1, X_2, X_3"\}]$

$\text{Plot}[\{xa_1, xa_2, xa_3\}, \{t, 0, m \cdot h\}, \text{PlotPoints} \rightarrow 200,$

$\text{AxesLabel} \rightarrow \{t, "xa_1, xa_2, xa_3"\}]$

$\text{ListPlot}[\text{Table}[\{X_2[[n]], Xp_2[[n]]\}, \{n, 200\}],$

$\text{AxesLabel} \rightarrow \{"X_2[[n]]", "Xp_2[[n]]"\}]$

$\text{ParametricPlot}[\{xa_2, xpa_2\}, \{t, 0, m \cdot h\},$

$\text{AxesLabel} \rightarrow \{"xa_2", "xpa_2"\}]$

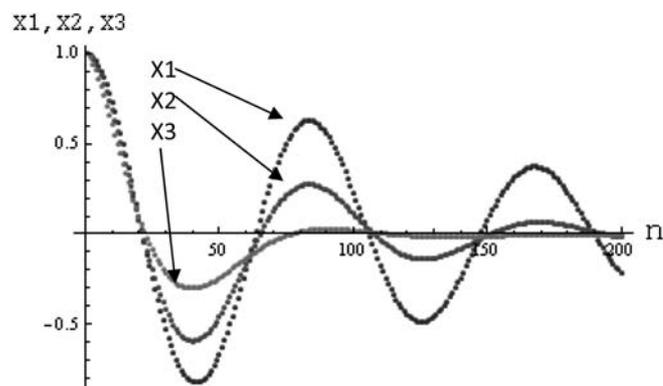


Рис. 1. Решения уравнения при различных дробных порядках в операционной области

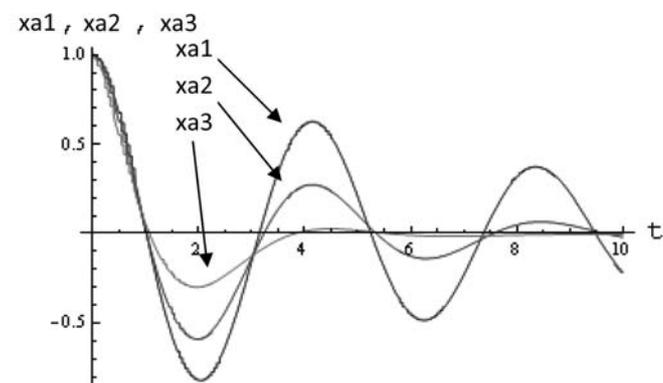


Рис. 2. Решения уравнения при различных дробных порядках в области временного аргумента

Для примеров 2 и 3 приводятся только результаты визуализации решений и фазовых портретов, так как программы различаются только числовыми значениями параметров.

Прокомментируем результаты визуализации решений уравнения дробного осциллятора при различных значениях дробного порядка дифференциального оператора, начальных условий и внешних возмущающих сил. На рис. 1, 5 и 8 показаны компоненты векторов S-преобразования решений уравнения дробного осциллятора при различных сочетаниях параметров уравнения. Здесь в качестве абсциссы графиков выступает порядковый номер компоненты. Рис. 2, 6 и 9 показывают результаты обратного S-преобразования, которые являются аппроксимациями решений уравнения при тех же сочетаниях параметров. Фазовые портреты решений в операционной области показаны на рис. 3, 7

и 10. Результат реконструкции фазового портрета для одного из вариантов решения приведен на рис. 4.

**Пример 2.**  $\alpha_1 = 1,9$ ;  $\alpha_2 = 1,75$ ;  $\alpha_3 = 1,5$ ;  $\omega = 1,5$ ;  $f_2 = 1$ ;  $x_{o2} = 0$ ;

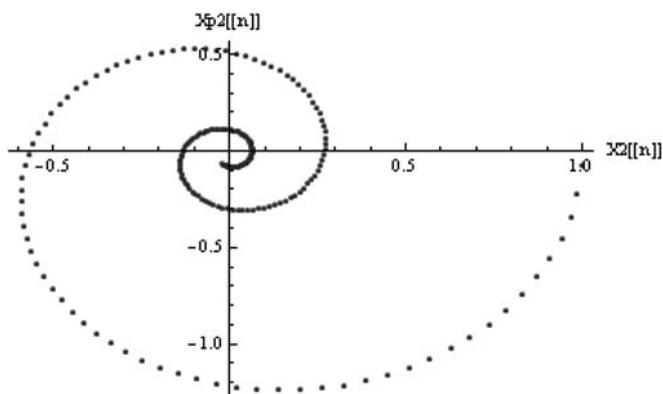


Рис. 3. Фазовый портрет модели дробного осциллятора в операционной области при  $\beta = 1,75$

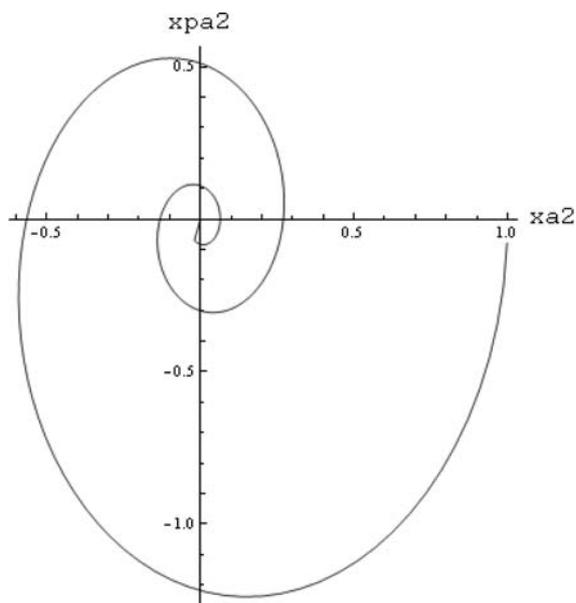


Рис. 4. Фазовый портрет модели дробного осциллятора в области временного аргумента при  $\beta = 1,75$

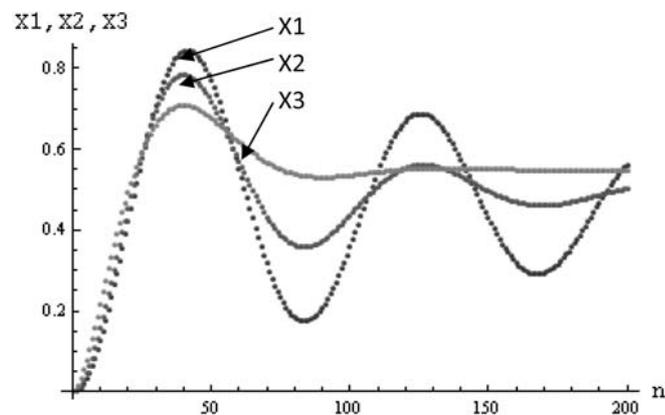


Рис. 5. Решения уравнения при различных дробных порядках в операционной области

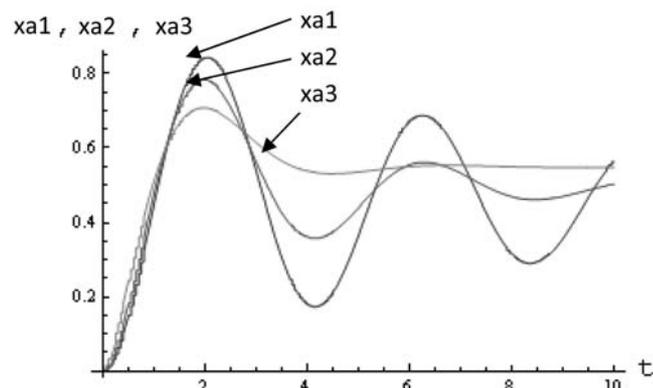


Рис. 6. Решения уравнения при различных дробных порядках в области временного аргумента

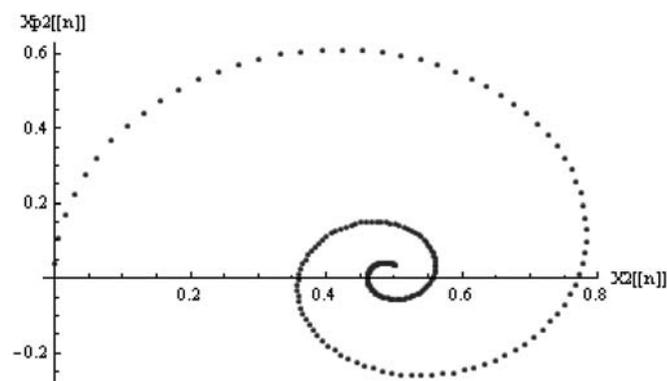


Рис. 7. Фазовый портрет модели дробного осциллятора в операционной области при  $\beta = 1,75$

**Пример 3.**  $\alpha_1 = 1,9$ ;  $\alpha_2 = 1,75$ ;  $\alpha_3 = 1,5$ ;  
 $\omega = 1,5$ ;  $f_3 = \sin(\omega t)$ ;  $x_{03} = 0$ .

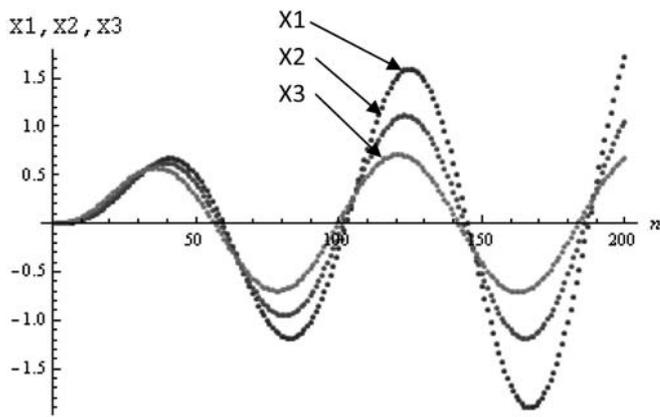


Рис. 8. Решения уравнения при различных дробных порядках в операционной области

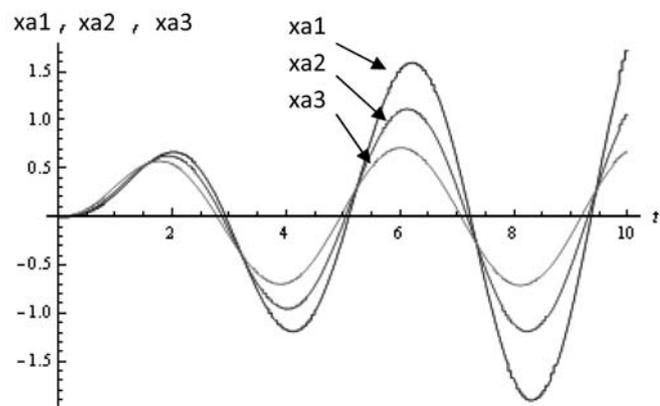


Рис. 9. Решения уравнения при различных дробных порядках в области временного аргумента

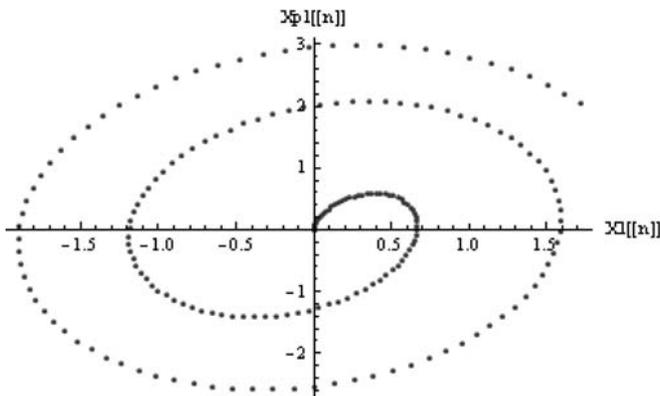


Рис. 10. Фазовый портрет модели дробного осциллятора в операционной области при  $\beta = 1,75$

## Заключение

В данной работе аппроксимационно-операционный метод S-преобразования, в основе которого лежит аппроксимация сигналов обобщенными полиномами по некоторой системе базисных функций, применен к решению дифференциального уравнения дробного порядка — математической модели дробного осциллятора. Необходимо отметить следующие особенности метода S-преобразования. Первая особенность заключается в том, что каждая система базисных функций порождает свою разновидность операционного исчисления. Вторая особенность — приближенный, аппроксимационный характер получаемых результатов. Третья — решения уравнений определены на предварительно заданном интервале изменения аргумента.

Существенными преимуществами метода являются простота обратного преобразования, применимость к математическим моделям динамических систем как целого, так и дробного порядков интегродифференциальных операторов. Сравнительный анализ результатов, полученных в ходе вычислительных экспериментов, показывает эффективность метода и позволяет рекомендовать его к применению при исследовании колебательных процессов во фрактальных средах.

## Список литературы

1. Гончаров В. Л. Теория интерполирования и приближения функций. М.: ГНИИЛ, 1954. 327 с.
2. Крауфорд Ф. Курс физики. Волны. М.: Наука, 1974. 528 с.
3. Васильев В. В., Симак Л. А. Полиномиальная аппроксимация сигналов как операционное исчисление: применение к моделированию динамических систем дробного порядка // Информационные технологии. 2012. № 3. С. 55–60.
4. Учайкин В. В. Метод дробных производных. Ульяновск: АРТИШОК, 2008. 512 с.
5. Jiang Z. H., Schaufelberger W. Block Pulse Functions and Their Applications in Control Systems, Springer-Verlag, 1992. 237 p.
6. Shaher M., Rabha W. Analytical Solutions of a Fractional Oscillator by the Decomposition Method // Int. Journal of Pure and Applied Mathematics. 2007. Vol. 37, N 1. P. 119–131.
7. Oldham K. B., Spanier J. The Fractional Calculus // Academic Press, 1974. 234 p.
8. Podlubny I. Fractional Differential Equations // Academic Press, 1999. 340 p.
9. Wolfram S. The Mathematica book. Cambridge University Press, 1996. 1403 p.

**П. В. Казаков**, канд. техн. наук, доц.,  
e-mail: pvk\_mail@list.ru,  
Брянский государственный  
технический университет

## Модель динамического изменения размера популяции для генетических алгоритмов многокритериальной оптимизации

*Рассматривается новый способ управления размером популяции в генетических алгоритмах многокритериальной оптимизации. Он отличается универсальностью и может быть реализован в любых эволюционных алгоритмах этого класса. Приводится сравнительный анализ эффективности использования динамического размера популяции при решении задач многокритериальной оптимизации разной сложности.*

**Ключевые слова:** многокритериальная оптимизация, принципы Парето, граница Парето, многокритериальные генетические алгоритмы, динамический размер популяции

### Введение

Многие возникающие в теории и на практике оптимизационные задачи предполагают минимизацию либо максимизацию более одного критерия, т. е. являются многокритериальными [1, 2]. Известно, что в процессе многокритериальной оптимизации (МО) необходимо одновременно оптимизировать множество несоизмеримых и конфликтующих критериев. Это существенно усложняет, а в ряде случаев делает невозможным определение глобально оптимальных решений. Также известно, что результатом многокритериальной оптимизации является решение, принадлежащее множеству Парето [1, 2]. Поэтому в настоящее время одним из подходов, позволяющих повысить качество решения задач многокритериальной оптимизации, является разделение этого процесса на два этапа. Вначале находится множество Парето-оптимальных решений, из которых затем выбирается одно (несколько), удовлетворяющее субъективным предпочтениям исследователя [2]. Тем самым, точность решения задачи МО, прежде всего, зависит от результатов выполнения первого этапа определения множества Парето. Его составляют так называемые недоминируемые решения, которые имеют значения по всем критериям не хуже и хотя бы по одному лучше, чем остальные решения. В пространстве критериев недоминируемые решения образуют границу Парето, играющую важную роль при управлении работой оптимизационного алгоритма и оценке качества

его результата. В общем случае точное определение множества Парето в задаче МО остается нерешенной проблемой, прежде всего из-за теоретически бесконечного числа, а также непредсказуемого расположения недоминируемых решений в пространстве поиска (как переменных, так и критериев). Кроме этого, важным ограничивающим фактором является высокая вычислительная сложность процедуры построения множества Парето. Все это требует от численных методов МО развитых средств управления процессом оптимизации, возможности всестороннего исследования поискового пространства, а также определения множества Парето-оптимальных решений за один запуск оптимизационного алгоритма. В настоящее время названные возможности наиболее эффективно реализуются генетическими алгоритмами [3], а точнее, их специальными версиями для многокритериальной оптимизации [4–7].

Генетические алгоритмы многокритериальной оптимизации (МГА) образуют отдельную группу методов эволюционных вычислений и отличаются от обычных (однокритериальных) генетических алгоритмов принципами оценки решений и управления популяцией. Последнее является наиболее важным в работе МГА, так как определяет результирующее множество и число найденных оптимальных решений. Эффективное управление популяцией подразумевает сохранение баланса между уже найденными хромосомами, кодирующими оптимальные решения, и хромосомами, предполагающими их дальнейшее изменение для продолжения исследования поискового пространства. При этом в составе популяции должно поддерживаться разнообразие в течение всего процесса оптимизации. Очевидно, что разные участки пространства поиска имеют отличающиеся по сложности топологии, требующие для их исследования различное число хромосом. Однако традиционное использование в МГА фиксированного размера популяции не позволяет изменять плотность распределения хромосом по поисковому пространству без нарушения их баланса в популяции и сохранения ее разнообразия. Данное ограничение может быть снято при использовании в МГА изменяемого размера популяции в зависимости от характера поиска и свойств уже найденных решений. Такая модификация позволит расширить функциональные возможности генетического алгоритма и снизить вероятность его схождения к локально-оптимальным решениям.

В статье предлагается оригинальная модель управления численностью популяции в процессе работы МГА. Эта модель может быть интегрирована в любой МГА, тем самым расширив его возможности адаптации к пространству поиска. Для проверки эффективности предлагаемая модификация МГА

исследуется применительно к двум основным генетическим алгоритмам многокритериальной оптимизации — SPEA2 [8] и NSGA-II [9]. Для этого названные алгоритмы (оригинальные и модифицированные) используются для решения тестовых задач МО с различным числом критериев. Полученные результаты оцениваются по набору показателей качества решений МО. В конце статьи приводится их описание, а также выводы по ним.

## 1. Принципы управления динамическим размером популяции

Для любого эволюционного алгоритма размер популяции является одним из главных параметров. От его значения зависит как скорость, так и точность поиска: недостаточный размер популяции ограничивает объем исследуемого пространства поиска, приводит к быстрому схождению алгоритма только к локально оптимальным решениям; слишком большой — замедляет поиск, снижает разнообразие популяции и эффект от оператора отбора. Проведенные исследования по определению универсального и оптимального размера популяции не позволили назвать его точное значение [4, 10]. Конкретные значения размера популяции зависят от особенностей решаемой задачи и часто определяются эмпирически из некоторого интервала его возможных значений. В связи с этим были предложены некоторые способы управления размером популяции динамически, т. е. непосредственно во время работы генетического алгоритма [10, 11].

### 1.1. Управление размером популяции в МГА

Для задач МО значение размера популяции во многом становится определяющим, так как в ней теперь необходимо хранить все найденные оптимальные решения, а также "материал" для их улучшения и увеличения числа решений. Для решения проблемы поиска баланса между такими частями популяции могут быть использованы как Парето-архив [4, 8] (для хранения всех найденных недоминируемых решений), так и способы динамического изменения размера популяции. Для последнего случая среди МГА существуют отдельные оригинальные версии. Это коэволюционный алгоритм CO-МОЕА [12], а также алгоритмы ДМОЕА (Dynamic Multi-Objective Evolutionary Algorithm) [13] и ИМОЕА (Incrementing Multi-Objective Evolutionary Algorithm) [14]. В каждом из них размер популяции (популяций) зависит от динамики процесса поиска, но механизмы изменения его значения имеют различные принципы реализации. В ДМОЕА рост или усечение популяции определяются на основе ряда качественных индикаторов. Их значения постоянно обновляются в зависимости от уровня доминирования соответствующих решений популяции

и ее разнообразия. Как отмечают авторы алгоритма, это позволяет в определенный момент его работы найти оптимальный для решаемой задачи размер популяции. В ИМОЕА размер популяции изменяется в зависимости от свойств исследуемой поверхности пространства критериев и плотности ее покрытия решениями из текущей популяции. Для этого используются специальные методы математической статистики. Также в этом алгоритме доступно интерактивное управление размером популяции.

В целом, используемые в названных МГА принципы управления размером популяции оказываются зависимыми от ряда других составляющих работы этих алгоритмов: схемы оценки решений, генетических операторов, а также от подбора значений для дополнительных управляющих параметров. Это значительно затрудняет перенос этих принципов в любой другой МГА без влияния на его базовый функционал. Вместе с тем, наличие у МГА возможности изменять в процессе работы размер популяции позволяет увеличить его способность самоадаптации к исследуемому поисковому пространству и, тем самым, к решаемой задаче МО. Поэтому представляет интерес использование доступных для эволюционных алгоритмов других моделей управления численностью популяции. Одна из них предложена в работах [10, 11] и предназначена для однокритериальных ГА (GAVaPS — Genetic Algorithm with Varying Population Size), однако после соответствующих изменений она может быть применена и для МГА.

### 1.2. Алгоритм самоадаптации размера популяции

Модель динамического изменения размера популяции GAVaPS отличается возможностью интеграции в любой генетический алгоритм и использует концепцию хромосомы, имеющей возраст и фиксированное время жизни. Возраст хромосомы  $age(c_p, t)$ ,  $i = 1, \dots, n_p(t)$ , где  $n_p(t)$  — размер популяции в поколении  $t$ , совпадает с числом поколений, в которых она остается в популяции, и зависит только от значения пригодности хромосомы. В GAVaPS в явном виде не используется оператор отбора, что позволяет соответствующим хромосомам большее число поколений оставаться в популяции. С одной стороны, это приводит к изменению численности популяции, в частности, ее росту, с другой — к сохранению в ней множества Парето-оптимальных решений, а именно — реализации стратегии элитизма для МГА. Не использование оператора отбора дает возможность каждой хромосоме участвовать в генетических операторах с одинаковой вероятностью независимо от значения ее пригодности, при этом, в частности, в кроссинговере могут участвовать хромосомы из разных поколений. Для размещения получаемых в результате новых хромосом

на каждом поколении  $t$  происходит увеличение размера популяции на величину

$$n_p^+(t) = n_p(t)\rho,$$

где  $\rho \in (0, 1]$  — коэффициент репродукции.

Кроме текущего возраста  $s$  каждой хромосомой связывается параметр время жизни (Life Time)  $LT(c_i, t)$ ,  $i = 1, \dots, n_p(t)$ . Его значение ограничивает число поколений существования хромосомы. В модели GAVaPS значение  $LT(c_i, t)$ ,  $i = 1, \dots, n_p(t)$ , вычислялось для каждой новой хромосомы как функция от ее пригодности и после этого не изменялось. Однако в МГА каждая хромосома оценивается только относительно текущей популяции, т. е. у одной и той же хромосомы в разных поколениях может быть разное значение пригодности. Поэтому значение параметра  $LT(c_i, t)$ ,  $i = 1, \dots, n_p(t)$ , должно меняться каждый раз после оценки хромосомы, для чего предлагаются следующие зависимости  $\forall c_i$ ,  $i = 1, \dots, n_p(t)$ :

$$\left\{ \begin{array}{l} LT(c_i, t) = age(c_i, t) + \text{round}\left(\frac{f(c_i, t) - f(c_i, t)^{\min}}{f(c_i, t)^{\max} - f(c_i, t)^{\min}}\right), \\ \text{если } \text{rank}(c_i, t) = 1; \\ LT(c_i, t) = \frac{1 + \max\{\text{rank}(c_i, t)\}}{2} + (\text{rank}(c_i, t - 1) - \\ - \text{rank}(c_i, t))\left(\frac{f(c_i, t) - f(c_i, t)^{\min}}{f(c_i, t)^{\max} - f(c_i, t)^{\min}}\right), \text{ иначе.} \end{array} \right. \quad (1)$$

Здесь  $\text{rank}(c_i, t)$ ,  $i = 1, \dots, n_p(t)$ , — ранг Парето хромосомы в поколении  $t$ . У недоминируемых хромосом ранг равен 1. Значение  $f(c_i, t)$ ,  $i = 1, \dots, n_p(t)$ , определяет пригодность каждой хромосомы, предполагается, что чем большее ее значение, тем она лучше. Параметры  $f(c_i, t)^{\min}$ ,  $f(c_i, t)^{\max}$ ,  $i = 1, \dots, n_p(t)$ , представляют собой соответственно минимальную и максимальную пригодности среди хромосом поколения  $t$ . Для начальной популяции ( $t = 0$ ) устанавливается, что

$$LT(c_i, t) = \text{rank}(c_i, t), \quad i = 1, \dots, n_p(t).$$

Система (1) содержит два выражения для вычисления времени жизни хромосом популяции. В общем случае оно зависит только от ранга Парето хромосомы и значения ее пригодности. Решения, представляемые хромосомами с рангом Парето, равным единице, относительно данной популяции являются оптимальными и, следовательно, важно их в ней сохранить. Поэтому для таких хромосом время жизни вычисляется по отдельной (первой) зависимости. Согласно ей каждая недоминируемая хромосома с пригодностью выше средней гарантированно переходит в следующее поколение. Это достигается непосредственным увеличением времени жизни хромосомы относительно ее текущего возраста.

Для остальных хромосом используется вторая зависимость. В ней при определении времени жизни учитываются такие параметры, как максимальный ранг Парето текущей популяции, пригодность хромосомы, а также изменение ее ранга Парето в сравнении с предыдущим поколением. Последнее дает возможность соответствующим хромосомам увеличить время своего пребывания в популяции, если их ранг улучшился (уменьшился). Также здесь возможна и обратная ситуация, так как названные параметры не влияют на постоянный рост возраста хромосомы. Использование параметра  $\max\{\text{rank}(c_i, t)\}$ ,  $i = 1, \dots, n_p(t)$ , предназначено для поддержания разнообразия популяции на начальных этапах работы алгоритма, когда происходит наиболее активное распределение ее членов по пространству поиска. В этих условиях ранг Парето хромосом популяции может принимать максимальное значение за все поколения. Таким образом, время жизни отдельных хромосом в начале процесса поиска будет увеличено. Веса коэффициентов слагаемых обеих выражений системы (1) были определены опытным путем на основе анализа динамики изменения значений определенных параметров популяции (рис. 1–3).

Известно, что ранг Парето во всех МГА вычисляется одинаково [5], в то время как пригодность хромосомы зависит от используемого в алгоритме аналитического представления fitness-функции. В ней могут учитываться различные показатели качества решения, например, особенности его расположения на границе Парето, плотность его соседства на ней с другими точками и т. п. Поэтому хромосомы с одинаковым рангом, но разной пригодностью должны иметь различное время жизни. В предлагаемой модели его вычисления это достигается тем, что в следующее поколение преимуще-

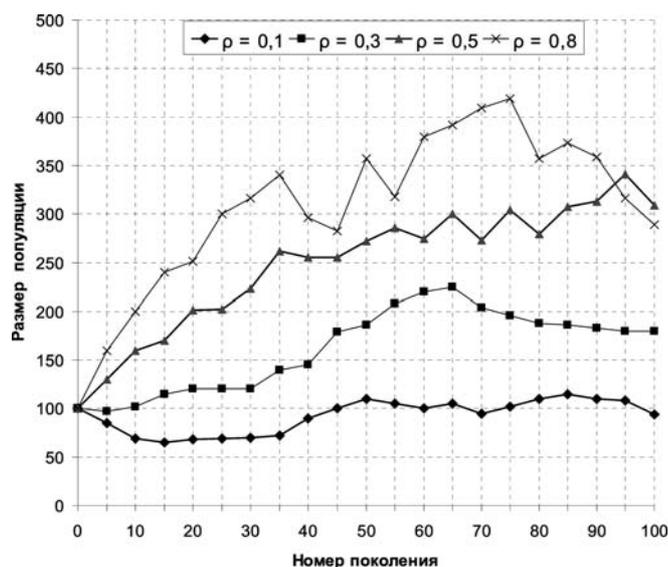


Рис. 1. Изменение размера популяции для разных значений коэффициента репродукции

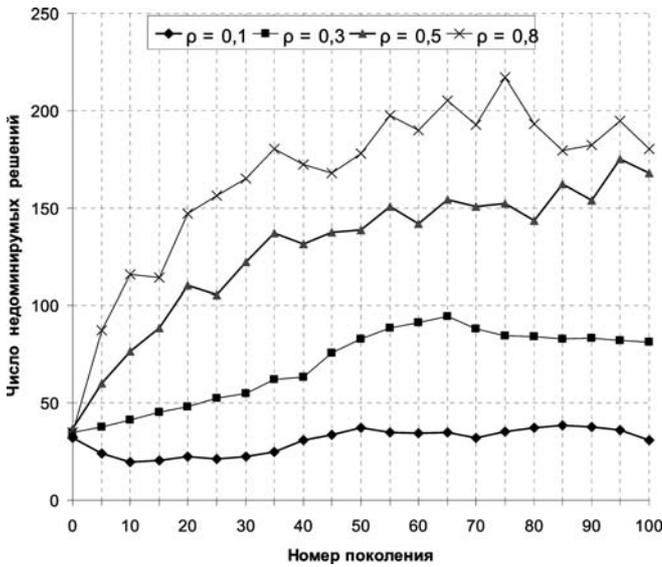


Рис. 2. Изменение числа недоминируемых решений для разных значений коэффициента репродукции

ственно переходят те хромосомы, у которых ранг Парето равен единице, либо он улучшился, а значение их пригодности при этом выше среднего. Также предлагаемая модель вычисления времени жизни хромосом отличается достаточно низкой вычислительной сложностью, не использует не относящихся к базовому МГА специальных переменных и функций. Для МГА это важный фактор, так как параметр  $LT(c_i, t)$ ,  $i = 1, \dots, n_p(t)$ , должен вычисляться для каждой хромосомы каждого поколения.

После вычисления значения  $LT(c_i, t)$ ,  $i = 1, \dots, n_p(t)$ , популяция сокращается посредством удаления из нее всех хромосом, у которых  $age(c_i, t) > LT(c_i, t)$ ,  $i = 1, \dots, n_p(t)$ . Таким образом, общий алгоритм работы МГА с данной моделью изменения размера популяции (MOGA+SAPS — MultiObjective Genetic Algorithm with Self-Adaptive Population Size) состоит в следующем.

**Алгоритм** MOGA+SAPS ( $n_p(t=0)$ ,  $\rho$ ,  $\{P\}$ ),

$n_p(t=0)$  — начальный размер популяции;

$\rho$  — коэффициент репродукции;

$\{P\}$  — множество управляющих параметров базового МГА.

**Шаг 1.**  $t = 0$ . Создать начальную популяцию размера  $n_p(t)$  и вычислить ранг Парето каждой хромосомы:

$$\forall c_i, i = 1, \dots, n_p(t): LT(c_i, t) = \text{rank}(c_i, t); \text{age}(c_i, t) = 0.$$

**Шаг 2.**  $t = t + 1$ . Изменить размер популяции на величину

$$n_p^+(t) = n_p(t)\rho; n_p(t) = n_p(t) + n_p^+(t).$$

**Шаг 3.** Применить генетические операторы базового МГА к хромосомам популяции. Увеличить

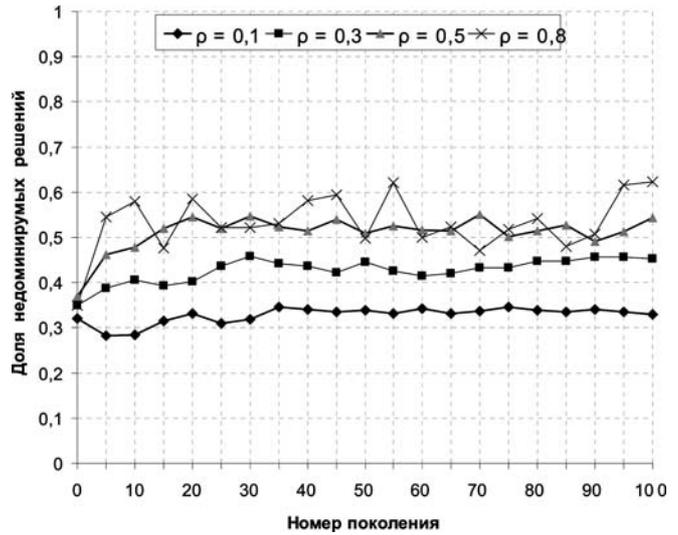


Рис. 3. Изменение доли недоминируемых решений для разных значений коэффициента репродукции

возраст "старых" хромосом, установить всем новым хромосомам нулевой возраст:

$$\forall c_i, i = 1, \dots, n_p(t) - n_p^+(t): \text{age}(c_i, t) = \text{age}(c_i, t) + 1;$$

$$\forall c_k, k = n_p(t) - n_p^+(t) + 1, \dots, n_p(t): \text{age}(c_k, t) = 0.$$

**Шаг 4.** Используя систему (1), вычислить для каждой хромосомы время жизни.

**Шаг 5.** Изменить размер популяции на величину

$$n_p^-(t) = |\{c_k, k = 1, \dots, n_p(t)\} | \text{age}(c_k, t) > LT(c_k, t);$$

$$n_p(t) = n_p(t) - n_p^-(t).$$

**Шаг 6.** Если выполнено условие окончания работы МГА, то закончить. Иначе перейти к шагу 2.

В приведенном алгоритме в качестве базового может быть использован любой МГА, однако достижение наибольшего эффекта возможно от применения изменяемого размера популяции в МГА, где не используется Парето-архив, а все найденные недоминируемые решения остаются в популяции. Сейчас, среди наиболее известных, это NSGA-II, который в сочетании с модификацией SAPS (NSGA-II+SAPS) может быть использован для анализа влияния ее основного управляющего параметра (коэффициент репродукции) на эффективность поиска.

### 1.3. Выбор значения коэффициента репродукции

Для анализа влияния различных значений коэффициента репродукции на изменение размера и состава популяции был проведен ряд экспериментов. В них алгоритм применялся для решения тестовой задачи DTLZ1 [15, 16] (число критериев  $m = 2$ ), при этом использовалось множество различных значений параметра  $\rho = \{0,1; 0,3; 0,5; 0,8\}$ . Для каж-

дого из них задавались одинаковые размер начальной популяции  $n_p(t = 0) = 100$ , число поколений, а также значения вероятностей операторов кроссинговера и мутации. В процессе поиска множества Парето оценивались динамика изменения размера популяции (см. рис. 1), число (см. рис. 2) и доля (см. рис. 3) в ней недоминируемых решений. Сравнительный анализ графиков, представленных на этих рисунках позволяет сделать следующие выводы.

1. Для малых значений коэффициента репродукции ( $\rho < 0,3$ ) не наблюдается заметного роста численности популяции. Она сначала снижается, далее растет, но достигает и превышает свой начальный размер, только начиная со второй половины поколений. При этом доля недоминируемых решений в популяции относительно высока (более 30 %) и почти не изменяется в течение всего поиска. Это позволяет сделать вывод о его быстрой стагнации из-за недостаточного разнообразия популяции — преобладания в ней одинаковых, постоянно повторяющихся хромосом.

2. С ростом значения коэффициента репродукции ( $\rho \geq 0,3$ ) у популяции заметен рост численности уже в первой трети поколений. При достижении  $\rho = 0,5$  популяция, а также число в ней недоминируемых решений быстро удваиваются. В последней трети поколений число и доля недоминируемых решений для  $\rho = 0,3$  практически не изменяется, что свидетельствует о схождении алгоритма в целом. В сравнении с этим изменение числа и доли недоминируемых решений при  $\rho = 0,5$  в этот интервал времени все еще подвержено заметным колебаниям, не позволяющим сделать окончательный вывод о схождении алгоритма к заданному числу поколений.

3. Дальнейшее увеличение значения коэффициента репродукции ( $\rho > 0,5$ ) привело к быстрому росту численности популяции в самом начале работы МГА. Впоследствии динамика изменения размера популяции и число недоминируемых решений в ней принимает хаотичный характер. Такое свойственно исключительно применению оператора мутации для получения новых членов популяции. Соотношение размера популяции и числа недоминируемых решений в ней носит преимущественно случайный характер. Это позволяет сделать вывод о существенных проблемах схождения МГА с такими значениями коэффициента репродукции, причем не только в условиях решаемой задачи МО.

Таким образом, для решаемой тестовой задачи наилучший баланс между скоростью роста популяции и числом недоминируемых решений в ней достигается при  $\rho \in [0,2, \dots, 0,4]$ . В то же время решение рассматриваемой задачи МО при тех же условиях экспериментов, но с большим числом критериев ( $m = 8$ ) показали оправданность использования коэффициента репродукции  $\rho > 0,5$  ( $\rho = 0,7$ ). Обосновать это можно нелинейным увеличением скорости заполнения популяции недоминируемыми

решениями при росте числа оптимизируемых критериев. Тем самым, только возможность быстрого роста численности популяции в моменты времени процесса оптимизации, когда требуется увеличение разнообразия популяции, позволяет сохранить баланс между исследованными и неисследованными участками пространства поиска.

## 2. Проведение испытаний модификации SAPS

Для оценки эффективности использования предложенной модели самоадаптации размера популяции был проведен ряд экспериментов. Они заключались в решении набора тестовых задач МО с различным числом критериев. Это позволило оценить влияние модификации SAPS на сохранение эффективности базовых МГА при росте сложности решаемых задач. В качестве таких МГА использовались наиболее известные SPEA2 (Strength Pareto Evolutionary Algorithm) [8] и NSGA-II (Nondominated Sorting Genetic Algorithm) [9]. Эти МГА отличаются относительно простой реализацией и настройкой, а также различными принципами поиска решений. В частности, SPEA2, в отличие от NSGA-II, для сохранения найденных в процессе поиска недоминируемых решений и управления ими использует Парето-архив. Оба этих МГА имеют реализации в ряде коммерческих и свободных библиотек программирования для решения задач оптимизации, а также являются обязательными участниками сравнительных испытаний с другими МГА.

В процессе МО оценка качества результатов, полученных МГА, определяется не только точностью (близостью к глобально оптимальным решениям) решений из множества Парето, но и по ряду других показателей, связанных с особенностями найденной границы Парето. Очевидно, что образующие ее точки могут иметь произвольные численность и расположение в пространстве критериев. Поэтому анализ результатов МО и эффективности самого МГА должен осуществляться на основе комплексной оценки по целому набору показателей качества, в частности, включающему:

- минимальное отличие результирующего множества Парето-оптимальных решений от глобального множества Парето, содержащего недоминируемые решения из всей области допустимых значений переменных оптимизаций;
- максимальные равномерность и плотность распределения решений вдоль найденной МГА границы Парето;
- максимальную протяженность найденного множества Парето;
- минимальное время поиска решений.

### 2.1. Индикаторы эффективности работы МГА

В результате решения задачи МО качество найденных недоминируемых решений, прежде всего,

может быть оценено на основе анализа соответствующей границы Парето. В общем случае для этого используется вычисление значений специальных индикаторов [17], позволяющих оценить различные свойства найденного МГА множества Парето  $A$  в пространстве решений или критериев ( $A_Y$ ). Среди известных индикаторов был выбран набор, позволяющий количественно оценить вышеназванные показатели качества результатов МГА и в совокупности требующий относительно небольшой объем вычислений.

*Индикатор  $I_{ONVG}$*  (Overall Nondominated Vector Generation) определяет мощность найденного множества Парето  $A$ :  $I_{ONVG}(A) = |A|$ , где  $|A|$  — число элементов множества  $A$ . Более высокое значение индикатора соответствует лучшему результату.

*Индикатор  $I_S$*  (Spacing) используется для оценки равномерности распределения решений вдоль границы Парето и вычисляется следующим образом:

$$I_S(A_Y) = \frac{\sum_{i=1}^{|A|-1} |d_i - \bar{d}|}{(|A|-1)\bar{d}},$$

где  $d_i$  — расстояние между каждой парой соседних решений, предварительно упорядоченных по некоторому критерию;  $\bar{d}$  — среднее всех значений  $d_i$ .

Данный индикатор принимает значения  $I_S \geq 0$ . Значение  $I_S$ , близкое к нулю, свидетельствует о достаточно равномерном распределении решений вдоль границы Парето.

*Индикатор  $I_{DE}$*  (Dimensions Extent) позволяет оценить максимальную протяженность границы Парето по каждой из размерностей. Чем больше его значение, тем шире диапазон изменения каждого критерия границы Парето

$$I_{DE}(A_Y) = \sqrt{\sum_{i=1}^m \max\{|f_i(a) - f_i(b)| \mid a, b \in A\}},$$

где  $m$  — число критериев.

*Индикатор  $I_{GD}$*  (Generational Distance) позволяет оценить степень близости некоторой границы Парето  $A_Y$  и заданной эталонной границы  $B_Y$  множества Парето  $B$ :

$$I_{GD}(A_Y, B_Y) = \frac{1}{|A|} \sum_{a \in A_Y} \{\|a - b\| \mid b \in B_Y\}.$$

Таким образом,  $I_{GD}$  вычисляется как сумма расстояний в пространстве критериев между каждым вектором  $a \in A_Y$  и ближайшим к нему вектором  $b \in B_Y$ . Меньшее значение  $I_{GD}$  соответствует лучшей оценке.

*Индикатор  $I_{OT}$*  (Overall Time Computing) ориентирован исключительно на оценку временной эффективности МГА при определении множества Парето. Под ней понимается общее время вычислений (часто в сравнении с другими аналогичными

показателями), затрачиваемое алгоритмом на решение задачи. Очевидно, что меньшее значение этого индикатора соответствует лучшему показателю. При сравнении результатов работы различных МГА алгоритм с лучшим значением индикатора  $I_{OT}$ , как правило, имеет более эффективную внутреннюю реализацию с точки зрения интенсивности использования процессорного времени, например, задействует минимальное число операций выполнения генетических операторов для получения нового поколения. В итоге это может предотвратить непредсказуемый рост времени вычислений при увеличении вычислительной сложности задачи МО, прежде всего, из-за увеличения числа критериев (переменных) оптимизации.

## 2.2. Методика проведения испытаний

Для каждого МГА должен быть определен набор значений управляющих параметров. В данном случае это выполнялось как в соответствии с рекомендациями [4], так и с использованием ранее полученных экспериментальных результатов. Как уже отмечалось, в работе МГА ключевую роль играет размер популяции. В работе [4] были определены соотношения между числом критериев, размером популяции и максимальным числом недоминируемых решений в ней. Следуя полученным рекомендациям, были выбраны значения размера популяции и Парето-архива для разного числа критериев (табл. 1).

Время работы МГА во всех случаях ограничивалось только числом поколений. Относительно выбора этого числа не существует определенных рекомендаций для МГА. Вместе с тем, можно предположить о наличии прямой зависимости между числом поколений МГА и численностью решений найденного множества Парето. Однако экспериментально показано, что простое увеличение времени работы МГА не гарантирует роста недоминируемых решений, а в случае мультимодальности отдельных критериев приводит к снижению мощности итогового множества Парето [4]. Поэтому в данном случае значения числа поколений выбирались с точки зрения сохранения такой пропорции с размером популяции, которая была бы достаточной для возникновения у процесса оптимизации при-

Таблица 1

Значения параметров работы МГА, зависящих от числа критериев

Число критериев	Размер популяции/ (с использованием SAPS)	Размер Парето-архива (доля от размера популяции)	Число поколений	Коэффициент репродукции
2	100/(100)	28 (0,28)	300	0,3
4	250/(100)	113 (0,45)	500	0,4
6	400/(100)	208 (0,52)	700	0,6
8	600/(100)	360 (0,6)	1000	0,7

знаков схождения. Принимая во внимание, что генетические алгоритмы имеют стохастическую составляющую, для объективности испытаний МГА запускались по 30 раз для  $m = \{2, 4\}$ , а учитывая резкий рост времени поиска — по 10 раз для  $m = \{6, 8\}$ . Для всех МГА и решаемых задач использовались турнирный отбор, одноточечный кроссинговер и одноточечная мутация. Значения вероятностей операторов кроссинговера и мутации ( $p_c, p_m$ ) выбирались из заданных интервалов  $p_c \in [0,7, 0,9]$ ,  $p_m \in [0,001, 0,01]$ /бит соответственно. Для этого при решении каждой из задач, но только для  $m = 2$ , определялись значения  $p_c, p_m$ , при которых SPEA2, NSGA-II достигали лучшего результата по индикатору  $I_{GD}$ . Найденные таким образом значения вероятностей для каждого МГА использовались во всех остальных случаях. Это позволило упростить процедуру исследования МГА без потери объективности полученных результатов.

### 3. Полученные результаты

Результаты решения всех задач приведены в табл. 2. Значения всех индикаторов усреднены по выполненному числу запусков МГА. Значения показателей, которые удалось улучшить, используя модификацию SAPS, выделены полужирным.

Как видно из таблицы, основной эффект от использования модификации SAPS в МГА связан с

увеличением во всех случаях числа найденных Парето-оптимальных решений. Это, в свою очередь, повлияло на улучшение значений ряда других индикаторов, в частности  $I_{DE}$ , предназначенного для оценки протяженности границы Парето. В отдельных случаях улучшились результаты по индикаторам  $I_{GD}$  и  $I_S$ , что, вероятно, вызвано увеличением плотности заполнения границы Парето большим числом решений, чем у оригинальных SPEA2 и NSGA-II. В то же время ни одна из модифицированных версий этих МГА не оказалась лучшей во всех задачах. Однако заметим, что в двух из них (DTLZ1, DTLZ6) алгоритм NSGA-II+SAPS полностью превзошел SPEA2+SAPS. Также обе версии NSGA-II во всех экспериментах оказались быстрее своего конкурента. Причем с ростом числа критериев разница во времени между SPEA2 и NSGA-II становится существенной. Вместе все это позволяет сделать вывод о предпочтительности использования предложенной модели динамического изменения размера популяции именно в NSGA-II.

### Заключение

Предложенная модель адаптивного управления размером популяции МГА в процессе МО позволяет обеспечить не только поддержание разнообразия популяции за счет постоянного обновления времени жизни хромосом, но и увеличить число найденных Парето-оптимальных решений. Это особенно важно, когда их число ограничивается первоначально заданным фиксированным размером популяции. Так, использование созданной модификации SAPS в генетических алгоритмах многокритериальной оптимизации SPEA2 и NSGA-II позволило достичь увеличения числа найденных Парето-оптимальных решений в сочетании с улучшением качества связанной с ними границы Парето. Рост временной сложности при интеграции в МГА механизмов динамического управления численностью популяции зависит от сложности топологии пространства критериев, а также выбора значения для единственного дополнительного управляющего параметра (коэффициента репродукции).

Дальнейшее развитие предложенной модели динамического управления размером популяции может быть связано с учетом ограничений на переменные и критерии оптимизации при вычислении времени жизни хромосом. Это может быть достигнуто посредством определения зависимости между длительностью существования хромосомы в популяции и степенью соответствия связанного с ним решения ограничения математической модели. Практическая реализация этого позволит расширить возможности применения МГА с модификацией SAPS в различных прикладных задачах МО, например, при проектировании технических систем.

Таблица 2

Результаты исследования эффективности модификации SAPS

Задача	Индикатор	SPEA2	NSGA-II	SPEA2 + SAPS	NSGA-II + SAPS
DTLZ1 ( $m = 2$ )	$I_{ONVG}$	28	35	28	<b>66</b>
	$I_{GD}$	0,071	0,067	<b>0,068</b>	<b>0,062</b>
	$I_S$	0,186	0,223	0,188	<b>0,214</b>
	$I_{DE}$	0,971	0,964	0,968	<b>0,973</b>
	$I_{OT}$	6,3	4,8	9,3	7,4
DTLZ2 ( $m = 4$ )	$I_{ONVG}$	113	138	<b>113</b>	<b>179</b>
	$I_{GD}$	5,732	6,124	<b>5,441</b>	6,132
	$I_S$	0,133	0,124	<b>0,121</b>	0,125
	$I_{DE}$	1,714	1,847	<b>1,793</b>	<b>1,892</b>
	$I_{OT}$	191,4	31,6	267,8	48,5
DTLZ3 ( $m = 6$ )	$I_{ONVG}$	188	207	<b>208</b>	<b>236</b>
	$I_{GD}$	226,953	310,603	231,762	314,547
	$I_S$	0,328	0,287	<b>0,297</b>	<b>0,276</b>
	$I_{DE}$	2,137	2,312	<b>2,167</b>	<b>2,376</b>
	$I_{OT}$	6217,3	698,7	10342,7	1095,4
DTLZ6 ( $m = 8$ )	$I_{ONVG}$	319	371	<b>360</b>	<b>401</b>
	$I_{GD}$	16,623	11,237	<b>14,547</b>	<b>11,142</b>
	$I_S$	0,251	0,201	<b>0,247</b>	<b>0,196</b>
	$I_{DE}$	1,862	1,913	<b>2,034</b>	<b>2,173</b>
	$I_{OT}$	76784,3	2843,8	128132,5	4546,4

## Список литературы

1. **Соболь И. М., Статников Р. Б.** Выбор оптимальных параметров в задачах со многими критериями. 2-е изд., перераб. и доп. М.: Дрофа, 2006. 175 с.
2. **Ногин В. Д.** Принятие решений в многокритериальной среде: количественный подход. 2-е изд., испр. и доп. М.: ФИЗМАТЛИТ, 2004. 176 с.
3. **Гладков Л. А., Курейчик В. В., Курейчик В. М.** Генетические алгоритмы. М.: ФИЗМАТЛИТ, 2006. 320 с.
4. **Deb K.** Multi-Objective Optimization Using Evolutionary Algorithms. Wiley, 2009. 536 p.
5. **Казаков П. В.** Генетические алгоритмы многокритериальной оптимизации. Обзор // Информационные технологии. 2011. № 9. С. 2—8.
6. **Карпенко А. П., Овчинников В. А., Семенихин А. С.** Программная система PRADIS//FRONT для построения множества Парето в задаче многокритериальной оптимизации динамических систем с использованием параллельного генетического алгоритма // Информационные технологии. 2009. № 8. С. 27—33.
7. **Карпенко А. П., Митина Е. В., Семенихин А. С.** Когенетический алгоритм Парето-аппроксимации в задаче многокритериальной оптимизации // Информационные технологии. 2013. № 1. С. 22—32.
8. **Zitzler E., Laumanns M., Thiele L.** SPEA2: Improving the Strength Pareto Evolutionary Algorithm // Proc. of the EUROGEN 2001. Evolutionary Methods for Design, Optimization and Control with Applications to Industrial Problems. 2002. P. 95—100.
9. **Deb K., Pratap A., Agarwal S., Meyarivan T.** A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II // IEEE Transactions on Evolutionary Computation. 2002. Vol. 6, N 2. P. 182—197.
10. **Michalewicz Z.** Genetic Algorithms + Data Structures = Evolutionary Programs. New York: Springer, 1996. 387 p.
11. **Гладков Л. А., Зинченко Л. А., Курейчик В. В., Курейчик В. М., Лебедев Б. К., Нужнов Е. В., Сорокин С. Н.** Методы генетического поиска. Таганрог: Изд-во ТРТУ, 2002. 122 с.
12. **Coello Coello A., Reyes Sierra M.** Coevolutionary Multi-objective Evolutionary Algorithm // Proc. of the 2003 Congress on Evolutionary Computation (CEC'2003). 2003. Vol. 1. P. 482—489.
13. **Yen G. G., Lu H.** Dynamic multiobjective evolutionary algorithm: adaptive cell-based rank and density estimation // IEEE Transactions on Evolutionary Computation. 2003. Vol. 7, N 3. P. 253—274.
14. **Tan K. C., Lee T. H., Khor E. F.** Evolutionary algorithms with dynamic population size and local exploration for multiobjective optimization // IEEE Transactions on Evolutionary Computation. 2001. Vol. 5, N 6. P. 565—588.
15. **Deb K., Thiele L., Laumanns M., Zitzler E.** Scalable Test Problems for Evolutionary Multi-Objective Optimization // Evolutionary Multiobjective Optimization. Theoretical Advances and Applications. New York: Springer, 2005. P. 105—145.
16. **Казаков П. В.** Оценка эффективности генетических алгоритмов многокритериальной оптимизации. Часть 2. // Информационные технологии. 2012. № 9. С. 42—46.
17. **Казаков П. В.** Оценка эффективности генетических алгоритмов многокритериальной оптимизации. Часть 1. // Информационные технологии. 2012. № 8. С. 2—6.

УДК 519.856.2

**В. В. Федосов**, канд. техн. наук, доц.,  
e-mail: vl.fdsv@gmail.com, Москва,

**А. В. Федосова**, канд. физ.-мат. наук, доц.,  
e-mail: afedosova@unal.edu.co,

Национальный университет, Богота, Колумбия  
(Universidad Nacional de Colombia)

## Стохастический поиск защитных нормативов объектов (зон), размещенных на общей территории с группой источников промышленных выбросов

*Предложены модель, алгоритм и получено численное решение задачи поиска защитных нормативов объектов (зон), размещенных на общей территории с группой источников выбросов. Объекты (зоны) имеют произвольные границы и численность. Источники выбросов различны по мощности, размещению и составу выбросов. Алгоритм исключает из рассмотрения промзоны источников выбросов.*

*В MATLAB разработан программный комплекс, обеспечивающий построение карт общего и покомпонентного загрязнения территории и точек, определяющих нормативы ОЗ в их границах. Комплекс позволяет моделировать проектные варианты размещения ОЗ на промышленных территориях.*

**Ключевые слова:** вредные выбросы, компоненты выбросов, нормативы загрязняющей нагрузки, зоны территории, стохастический алгоритм

## Введение

Совместное размещение группы источников вредных выбросов и группы объектов (зон) (ОЗ), защищенных экологическими нормативами, требует их оптимального соответствия. При любых изменениях параметров или характеристик участников необходимы средства оперативного моделирования для принятия обоснованных решений по оценке, выбору или отклонению возможных вариантов развития инфраструктуры территории.

При изучении промышленных выбросов в виде дисперсных или капельных частиц, химических веществ или их смесей рассматривают одно-, двух- [1, 2] или трехмерные модели [3—5, 9], различное описание факела выбросов [1—3, 5, 6, 9], асимметрию и турбулентность потока [5, 7], воздействие ветра [5], компонентный и фазовый состав, динамику [8, 9], воздействие температур [3, 5], местный рельеф [8], сезонные погодные условия и др. Реальные сложности описания выбросов ведут к быстрому нарастанию числа переменных системы, принятых допущений, привлеченных коэффициентов. При этом теряется инженерная применимость подходов. Проектировщик не в состоянии использовать их для моделирования, выделять из массива данных и оценивать влияние базовых параметров. Понимание адекватности как насыщенности моделей усложнениями часто ограничивает такие разработки только авторским применением.

Модели, отдающие приоритет экологии территории, ведут к ограничению избыточных мощно-

стей источников выбросов [1–3]. В работе рассмотрена противоположная задача — поиск допустимых нормативов ОЗ, которые сохраняют исходный промышленный потенциал территории, представленный группой источников выбросов. Она характерна для развивающихся или находящихся в стадии освоения территорий с уже сформировавшейся антропогенной нагрузкой, когда свободные площади необходимо отвести под непромышленную инфраструктуру в составе различных по перечню, площадям, размещению и экологическим нормативам ОЗ.

Моделирование, программируемые эксперименты выполнены в MATLAB. Предложенный подход совершенствует контроль и прогнозирование загрязнения территорий, совмещающих промышленную и экологическую инфраструктуру.

### Постановка задачи

На практике загрязняющие выбросы источников представляют данными мониторинга, картами прогнозов либо допустимыми функциональными приближениями. В любом из таких случаев наложения выбросов от группы источников приводят к появлению достаточно сложных, возможно многоэкстремальных, распределений загрязнений, определяемых позиционированием точек территории. При учете многокомпонентных выбросов число распределений соответственно возрастает. Обнаружение и анализ локальных максимумов таких распределений, с которыми связаны нормативы ОЗ, может потребовать емких вычислений. В сравнении с переборными и сеточными для задач такого типа предпочтительны стохастические алгоритмы. В частности, хорошо зарекомендовали себя алгоритмы полубесконечной оптимизации (*SIP, Semi-Infinite Programming*) [1–3, 10], элементы которых применены в работе.

Базовые параметры множества источников выбросов точечного типа включали: мощности, координаты размещения, переменные наборы компонентов выбросов, множества ОЗ — нормативы, произвольные геометрия и размещение. Рассмотрены загрязнения от источников при симметричном разбросе, описываемым функциями параболического типа.

### Исходные данные эксперимента

Параметры источников	Номера источников														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$H$ , у. е.	157	175	58	29	154	190	150	124	140	166	98	27	118	73	81
$i(1)$	575	128	315	210	395	570	19	339	81	80	347	504	451	533	478
$i(2)$	29	213	247	242	282	148	25	276	241	108	241	263	191	211	135
$k = 1$	0	1	0	1	0	1	1	0	1	1	0	1	1	0	1
$k = 2$	1	1	1	1	1	1	1	1	1	0	1	0	0	0	1
$k = 3$	0	0	0	1	1	1	1	1	0	1	0	1	0	1	0
$k = 4$	1	0	0	0	0	1	0	1	1	0	1	1	1	1	1
$k = 5$	1	1	1	1	0	0	1	0	1	0	1	1	1	1	0
$k = 6$	1	1	1	0	1	1	0	0	1	1	0	0	1	0	1
$k = 7$	0	1	0	1	0	0	1	1	0	0	1	0	1	0	1

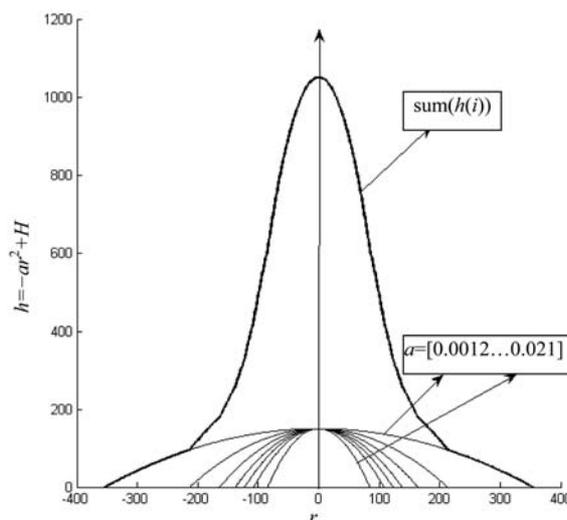


Рис. 1. Суммарное загрязнение территории одним источником в результате наложения многокомпонентного состава выбросов

Решение получено для многокомпонентных выбросов единой физической природы, например полидисперсного состава. Это подразумевает различную дальность разброса в зависимости от мощности источника и размеров фракций, но общую по компонентам суммарную загрязненность областей территории.

На рис. 1 показаны возможные связи величин и зон разброса загрязнений различного дисперсного состава выбросов при  $h = -ar^2 + H$ , где  $h$  — значение точечного загрязнения на удалении  $r$  от источника условной мощностью выброса  $H = 150$  у. е. Коэффициент параболы  $a$  связывается с конкретной фракцией частиц  $\delta$  в составе выбросов и определяет следующее сочетание направлений векторов:  $a \downarrow, \delta \downarrow$ . Выделенной линией показано распределение суммарных загрязнений от источника, которая близка к функции Гаусса, распространенной для такого класса задач [3, 5], соответствует полному набору компонентов в составе выбросов ( $K = 7$ ). Переменный состав выбросов источников учитывали введением в модель матрицы (наборов) компонентов (1/0) (см. таблицу). Наборы компонентов

приводят к вариациям суммирующего распределения по каждому источнику.

В задаче поиска допустимых нормативов ОЗ базовое ограничение принимает вид:

$$g = \left[ \sum_{n=1}^N \sum_{k=1}^K \begin{cases} \text{kod}(k, n)[-a(k)r^2 + H(n)], \\ \text{если } r < R(k, n) \text{ и } \text{kod}(k, n) = 1 \\ 0, \\ \text{если } r \geq R(k, n) \text{ или } \text{kod}(k, n) = 0 \end{cases} \right] + \lambda(\zeta(s)) - (x(\omega(s)) + \text{Norm}(\omega(s)) \leq 0, \quad (1)$$

где  $s$  — координаты случайной точки из бесконечного множества  $d$  территории;  $n$  — номер источника [1,  $N$ ];  $\text{kod}(k, n)$  — код наличия компонента  $k$  в составе выброса источника  $n$ ;  $H(n)$  — мощность источника в точке размещения с координатами  $t$ ;  $R(k, n)$  — радиус действия источника  $n$  для компонента  $k$ ;  $\zeta(s)$  — номер зоны начального загрязнения;  $\lambda(\zeta(s))$  — начальное загрязнение в  $s$ ;  $\omega(s)$  — номер ОЗ;  $\text{Norm}(\omega(s))$  — норматив загрязнения для точки  $s$  зоны  $\omega(s)$ ,  $x(\omega(s))$  — изменение (+/-) норматива ОЗ,

$$\begin{aligned} x(\omega(s)) + \text{Norm}(\omega(s)) = \\ = \begin{cases} \text{MAX}(\mathbf{Y}(\omega(\text{smax}))), \text{ если } \mathbf{Y} \notin \emptyset; \\ 0, \text{ если } \mathbf{Y} \in \emptyset, \end{cases} \quad (2) \end{aligned}$$

где  $\mathbf{Y}(\omega(\text{smax}))$  — вектор обнаруженных локальных максимумов загрязнений, отсортированных по принадлежности каждому ОЗ.

Рассмотренная модель учитывает:

- произвольные границы ОЗ (за счет полубесконечной трактовки задачи: конечное число источников и ОЗ; бесконечное число точек территории);
- переменные наборы компонентов выбросов источников;
- различную дальность разброса компонентов выбросов;
- суммарное загрязнение территории всеми компонентами выбросов;
- произвольное расположение источников выбросов;
- переменные мощности источников выбросов;
- различные уровни начального загрязнения областей территории;
- вывод из рассмотрения промзон.

Промзоны непосредственно окружают точки выброса, имеют потенциально высокие уровни и собственные нормативы, не связанные с нормативами ОЗ.

### Алгоритм решения задачи

В реальности (в особенности по данным мониторинга) рассчитывать на непрерывность функций загрязнения территории не приходится. При прогнозах или функциональном задании разброса непрерывность общей функции загрязнения может иметь ограниченные области территории. Дополнительные усложнения ограничений (начальные загрязнения, выключение промзон) приводят к появлению разрывов. Разрывы функций увеличива-

ют число локальных максимумов, причем часто они сосредоточены в точках таких разрывов. Однако стохастические алгоритмы продолжают работать с функциями, локально теряющими непрерывность (выявляют точки невыполнения ограничений и далее координаты ближайших локальных максимумов невыполнения (ЛМН). Координаты и значение ЛМН обнаруживают средствами поиска min/max функции ограничения (1) нескольких переменных. Точкой старта является очередная случайная точка невыполнения ограничения.

Структура алгоритма:

**инициализация**

$d$  (размеры территории);

$\text{Norm}0$  (вектор начальных нормативов ОЗ);

$po, \text{lim}$  (счетчик и лимит попыток);

$H, t, \text{kod}()$  (параметры источников);

$x()$  (вектор изменения нормативов);

карта начального загрязнения территории;

$P_z$  (карта промзон);

$LMN$  (матрица ЛМН);

$O_z$  (количество ОЗ);

$eps = 1$  (критерий остановки);

**пока**  $eps \neq 0$

$x() = 0$

$po = 0$

$\text{Norm}1() = \text{Norm}()$  (буфер);

$it = it + 1$  (итерация)

**пока**  $po \leq \text{lim}(it)$  (контроль лимитов)

$po = po + 1$

$s = \text{rand}(d)$  (случайная точка)

**если**  $s \notin P_z$  (вывод промзон)

**если**  $g > 0$  (нарушение ограничения)

$[s_{\text{max}}, z_{\text{max}}] = \text{fminsearch}(@s)(-zf(s), s)$   
(координаты и значение ЛМН)

**если**  $s_{\text{max}} \notin P_z$  (вывод промзон)

**если**  $\omega(s_{\text{max}}) = \omega(s)$  (рамки ОЗ)

$LMN = [LMN: s_{\text{max}}, z_{\text{max}}, \omega(s_{\text{max}})]$  (стек ЛМН)

**конец**

**конец**

**конец**

**конец**

**конец**

сортировка  $LMN$  по  $\omega(s_{\text{max}})$

**для**  $i = 1:O_z$  (номер ОЗ)

**если**  $\text{max}(z_{\text{max}}(i)) > \text{Norm}(i)$

$\text{Norm}(i) = \text{max}(z_{\text{max}}(i))$  (выборка)

**конец**

**конец**

$x() = \text{Norm}1() - \text{Norm}()$  (вектор изменения нормативов)

$eps = \text{round}(\sum_{i=1}^{O_z} \text{Norm}() - \sum_{i=1}^{O_z} \text{Norm}1())$  (критерий

остановки)

**конец**

Произвольное размещение ОЗ предполагает их единичную (сквозную) или групповую нумерации. При групповой решение определяется наиболее критичным ОЗ по группе, что может понизить долю экологически значимой площади территории. Сквозная может дать неодинаковое решение для ОЗ с равными стартовыми нормативами и таким образом увеличивает первоначальное число ОЗ в ходе поиска решения. При значительной разности ОЗ по территории групповая нумерация нецелесообразна. Допустимо сочетать оба принципа нумерации в зависимости от целей моделирования.

Старт алгоритма возможен с нулевого вектора нормативов ОЗ, хотя, как правило, уже известен некоторый вектор, диктуемый конкретными границами и расположением ОЗ территории. Оптимальное решение допускает как ослабление (+), так и усиление (–) нормативов отдельных ОЗ.

Поскольку на очередной итерации численность необнаруженных ЛМН и, следовательно, вероятность их обнаружения уменьшаются, эффективно применять растущий с итерациями лимит попыток обнаружения точек невыполнения ограничений.

### Численные эксперименты

Выше в таблице были представлены исходные данные эксперимента ( $N = 15$ ),  $t()$  — координаты источников, матрица ( $K \times M$ ) наборов компонентов в выбросах источников.

На рис. 2 (см. третью сторону обложки) представлены области начального суммарного по компонентам загрязнения территории (10 зон, четыре типа), а также очерчены границы промзон источников. В работе [1] границы промзон функционально зависели от мощностей источников. В реальности границы заданы либо увеличены за счет прилегающей территории по данным мониторинга. В границах промзон действуют иные нормативы, связанные с мощностями и технологией работы источников. На промзоны нормативы ОЗ не распространяются и поэтому должны быть выведены из рассмотрения. Области промзон в алгоритме представлены отдельной матрицей. Загрязнения в четырех типах зон начального загрязнения представлены вектором (первый тип — отсутствие загрязнения), конфигурации зон — отдельной матрицей.

Принцип нумерации ОЗ определяет размерность вектора решения. Алгоритм работает с групповой или автономной нумерациями ОЗ. При групповой (особенно при значительных количествах ОЗ в группах) существуют риски присвоения ОЗ завышенных нормативов по наихудшему в группе и, таким образом, проигрышам в экологии использования территории. По этим соображениям сравниваемые ниже карты размещения ОЗ имели автономную нумерацию.

Назначенные параметры, значения скаляров, векторов, матриц соответствуют полному покрытию территории выбросами источников.

На рис. 3 (см. четвертую сторону обложки) показана карта загрязнений территории с учетом начальной нагрузки и выбросов источников для компонента  $k = 7$ . Цифрами обозначены значения мощностей источников. Уже видны формирующиеся центры максимальных загрязнений (до ~350 у. е.), но при этом еще остаются свободные от загрязнения области. При учете загрязнений по всем компонентам выбросов загрязнения резко возрастают (до ~2200 у. е.) при полном покрытии территории (рис. 4, см. четвертую сторону обложки).

Такие представления достаточно наглядно дифференцируют загрязнение территории заданным множеством источников и матрицей наборов компонентов выбросов, но уровни загрязнения ограничены собственными зонами и, следовательно, не дают понимания о размещении ОЗ с иными конфигурациями, различными повторяемостью и защитными нормативами. Задачу решает предложенный алгоритм.

Группировка источников территории по отдельным компонентам выбросов может не покрывать всю территорию (см. рис. 3). Экологические выигрышные зоны видны просто сразу. Зоны пиковых нагрузок имеют порядок значений, соответствующий наборам (из матрицы). Просматриваются соответствия нагруженных зон наборам (из матрицы). Так источник 7 (150 у. е.), имеющий компонент 7, в выбросах на карте отмечен характерным эпицентром. Напротив, отсутствие компонента 7 в выбросах источника 1 (157 у. е.) оставило ненагруженной зону вокруг его места расположения.

Общая карта нагрузок по всем компонентам имеет уже иной порядок величин и покрывает всю территорию. Отмечено появление нескольких возвышений, причем необязательно совпадающих с непосредственными координатами источников выбросов. Но в целом геометрия нагрузок соответствует облаку размещения источников.

Однако полевые рисунки дают только зонную ориентировку. Алгоритм же работает на точечном уровне. Прямое сопоставление найденных нормативов ОЗ просматривается только для min/max значений. Таким образом, выявляются предельно жесткие нормативы. Итак, алгоритм на точечном уровне и эффективно (экономично) назначает нормативы площадным ОЗ.

Особенностью поиска является смещение координат обнаруженного ЛМН относительно первоначальной точки нарушения ограничения. В зависимости от направления и размеров смещения ЛМН может сохранять принадлежность данному ОЗ либо позиционироваться на сопредельном и даже более отдаленном ОЗ. Для приграничных ОЗ возможно покидание ЛМН учетной территории  $d$ .

Алгоритм отклоняет ЛМН, сменившие учетный номер ОЗ или покинувшие территорию.

Исключение промзон нарушает непрерывность (гладкость) функции общего загрязнения, а точки разрывов при максимизации могут быть идентифицированы как ЛМН. В результате число ЛМН возрастает.

На рис. 5 (см. четвертую сторону обложки) показана карта 1 размещения и найденные алгоритмом допустимые нормативы ОЗ территории. Результаты получены с помощью выявленных ЛМН для каждого ОЗ. Значения нормативов ОЗ достаточно дифференцированы. Конечно, точная и прямая трактовка уровней нормативов вряд ли возможна. Однако некоторые соответствия величин, связанных с размерами и расположением ОЗ, просматриваются. Наименее экологично расположение ОЗ 2, ОЗ 4 и ОЗ 5, точки которых приближены к зонам максимальных выбросов территории. Напротив, лучшим признано размещение ОЗ 8 и ОЗ 12, действительно размещенных в местах меньших выбросов. Близки к ним ОЗ 14 и ОЗ 15. Неожидан достаточно низкий норматив ОЗ 6 и высокий у ОЗ 7. Видимо сыграли роль удаленные промзоны или отклоненные ЛМН. Приемлемо ожидаемыми оказались средние нормативы остальных ОЗ.

Алгоритм указывает общую направленность выбора нормативов в условиях принятых правил и сделанных назначений. Последние, безусловно, могут быть уточнены или дополнены новыми требованиями к ЛМН, тщательности их обнаружения и условиям включения в массив выбора. Желателен при этом учет особенностей конкретных функций разброса загрязнений. В работе для обеих карт сознательно выбраны наиболее сложные примеры размещения ОЗ непосредственно по промзонам с источниками. В реальности обычно ОЗ размещают отдельно на заранее выигрышных участках территории.

Достоверность найденных нормативов ОЗ подтверждается перекрестным сравнением координат максимальных по значениям ЛМН обеих карт (показаны ■). Так близкие по расположению точки максимумов объектов 17, 1 (карта 1) и объекта 3 (рис. 6, карта 2, см. четвертую сторону обложки) определены также близким рядом нормативов: 2337, 2482, 2353. Близки по расположению и нормативам точки ОЗ 4 (карта 1, 3807) и ОЗ 11 (карта 2, 3833). Для ОЗ 7 в обеих картах нормативные точки также недалеко (3098, 3138). Похожи по расположению и нормативам ОЗ 14 (карта 1, 1957) и ОЗ 17 (карта 2, 2077); ОЗ 12 (карта 1, 1641) и ОЗ 9 (карта 2, 1893); ОЗ 10 (карта 1, 2276) и ОЗ 8 (карта 2, 2131); ОЗ 6 и ОЗ 10 (карта 2, 2889, 2953).

Приведенные сравнения дают понимание работы алгоритма — фактическая аппроксимация имеющейся функции общего загрязнения территории значениями допустимых нормативов ОЗ, плоских в заданных границах. В отличие от рассмотренных

карт с прямоугольными ОЗ, алгоритм работает с любыми иными произвольными границами ОЗ.

Полученные данные могут быть интерпретированы производными критериями для оценки общей ситуации по территории. Например, по взвешен-

ной оценке  $\min \rightarrow \sum_{i=1}^{Oz} (Norm(i)/A(i))$ , где  $Norm(i)$  —

допустимый норматив  $i$ -го ОЗ;  $A(i)$  — доля площади, занимаемая  $i$ -м ОЗ ( $\sum_{i=1}^{Ox} A(i) = 1$ ), имеем для кар-

ты 1  $\min 1 = 7,8412e + 005$ , для карты 2  $\min 2 = 8,3944e + 005$ . То есть при одинаковой антропогенной нагрузке территории карта 1 является более удачным размещением ОЗ по связке двух параметров (экологические нормативы — площади).

Алгоритм работает на точечном уровне и обнаруживает координаты максимального из ЛМН ОЗ. Найденный норматив ОЗ принципиален, однако для лица, принимающего решение, может рассматриваться как рекомендательный. В случае его неприемлемости или несоответствия планам использования территории конфигурация или размеры такого ОЗ могут быть изменены в сторону сокращения (стянуты) вокруг максимального ЛМН, а высвободившаяся площадь передана соседним ОЗ с более экологичными нормативами. Иная конфигурация ОЗ должна быть заново подвергнута моделированию.

Предложенный алгоритм учитывает достаточный набор параметров промышленной инфраструктуры территории, экономичен по вычислительным ресурсам, легко выходит на обобщения в зависимости от формируемых вопросов к учтенной информационной базе модели, открыт для дополнений или модификаций.

## Заключение

Получено численное решение полубесконечной задачи поиска допустимых защитных нормативов ОЗ при воздействии группы источников выбросов в составе единой территории.

Применен стохастический алгоритм, ставящий в соответствие вектору нормативов максимальный из обнаруженных в полях ОЗ локальных максимумов с нарушенными ограничениями модели.

Информационная база модели содержит параметры антропогенной нагрузки территории, геометрию и размещение нормативных объектов (зон), что делает алгоритм инструментом сравнения различных карт размещения ОЗ по сформулированным критериям.

## Список литературы

1. Федосов В. В., Федосова А. В. Полубесконечная модель ограничения вредных выбросов предприятий на территориях со смешанным ландшафтом // Вестник компьютерных и информационных технологий. 2011. № 8. С. 14–22.

2. Fedosova A., Kafarov V., Mahecha Bohórquez D. P. Solucion Numerica del Problema de Control de Contaminacion del Aire // Colombian Journal of Computation (RCC). 2003. Vol. 4, N 2. P. 21—28.
3. Ismael A., Vaz F., Ferreira F. C. Air pollution control with semi-infinite programming. Applied Mathematical Modelling // ELSEVIER. 2009. N 33. P. 1957—1969.
4. Антропов К. М., Казмер Ю. И., Вараксин А. Н. Описание пространственного распределения загрязнения атмосферного воздуха промышленного центра методом Land Use Regression (обзор) // Экологические системы и приборы. 2010. № 1. С. 28—41.
5. Замай С. С., Якубайлик О. Э. Модели оценки и прогноза загрязнения атмосферы промышленными выбросами в информационно-аналитической системе природоохранных служб крупного города: учеб. пособие. Красноярск: Изд-во Краснояр. гос. ун-та, 1998. 109 с.
6. Крылова А. Д., Рапуга В. Ф., Суторихин И. А. Планирование и анализ подфакельных наблюдений концентрации при-

- меси в атмосфере // Метеорология и гидрология. 1993. № 5. С. 5—13.
7. Аверина Т. А. О математической модели рассеяния частиц в неоднородном турбулентном конвективном граничном слое // Тр. Всерос. науч. конференции "Водные и экологические проблемы Сибири и Центральной Азии". Барнаул: СО РАН, 2012. Т. 1. С. 3—7.
8. Баранова М. Е., Гаврилов А. С. Методы расчетного мониторинга загрязнения атмосферы мегаполисов // Естественные и технические науки. 2008. № 4. С. 221—225.
9. Аргучинцев В. К., Аргучинцева А. В. Моделирование мезомасштабных гидротермодинамических процессов и переноса антропогенных примесей в атмосфере и гидросфере региона оз. Байкал. Иркутск: Изд. Иркутского гос. университета, 2007. 255 с.
10. Завриев С. К., Новикова Н. М., Федосова А. В. Стохастический алгоритм решения выпуклых задач полубесконечной оптимизации с ограничениями равенствами и неравенствами // Вестник Московского университета. Сер. 15. "Вычислительная математика и кибернетика". 2000 № 4. С. 30—35.

УДК 519.6(075.8)+536.24

**К. В. Максименко-Шейко<sup>1</sup>**,

канд. физ.-мат наук, ст. науч. сотр.,

**А. В. Толок<sup>2</sup>**, д-р техн. наук, проф., зав. каф.,

**Т. И. Шейко<sup>1</sup>**, д-р техн. наук, проф., зав. отд.,

e-mail: sheyko@ipmach.kharkov.ua

<sup>1</sup>Институт проблем машиностроения

им. А. Н. Подгорного НАН Украины

<sup>2</sup>ФГБОУ ВПО Московский государственный

технологический университет "Станкин"

## Сопряженная задача конвективного теплообмена в топливной кассете ТВЭЛов

*Рассмотрены сопряженные краевые задачи теплообмена для случаев, когда вязкая несжимаемая жидкость движется по каналам неканонического сечения, обтекая пучок стержней. Исследовано влияние вида упаковки на распределение скорости и температуры. Для решения использовалась теория R-функций в сочетании с вариационным методом Рунца. Рассмотрены различные упаковки ТВЭЛов. Каждая упаковка содержит 91 стержень, и соответствующие уравнения построены с использованием новых конструктивных средств теории R-функций.*

**Ключевые слова:** теория R-функций, метод Рунца, сопряженная задача, теплообмен, топливная кассета, треугольная упаковка, циклическая симметрия

### Введение

В активной зоне современных ядерных энергетических установок ядерное топливо сконцентрировано в тепловыделяющих элементах (ТВЭЛх). Одна из основных задач расчета реактора — определение поля температур, так как к ТВЭЛам предъ-

являются очень высокие требования в отношении их надежности. Выход из строя нескольких ТВЭЛов, а в реакторе их тысячи, может привести к аварийной ситуации. В этом случае появляется необходимость в решении соответствующих задач переноса в сопряженной постановке, так как эти решения позволяют проводить исследования реальных теплообменных процессов, где существенно проявляется взаимное влияние движущейся жидкости, стенок канала, ТВЭЛов и др.

Рассмотрим сопряженные краевые задачи теплообмена для случаев, когда вязкая несжимаемая жидкость движется по каналу неканонического сечения, обтекая пучок стержней. Предполагается, что физические свойства жидкости постоянны, ламинарное течение гидродинамически стабилизировано, процесс теплообмена стационарен. На внешней поверхности канала могут быть заданы условия первого, второго, третьего рода. Предполагается, что изменение плотности теплового потока вдоль оси канала за счет аксиальной теплопроводности пренебрежимо мало по сравнению с изменением теплового потока, обусловленного конвекцией. Предполагается также, что стенки трубы и внутренние стержни выполнены из изотропного материала, причем теплопроводность последнего можно считать постоянной в рассматриваемом интервале температур [1—3].

Наличие внутренних источников теплоты в элементах реактора усложняет как уравнение теплопроводности, так и методы его решения. В работах [1, 3] рассматривался теплообмен в пучке продольно обтекаемых бесконечных цилиндров. Предполагалась симметрия температурного поля в силу симметрии системы, и рассматривалась лишь область *OABC* (рис. 1).

**Цель работы** — совершенствование конструктивных средств и алгоритмов метода R-функций

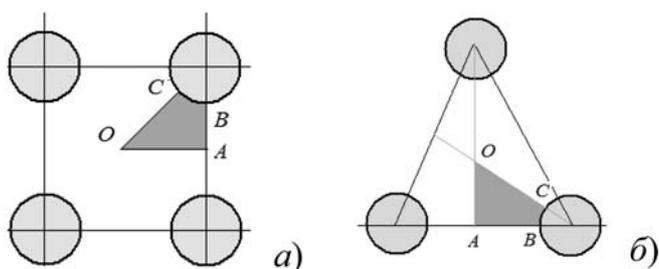


Рис. 1. Расположение рабочих каналов в активной зоне реактора: а – коридорное; б – шахматное

(функций Рвачева) для математического и компьютерного моделирования сопряженной задачи конвективного теплообмена в решетках ТВЭЛов и исследование влияния вида упаковки на распределение скорости и температуры.

### Основная часть

Основная система уравнений, описывающая процесс теплообмена в потоке вязкой жидкости при постоянных физических свойствах жидкости и температуры, имеет вид

$$\begin{cases} \frac{DT}{D\tau} = a\Delta T + \frac{q_V}{\rho c_p} + \frac{\mu\Phi}{\rho c_p}, \\ \frac{DV}{D\tau} = -\frac{1}{\rho}\nabla p + \nu\Delta V, \\ \operatorname{div}V = 0, \end{cases}$$

где  $\frac{D}{D\tau} = \frac{\partial}{\partial\tau} + (V \cdot \nabla)$  — субстанциальная (или полная) производная;  $\mu\Phi$  — диссипативная функция [2, 3];  $a = \frac{\lambda}{\rho c_p}$  — коэффициент температуропроводности;  $c_p$  — теплоемкость среды;  $q_V$  — мощность внутренних источников теплоты.

В случае стационарных процессов температура тела не зависит от времени, и уравнение теплопроводности при продольном обтекании ТВЭЛов принимает вид

$$V_z \frac{\partial T}{\partial z} = a\Delta T + \frac{q_V}{\rho c_p},$$

а математическая модель поля скоростей при ламинарном течении имеет вид  $\Delta V_z = -\frac{\Delta P}{\mu l} = -C$ , где  $\Delta P$  — постоянное вдоль трубы падение давления на произвольно выбранном участке длины  $l$ ;  $\mu$  — коэффициент вязкости [7].

В области тепловой стабилизации, когда  $\frac{\partial T}{\partial z} = \text{const}$ , получим

$$-\operatorname{div}(\lambda\nabla T) = q_V - V_z C_1.$$

Запись этого уравнения предполагает, что фактически происходит разделение переменных: температура и компонента скорости  $V_z$  ищутся только как функции координат  $x, y$ . Дифференциальные инварианты являются двумерными.

Таким образом математическая модель теплообмена при ламинарном движении жидкости по каскете с ТВЭЛами сводится к системе уравнений в 2D

$$\begin{cases} \Delta V_z = -C \text{ в } \Omega_b \cap \overline{\Omega_{tv}} \\ -\operatorname{div}(\lambda_i \nabla T_i) = F_i \text{ в } \Omega_b, \end{cases} \quad (1)$$

где  $\begin{cases} F_1 = -V_z C_1 \text{ в } \Omega_b \cap \overline{\Omega_{tv}}; \\ F_2 = q_V, q_V = \text{const в } \Omega_{tv} \end{cases}$ ;  $\Omega_b$  — область шести-

гранного кожуха;  $\Omega_{tv}$  — область ТВЭЛов;  $\Omega_b \cap \overline{\Omega_{tv}}$  — область, в которой происходит течение (канал с ТВЭЛами).

Граничные условия имеют вид:

$$\begin{aligned} V_z|_{\partial\Omega_b \cap \overline{\Omega_{tv}}} = 0, \quad \frac{\partial T}{\partial n} + hT|_{\partial\Omega_b} = 0, \quad T_1|_{\partial\Omega_{tv}} = T_2|_{\partial\Omega_{tv}}, \\ \lambda_1 \frac{\partial T_1}{\partial n} \Big|_{\partial\Omega_{tv}} = \lambda_2 \frac{\partial T_2}{\partial n} \Big|_{\partial\Omega_{tv}}, \end{aligned}$$

где  $h$  — коэффициент теплообмена;  $\lambda_i$  — коэффициент теплопроводности, а символом  $\partial$  здесь и далее обозначены границы соответствующих областей, уравнения которых строятся с использованием теории  $R$ -функций [4, 5].

При построении в работе используются  $R$ -операции системы  $\{R_0\}$ :

$$\begin{cases} f_1 \wedge_0 f_2 \equiv f_1 + f_2 - \sqrt{f_1^2 + f_2^2} \\ f_1 \vee_0 f_2 \equiv f_1 + f_2 + \sqrt{f_1^2 + f_2^2} \\ \bar{f}_1 \equiv -\bar{f}_1. \end{cases}$$

Рассмотрим типовую конструктивную схему реактора, активная зона которого собирается из большого числа топливных каскет [2, 3, 9, 10]. Каскеты представляют собой шестигранные кожухи, в которых размещены тепловыделяющие элементы. Построим уравнение топливной каскеты с 91 ТВЭЛом и раздвинутой треугольной упаковкой, которую иногда называют шахматной. Заметим, что при применении обычной методики, используемой в теории  $R$ -функций, мы получим в результате 93  $R$ -операции в уравнении. Громоздкая формула приведет не только к увеличению времени счета, но и, возможно, к некоторому нарушению симметрии вследствие неассоциативности  $R$ -операций. Поэтому для построения уравнения границы  $\omega_b$  шестигранного кожуха воспользуемся методикой, разработанной в работах [4–6]. Рассмотрим уравнение прямой  $\sigma \equiv R_b - x \geq 0$ , где  $R_b$  — радиус окружности, вписанной в правильный шестиугольник, и периоди-

ческую функцию  $\mu_v = \frac{4}{3\pi} \sum_k (-1)^{k+1} \frac{\sin[(2k-1)3\theta]}{(2k-1)^2}$ .

В результате получим  $\omega_b \equiv R_b - r \cos \mu_v = 0$ , где  $r = \sqrt{x^2 + y^2}$ ,  $\theta = \arctg \frac{y}{x}$ . Для построения треугольной

упаковки ТВЭЛов зададим  $f_1 = R_{tv}^2 - \mu_x^2 - \mu_y^2 \geq 0$ , где  $R_{tv}$  — радиус ТВЭЛа,

$$\mu_x = \frac{4h_x}{\pi^2} \sum_k (-1)^{k+1} \frac{\sin\left[(2k-1)\frac{\pi x}{h_x}\right]}{(2k-1)^2},$$

$$\mu_y = \frac{4h_y}{\pi^2} \sum_k (-1)^{k+1} \frac{\sin\left[(2k-1)\frac{\pi y}{h_y}\right]}{(2k-1)^2},$$

и  $f_2 = R_{tv}^2 - \mu_{x1}^2 - \mu_{y1}^2 \geq 0$ ,

$$\text{где } \mu_{x1} = \frac{4h_x}{\pi^2} \sum_k (-1)^{k+1} \frac{\sin\left[\frac{(2k-1)\pi(x-h_x/2)}{h_x}\right]}{(2k-1)^2},$$

$$\mu_{y1} = \frac{4h_y}{\pi^2} \sum_k (-1)^{k+1} \frac{\sin\left[\frac{(2k-1)\pi(y-h_y/2)}{h_y}\right]}{(2k-1)^2},$$

$h_x, h_y$  — шаг трансляции по осям  $Ox, Oy$  соответственно. Тогда уравнение топливной кассеты имеет вид

$$\omega \equiv \omega_b \wedge_0 \bar{\omega}_{tv} \geq 0, \quad \omega_{tv} \equiv (f_1 \vee_0 f_2) \geq 0.$$

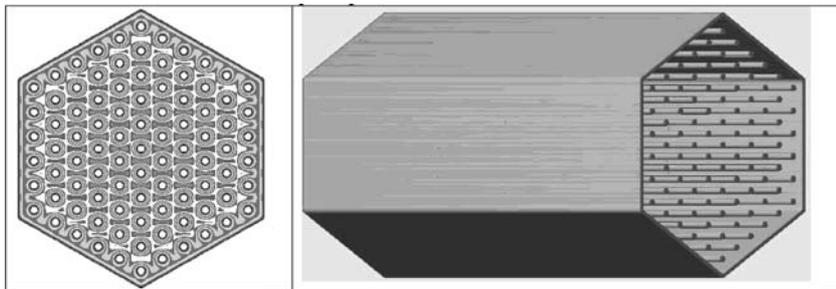


Рис. 2. Топливная кассета с 91 ТВЭЛами, расположенными по шахматной схеме

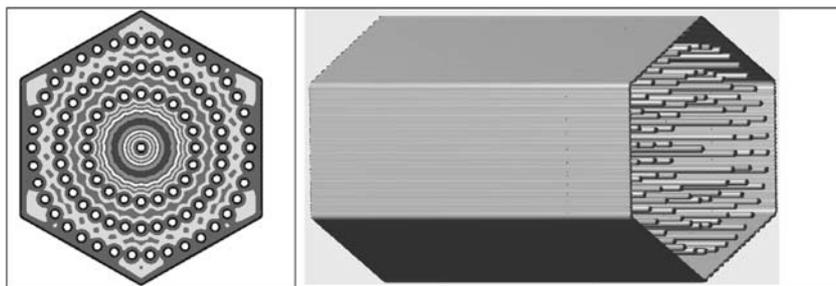


Рис. 3. Топливная кассета с 91 ТВЭЛом при трансляции с циклической симметрией и центральным ТВЭЛом

На рис. 2 приведена картина линий уровня функции  $\omega(x, y)$  при следующих значениях буквенных параметров:  $R_{tv} = 0,2$ ;  $h_x = 2,32$ ;  $h_y = 1,35$ ;  $R_b = 6,7$ , и общий вид кассеты в 3D. Следует отметить, что при построении уравнения кассеты по новой методике  $R$ -операции использовались лишь 2 раза.

Остановимся также на построении функции  $\omega(x, y) \equiv \omega_b \wedge_0 \bar{\omega}_{tv} \geq 0$ , когда ТВЭЛ транслируется с циклической симметрией  $n_d$  раз вдоль окружности радиуса  $R$ ,  $n_b$  раз вдоль окружности радиуса  $R_1$  и  $n_c$  раз вдоль окружности радиуса  $R_2$ .

Для построения уравнения границы ТВЭЛа, транслированного с циклической симметрией  $n_d$  раз вдоль окружности радиуса  $R$ , воспользуемся функцией  $\omega_o \equiv \frac{1}{2R_{tv}} (R_{tv}^2 - (x - R)^2 - y^2)$  и формулой

$$\mu_d = \frac{8}{n_d \pi} \sum_k (-1)^{k+1} \frac{\sin\left[(2k-1)\frac{n_d \theta}{2}\right]}{(2k-1)^2}.$$

В результате получим  $\omega_{tv1} \equiv \frac{1}{2R_{tv}} (R_{tv}^2 - (r \cos \mu_d - R)^2 - (r \sin \mu_d)^2) \geq 0$ .

Для построения уравнения границы ТВЭЛа, транслированного с циклической симметрией  $n_b$  раз вдоль окружности радиуса  $R_1$ , воспользуемся

$$\text{функцией } \omega_{o1} \equiv \frac{1}{2R_{tv}} (R_{tv}^2 - (x - R_1)^2 - y^2)$$

и формулой  $\mu_b = \frac{8}{n_b \pi} \sum_k (-1)^{k+1} \frac{\sin\left[(2k-1)\frac{n_b \theta}{2}\right]}{(2k-1)^2}$ .

$$\text{Тогда } \omega_{tv2} \equiv \frac{1}{2R_{tv}} (R_{tv}^2 - (r \cos \mu_b - R_1)^2 - (r \sin \mu_b)^2) \geq 0.$$

Для построения уравнения границы ТВЭЛа, транслированного с циклической симметрией  $n_c$  раз вдоль окружности радиуса  $R_2$ , воспользуемся функцией

$$\omega_{o2} \equiv \frac{1}{2R_{tv}} (R_{tv}^2 - (x - R_2)^2 - y^2)$$

и формулой

$$\mu_c = \frac{8}{n_c \pi} \sum_k (-1)^{k+1} \frac{\sin\left[(2k-1)\frac{n_c \theta}{2}\right]}{(2k-1)^2}.$$

Тогда  $\omega_{tv3} \equiv \frac{1}{2R_{tv}} (R_{tv}^2 - (r \cos \mu_c - R_2)^2 - (r \sin \mu_c)^2) \geq 0$ .

Таким образом уравнение границы кассеты с 91 ТВЭЛом имеет вид  $\omega \equiv (\omega_b \wedge_0 \bar{\omega}_{tv1} \vee_0 \omega_{tv2} \vee_0 \omega_{tv3}) \geq 0$ , при  $n_d = 38$ ,

$R = 6, n_b = 32, R_1 = 4,5, n_c = 21, R_2 = 3$  и является семипараметрическим  $(n_b, n_d, n_c, R, R_1, R_2, R_{iv})$  семейством кривых (рис. 3). Следует отметить, что  $R$ -операции использовались лишь 3 раза. При наличии центрального ТВЭЛа получим  $\omega \equiv \omega_b \wedge_0 \omega_{iv1} \vee_0 \omega_{iv2} \vee_0 \omega_{iv3} \vee_0 \frac{1}{2R_{iv}} (R_{iv}^2 - x^2 - y^2) \geq 0$  при  $n_c = 20$ .

Зная уравнения кассеты и ТВЭЛов, мы можем постановку задачи (1) переписать в виде  $\begin{cases} \Delta V_z = -C \\ -\text{div}(\lambda \nabla T) = F \end{cases}$ , с граничными условиями

$$V_z|_{\partial\Omega_b \cap \overline{\partial\Omega_{iv}}} = 0, \quad \frac{\partial T}{\partial n} + hT|_{\Omega_b} = 0, \quad T_1|_{\partial\Omega_{iv}} = T_2|_{\partial\Omega_{iv}},$$

$$\lambda_1 \frac{\partial T_1}{\partial n} \Big|_{\partial\Omega_{iv}} = \lambda_2 \frac{\partial T_2}{\partial n} \Big|_{\partial\Omega_{iv}},$$

где  $\lambda = \lambda_1 \frac{1 - \text{sign}\omega_{iv}}{2} + \lambda_2 \frac{1 + \text{sign}\omega_{iv}}{2}$ ,

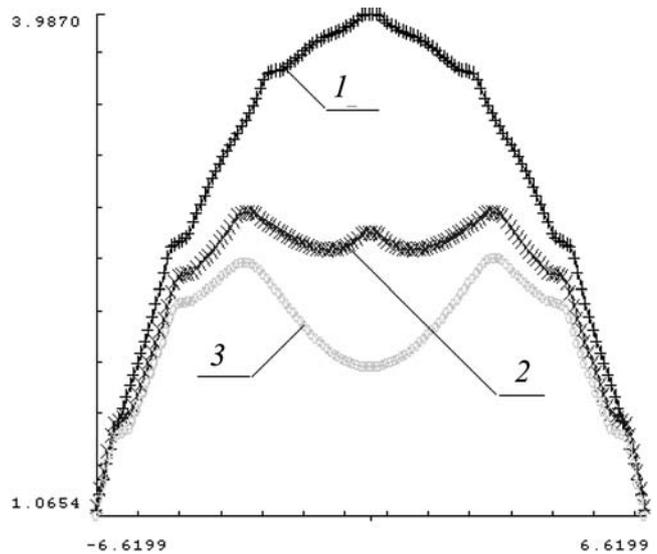
$$F = -V_z C_1 \frac{1 - \text{sign}\omega_{iv}}{2} + q_V \frac{1 + \text{sign}\omega_{iv}}{2}, \quad q_V = \text{const.}$$

Для решения использовался метод  $R$ -функций в сочетании с вариационным методом Ритца. Структура решения задачи о ламинарном течении при продольном обтекании ТВЭЛов жидкостью имеет вид  $V_z = \omega \Phi_1$ , где  $\omega(x, y) \equiv \omega_b \wedge_0 \bar{\omega}_{iv} \geq 0$  — уравнение границы области течения жидкости. Неопределенную компоненту  $\Phi_1 = \sum_{i=1}^N c_{ik} \phi_{ik}(x, y)$  будем отыскивать, минимизируя функционал  $I = \int_{\Omega \cap \overline{\partial\Omega_{iv}}} ((\nabla V_z - 2CV_z)d\Omega$ .

Заметим, что решение  $V_z$  получаем в аналитическом виде и используем без какой-либо дальнейшей обработки (аппроксимации, интерполяции). Поэтому полученное распределение скорости мы подставляем в правую часть уравнения теплопроводности. Структура решения задачи определения температурного поля использовалась как точно удовлетворяющая граничным условиям на  $\partial\Omega_b$   $u = \Phi_2 + \omega_b(-D_1\Phi_2 + h\Phi_2)$ , так и в виде  $T = \Phi_2$ , где, как и ранее,  $\Phi_2 = \sum_{i=1}^N d_{ik} \phi_{ik}(x, y)$ . Здесь следует от-

метить, что граничные условия  $\frac{\partial T}{\partial n} + hT|_{\partial\Omega_b} = 0$  и  $\lambda_1 \frac{\partial T_1}{\partial n} \Big|_{\partial\Omega_{iv}} = \lambda_2 \frac{\partial T_2}{\partial n} \Big|_{\partial\Omega_{iv}}$  являются естественными

и следуют из функционала Ритца  $I = \int_{\Omega} (\lambda(\nabla T)^2 - 2FT)d\Omega + \int_{\partial\Omega_b} hT^2 d\Omega_b$ . В качестве аппроксима-



**Рис. 7. Графики поля температур для различных упаковок ТВЭЛов в сечении  $y = 0$ :**  
 1 — шахматная схема; 2 — циклическая симметрия с центральным ТВЭЛом; 3 — циклическая симметрия без центрального ТВЭЛа

ционных средств  $\phi_{ik}(x, y)$  использовались кубические сплайны Шенберга при  $N = 6400, 10000$ . Вычислительные эксперименты проводили в условиях эксплуатации системы ПОЛЕ, разработанной в отделе прикладной математики и вычислительных методов ИПМаш НАН Украины [8]. На рис. 4—6 (см. вторую сторону обложки) приведены результаты исследований для различных упаковок ТВЭЛов. Каждая упаковка содержит 91 стержень при всех прочих равных условиях  $\lambda_1 = 1, \lambda_2 = 10, h = 1, q_1 = 4$ . Меняя значения буквенных параметров, получим различные распределения исследуемых полей.

На рис. 7 приведен сравнительный анализ графиков поля температур для различных упаковок ТВЭЛов.

Анализируя результаты, можно сделать вывод, что наличие стержней в центральной зоне приводит к более высокой температуре. Поэтому, меняя конструктивный характер упаковки и типы симметрии, можно регулировать характер течения и распределение температуры по кассете, добываясь необходимого значения, обусловленного техническим заданием. Анализ характера распределения скорости и температуры позволяет сделать вывод, что рассмотрение поля скоростей не для всей области течения при шахматной упаковке ТВЭЛов (см. рис. 4), а лишь для ячейки в случае ее достаточной удаленности от границы, является целесообразным. Однако поле температур при этом будет весьма далеким от действительности, о чем свидетельствуют результаты, полученные для кассеты в целом. То есть отыскание поля температур необходимо проводить во всей области течения.

## Выводы

Показано, что метод  $R$ -функций является эффективным методом решения задач расчета физических полей в элементах конструкций ядерных энергетических установок сложной формы. Разработанные конструктивные средства построения уравнений границ областей с трансляционным и циклическим типами симметрии позволили существенно уменьшить число операций с последующей автоматизацией этого процесса, а следовательно, уменьшить и время решения задач. При этом следует отметить, что метод  $R$ -функций дает возможность точно удовлетворить граничным условиям. Проведенные исследования позволяют конструкторам выбирать те или иные виды упаковок в зависимости от технических требований. При этом существенным является расчет температурного поля для кассеты в целом. Математическое моделирование и связанный с ним компьютерный эксперимент незаменимы в тех случаях, когда натурный эксперимент невозможен или затруднен по тем или иным причинам. Кроме того, работа с математической моделью процесса и вычислительный эксперимент дают возможность безболезненно, относительно быстро и без существенных затрат исследовать свойства и течение процесса в различных ситуациях. В то же время вычислительные эксперименты с моделями объектов позволяют, опираясь на современные численные методы, подробно и глубоко их изучать. Достоверность аналитической идентификации гео-

метрических объектов подтверждена их визуализацией. Достоверность методов расчета, результатов и выводов подтверждена сравнением со сведениями, известными в литературе, анализом численной сходимости решений и вычислением невязки.

## Список литературы

1. Слесаренко А. П., Котульский Д. А. Регионально-аналитический и вариационные методы в решении сопряженных задач конвективного теплообмена // Теплообмен ММФ-2000: Труды IV Минского международного форума (Беларусь, Минск, май 2000). Минск: ИТМО АН Беларуси, 2000. Т. 3. С. 135–142.
2. Петухов Б. С. Теплообмен и сопротивление при ламинарном течении жидкости в трубах. М.: Энергия, 1967. 412 с.
3. Петухов Б. С., Генин Л. Г., Ковалев С. А. Теплообмен в ядерных энергетических установках. М.: Атомиздат, 1974. 367 с.
4. Рвачев В. Л. Теория  $R$ -функций и некоторые ее приложения. Киев: Наукова думка, 1982. 552 с.
5. Максименко-Шейко К. В.  $R$ -функции в математическом моделировании геометрических объектов и физических полей. Харьков: Изд. ИПМаш НАН Украины, 2009. 306 с.
6. Максименко-Шейко К. В., Мацевитый А. М., Толк А. В., Шейко Т. И.  $R$ -функции и обратная задача аналитической геометрии в трехмерном пространстве // Информационные технологии. 2007. № 10. С. 23–32.
7. Лойцянский Л. Г. Механика жидкости и газа. М.: Наука, 1970. 904 с.
8. Рвачев В. Л., Шевченко А. Н. Проблемно-ориентированные языки и системы для инженерных расчетов. Киев: Техніка, 1988. 197 с.
9. Chen B. C., Todreas N. E. Prediction of Coolant Temperature Field in The Breeder Reaction Including Interassembly Heat Transfer // Nuclear Engineering and Design. 1975. Vol. 35, N 3. P. 423–440.
10. Dwyer O. E., Berry H. C., Hlavac P. J. Heat transfer to liquid metals flowing turbulently and longitudinally through closely spaced rod bundles. Pt. 1–2 // Nucl. Sci. Engng. 1972. Vol. 23. P. 273–308.

## ИНФОРМАЦИЯ



## VIII Международная научно-практическая конференция Объектные системы — 2014

10–12 мая 2014, г. Ростов-на-Дону

Конференция посвящена принципам проектирования, реализации и сопровождения объектных систем и включает в себя обсуждение широкого круга проблем.

### Секции научной конференции

- Графические нотации, используемые при объектном проектировании ИС
- Принципы объектного проектирования информационных систем
- Инструменты объектного моделирования
- Теория объектно-ориентированного программирования
- Методы (шаблоны) объектно-реляционного отображения
- Реализация и использование объектных расширений в реляционных СУБД
- Проектирование, разработка и реализация распределенных систем
- Типовые реализации КИС с применением объектных технологий
- Принципы организации и реализации объектных баз данных
- Проблемы реализации объектных СУБД
- Имитационное моделирование объектных систем
- Темпоральные объектные системы
- Проблемы изучения (преподавания) объектных технологий в вузе

Сайт конференции <http://objectsystems.ru>

УДК 004.042

**А. А. Востриков,**

канд. техн. наук, доц., зам. директора,  
НИИ информационно-управляющих систем  
Санкт-Петербургского национального  
исследовательского университета информационных  
технологий, механики и оптики,  
e-mail: [vostrikov@mail.ru](mailto:vostrikov@mail.ru)

## О матрицах Адамара—Мерсенна и маскировании изображений

*Рассматривается вопрос существенного расширения базиса ортогональных матриц, предлагается процедура маскирования изображений с использованием  $M$ -матриц как ортогонального базиса, построенного на числах последовательности Мерсенна.*

**Ключевые слова:** ортогональные матрицы,  $M$ -матрицы, матрицы Адамара, матрицы Адамара—Мерсенна, числа Мерсенна

### Введение

В современном мире защита информации от несанкционированного доступа и подмены имеет огромное значение, особенно защита видеоинформации в сетях общего пользования [1—3]. Разработано множество достаточно безупречных систем защиты, которые успешно применяют на практике. Но большинство традиционных систем нельзя напрямую использовать для защиты цифрового видео в системах реального времени, поскольку они базируются на алгоритмах шифрации и требуют значительных вычислительных затрат.

В работе [1] рассмотрены матричные методы кодирования видеоинформации на основе базисов ортогональных преобразований, дальнейшее развитие и исследование которых протекает достаточно интенсивно на множестве  $M$ -матриц [4—8]. В работе [6] введено понятие матриц Мерсенна, дополняющее множество матриц Адамара на классе ортогональных матриц нечетных порядков, равных числам последовательности Мерсенна. Важно то, что согласно гипотезе, сформулированной Н. А. Балониным [9], множества матриц Адамара и Мерсенна равновеликие.

Число ортогональных матриц Мерсенна не уступает числу целочисленных матриц Адамара, но

значения коэффициентов двухуровневых матриц Мерсенна — вещественные числа, что позволяет лучше защитить видеоинформацию.

Современное состояние процессоров цифровой обработки сигналов, характеризующееся увеличением производительности и структурной ориентацией на выполнение операции свертки в формате вещественных чисел, позволяет использовать достаточно сложные базисы, включая и многоуровневые  $M$ -матрицы [4, 5].

### Матрицы Адамара—Мерсенна

В работе [8] дано определение уровней матриц — значений, которым равны элементы матрицы. Введение определения уровней дало возможность интерпретировать математические объекты окрашенными плоскими и объемными графическими портретами матриц, позволяющими выявлять неизвестные ранее закономерности и взаимосвязи ортогональных матриц.

Последовательность чисел Мерсенна, задаваемая формулой  $n = 2^k - 1$ , начинается с чисел 1, 3, 5, 15, 31, ... и принадлежит подмножеству чисел вида  $4k - 1$ . Двухуровневая матрица Адамара—Мерсенна третьего порядка имеет вид

$$\mathbf{M}_3 = \begin{pmatrix} a & -b & a \\ -b & a & a \\ a & a & -b \end{pmatrix}.$$

Для итерационного получения ортогональных двухуровневых матриц Адамара—Мерсенна последующих порядков из предыдущих на основе формулы Сильвестра строят сначала четырехуровневые матрицы Адамара—Эйлера

$$\mathbf{E}_n = \begin{pmatrix} \mathbf{M}_{n/2} & \mathbf{M}_{n/2} \\ \mathbf{M}_{n/2} & -\mathbf{M}_{n/2} \end{pmatrix},$$

где  $\mathbf{M}_{n/2}$  — двухуровневая матрица Адамара—Мерсенна вдвое меньшего нечетного порядка, имеющая два уровня  $\{a = 1, -b\}$  [8]. При этом преобразовании число уровней ввиду инверсии двухуровневой матрицы Адамара—Мерсенна удваивается, что демонстрируется объемными портретами матриц на рис. 1 (см. третью сторону обложки).

На этом, однако, получение ортогональных матриц не прекращается. Матрица Адамара—Эй-

лера пересчитывается в матрицу Адамара—Мерсенна дополнением ее строкой и столбцом (кайма) в виде

$$\mathbf{M}_{n+1} = \begin{pmatrix} -\lambda & \mathbf{e}^T \\ \mathbf{e} & \mathbf{E}_{2n}^* \end{pmatrix},$$

где  $\lambda = -a$  — собственное число, а  $\mathbf{e}$  — собственный

вектор "сопряженной" матрицы  $\mathbf{E}_{2n}^* = \begin{pmatrix} \mathbf{M}_{n/2} & \mathbf{M}_{n/2} \\ \mathbf{M}_{n/2} & \mathbf{M}_{n/2}^* \end{pmatrix}$ ;

$\mathbf{M}_{n/2}^*$  получается из матрицы Мерсенна соответствующего порядка взаимной заменой элементов  $a = 1$  и  $-b$ , причем первую половину отличных от  $a$  коэффициентов собственного вектора составляют элементы  $-b$ ,  $b = 1/2$  при  $n = 3$ , при  $n > 3$  уровень

$b = \frac{q - \sqrt{4q}}{q - 4}$ ,  $q = n + 1$  (порядок матрицы Адамара).

Из приведенных выше формул видно, что такая последовательность действий позволяет снова вернуться к двухуровневому варианту матрицы Адамара—Мерсенна, представленной на рис. 2 (см. третью сторону обложки), порядок которой соответствует следующему числу в последовательности Мерсенна.

### О гипотезе Н. А. Балонины

В отличие от матриц Адамара, матрицы Адамара—Мерсенна — матрицы нечетного порядка. В некотором смысле они более первичны, поскольку начинаются с единицы. У этой аналогии есть и своя более глубокая интерпретация. Напомним, что матрица Адамара — квадратная двухуровневая матрица  $\mathbf{H}_n$  порядка  $n$ , состоящая из чисел  $\{+1, -1\}$ , столбцы которой ортогональны

$$\mathbf{H}_n^T \mathbf{H}_n = n\mathbf{I},$$

где  $\mathbf{I}$  — единичная матрица.

В отличие матриц Адамара—Эйлера из элементарной матрицы Адамара—Мерсенна, равной единице, по правилу Сильвестра получается только матрица Адамара второго порядка, далее итерации продолжают с нею.

Еще Адамар заметил, что порядки таких матриц слишком разнесены, причем расстояние между порядками соседних матриц интенсивно увеличивается. Он привел примеры двухуровневых ортогональных матриц с единичными значениями модулей элементов промежуточных порядков  $n = 12$  и  $n = 20$ , что, собственно, и дало этим матрицам название. Согласно гипотезе Адамара существуют матрицы Адамара порядков  $4k$ , где  $k$  — целое.

Гипотеза Н. А. Балонины [9] состоит в утверждении о существовании матриц Мерсенна порядков  $4k - 1$ .

Матрицы порядков  $n = 11$  [8] и  $n = 19$  с признаками матриц Адамара—Мерсенна найдены в обоснование равновеликости множеств матриц Адамара и Адамара—Мерсенна, называемых для большей простоты просто матрицами Мерсенна.

При видимой симметрии обеих гипотез существуют различия. В отличие от матриц Мерсенна, матрицы Адамара строят на основе квадратных трехуровневых матриц Белевича  $\mathbf{C}_n$  порядка  $n$ , состоящих из чисел  $\{1, 0, -1\}$ , столбцы которых ортогональны:

$$\mathbf{C}_n^T \mathbf{C}_n = (n - 1)\mathbf{I},$$

а нулевые элементы сосредоточены на диагонали.

Существование матриц Белевича определяется критерием Эйлера (и его интерпретациями, на случай непростых чисел) о разложимости множителя  $n - 1$  на сумму двух квадратов чисел.

Когда матрицы Белевича не существуют, а это порядки 22, 34, 58 и т. п., существуют матрицы Адамара—Эйлера или просто матрицы Эйлера, если следовать логике упрощения их наименования. Матрица Эйлера порядка  $n = 22$  получается, например, на основе упомянутой выше матрицы Мерсенна порядка  $n = 11$ , которую и предложил Балонин. Тем самым, заполнены пропуски в ряду ортогональных трехуровневых матриц Белевича, что имеет существенное теоретическое и практическое значение.

Очевидно, матрицы Адамара—Мерсенна порядков  $4k - 1$  значительно расширяют базис ортогональных матриц для защиты информации в упомянутых прикладных задачах.

### Маскирование изображений

Типичный тракт обработки изображения в процедурах матричного преобразования изображения предусматривает несколько этапов:

- дискретное преобразование Фурье, обеспечивающее получение спектрального разложения;
- применение фильтра низких частот, устраняющего высокочастотную часть спектра;
- статистическая обработка по Хаффману, устраняющая избыточность.

В процедуре маскирования все представленные этапы предлагается сохранить, заменив матрицу дискретного преобразования Фурье оригинальной матрицей ортогонального базиса Адамара—Мерсенна (сильвестровой последовательности матриц).

Кроме дискретности уровней элементов матрицы, как у матрицы Адамара, не менее важную роль играет оригинальность базиса, обеспечивающая скрытность получаемых преобразованных данных, что делает матрицы с дробно-иррациональными и

иррациональными значениями уровней удобными при выполнении процедуры маскирования изображений на цифровых устройствах.

В отличие от традиционных базисов в задаче маскирования большое значение имеют качества, обретаемые из экстремальных свойств базисных наборов. Матрицы Адамара и близкие к ним матрицы Мерсенна оптимальны в смысле нейтрализации последствий воздействия точечных помех при передаче по каналам связи.

Не менее важны при этом и рекурсивные процедуры увеличения порядка матрицы. Дополнительный аргумент рациональности использования базисов, построенных на последовательностях чисел [4–8], состоит в том, что алгоритм построения их фрактален и матрицы при определенном устройстве алгоритма обладают повышенной чувствительностью к изменению разрядной сетки процессора и начальных данных.

Рассмотренная схема вычислений позволяет, во-первых, сохранить принципиальную возможность сжатия маскируемой информации, например, адаптацией процедуры фильтрации к структурным особенностям базиса. Во-вторых, необычная матрица и ключ маскирования в виде вектора перестановки строк и столбцов, неизвестные третьей стороне, способствуют надежному сокрытию видеоизображения от перехвата и подмены, что показало исследование программной реализации маскиратора [10] и демаскиратора изображений [11].

### Заключение

В процессе поиска для алгоритмов маскирования изображений оригинальных ортогональных матриц нечетных порядков, близких к матрицам Адамара по свойствам, выделен предпочтительный класс двухуровневых матриц, называемых матрицами Мерсенна. Порядки этих матриц равны числам Мерсенна вида  $2^k - 1$ , их состав значительно расширяется гипотезой Балонина Н. А. Элементы матриц Мерсенна с ростом значений целочисленного аргумента  $k$  стремятся к значениям  $\{1, -1\}$ , как у матриц Адамара.

Инструментом для поиска новых матриц являлся программный комплекс [12], который может быть

использован при усложнении задачи маскирования видеоизображений, заключающемся в том, что матрица ортогонального преобразования не вычисляется заранее, а является результатом работы алгоритма. По открытому каналу в качестве ключа передаются только настройки для ее вычисления.

Практическое применение рассмотренных в работе матриц целесообразно в задачах повышения степени помехоустойчивости и защищенности при передаче информации, поскольку матричные методы предполагают эффективную реализацию в современных микропроцессорных структурах, ориентированных на цифровую обработку информации.

### Список литературы

1. **Мироновский Л. А., Слаев В. А.** Стрип-метод преобразования изображений и сигналов. СПб.: Политехника, 2006. 163 с.
2. **Ерош И. Л., Сергеев А. М., Филатов Г. П.** О защите цифровых изображений при передаче по каналам связи // Информационно-управляющие системы. 2007. № 5(30). С. 20–22.
3. **Литвинов М. Ю., Беззатеев С. В., Трояновский Б. К., Филатов Г. П.** Выбор алгоритма преобразования, обеспечивающего изменение структуры изображения // Информационно-управляющие системы. 2006. № 6(25). С. 2–6.
4. **Балонин Н. А., Сергеев М. Б.** М-матрицы // Информационно-управляющие системы. 2011. № 1. С. 14–21.
5. **Балонин Ю. Н., Сергеев М. Б.** М-матрица 22-го порядка // Информационно-управляющие системы. 2011. № 5. С. 87–90.
6. **Балонин Н. А., Сергеев М. Б., Мироновский Л. А.** Вычисление матриц Адамара—Мерсенна // Информационно-управляющие системы. 2012. № 5. С. 92–94.
7. **Балонин Н. А., Сергеев М. Б., Мироновский Л. А.** Вычисление матриц Адамара—Ферма // Информационно-управляющие системы. 2012. № 6. С. 90–93.
8. **Балонин Н. А., Сергеев М. Б.** О двух способах построения матриц Адамара—Эйлера // Информационно-управляющие системы. 2013. № 1. С. 7–10.
9. **Ресурс** Интернет <http://mathscinet.ru/ebook/merenne/> Последнее обращение 28.01.2013.
10. **Балонин Н. А.** О существовании матриц Мерсенна 11-го и 19-го порядков // Информационно-управляющие системы. 2013. № 2. С. 89–90.
11. **Свидетельство** о государственной регистрации программы для ЭВМ № 20126188124 от 27 сентября 2012 г. "Программа "Маскиратор изображений" / Software "Images masking" / Сергеев М. Б., Балонин Н. А., Балонин Ю. Н., Востриков А. А.
12. **Свидетельство** о государственной регистрации программы для ЭВМ № 2012618700 от 24 сентября 2012 г. "Программа "Демаскиратор изображений" / Software "Images recovering" / Сергеев М. Б., Балонин Н. А., Балонин Ю. Н., Востриков А. А.
13. **Свидетельство** о государственной регистрации программы для ЭВМ № 2012614356 от 16 мая 2012 г. "Программа поиска М-матриц" / Сергеев М. Б., Балонин Н. А., Балонин Ю. Н.

**Е. А. Беляева**, аспирант,  
Национальный исследовательский  
ядерный университет МИФИ,  
e-mail: belyaeva.eli@gmail.com

## **Сравнительная оценка аппаратно-программных модулей доверенной загрузки на основе классификации их функциональных возможностей**

*Предлагается научно обоснованный подход к получению сравнительной оценки АПМДЗ, основывающийся на алгоритме получения комплексного показателя для оценки эффективности защиты от несанкционированного доступа (НСД), обеспечиваемой каждым модулем, методике сравнительного оценивания АПМДЗ по критерию "эффективность—стоимость", а также классификации функциональных возможностей модулей.*

**Ключевые слова:** функциональная возможность, классификация, несанкционированный доступ, методика, алгоритм

### **Введение**

Обеспечение национальных интересов государства в информационной сфере, как определено в Доктрине информационной безопасности Российской Федерации, включает в себя защиту информационных ресурсов от НСД, обеспечение безопасности информационных и телекоммуникационных систем как уже развернутых, так и создаваемых на территории России [1].

Развитие методов и средств защиты информационных ресурсов от НСД происходило параллельно с развитием технологий обработки информации, что обуславливалось их тесной взаимосвязью. Были сформированы основные подходы к созданию и разработке методов и средств защиты информации от НСД, которые предполагают использование специально создаваемых аппаратных, программных и программно-аппаратных средств защиты информации [2, 3].

К настоящему времени разработано большое количество различных аппаратных, программных и программно-аппаратных средств защиты информации от НСД, с помощью которых может быть обеспечена защищенность информации в информационных и телекоммуникационных системах. Множество и разнообразие классов средств защиты определяется прежде всего способами воздействия на дестабилизирующие факторы или порождающие их причины, причем воздействия в на-

правлении, способствующем повышению значений показателей защищенности или, по крайней мере, сохранению прежних, ранее достигнутых их значений [4—6].

Рассматриваемый в данной статье наиболее широко применяемый класс аппаратно-программных средств защиты информации — аппаратно-программных модулей доверенной загрузки (АПМДЗ) сочетает в себе возможности и достоинства как аппаратных, так и программных средств защиты.

Основным назначением АПМДЗ является обеспечение защиты от НСД к информации ограниченного доступа путем реализации надежного контроля доступа пользователей к техническим средствам автоматизированных рабочих мест и выполнения доверенной загрузки операционной системы (ОС).

Каждый АПМДЗ имеет много функциональных возможностей в части обеспечения защиты от НСД к информации. Некоторые функциональные возможности являются обязательными для реализации в АПМДЗ согласно требованиям руководящих документов ФСБ России. Такие функциональные возможности являются основными.

Благодаря научно-техническому прогрессу в области микроэлектроники, схемотехники в одном АПМДЗ можно сосредоточить функциональные возможности, состав которых в несколько раз перекрывает требования ФСБ России. Данные функциональные возможности называются дополнительными, и их наличие не регламентировано руководящими документами. Дополнительные функциональные возможности АПМДЗ позволяют повысить эффективность защиты обрабатываемой в автоматизированных системах информации от НСД, а также оценить возможность применения АПМДЗ в системе с заданными параметрами, не регламентируемыми требованиями руководящих документов.

Проведенный автором анализ результатов исследований специалистов в области аппаратно-программных средств защиты информации позволяет констатировать, что проблемы оценивания АПМДЗ по основным и дополнительным функциональным возможностям являются актуальными для теории и практики информационной безопасности и требуют дальнейшего развития. Значительное число публикаций посвящено общим вопросам аппаратно-программной защиты информации от НСД.

Различные методические аспекты аппаратно-программной защиты информации отражены в большом числе работ известных специалистов в этой области (В. А. Петров, А. П. Дураковский, В. И. Ярочкин, Ю. Н. Лаврухин, Ю. Г. Бугров и др.). В этих трудах разработан ряд основных положений организации защиты от НСД, применения аппаратно-программных средств защиты информации от НСД, рассмотрен понятийный аппарат, сущность и осо-

бенности аппаратно-программных средств защиты информации.

Вместе с тем, автору хотелось бы отметить отсутствие серьезных, по его мнению, работ по отдельным вопросам оценивания аппаратно-программных средств защиты информации, которое не позволяет решить одну из важнейших проблем создания защиты информации от НСД — проблему научного обоснованного выбора аппаратно-программного средства защиты информации для создания автоматизированной системы в защищенном исполнении.

Это обусловило глубокую проработку вопросов выявления и классификации функциональных возможностей АПМДЗ, формирования на их основе классификационных признаков для оценивания АПМДЗ, разработки алгоритма получения показателя оценки эффективности обеспечиваемой защиты от НСД и методики выбора АПМДЗ с наилучшими характеристиками обеспечения защиты от НСД.

### Постановка задачи

С учетом изложенного выше, проблема обеспечения научно обоснованного выбора АПМДЗ для решения задач по защите от НСД может быть решена путем разработки методики сравнительной оценки функциональных возможностей АПМДЗ, которая в содержательном виде может быть сформулирована следующим образом.

Применительно к заданным условиям исследования АПМДЗ и классификации функциональных возможностей АПМДЗ разработать научно обоснованную методику получения комплексной оценки функциональных возможностей АПМДЗ, позволяющую выявить АПМДЗ с наилучшими характеристиками обеспечения защиты информации от НСД.

В целях формализации задачи и способов ее решения в соответствии со сформулированной содержательной постановкой, обозначим:

- $A = \{A_i\}$ ,  $i = 1, 2, \dots, |A|$  — множество АПМДЗ;
- $S = \{S_t\}$ ,  $t = 1, 2, \dots, |S|$  — множество стоимостей АПМДЗ.

Каждый АПМДЗ  $A_i \in A$  характеризуется:

- $V_{main} = \{V_{main}^j\}$ ,  $j = 1, 2, \dots, |V_{main}|$  — множеством основных функциональных возможностей;
- $V_{add} = \{V_{add}^l\}$ ,  $l = 1, 2, \dots, |V_{add}|$  — множеством дополнительных функциональных возможностей;
- $E(A_i) = f(V_{main}, V_{add}, S)$  — критерием эффективности АПМДЗ  $A_i$  в части обеспечения защиты обрабатываемой информации от НСД.

**Найти** такие  $A' \in A$ , при которых  $E(A') = f(V_{main}(A'), V_{add}(A'), S(A')) = \max(E(A_i))$ .

Сформулированную задачу целесообразно разделить на следующие подзадачи:

- анализ основных и дополнительных функциональных возможностей АПМДЗ в части защиты от НСД;
- выделение классификационных признаков для оценки функциональных возможностей АПМДЗ;
- разработка алгоритма получения значений показателей для оценки функциональных возможностей АПМДЗ;
- разработка методики сравнительной оценки функциональных возможностей АПМДЗ.

Далее предлагаются способы решения поставленной задачи.

### Классификация функциональных возможностей АПМДЗ

В современных АПМДЗ реализуется большое число защитных механизмов, к которым относятся [7]:

- идентификация и аутентификация пользователей;
- разграничение доступа к файлам, директориям, дискам;
- контроль целостности программных средств и данных;
- обеспечение доверенной загрузки ОС;
- криптографическое преобразование информации;
- регистрация событий, критичных для защиты от НСД.

Каждый из используемых в АПМДЗ защитных механизмов реализуется с помощью основных и дополнительных функциональных возможностей.

**Функциональные возможности АПМДЗ.** Одной из научных задач в области оценивания эффективности реализации защитных механизмов АПМДЗ является классификация основных и дополнительных функциональных возможностей АПМДЗ.

АПМДЗ имеют большое количество функциональных возможностей в части защиты от НСД, которые предлагается классифицировать следующим образом:

- основные функциональные возможности, обязательные для реализации в АПМДЗ согласно требованиям руководящих документов по защите информации Российской Федерации;
- дополнительные функциональные возможности АПМДЗ, позволяющие повысить защищенность обрабатываемой информации от НСД, и реализация которых не регламентирована руководящими документами.

В ходе изучения АПМДЗ был выявлен ряд основных функциональных возможностей в части защиты от НСД, присущих большинству из существующих АПМДЗ, а именно:

- контроль целостности внутренней памяти АПМДЗ;

- контроль целостности оперативной памяти ЭВМ;
- контроль целостности файлов и реестра ОС;
- разграничение реакции АПМДЗ на события нарушения контроля целостности;
- использование для контроля целостности сертифицированного ФСБ России алгоритма;
- очистка памяти, используемой АПМДЗ;
- проверка работоспособности основных рабочих узлов АПМДЗ;
- наличие функций периодического самотестирования АПМДЗ;
- использование пароля, как дополнительного фактора аутентификации;
- регистрация попыток НСД к ресурсам системы;
- регистрация изменения настроек АПМДЗ;
- регистрация фактов смены пароля;
- регистрация событий контроля целостности;
- и другие.

В ходе изучения АПМДЗ также выявлен ряд дополнительных функциональных возможностей в части защиты от НСД, а именно:

- диагностика неисправностей АПМДЗ;
- удаленное администрирование АПМДЗ в режиме реального времени и в отложенном режиме;
- удаленное управление АПМДЗ с любого рабочего места;
- проверки целостности всей информации, записываемой в АПМДЗ при удаленном доступе;
- автоматическое формирование списка программных модулей для контроля;
- контроль целостности журналов файловых систем и проверка типов поддерживаемых файловых систем;
- проверка типов поддерживаемых ОС;
- поддержка (U)EFI BIOS;
- электростатическая защита интерфейса считывателя;
- защита от внешних высокоэнергетических электрических воздействий;
- возможность подключения датчиков;
- и другие.

**Классификационные признаки для оценивания АПМДЗ.** Для классификации функциональных возможностей АПМДЗ предлагаются следующие классификационные признаки, характеризующие влияние функциональных возможностей на эффективность защиты информации от НСД:

- *Учет и регистрация критических событий.* Определяет состав основных функциональных возможностей АПМДЗ по регистрации критических событий в части защиты обрабатываемой информации от НСД. Регистрация критических событий позволяет определять объект, с которым произошло критическое для защиты от НСД событие, пользователя и другие данные.
- *Обеспечение целостности ПО АПМДЗ, объектов ОС и объектов пользователей.* Определяет состав

основных функциональных возможностей АПМДЗ в части обеспечения целостности программного обеспечения АПМДЗ, целостности объектов, используемых АПМДЗ для загрузки ОС, а также объектов, используемых пользователями при работе с обрабатываемой информацией.

- *Защищенность данных АПМДЗ.* Определяет состав программных и аппаратных функциональных возможностей для реализации защиты информации: возможность очистки памяти ЭВМ, используемой АПМДЗ, наличие функций периодического самотестирования АПМДЗ для проверки работоспособности основных узлов и другие.
- *Управление доступом пользователей.* Определяет состав основных функциональных возможностей АПМДЗ в части реализации аутентификации/идентификации пользователя в системе.
- *Дополнительные аппаратные функции АПМДЗ.* Определяет состав дополнительных функциональных возможностей по проверке работоспособности АПМДЗ в целях выявления сбоев/отказов в работе АПМДЗ.
- *Реализация удаленного управления.* Определяет состав дополнительных функциональных возможностей в части реализации удаленного управления АПМДЗ: возможность удаленного администрирования АПМДЗ; возможность удаленного управления с любого рабочего места и другие.
- *Дополнительные функции обеспечения целостности.* Определяет состав дополнительных функциональных возможностей в части реализации контроля целостности — автоматическое формирование списка программных модулей для контроля; возможность задания режима контроля журналов файловых систем и другие.
- *Функции взаимодействия с ОС.* Определяет состав дополнительных функциональных возможностей в части работы с ОС: поддерживаемые ОС; возможность принудительной загрузки ОС с определяемого администратором носителя информации и другие.

Полученные автором в результате анализа классификационные признаки функциональных возможностей АПМДЗ позволяют решить следующие задачи:

- провести сравнительный анализ АПМДЗ с основными и дополнительными функциональными возможностями;
- обеспечить получение исходных данных для оценивания АПМДЗ по степени эффективности защиты от НСД.

#### **Получение значений показателей для оценки функциональных возможностей АПМДЗ**

В рамках каждого классификационного признака для оценки функциональных возможностей (как основных, так и дополнительных) предлагается проводить ранжирование оценок функциональных

возможностей на основе метода Черчмена—Акоффа (последовательное сравнение). Выбор данного метода обуславливается следующим [8]:

- поскольку характер анализируемой информации таков, что целесообразно получить численные оценки функциональных возможностей АПМДЗ, необходимо использовать метод численной оценки;
- данный метод позволяет последовательно корректировать оценки, указанные экспертами, в целях выявления наиболее предпочтительной функциональной возможности.

**Алгоритм получения значений показателей для оценки функциональных возможностей АПМДЗ.**

Рассмотрим разработанный алгоритм получения значения показателя для оценки функциональных возможностей АПМДЗ. Данный алгоритм, в зависимости от типа оцениваемых функциональных возможностей (основные или дополнительные), включает в себя следующие этапы.

*Этап 1.* Выбор АПМДЗ для оценивания.

Выбирается модуль, оценку которого требуется получить.

*Этап 2.* Выбор типа функциональных возможностей для оценивания.

Определяются наборы функциональных возможностей АПМДЗ для оценивания: оценивание по основным функциональным возможностям; оценивание по дополнительным функциональным возможностям; комплексное оценивание по всем функциональным возможностям АПМДЗ.

*Этап 3.* Оценивание АПМДЗ в соответствии с выбранными типами функциональных возможностей.

На данном этапе происходит непосредственное оценивание каждым из экспертов выбранного АПМДЗ, состоящее из следующих действий:

- 1) постановка оценки для каждой из функциональных возможностей в соответствии с заданными классификационными признаками;
- 2) постановка рангов функциональных возможностей в рамках классификационного признака;
- 3) корректировка выставленных ранее рангов функциональных возможностей в соответствии с выбранным методом корректировки;
- 4) вычисление показателя оценки АПМДЗ по каждому из классификационных признаков;
- 5) постановка ранга предпочтительности оценки АПМДЗ по каждому из классификационных признаков;
- 6) корректировка ранга предпочтительности оценки АПМДЗ по каждому из классификационных признаков;
- 7) вычисление показателя оценки АПМДЗ по типу функциональных возможностей в соответствии с оценками, полученными по классификационным признакам.

Рассмотрим часть алгоритма получения значения показателя для оценки функциональных возможностей АПМДЗ, касающуюся оценивания набора основных функциональных возможностей  $V_{main}(A)$ , выбранного для оценивания АПМДЗ  $A$ :

*Шаг 1.* Для каждого классификационного признака  $K_{main}^i$  оценки основных функциональных возможностей АПМДЗ  $A$  выполняются ( $i$  от 1 до  $|K_{main}|$ ):

- а) постановка показателя оценки  $E(V_{K_{main}^i}^l)$  для каждой основной функциональной возможности АПМДЗ  $A$ , удовлетворяющей классификационному признаку  $K_{main}^i$  ( $l$  от 1 до  $|V_{K_{main}^i}|$ );
- б) постановка ранга  $\varphi(V_{K_{main}^i}^l)$  для каждой основной функциональной возможности АПМДЗ  $A$ , удовлетворяющей классификационному признаку  $K_{main}^i$  ( $l$  от 1 до  $|V_{K_{main}^i}|$ );
- в) корректировка ранга  $\varphi(V_{K_{main}^i}^l)$  для каждой основной функциональной возможности АПМДЗ  $A$ , удовлетворяющей классификационному признаку  $K_{main}^i$  ( $l$  от 1 до  $|V_{K_{main}^i}|$ );
- г) вычисление показателя  $E(K_{main}^i)$  оценки основных функциональных возможностей АПМДЗ  $A$  по классификационному признаку  $K_{main}^i$  на основе полученных ранее данных:

$$E(K_{main}^i) = \sum_{l=1}^{|V_{K_{main}^i}|} E(V_{K_{main}^i}^l) \varphi(V_{K_{main}^i}^l);$$

- д) постановка ранга  $\psi(K_{main}^i)$  для оценки по каждому из классификационных признаков;
- е) корректировка ранга  $\psi(K_{main}^i)$  для оценки по каждому из классификационных признаков.

*Шаг 2.* Вычисление показателя  $E(K_{main})$  оценки основных функциональных возможностей АПМДЗ  $A$  по всем классификационным признакам  $K_{main}^i$  на основе полученных ранее данных:

$$E(K_{main}) = \sum_{l=1}^{|K_{main}|} E(K_{main}^i) \psi(K_{main}^i).$$

Получим

$$\begin{aligned}
 E(K_{\text{main}}) &= \sum_{i=1}^{|K_{\text{main}}|} E(K_{\text{main}}^i) \psi(K_{\text{main}}^i) = \\
 &= \sum_{i=1}^{|K_{\text{main}}|} \psi(K_{\text{main}}^i) \left( \sum_{l=1}^{|V_{K_{\text{main}}^i}|} E(V_{K_{\text{main}}^i}^l) \varphi(V_{K_{\text{main}}^i}^l) \right) = \\
 &= \sum_{i=1}^{|K_{\text{main}}|} \sum_{l=1}^{|V_{K_{\text{main}}^i}|} E(V_{K_{\text{main}}^i}^l) \varphi(V_{K_{\text{main}}^i}^l) \psi(K_{\text{main}}^i).
 \end{aligned}$$

Рассмотрим часть алгоритма получения значения показателя для оценки функциональных возможностей АПМДЗ, касающуюся оценивания набора дополнительных функциональных возможностей  $V_{\text{add}}(A)$ , выбранного для оценивания АПМДЗ  $A$ :

*Шаг 1.* Для каждого классификационного признака  $K_{\text{add}}^j$  оценки дополнительных функциональных возможностей выполняется ( $j$  от 1 до  $|K_{\text{add}}|$ ):

- постановка показателя оценки  $E(V_{K_{\text{add}}^j}^t)$  для каждой дополнительной функциональной возможности АПМДЗ  $A$ , удовлетворяющей классификационному признаку  $K_{\text{add}}^j$  ( $t$  от 1 до  $|V_{K_{\text{add}}^j}|$ );
- постановка ранга  $\varphi(V_{K_{\text{add}}^j}^t)$  для каждой дополнительной функциональной возможности АПМДЗ  $A$ , удовлетворяющей классификационному признаку  $K_{\text{add}}^j$  ( $t$  от 1 до  $|V_{K_{\text{add}}^j}|$ );
- корректировка ранга  $\varphi(V_{K_{\text{add}}^j}^t)$  для каждой дополнительной функциональной возможности АПМДЗ  $A$ , удовлетворяющей классификационному признаку  $K_{\text{add}}^j$  ( $t$  от 1 до  $|V_{K_{\text{add}}^j}|$ );
- вычисление показателя  $E(K_{\text{add}}^j)$  оценки дополнительных функциональных возможностей АПМДЗ  $A$  по классификационному признаку  $K_{\text{add}}^j$  на основе полученных ранее данных:

$$E(K_{\text{add}}^j) = \sum_{t=1}^{|V_{K_{\text{add}}^j}|} E(V_{K_{\text{add}}^j}^t) \varphi(V_{K_{\text{add}}^j}^t);$$

- постановка ранга предпочтительности  $\psi(K_{\text{add}}^j)$  для оценки по каждому из классификационных признаков;

- корректировка ранга предпочтительности  $\psi(K_{\text{add}}^j)$  для оценки по каждому из классификационных признаков.

*Шаг 2.* Вычисление показателя  $E(K_{\text{add}})$  оценки дополнительных функциональных возможностей АПМДЗ  $A$  по всем классификационным признакам  $K_{\text{add}}^j$  на основе полученных ранее данных:

$$E(K_{\text{add}}) = \sum_{j=1}^{|K_{\text{add}}|} E(K_{\text{add}}^j) \psi(K_{\text{add}}^j).$$

При подстановке формул получим

$$\begin{aligned}
 E(K_{\text{add}}) &= \sum_{j=1}^{|K_{\text{add}}|} E(K_{\text{add}}^j) \psi(K_{\text{add}}^j) = \\
 &= \sum_{j=1}^{|K_{\text{add}}|} \psi(K_{\text{add}}^j) \left( \sum_{t=1}^{|V_{K_{\text{add}}^j}|} E(V_{K_{\text{add}}^j}^t) \varphi(V_{K_{\text{add}}^j}^t) \right) = \\
 &= \sum_{j=1}^{|K_{\text{add}}|} \sum_{t=1}^{|V_{K_{\text{add}}^j}|} E(V_{K_{\text{add}}^j}^t) \varphi(V_{K_{\text{add}}^j}^t) \psi(K_{\text{add}}^j).
 \end{aligned}$$

Комплексное оценивание функциональных возможностей выбранного АПМДЗ  $A$  состоит из последовательного получения показателей оценок для основных и дополнительных функциональных возможностей и вычисления единого показателя комплексной оценки АПМДЗ. Поэтому формула для вычисления показателя  $E(K)$  комплексной оценки функциональных возможностей для АПМДЗ  $A$  выглядит следующим образом:

$$\begin{aligned}
 E(K) &= E(K_{\text{main}}) + E(K_{\text{add}}) = \\
 &= \sum_{l=1}^{|V_{K_{\text{main}}}|} E(V_{K_{\text{main}}}^l) \varphi(V_{K_{\text{main}}}^l) \psi(K_{\text{main}}^l) + \\
 &+ \sum_{j=1}^{|K_{\text{add}}|} \sum_{t=1}^{|V_{K_{\text{add}}^j}|} E(V_{K_{\text{add}}^j}^t) \varphi(V_{K_{\text{add}}^j}^t) \psi(K_{\text{add}}^j).
 \end{aligned}$$

**Получение комплексной оценки функциональных возможностей АПМДЗ.** В соответствии с алгоритмом получения значений показателей для оценки функциональных возможностей АПМДЗ, представленным выше, и с учетом полученных формул показатель  $E_{\text{effective}}^m(A)$  комплексной оценки функциональных возможностей АПМДЗ  $A$ , полученный

на основе оценок  $m$ -го эксперта, вычисляется по следующей формуле:

$$E_{effective}^m(A) = \sum_{i=1}^{|K_{main}|} \psi(K_{main}^i) E(K_{main}^i) + \sum_{j=1}^{|K_{add}|} \psi(K_{add}^j) E(K_{add}^j) = \sum_{i=1}^{|K_{main}|} \sum_{l=1}^{|V_{K_{main}^i}|} \varphi(V_{K_{main}^i}^l) E(V_{K_{main}^i}^l) \psi(K_{main}^i) + \sum_{j=1}^{|K_{add}|} \sum_{n=1}^{|V_{K_{add}^j}|} \varphi(V_{K_{add}^j}^n) E(V_{K_{add}^j}^n) \psi(K_{add}^j),$$

где  $E_{effective}(A)$  — показатель комплексной оценки эффективности реализации функциональных возможностей АПМДЗ  $A$  в части обеспечения защиты

$$E_{effective}(A) = \frac{\sum_{m=1}^{|M|} E_{effective}^m R_m}{\sum_{m=1}^{|M|} R_m} = \frac{\sum_{m=1}^{|M|} R_m \left( \sum_{i=1}^{|K_{main}|} \sum_{l=1}^{|V_{K_{main}^i}|} \varphi(V_{K_{main}^i}^l) E(V_{K_{main}^i}^l) \psi(K_{main}^i) + \sum_{j=1}^{|K_{add}|} \sum_{n=1}^{|V_{K_{add}^j}|} \varphi(V_{K_{add}^j}^n) E(V_{K_{add}^j}^n) \psi(K_{add}^j) \right)}{\sum_{m=1}^{|M|} R_m},$$

где  $R$  — ранг предпочтительности оценок, полученных  $m$ -м экспертом;  $M$  — множество экспертов, входящих в экспертную группу;  $E_{effective}^m(A)$  — показатель комплексной оценки функциональных

от НСД;  $K_{main}$  — набор классификационных признаков оценки основных функциональных возможностей АПМДЗ  $A$ ;  $K_{add}$  — набор классификационных признаков оценки дополнительных функциональных возможностей АПМДЗ  $A$ ;  $E(K)$  — показатель оценки эффективности реализации функциональных возможностей по классификационному признаку  $K$  АПМДЗ  $A$ ;  $E(V)$  — показатель оценки эффективности реализации функциональной возможности  $V$  АПМДЗ  $A$ ;  $\varphi(V)$  — ранг предпочтительности функциональной возможности  $V$ ;  $\psi(K)$  — ранг предпочтительности классификационного признака  $K$ .

Показатель  $E_{effective}(A)$  комплексной оценки функциональных возможностей АПМДЗ  $A$  вычисляется по следующей формуле (в соответствии с оценками, полученными от всех экспертов, входящих в экспертную группу):

возможностей АПМДЗ, полученный на основании оценок  $m$ -го эксперта.

Показатель комплексной оценки АПМДЗ по критерию "эффективность—стоимость" вычисляется по следующей формуле (для оценок, полученных от всех экспертов, входящих в экспертную группу):

$$E(A) = \frac{E_{effective}(A)}{S(A)} = \frac{\sum_{m=1}^{|M|} E_{effective}^m R_m}{S(A) \sum_{m=1}^{|M|} R_m} = \frac{\sum_{m=1}^{|M|} R_m \left( \sum_{i=1}^{|K_{main}|} \sum_{l=1}^{|V_{K_{main}^i}|} \varphi(V_{K_{main}^i}^l) E(V_{K_{main}^i}^l) \psi(K_{main}^i) + \sum_{j=1}^{|K_{add}|} \sum_{n=1}^{|V_{K_{add}^j}|} \varphi(V_{K_{add}^j}^n) E(V_{K_{add}^j}^n) \psi(K_{add}^j) \right)}{S(A) \sum_{m=1}^{|M|} R_m}.$$

### Методика сравнительной оценки функциональных возможностей АПМДЗ

Предлагаемая методика сравнительной оценки функциональных возможностей АПМДЗ состоит из выполнения следующих этапов.

**Этап 1.** Выбор множества  $A = \{A_i\}$ ,  $i \in \overline{1, |A|}$  АПМДЗ для сравнения.

На данном этапе определяется набор модулей, которые требуется сравнить для выявления модуля с наилучшими характеристиками.

**Этап 2.** Выбор группы экспертов для оценивания заявленного множества АПМДЗ:  $M = \{M_m\}$ ,  $m \in \overline{1, |M|}$ .

**Этап 3.** Ранжирование группы экспертов. Выбранная группа экспертов ранжируется в соответствии с предпочтительностью мнения каждого из экспертов: каждому эксперту  $M_m$ ,  $m \in \overline{1, |M|}$  ставится в соответствие ранг предпочтительности.

**Этап 4.** Выдача анкет. Каждому эксперту  $M_m$ ,  $m \in \overline{1, |M|}$  выдается одна анкета, содержащая:

- список функциональных возможностей, сгруппированных в соответствии с классификационными признаками для оценки, с проставлением им показателя оценки и ранга предпочтительности;
- список классификационных признаков оценки функциональных возможностей с проставлением им ранга предпочтительности.

*Этап 5.* Заполнение анкет экспертами.

Каждый эксперт  $M_m$ ,  $m \in \overline{1, |M|}$  заполняет выданную ему анкету:

— выставляя оценку каждой функциональной возможности;

— задавая и корректируя ранг каждой функциональной возможности;

— задавая и корректируя ранг предпочтительности каждого классификационного признака для оценки.

*Этап 6.* Сбор анкет экспертов для дальнейшей обработки.

*Этап 7.* Подсчет показателя комплексной оценки функциональных возможностей АПМДЗ.

Согласно полученным анкетам с экспертными оценками для каждого АПМДЗ  $A_i$ ,  $A = \{A_i\}$ ,  $i \in \overline{1, |A|}$ , вычисляют:

а)  $E_{effective}^m(A_i)$  — показатель комплексной оценки функциональных возможностей АПМДЗ  $A_i$ , полученный  $m$ -м экспертом;

б)  $E_{effective}(A_i)$  — показатель комплексной оценки функциональных возможностей АПМДЗ  $E_{effective}(A_i)$  в соответствии с оценками, полученными от всех экспертов;

в) показатель комплексной оценки по критерию "эффективность—стоимость"  $E(A_i)$ .

*Этап 8.* Из всех  $A_i \in A$  выбирается  $A^*$ :  $E(A^*) = \max_{A_i \in A} (E(A_i))$ .

### Заключение

Автором разработан и предложен алгоритм, позволяющий получить численные показатели оценки функциональных возможностей АПМДЗ в части обеспечения защиты обрабатываемой информации от НСД. Данный алгоритм предоставляет возмож-

ность получить показатель оценки защищенности, обеспечиваемой с помощью АПМДЗ. Алгоритм позволяет получать как показатель комплексной оценки по основным и дополнительным функциональным возможностям, так и показатель оценки по отдельным классификационным признакам АПМДЗ для ситуаций, когда требуется оценить одну из подсистем АПМДЗ или реализацию заданного набора функциональных возможностей.

Методика сравнительной оценки функциональных возможностей АПМДЗ, основанная на классификации функциональных возможностей и критерии "эффективность—стоимость", позволяет из совокупности АПМДЗ выявить модуль с наилучшими характеристиками обеспечения защиты информации от НСД.

### Список литературы

1. Доктрина информационной безопасности РФ // Российская газета. 28.09.2000.
2. Малюк А. А., Пазизин С. В., Пригожин Н. С. Введение в защиту информации в автоматизированных системах. М.: Горячая линия-Телеком, 2001. 148 с.
3. Михайлов С. Ф., Петров В. А., Тимофеев Ю. А. Информационная безопасность. Защита информации в автоматизированных системах. Основные концепции: учеб. пособие. М.: Изд-во МИФИ, 1995. 112 с.
4. Герасименко В. А. Защита информации в автоматизированных системах обработки данных: в 2 кн. Кн. 1. М.: Энергоатомиздат, 1994. 400 с.
5. Герасименко В. А. Защита информации в автоматизированных системах обработки данных: в 2 кн. Кн. 2. М.: Энергоатомиздат, 1994. 176 с.
6. Герасименко В. А., Малюк А. А. Основы защиты информации: учебник для высших учебных заведений Министерства общего и профессионального образования РФ. М.: Изд-во МИФИ, 1997. 538 с.
7. Мельников В. П., Клейменов С. А., Петраков А. М. Информационная безопасность и защита информации. 3-е изд., стер. М.: Академия, 2008. 336 с.
8. Волкова В. Н. Системный анализ и принятие решений. Словарь-справочник: учеб. пособие для вузов / Под ред. В. Н. Волковой, В. Н. Козлова. М.: Высш. школа, 2004. 616 с.

## ИНФОРМАЦИЯ

### Международная конференция по компьютерной лингвистике

### «Диалог—2014»

*Конференция пройдет в начале июня 2014 года в Подмосковье.*

Конференция является традиционным местом встречи и интенсивного обмена идеями между специалистами в области лингвистики, коммуникации, представления знаний, автоматической обработки текста.

Практической целью этого междисциплинарного общения является решение широкого спектра задач автоматической обработки естественного языка. Научная цель заключается в том, чтобы получить теоретические и языковые описания той степени полноты и эксплицитности, которая позволила бы решать вышеуказанные задачи. В этой ориентации инженерной практики на адекватные лингвистические и коммуникативные модели и состоит специфика «Диалога», отличающая его от большинства конференций по компьютерной лингвистике.

Сайт конференции: <http://www.dialog-21.ru>

**А. В. Еременко**, канд. техн. наук, доц.,  
e-mail: nexus-@mail.ru

**А. Е. Сулавко**, аспирант,  
e-mail: sulavich@mail.ru

Федеральное государственное бюджетное  
образовательное учреждение высшего  
профессионального образования  
"Сибирская государственная  
автомобильно-дорожная академия"  
(СибАДИ), г. Омск

## Исследование алгоритма генерации криптографических ключей из биометрической информации пользователей компьютерных систем<sup>1</sup>

*Рассмотрена проблема влияния человеческого фактора на защищенность криптографических ключей. Предложено использовать биометрические характеристики человека в качестве исходного материала для выработки ключевых последовательностей в целях предотвращения несанкционированного доступа к секретным ключам. Исследованы преобразования, позволяющие получить достоверно воспроизводимую, случайную, равномерно распределенную двоичную последовательность из псевдослучайных данных, и рассмотрен процесс создания криптографического ключа из динамических характеристик рукописных паролей.*

**Ключевые слова:** динамика подписи, помехоустойчивое кодирование, исправляющая способность, управление доступом, корреляция функций, идентифицирующие признаки, нечеткие данные, ключевая последовательность

### Введение

Функционирование современного общества без применения криптографических методов в различных сферах человеческой деятельности не представляется возможным. Достижения современной криптографии используются в стратегических компонентах мировой экономики: электронной коммерции, электронном правительстве, телекоммуникации и др. Один из самых сложных вопросов при эксплуатации криптосистем — сохранение криптографических ключей в тайне. При реализации не-

<sup>1</sup> Работа выполнена в рамках государственного задания на 2011—2014 годы на выполнение поисковых научно-исследовательских работ для государственных нужд по теме № 8.2018.2011 "Интеллектуализация транспортных систем: модели скрытой дистанционной идентификации водителей транспортных средств и их психоэмоционального состояния".

обходимых мер по защите информационных систем нельзя упускать из виду человеческий фактор. Имеющиеся оценки потерь от недостаточного внимания к данной проблеме впечатляют (1 трлн долл. в год) [1]. Анализ сложившейся ситуации говорит о том, что существующие технологии предотвращения доступа к информационным ресурсам неавторизованных лиц требуют усовершенствования.

Для решения проблемы допуска пользователей информационных систем (далее пользователей) к защищаемым ресурсам разрабатывают биометрические технологии идентификации личности [2]. В настоящей работе рассмотрен иной подход к защите информации с учетом человеческого фактора, основанный на симбиозе криптографии и биометрии, согласно которому секретные ключи пользователей больше не хранятся в системе, а генерируются каждый раз при обращении к криптографическому сервису на основе динамических характеристик рукописных паролей [3—5]. Это позволяет решить проблему хранения секретных ключей пользователей и выбора слабых паролей для доступа к ним. Динамика подписи является перспективной биометрической характеристикой для использования в системах контроля доступа. Динамический портрет личности практически не поддается копированию и остается тайным для окружающих сотрудников в отличие от статических признаков, для которых разработаны методы по созданию муляжей. Основной проблемой при генерации ключевых последовательностей по биометрическим данным (динамике подписи) является техническая невозможность получения одинаковых образов биометрических характеристик при их повторном вводе пользователем, что связано со стохастическим характером данного процесса. На результаты данного процесса влияют следующие факторы, несущие в себе неопределенность: ситуационные изменения в психофизиологическом состоянии человека; изменения его почерка в течение жизни, они приводят к проблеме несовпадения биометрических образов одного и того же человека, и генерация ключа при этом становится затруднительной. Решение обозначенной проблемы выглядит возможным с использованием помехоустойчивых кодов, исправляющих ошибки. Исследованию этого вопроса и посвящена настоящая работа.

### 1. Алгоритм генерации криптографических ключей из биометрической информации пользователей

В общем виде процесс преобразования биометрических данных в криптографические ключи можно представить следующим образом. В начале биометрические данные пользователей оцифровываются. Чтобы использовать эти данные в качестве ключевого материала, они должны быть преобразованы в достоверно воспроизводимую, случайную,

равномерно распределенную двоичную последовательность. Однако на этапе преобразования биометрических данных в ключевой материал возникает ряд сложностей. Это связано с тем, что биометрические данные неточно воспроизводимы и не имеют равномерного распределения вероятностей. Зарубежными учеными был предложен метод такого преобразования, пригодный для практической реализации, который получил название метода "нечетких экстракторов" (*fuzzy extractors*) [6]. "Нечеткий экстрактор" — это криптографический примитив, состоящий из двух алгоритмов, один из которых описывает процедуру генерации, другой — процедуру восстановления ключа из биометрических данных.

Пусть  $M$  — множество  $n$ -битных двоичных векторов. Будем обозначать их  $M = \{0, 1\}^n$ . Вероятностная процедура генерации  $Gen$  получает на вход представленные в двоичном виде биометрические данные  $\omega \in M$ , а на выходе дает извлеченную последовательность  $K \in \{0, 1\}^l$  и вспомогательную открытую последовательность  $P \in \{0, 1\}^*$ . Таким образом, при первоначальном считывании биометрических данных из них извлекается равномерно распределенная последовательность, которая может служить секретным ключом, и вырабатывается открытый шаблон, который сохраняется для последующего использования. Длина ключа может быть задана в виде параметра.

В качестве входных данных детерминированная процедура восстановления  $Rep$  использует биометрические данные  $\omega^* \in M$  и последовательность бит  $P \in \{0, 1\}^*$ . Эта процедура позволяет восстанавливать  $K$  из соответствующей открытой последовательности  $P$ . Пусть  $dis(\omega, \omega^*)$  — расстояние Хемминга между двоичными векторами  $\omega, \omega^*$ , а  $t$  — кратность исправляемых ошибок. Свойство корректности нечетких экстракторов состоит в том, что если  $dis(\omega, \omega^*) \leq t$  и  $K, P$  сгенерированы с помощью процедуры  $Gen$ , т. е.  $(K, P) \leftarrow Gen(\omega)$ , то  $Rep(\omega^*, P) = K$ . Если же  $dis(\omega, \omega^*) > t$ , то неизвестно, будет ли обратно получена  $K$  на выходе  $Rep$  или нет. Последовательность  $\omega^*$  будет правильно декодирована в последовательность  $\omega$  в том случае, если расстояние до  $\omega$  является минимальным, т. е.  $\omega$  является ближайшей последовательностью к  $\omega^*$ , и  $\omega^*$  находится внутри шара с центром  $\omega$  и радиусом  $t$ . Таким образом, при повторном считывании биометрических данных, достаточно схожих с первоначальными, секретный ключ может быть безошибочно воспроизведен на основе введенных данных и сохраненного открытого шаблона. Кратность исправляемых ошибок  $t$  является регулируемым параметром [7].

## 2. Экспериментальная проверка алгоритма генерации ключей из динамических характеристик рукописных паролей

В целях подтверждения возможности использования представленного выше нечеткого (размытого) экстрактора (рис. 1) для генерации криптографи-

ческих ключей на основе динамики подписи необходимо реализовать обе части данного метода: секретный набросок и сильный экстрактор, а также экспериментально определить вероятность генерации правильной ключевой последовательности. Роль секретного наброска заключается в создании вспомогательного сообщения на основе исходного такого, чтобы для двух близких в какой-нибудь метрике исходных сообщений получалось в точности одно и то же вспомогательное сообщение (ключевая последовательность). Затем к этому вспомогательному сообщению применяется сильный экстрактор, создавая короткую строку с однородным распределением вероятности символов в заданной позиции строки, которая пригодна для использования в качестве криптографического ключа или может служить параметром сравнения при идентификации.

Условимся считать, что исходное сообщение — бинарная строка, а используемая метрика — метрика Хемминга. Также введем корректирующий ошибки код [8, 9] на пространстве сообщений так, что минимальное расстояние между двумя кодами больше или равно удвоенному максимальному расстоянию, допустимому для двух считающихся близкими исходных сообщений плюс единица:  $d = 2t + 1$ , где  $d$  — минимальное расстояние для кода;  $t$  — максимальное расстояние для двух близких сообщений.

Побитное "отрицающее или", выполненное с одним из кодов и образцовым исходным сообщением, даст некое сообщение, которое будем называть открытой строкой (сообщением). "Отрицающее или" обладает свойством изометрии для метрики Хемминга. Поэтому применение "отрицающего или" к тестируемому исходному сообщению и открытому сообщению, а также последующее декодирование приведет к восстановлению одного из



Рис. 1. Алгоритм Fuzzy Extractor

кодов. Таким образом, "отрицающее или" с открытой строкой и процедура декодирования для корректирующего ошибки кода могут быть рассмотрены как секретный набросок.

В качестве сильного экстрактора может быть применен алгоритм Рабина—Карпа:

$$H = c_1 a^{k-1} + c_2 a^{k-2} + c_3 a^{k-3} + \dots + c_k a^0,$$

где  $c_1, \dots, c_k$  — составные части строки;  $a$  — некая константа.

Все операции целесообразно выполнять по модулю некоего простого числа  $n$ .

Другой вариант сильного экстрактора — сдвиговый хеш:

$$H \leftarrow H \oplus S^k(h(c_1)) \oplus h(c_{k+1}),$$

где  $h()$  — биективная функция;  $S^k()$  — циклический побитовый сдвиг порядка  $k$ ;  $\oplus$  — побитовое "отрицающее или".

Суммируя сказанное выше, для проведения экспериментов был составлен следующий алгоритм.

1. Получить биометрические данные пользователя.
2. Составить из биометрических данных бинарные строки. В качестве биометрических характеристик было использовано 15 коэффициентов корреляции между функциями координат положения пера от времени, функцией давления и их производными [10, 11]. Значение каждого признака округлялось до сотых и преобразовалось в целочисленный вид (например, 0,97342  $\rightarrow$  97  $\rightarrow$  1100001).
3. Инициализировать корректирующий код длиной, соответствующей длине бинарных строк, и корректирующий заданное количество ошибок.
4. Провести инициализацию базы данных образцов. Процедура создания нового образца включает в себя:
  - ввод имени образца;
  - ввод бинарных строк образца;
  - выбор случайного кода;
  - "отрицающее или" строк образца с выбранным кодом (получение открытой строки);
  - экстракцию выбранного кода (получение хеша);
  - сохранение имени образца, открытой строки, хеша в базе данных.

5. Проверка образца:
  - ввод имени образца;
  - ввод бинарной строки образца;
  - "отрицающее или" с открытой строкой;
  - декодирование результата пункта выше;
  - экстракция пункта выше;
  - сравнение с имеющимся в базе данных хешем.

Работоспособность программы можно проверить на одном из самых простых кодов — коде Хэмминга. Более интересным с практической точки зрения является код Боуза—Чоудхури—Хоквингхема (БЧХ), отличающийся возможностью построения кода с заранее определенными корректирующими свойствами. Ниже будут изложены алгоритмы обоих кодов.

Пусть  $x_3, x_5, x_6, x_7 \in Z_2$  — четырехбитное представление номера кода, тогда:

$$x_4 = x_5 + x_6 + x_7;$$

$$x_2 = x_3 + x_6 + x_7;$$

$$x_1 = x_3 + x_5 + x_7,$$

где  $(x_1, x_2, x_3, x_4, x_5, x_6, x_7)$  — семибитное представление кода.

Для исправления ошибки вычисляют следующее:

$$a = x_4 + x_5 + x_6 + x_7;$$

$$b = x_2 + x_3 + x_6 + x_7;$$

$$c = x_1 + x_3 + x_5 + x_7,$$

где  $(a, b, c)$  — номер бита, в котором произошла ошибка (равенство нулю означает отсутствие ошибки). Так генерируется и декодируется ошибка код Хэмминга.

Построение кода БЧХ может быть описано следующим образом:

- выбрать поле Галуа  $GF(q)$ , проще всего взять  $Z_p$ , где  $p$  — простое число;
  - выбрать длину  $n$  кода из условия  $n = (q^m - 1)/s$ , где  $m, s$  — целые положительные числа;
  - задать минимальное расстояние  $d$ ;
  - построить циклотомические классы элемента  $\beta = \alpha^s$  поля  $GF(q^m)$  над полем  $GF(q)$ , где  $\alpha$  — примитивный элемент  $GF(q^m)$ ;
  - выбрать  $\beta^{i_0}, \beta^{i_0+1}, \dots, \beta^{i_0+d-2}$  таким образом, чтобы суммарная длина циклотомических классов была минимальна;
  - вычислить порождающий полином  $g(x) = f_1(x)f_2(x)\dots f_n(x)$ , где  $f_i(x)$  — полином, соответствующий  $i$ -му циклотомическому классу;
  - умножить порождающий полином на информационный полином, т. е. полином, имеющий коэффициенты, задающие номер кода.
- Декодирование кода БЧХ происходит немного более сложным образом, чем в случае кода Хэмминга. Используют элементы порождающего поля для нумерации элементов принятого сообщения (см. таблицу).

Нумерация элементов сообщения

Значения	$r_0$	$r_1$	...	$r_{n-1}$
Локаторы позиций	1	$\alpha$	...	$\alpha^{n-1}$

Пусть принятое сообщение ассоциировано с полиномом  $r(x) = v(x) + e(x)$ , где многочлен ошибок определен как

$$e(x) = e_{j_1}x^{j_1} + e_{j_2}x^{j_2} + \dots + e_{j_v}x^{j_v},$$

где  $v \leq t_d$  — число ошибок в принятом слове.

Множества  $e_{j_1}, e_{j_2}, \dots, e_{j_v}$  и  $\alpha^{J_1}, \alpha^{J_2}, \dots, \alpha^{J_v}$  называют значениями и локаторами, соответственно, где  $e_j \in GF(q)$  и  $\alpha \in GF(q^m)$ .

Введем синдромы как значения принятого полинома  $r(x)$  в нулях порождающего многочлена кода:

$$\begin{aligned} S_1 &= r(\alpha^b) = e_{J_1} \alpha^{bJ_1} + e_{J_2} \alpha^{bJ_2} \dots e_{J_v} \alpha^{bJ_v}; \\ S_2 &= r(\alpha^{b+1}) = \\ &= e_{J_1} \alpha^{(b+1)J_1} + e_{J_2} \alpha^{(b+1)J_2} + \dots + e_{J_v} \alpha^{(b+1)J_v}; \\ &\vdots \\ S_{2t_d} &= r(\alpha^{b+2t_d-1}) = e_{J_1} \alpha^{(b+2t_d-1)J_1} + \\ &+ e_{J_2} \alpha^{(b+2t_d-1)J_2} + \dots + e_{J_v} \alpha^{(b+2t_d-1)J_v}; \end{aligned}$$

где  $2t_d = d - 1$  — максимальное число ошибок, корректируемых кодом.

Введем также полином локаторов ошибок:

$$\begin{aligned} \Lambda(x) &= (1 - xX_1)(1 - xX_2) \dots (1 - xX_t) = \\ &= \Lambda_t x^t + \Lambda_{t-1} x^{t-1} + \dots + \Lambda_1 x + 1. \end{aligned}$$

Коэффициенты последнего полинома связаны с синдромами системой линейных уравнений:

$$\begin{aligned} S_1 \Lambda_t + S_2 \Lambda_{t-1} + \dots + S_t \Lambda_1 &= -S_{t+1}; \\ S_2 \Lambda_t + S_3 \Lambda_{t-1} + \dots + S_{t+1} \Lambda_1 &= -S_{t+2}; \\ &\vdots \\ S_t \Lambda_t + S_{t+1} \Lambda_{t-1} + \dots + S_{2t-1} \Lambda_1 &= -S_{2t}. \end{aligned}$$

В программе было реализовано прямое решение этой системы, которое носит название алгоритма Питерсона—Горенштейна—Цирлера (ПГЦ).

После нахождения коэффициентов локаторов ошибок перебором по соответствующему порождающему полю ищутся корни этого многочлена. С помощью корней, как обратную величину корням, находят локаторы ошибок. А затем по локаторам находят позиции ошибок:

$$i_k = \log_{\beta} X_k.$$

Далее по формулам Форни находят значения ошибок:

$$\begin{aligned} e_{i_k} &= -\frac{\Omega(X_k^{-1})}{\prod_{l \neq k} (1 - X_l X_k)}; \\ \Omega(x) &= \Lambda(x) \left( 1 + \sum_{k=1}^{2t_d} S_k x^k \right). \end{aligned}$$

Зная значение и расположение ошибок, можно получать искомые кодовые последовательности.

### 3. Пример применения предложенного алгоритма

Предложенный алгоритм может быть использован при построении инфраструктуры облачного хранения данных пользователей с шифрованием на основе динамических характеристик рукописных паролей, либо других биометрических признаков. После получения секретного ключа и открытой строки к нему, пользователь зашифровывает данные на стороне клиента, размещает зашифрованные данные и открытую строку в облачном хранилище (рис. 2). Секретный ключ при этом уничтожается. Такое решение позволит пользователю всегда иметь доступ к данным при наличии подключения к Интернету и соответствующего программно-аппаратного обеспечения без снижения степени защищенности конфиденциальной информации (злоумышленник не сможет восстановить секретный ключ пользователя в том случае, если он получит доступ к данным и открытой строке). Вопрос надежного хранения секретных ключей с пользователя снимается.

На рис. 2  $v_x(t), v_y(t)$  — скорости изменения положения пера в плоскости планшета;  $p(t)$  — изменение давления пера на поверхность планшета;  $a(t)$  — наклон пера к плоскости графического планшета.

Описанный алгоритм (см. рис. 1) положен в основу программы эксперимента по исследованию выбранного в качестве прототипа системы генерации криптографических ключей по динамическим характеристикам рукописного пароля метода "нечетких экстракторов". С помощью разработанного программного модуля от 30 пользователей, имеющих разную стабильность почерка, в процессе эксперимента были собраны 1282 примера подписи. Из трех примеров подписи каждого пользователя генерировался ключ и открытая строка для него. Оставшиеся реализации подписей в дальнейшем использовались для извлечения биометрических данных. На вход программы подавали сгенерированные бинарные строки длиной 120 бит (15 признаков, занимающих в бинарном виде 15 байт). Суть эксперимента заключалась в проверке идентичности генерируемых ключей на основании воспроизводимых паролей у каждого отдельного пользователя и их несовпадения у разных пользователей. В результате было проведено 1192 опыта, в которых было зарегистрировано 107 ошибок: в 107 случаях не удалось восстановить исходные ключевые последовательности. Таким образом, вероятность генерации правильной ключевой последовательности составила 0,91. Ситуации совпадения ключевых последовательностей у разных пользователей не возникло, что объясняется стохастической природой процесса генерации ключей (вероятность случайного совпадения ничтожно мала). Для моделирования такой ситуации необходимо провести ряд опытов по подделке рукописных паролей, что является темой отдельного исследования.

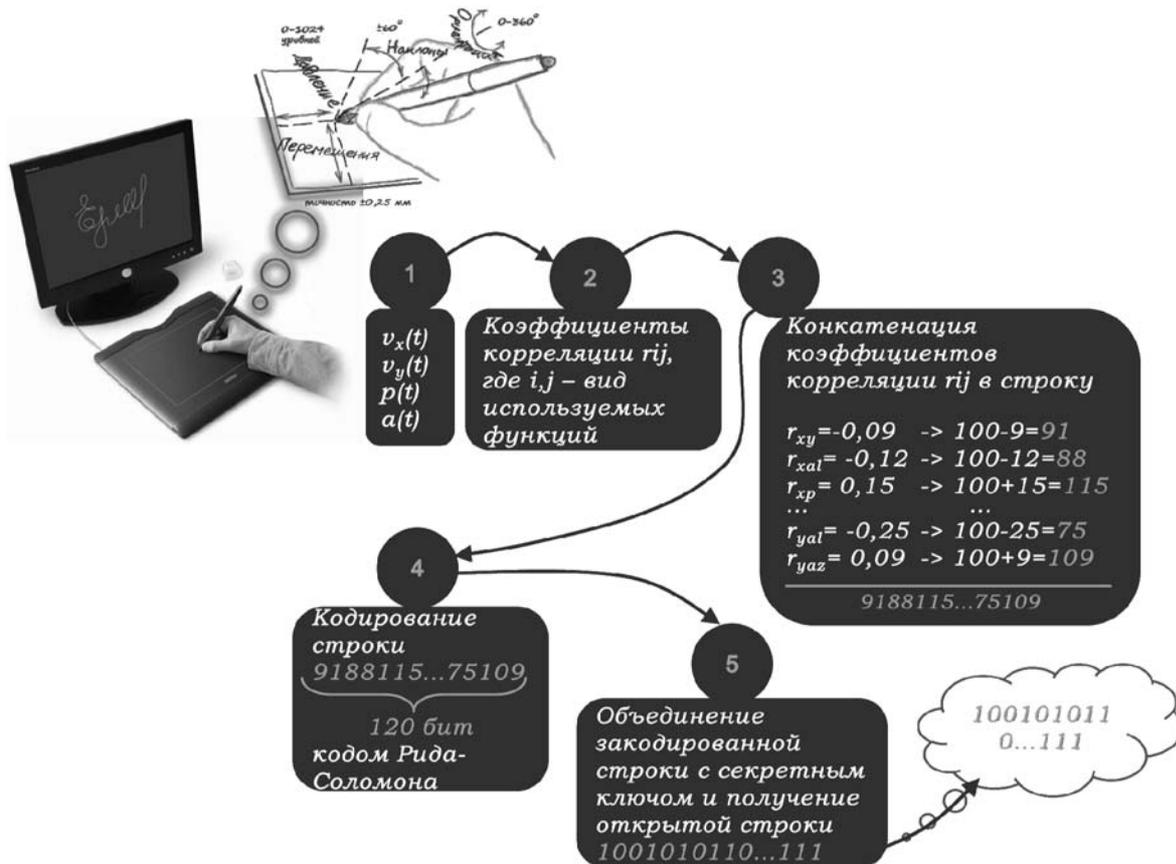


Рис. 2. Пример использования рассмотренного в статье алгоритма

## Заключение

Таким образом, в поставленном эксперименте подтверждена принципиальная возможность генерации криптографических ключей, отвечающих требованиям, предъявляемым к криптографическому материалу на основании биометрических характеристик, получаемых при воспроизведении парольных слов на графическом планшете. Оценен уровень ошибок восстановления исходной ключевой последовательности из динамики подписи пользователей, который на данном этапе развития биометрических систем можно охарактеризовать как "высокий". Дальнейшее направление работ должно заключаться в снижении ошибок генерации ключей за счет более глубокого исследования процесса написания слов на графическом планшете и поиска признаков, близких к детерминированным.

## Список литературы

1. Разработка комплексированной технологии оперативного выявления террористических угроз на магистральных продуктопроводах: Научно-технический отчет о выполнении 2 этапа Государственного контракта № П215 от 22 июля 2009 г. СибАДИ. Руководитель: Елифанцев Б. Н. Омск, 2011.
2. Сулавко А. Е. Технологии защиты от внутренних угроз информационной безопасности // Вестник Сибирской государственной автомобильно-дорожной академии. 2011. № 1(19). С. 45–51.

3. Еременко А. В. О синтезе криптографических ключей по динамическим характеристикам рукописного пароля // Материалы 64-й научно-технической конференции ГОУ "СибАДИ" в рамках Юбилейного Международного конгресса "Креативные подходы в образовательной, научной и производственной деятельности", посвященного 80-летию академии. Омск: СибАДИ, 2010. Кн. 2. С. 147–151.
4. Ellison C., Hall C., Milbert R., Schneier B. Protecting Keys with Personal Entropy // Future Generation Computer Systems. 2000. N 16. P. 311–318.
5. Hao F., Chan C. W. Private key generation from on-line handwritten signatures // Information Management & Computer Security, Issue 10. 2002. N 2. P. 159–164.
6. Dodis Y., Reyzin L., Smith A. Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy Data // Proc. from Advances in Cryptology. EuroCrypt. 2004.
7. Корнюшин П. Н., Гончаров С. М., Харин Е. А. Построение систем биометрической аутентификации с использованием генератора ключевых последовательностей на основе нечетких данных // Материалы 50-й всероссийской научной конференции. Владивосток: ТОВМИ, 2007. Т. 2. С. 112–115.
8. Питерсон У., Уэлдон Э. Коды, исправляющие ошибки: пер. с англ. / Под ред. Р. Л. Добрушиной, С. И. Самойленко. М.: Мир, 1976.
9. MacWilliams F. J., Sloane N. J. A. The Theory of Error-correcting Codes. North Holland. 1991.
10. Елифанцев Б. Н., Ложников П. С., Еременко А. В. Идентификация пользователя ПЭВМ по рукописному паролю: конкурентные возможности // Вестник СибАДИ. Омск: СибАДИ, 2011. № 2. С. 52–56.
11. Иванов А. И. Биометрическая идентификация личности по динамике подсознательных движений. Пенза: Изд-во ПГУ, 2000. 188 с.

УДК 004.93'11

А. А. Домунян, мл. науч. сотр.,  
ИПУ им. В. А. Трапезникова РАН,  
e-mail: inventor@pisem.net

## Метод распознавания жестов руки на основе позиционного расстояния

*Рассматривается задача распознавания жестов руки. В качестве предобработки сцены выбирается метод разностных изображений и оконтуривание объектов. Задача распознавания жестов разбивается на три задачи. Описывается способ выделения признаков, функции позиционного расстояния. Приводится метод обучения системы.*

**Ключевые слова:** распознавание жестов, выделение признаков, позиционное расстояние, системы с обучением

### Введение

Существует широкий круг технических и бытовых приложений, автоматизация которых сдерживается отсутствием удобной и дешевой вычислительной платформы. Так, например, замена кнопочных выключателей комнатного света на умные выключатели, управляемые с помощью жестов, требует использования такой относительно дорогой вычислительной платформы, как персональный компьютер. Очевидно, что подобный подход не может быть использован для разработки коммерчески приемлемых умных выключателей минимальной стоимости. Другим примером могут служить бесконтактные способы управления самыми различными устройствами — от аудиосистем автомобилей до детских игрушек. В этих случаях умные управляющие устройства должны понимать наборы самых разнообразных команд.

Создание вычислительной платформы, которая на порядок дешевле, чем персональный компьютер и мобильный телефон, открывает дорогу в мир простых, удобных и умных устройств, которые рано или поздно найдут широкое применение в повседневной жизни.

### Постановка задачи и ее особенности

В настоящее время распознаванием жестов занимаются многие ИТ-корпорации. Например, Майкрософт делает Microsoft Kinect — бесконтакт-

ный сенсорный контроллер, который позволяет пользователю взаимодействовать с ним без помощи контактного игрового контроллера через устные команды, позы тела и показываемые объекты или рисунки, в том числе с помощью жестов рук. Тем не менее, он достаточно дорог, велик по размерам и во время работы потребляет достаточно большую мощность, что не позволяет его использовать во многих простых и дешевых устройствах.

Большинство ведущихся сейчас зарубежных разработок по распознаванию жестов руки используют именно датчик Kinect. Тем самым они возлагают на него вычислительно сложную предобработку изображения, но тем самым поигрывают по цене, габаритным размерам и энергопотреблению более специализированным устройствам.

Среди отечественных разработок превалирует борьба за качество распознавания. Большинство алгоритмов пытаются выделить специфические признаки руки. Некоторые авторы разработок ищут пять пальцев и ладонь, смотрят на типичный цвет кожи, соотношение размеров, присущих руке. Тем самым они справляются с качеством распознавания руки вплоть до 80 % при достаточно малой области попадания руки в кадр. Но для этого используются мощные процессоры или ПЛИС. Мне же хотелось сделать простое, компактное и дешевое устройство, качество распознавания которого во многом будет зависеть от самого человека, показывающего жесты перед ним. То есть для необходимого качества распознавания необходимо показать жест перед камерой под правильным углом, при достаточном освещении и на правильном удалении. Это обычно выполняется. Действительно, устройство предназначено для управления им, и нет никакой необходимости ложного срабатывания от жеста человека, который находится далеко от него. Если человек хочет включить прибор, то он должен показать жест прибору, а не просто случайно махнуть рукой.

Целью работы являлось исследование математических и алгоритмических аспектов задач распознавания жестов руки, разработка математического обеспечения и алгоритмов, формулировка требований к аппаратным характеристикам платформы, зависящим от вычислительной сложности предлагаемых алгоритмов анализа видеoinформации и от существующей элементной базы. Значительная часть работы посвящена натурному тестированию разработанных алгоритмов в реальном масштабе времени.

## Выделение признаков

Важным этапом распознавания жестов является решение задачи выделения признаков, описывающих произвольные жесты. В качестве основной характеристики объекта используется его контур. *Первым этапом процедуры выделения признаков является оконтуривание.* Последовательность контуров достаточно четко описывает изменяющуюся во времени форму руки, о чем свидетельствует то обстоятельство, что человеку не представляет труда распознать жест по серии его контурных изображений.

Как отмечено в работе [1], фундаментальным способом оконтуривания является двумерное преобразование Фурье, которое позволяет выделить низкочастотную составляющую спектра для ее последующего удаления. Затем в результате применения обратного преобразования Фурье восстановленный исходный объект принимает вид контура, так как участкам плавного изменения яркости соответствует низкочастотная часть спектра, удаленная после прямого преобразования. Однако такой подход требует больших вычислительных ресурсов, объем которых имеет порядок  $O(N^4)$ ,  $N \times N$  — это размер изображения. Очевидно, что такая вычислительная нагрузка велика для работы с видеоклипами в реальном масштабе времени.

Другие фундаментальные способы оконтуривания связаны с использованием локальных операторов, рассмотренных в работе [2]. В этом случае объем вычислительной нагрузки пропорционален  $O(N^2) \times O(M^2)$ , где  $M \times M$  — это размер локального окна, сканирующего весь кадр. Недостаток этого подхода состоит в том, что контуры неподвижных объектов также будут выделены, а представляющая их информация поступит для обработки в последующие модули системы. Таким образом, неподвижные объекты сцены будут обрабатываться многократно, в то время как их можно вообще не обрабатывать. Действительно, рука, принимающая форму определенного жеста, проходит через ряд промежуточных положений и, следовательно, достаточно рассматривать только динамику ее движения. Кроме того есть возможность избавиться от сомножителя  $O(M^2)$ , если рассматривать только точки изображения без их локальных окрестностей. С учетом вышесказанного в работе выбрана стратегия анализа пар изображений, результатом которого является разностное изображение. При этом, если объект неподвижен, то разностное изображение не существует, так как оно представлено в этом случае пустым множеством точек. Разностное изображение формируется путем поточечного сравнения двух кадров и использования порога яркости  $T$ . Если разность яркостей двух точек с одинаковыми координатами  $(x, y)$  меньше заданного порога, то в разностном изображении будет отсутствовать точка с координатами  $(x, y)$ . Разностное изображение представляется

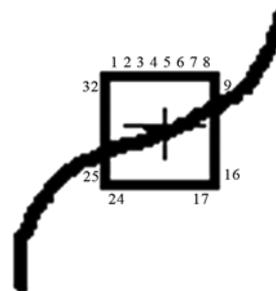


Рис. 1. Определение локального признака

неупорядоченным списком выделенных точек, т. е. списком координат  $(x_1, y_1), (x_2, y_2), \dots, (x_p, y_p)$ , длина которого меняется в зависимости от степени различия сравниваемых изображений.

*Второй этап выделения признаков начинается с прореживания разностного списка.* При прореживании в списке оставляется только каждая  $k$ -я точка (например  $k = 10$ ). Второй этап заканчивается определением угловой ориентации контурных точек по отношению к каждой точке списка. Для этого каждая точка списка окружается квадратной локальной рамкой. Точки рамки перенумеровываются от 0 до  $L - 1$ , где  $L$  — это число точек в рамке. Если рамка пересекает контур объекта в точке  $a$ , то признак описывается значением  $a$ . Если рамка пересекает контур объекта в точках  $a$  и  $b$ , то признак описывается значениями  $(a, b)$  и т. д. В примере, изображенном на рис. 1, рамка пересекает контур в точках 10 и 26. Значит, признак описывается множеством  $(10, 26)$ .

Очевидно, что выбранный способ представления контура является далеко не однозначным, так как разные контуры могут иметь одно описание. Однако экспериментально установлено, что использование данных признаков позволяет на последующих этапах довести решение поставленной задачи до конца.

Последовательность выделенных признаков является входной информацией для системы распознавания признаков, выходом которой является множество имен распознанных признаков и которая завершает этап выделения признаков.

### Распознавание динамических последовательностей контуров руки — жестов

Для распознавания жестов необходимо последовательно решить следующие задачи:

- распознавания локальных признаков, описывающих контур руки;
- распознавание отдельных контуров;
- распознавание динамических последовательностей контуров, представляющих жесты.

*Первая задача*, т. е. задача распознавания признаков, сводится к задаче поиска ближайшего соседа. Пусть  $F_n, n \in N$ , — это конечный набор конеч-

ных множеств. Для заданного конечного множества  $U$  требуется найти множество  $F_m$ , такое что:

$$D(F_m, U) = \min_{n \in N} D(F_n, U).$$

Здесь  $D(F_m, U)$  — это расстояние между множествами  $F_m$  и  $U$ .

Каждое множество  $F_n$ , а также множество  $U$  — это множества чисел  $\{a_1, a_2, \dots, a_p\}$ , значения которых равны локальным ориентациям контурных точек относительно центральной точки локальной рамки. Поскольку множества  $\{a_1, a_2, \dots, a_p\}$  могут иметь разное число элементов, то рассматриваемая задача не сводится к сравнению  $p$ -мерных векторов, а требует сравнения множеств переменной длины. Для этого используется метод позиционного расстояния, предложенный в работах [1–2].

Поскольку максимальное число угловых ориентаций, представляющих локальный признак, равно  $L$ , то каждое множество  $F_n$  можно взаимно однозначно представить бинарной последовательностью, содержащей  $L$  разрядов. Так, если  $L = 8$ , то набор чисел 0, 4 и 6 представляет последовательность 10001010. В общем случае нормализованное позиционное расстояние между двумя множествами имеет вид

$$D_p(A, B) = \sum_n (\&_k(a_n \text{ XOR } b_{n-k})) / \sum_n (a_n \text{ OR } b_{n-k}), \quad (1)$$

где  $|k| \leq R$ ,  $R < L$  и  $n = 1, 2, \dots, L$ ;  $R$  — радиус ступенчатой функции.

OR — операция побитовое ИЛИ (бинарная операция, действие которой эквивалентно применению логического ИЛИ к каждой паре битов, которые стоят на одинаковых позициях в двоичных представлениях операндов); XOR — побитовое исключающее ИЛИ; & — побитовое И.

Так, если  $A = 1000110000100100001$  и

$B = 0100001001000010010$ , то позиционное расстояние между  $A$  и  $B$  равно нулю;  $D_p(A, B) = 0$  при  $R = 2$ . Это означает, что смещение позиций точек внутри радиуса  $R = 2$  не влияет на позиционное расстояние между последовательностями и последовательности можно считать похожими. Действительно, рядом (не далее двух разрядов) с каждой битовой единицей из  $A$  есть единица из  $B$ , и наоборот.

Нормированное позиционное расстояние изменяется в интервале от 0 до 1. Если расстояние  $\rho$  между неизвестным входным признаком и ближайшим признаком из базы данных, содержащей на текущий момент  $N$  признаков, превышает порог  $T_1$ , то неизвестный признак получает имя  $N + 1$  и заносится в базу данных. Создаваемое новое имя по существу является именем нового класса, к которому будут отнесены все признаки, являющиеся соседями (с точностью до порога  $T_1$ ) признака, занесенного в базу данных. Если  $\rho \leq T_1$ , то входной признак идентифицируется именем его ближайшего соседа.

Таким образом, база признаков, вначале пустая, постоянно расширяется.

*Вторая задача* — задача распознавания отдельных контуров по совокупности описывающих их признаков ставится следующим образом. Входной информацией служит набор имен признаков, описывающих текущий контур. Если признак повторяется в контуре  $I$  несколько раз, то число повторений игнорируется. Обозначая контур, подлежащий идентификации, через  $U$ , а число контуров в базе контуров на текущий момент через  $C$ , мы приходим к необходимости минимизации следующего выражения:

$$D_T(I, U) \rightarrow \min_{I \in C},$$

где расстояние

$$D_T(I, U) = \frac{\sum_n (i_n \text{ XOR } u_n)}{\sum_n (i_n \text{ OR } u_n)} \quad (2)$$

вычисляется путем представления множеств  $I$  и  $U$  в виде бинарных последовательностей, аналогично тому, как это делалось при решении первой задачи. Как и при решении первой задачи, если расстояние  $\rho$  между неизвестным входным описанием контура и ближайшим контуром из базы данных, содержащих на текущий момент  $C$  контуров, превышает порог  $T_2$ , то неизвестный контур получает имя  $C + 1$  и заносится в базу данных. Создаваемое новое имя будет именем нового класса, к которому будут отнесены все контуры, являющиеся соседями (с точностью до порога  $T_2$ ) контура, занесенного в базу данных.

*Третья задача* — задача распознавания динамических последовательностей контуров, представляющих жесты, ставится во многом аналогично первой и второй задачам. При этом входной информацией служат имена контуров, идентифицированных на втором уровне. Однако при решении данной задачи необходимо учитывать порядок следования контуров в видеоклипе. Учет порядка обеспечивается путем распознавания последовательностей имен контуров с помощью множеств со взвешенными элементами и путем минимизации расстояния

$$D_{TW}(V, U) \rightarrow \min_{V \in C},$$

где

$$D_{TW}(V, U) = \frac{\sum_n \min(a_n, b_n)}{\sum_n \max(a_n, b_n)}, \quad (3)$$

$V$  пробегает имена клипов-кандидатов,  $U$  — имя идентифицируемого клина, а коэффициенты  $a_n, b_n$  — это веса элементов аппроксимирующих множеств.

## Обучение системы

Перед началом обучения фиксируются внешние классы клипов, число которых равно числу жестов, подлежащих распознаванию. Однако при выбранном значении порога третьего уровня  $T_3 < 1$  число внутренних классов, т. е. классов, автоматически создаваемых системой на третьем уровне, может превышать число внешних классов третьего уровня. Таким образом, несколько внешних классов могут соответствовать одному внутреннему классу, что учитывается с помощью функции соответствия  $f$ : внешний класс =  $f$ (внутренний класс). Замечу, что на 1-м и 2-м уровнях внешние классы отсутствуют.

### Программное обеспечение

Данная задача была реализована в виде программы на персональном компьютере. Минимизация расстояний проведена с помощью индексной системы, которая содержит:

- списки элементов;
- модуль записи, обновляющий списки элементов;
- модуль чтения, подсчитывающий активность элементов списков и выделяющий наиболее активные элементы;
- модуль обратной связи, сравнивающий наиболее активные элементы;
- модуль обратной связи, сравнивающий наиболее активные образы с входным образом.

В данной работе списки определяются динамически. Определяются три одномерных массива — массив адресов  $A[R]$ , массив имен  $N[G]$  и массив связей  $L[G]$ . Такой подход позволяет использовать память наиболее рационально, так как при этом не требуется делать предположений о неизвестных размерах списков.

При поступлении на вход адреса списка  $add$  из массива адресов  $A[R]$  считывается адрес первой связи  $p = A[add]$ , а адрес второй связи и адреса всех последующих связей  $p = A[p]$  считываются из массива связей вплоть до обнуления адреса связи. Каждая связь  $p$  не только итеративно указывает на следующую связь  $p = A[p]$ , но и указывает с помощью массива имен  $N[G]$  на имя, ассоциированное с ней  $n = N[p]$ . Последнее обстоятельство позволяет построить гистограмму активности имен. Таким образом, алгоритм построения активности при использовании одномерного модуля чтения имеет следующий вид:

```
p = A[add];
while (p > 0) {
  H[N[L[p]]] ++;
}
```

Работа модуля чтения завершается нахождением максимально активного элемента  $n$ .

Модуль записи построен аналогично модулю чтения с той разницей, что после обнуления связи

в результате итеративного процесса создается новая связь в массиве связей и по адресу, определяемому этой новой связью, создается новое имя в массиве имен. Значение этого имени определяется путем увеличения значения последнего созданного имени на единицу:

```
p = A[add];
while (p > 0) {
  p = L[p] ++;
}
L[p] = p + 1;
N[p] = n + 1.
```

Модуль обратной связи построен аналогично вышерассмотренному прямому модулю с той разницей, что адреса и имена меняются местами. В прямом модуле в качестве адресов выступают значения признаков образов, а имена элементов списков — это имена образов. В обратном модуле имена образов играют роль адресов списков, а в ячейки списков заносятся имена образов. Кроме того, при считывании информации из обратного модуля, т. е. при считывании значений признаков, гистограмма не строится, а множество извлеченных признаков, представляющее образ, сравнивается с текущим входным образом.

В целом процедура обновления прямых и обратных списков состоит в следующем:

- при поступлении на вход неизвестного образа осуществляется считывание информации из прямого модуля и выделяется множество наиболее активных имен образов с использованием гистограммы активности;
- для каждого найденного имени-кандидата с помощью модуля обратной связи извлекаются значения признаков образа-кандидата, которые сравниваются с признаками входного образа с использованием рассмотренных выше мер расстояния;
- если расстояние до ближайшего найденного множества входного образа не превышает пороговую величину, то входной образ идентифицируется, т. е. получает имя его ближайшего соседа. В противном случае обновляется прямой модуль, в который записываются копии нового имени по адресам, определяемым значениями признака входного образа, и обновляется обратный модуль, в который записываются значения признаков входного образа в список, определяемый новым созданным именем.

### Обучение и распознавание

Для тестирования системы были записаны три видеоклипа, показанные на рис. 2, 3 и 4. Каждый клип содержит от 400 до 600 кадров. При частоте 20 кадров в секунду один клип длится около 25 с. Из-за ограниченных размеров страницы на рисунках показаны только несколько фрагментов каж-



Рис. 2. Фрагменты первого жеста из видеоклипа 1, показанные в виде двух строчек по направлению справа налево



Рис. 3. Фрагменты второго жеста из видеоклипа 2, показанные в виде двух строчек по направлению справа налево



Рис. 4. Фрагменты третьего жеста из видеоклипа 3, показанные в виде двух строчек по направлению справа налево

дого жеста. В каждом клипе соответствующий жест повторяется 5...6 раз. При этом повторы разделены (1...2)-секундными интервалами. Очевидно, что человек не может абсолютно точно воспроизводить свои жесты. Поэтому первые три жеста в каждом клипе использовались для обучения системы, а оставшиеся жесты — для ее тестирования. Было установлено, что после обучения система оказалась способной правильно распознавать оставшиеся жесты. Были выбраны следующие пороги распознавания:  $T_f = 0,5$ ,  $T_i = 0,4$ ,  $T_v = 0,5$ . В результате выбора таких значений были идентифицированы 29, 99 и 17 внутренних классов на уровнях 1, 2 и 3 соответственно. Следует отметить, что при обучении порог распознавания клипа  $T_v$  был равным 0,5, а при распознавании — 0,99. Последний выбор повышает обобщающую способность системы на этапе распознавания.

#### Список литературы

1. Gonzalez R. C., Woods R. E. Digital Image Processing. Prentice Hall, 2008.
2. Tian Y. L., Kanade T., Cohn J. F. Recognizing action units for facial expression analysis // IEEE Trans. Pattern Anal. Mach. Intell. 2001. Vol. 23, N 2. P. 97—115.
3. Лазаренко А. В., Михайлов А. М. Алгоритм оптического распознавания движущихся объектов // Автоматика и телемеханика, 2011.
4. Михайлов А. М. Распознавание образов с помощью их индексирования // Автоматика и телемеханика, 2011.
5. Дуда Р., Харт П. Распознавание образов и анализ сцен. М.: Мир, 1976.

## ИНФОРМАЦИЯ



### VIII Международная научная конференция "Параллельные вычислительные технологии" (ПаВТ'2014)

31 марта—4 апреля 2014 г., Южный федеральный университет (г. Ростов-на-Дону)

**ПаВТ** — серия международных научных конференций, представляющих собой авторитетный и престижный форум в области применения параллельных вычислительных технологий в различных областях науки и техники.

Учредителями конференции являются Российская академия наук и Суперкомпьютерный консорциум университетов России.

В первый день работы конференции будет объявлена **20-я редакция списка Top50** самых мощных компьютеров СНГ.

**Тематика конференции** охватывает следующие основные направления (но не ограничивается ими):

- Технологии параллельных и распределенных вычислений
- Грид и облачные вычисления
- Перспективные многопроцессорные архитектуры
- Параллельные и распределенные системы баз данных
- Администрирование, мониторинг и тестирование многопроцессорных систем
- Вычислительная математика
- Вычислительная физика
- Вычислительная химия
- Гидро-газодинамика и теплообмен
- Высоконелинейные и быстротекущие процессы в задачах механики
- Биоинформатика
- Нанотехнологии
- Климат и экология
- Криптография
- Визуализация
- Компьютерная алгебра
- Суперкомпьютерные научно-образовательные центры.

В рамках конференции будет работать **индустриальная сессия**. На индустриальную сессию принимаются высококачественные презентации по инновационному коммерческому аппаратному и программному обеспечению, ориентированному на применение суперкомпьютерных и параллельных вычислительных технологий в различных областях науки и техники.

**Официальный сайт конференции: <http://ПаВТ.РФ>**

**ЖУРНАЛ В ЖУРНАЛЕ**



**НЕЙРОСЕТЕВЫЕ  
ТЕХНОЛОГИИ**

**№ 11**

**НОЯБРЬ**

**2013**

**Главный редактор:**

ГАЛУШКИН А. И.

**Редакционная коллегия:**

АВЕДЬЯН Э. Д.  
БАЗИЯН Б. Х.  
БЕНЕВОЛЕНСКИЙ С. Б.  
БОРИСОВ В. В.  
ГОРБАЧЕНКО В. И.  
ЖДАНОВ А. А.  
ЗЕФИРОВ Н. С.  
ЗОЗУЛЯ Ю. И.  
КРИЖИЖАНОВСКИЙ Б. В.  
КУДРЯВЦЕВ В. Б.  
КУЛИК С. Д.  
КУРАВСКИЙ Л. С.  
РЕДЬКО В. Г.  
РУДИНСКИЙ А. В.  
СИМОРОВ С. Н.  
ФЕДУЛОВ А. С.  
ЧЕРВЯКОВ Н. И.

**Иностранные  
члены редколлегии:**

БОЯНОВ К.  
ВЕЛИЧКОВСКИЙ Б. М.  
ГРАБАРЧУК В.  
РУТКОВСКИЙ Л.

**Редакция:**

БЕЗМЕНОВА М. Ю.  
ГРИГОРИН-РЯБОВА Е. В.  
ЛЫСЕНКО А. В.  
ЧУГУНОВА А. В.

**Гриняк В. М., Девятисильный А. С.**

Нейронечеткая обучаемая система распознавания воздушных объектов . . . . . 58

**Горбатков С. А., Белолипец И. И., Солнцев А. М.**

О состоятельности процедур регуляризации нейросетевых моделей налогового и финансового контроля . . . . . 64

**В. М. Гриняк**, канд. техн. наук, зав. каф.,  
e-mail: Viktor.Grinyak@vvsu.ru,  
Владивостокский университет  
экономики и сервиса,

**А. С. Девятисильный**,  
д-р техн. наук, проф., гл. науч. сотр.,  
Институт автоматизации и процессов управления  
ДВО РАН, г. Владивосток

## Нейронечеткая обучаемая система распознавания воздушных объектов\*

*Рассматривается проблема идентификации воздушных объектов информационно-измерительными системами, образуемыми на основе двухкоординатных РЛС кругового обзора. В основу предлагаемого подхода к решению задачи положена оценка высоты наблюдаемого объекта по измерениям дальности и азимута. В дальнейшем полученные оценки обрабатываются нейронечеткой системой, определяющей степень принадлежности объекта к классу воздушных. Описана конфигурация нейронечеткой сети, даются рекомендации по ее обучению. С помощью компьютерного моделирования показана конструктивность предлагаемого подхода для типичных ситуаций.*

**Ключевые слова:** управление движением судов, воздушный объект, радар, измерение, высота объекта, нейронечеткая система

### Введение

Современные береговые системы управления движением судов (СУДС) представляют собой наукоемкие и сложные в техническом и организационном отношении предприятия [1, 2], основное назначение которых — обеспечение безопасности движения в зоне ответственности. Информационной базой современных СУДС, обеспечивающей их навигационные функции, являются двухкоординатные радиолокационные станции (РЛС) кругового обзора, дополняемые средствами спутниковой навигации — транспондерами автоматической идентификационной системы (АИС).

Задача заблаговременного распознавания опасно сближающихся судов (одна из центральных функций СУДС) оформилась в настоящее время как особый раздел науки об управлении [3, 4]. Методологической основой распознавания опасного сближения судов является оценка параметров траекто-

рии движения каждого судна (координат, скоростей и т. д.) и их экстраполяция. Если суда идентифицированы как опасно сближающиеся, система управления движением генерирует тревожный сигнал и рекомендации по изменению траектории движения.

Присутствие над оживленной акваторией в зоне ответственности СУДС маловысотных низкоскоростных воздушных объектов (вертолетов) может в корне исказить представления о навигационной обстановке. Суть проблемы в том, что ошибочное заключение судоводителя или оператора СУДС о воздушной цели как о морской (когда их скорости движения сравнимы) способно привести к генерации ложной тревоги и ошибочным управленческим решениям. Обозначенная проблема частично решается применением АИС на воздушном объекте (информация АИС позволяет, в том числе, однозначно идентифицировать тип цели). Вместе с тем, транспондерами АИС оснащаются далеко не все воздушные объекты, допускающие полет над акваторией, что требует решения задачи селекции воздушных объектов путем расширения навигационных функций систем, образуемых на основе двухкоординатных радаров.

Проблема трехкоординатного наблюдения воздушных объектов двухкоординатными измерителями неоднократно привлекала внимание исследователей [5—13]. Так, в работах [5, 6] рассматривается способ определения высоты объекта по измерениям его дальности и азимута несколькими РЛС путем решения геометрической задачи.

В работе [7] была показана принципиальная возможность (хотя и с ограниченным эффектом) решения трехкоординатной задачи при использовании даже одного двухкоординатного радара. Сущность предлагаемой авторами методики состоит в использовании линейного динамического алгоритма оптимального оценивания, обрабатывающего измерение дальности и азимута объекта последовательно, по мере их поступления и основанном на дискретном фильтре Калмана [8, 9]. Если в системе имеется несколько радаров, то после отдельной обработки измерений каждой РЛС проводится дополнительная совместная обработка полученных оценок высоты объекта. При этом используется прямоугольная декартова система координат.

В работе [10] предложен метод определения высоты объекта по измерениям одного двухкоординатного радара путем совместной обработки методом наименьших квадратов измерений дальности и азимута, полученных на конечном интервале времени. Продемонстрировано качество решения задачи в зависимости от дальности объекта и точности измерений, а также даны теоретические оценки границ применимости метода.

\*Работа выполнена в рамках Государственного задания высшим учебным заведениям в части проведения научно-исследовательских работ, проект № 7.2104.2011

В работах [11, 12] показан результат решения задачи оценки высоты объекта при переходе к многопозиционному наблюдению (два или несколько радаров). Авторы доказывают перспективность совместной оценки координат объектов методом наименьших квадратов в сферической системе  $\varphi, \lambda, R$  — соответственно географические широта, долгота и расстояние от центра Земли до объекта. Использование сферической системы координат позволяет избежать проблемы сингулярности задачи в области малых высот и повысить устойчивость оценки высоты объекта. В работе [12] предлагаемый метод сравнивается с известным способом [7], демонстрируется большая эффективность метода [12].

Общей проблемой всех известных способов определения высоты наблюдаемого объекта двухкоординатными измерителями является нерегулярность оценок высоты. В работе [13] авторы предлагают дополнить задачу оценки высоты оценкой высотного диапазона объектов, используя идею алгоритмов с обучением. Похожий путь предлагается и в настоящей работе: исследуется возможность создания на базе двухкоординатных радаров информационной системы, обеспечивающей достоверную селекцию воздушных объектов с использованием идей, положенных в основу нейронечетких систем.

### Модельные представления и постановка задачи

Рассмотрим сферическую систему координат  $\varphi, \lambda, R$  (с учетом пространственной локальности рассматриваемой задачи за модель поверхности Земли принимается сфера).

Особенностью внешнего наблюдения, осуществляемого с помощью радаров, является отсутствие возможности непосредственного измерения сил и моментов, обуславливающих движение объекта. Поэтому при описании эволюции координат наблюдаемых объектов традиционно обращаются к кинематическим моделям следующего полиномиального вида:

$$\begin{aligned} \varphi_{k+1} &= \varphi_k + \sum_{i=1}^{n_\varphi} a_i^\varphi(k) T^i; \\ \lambda_{k+1} &= \lambda_k + \sum_{i=1}^{n_\lambda} a_i^\lambda(k) T^i; \\ R_{k+1} &= R_k + \sum_{i=1}^{n_R} a_i^R(k) T^i; \\ k &= \overline{1, m}, \end{aligned} \quad (1)$$

где  $\varphi_k, \lambda_k, R_k$  — значения соответствующих координат объекта в момент времени  $t_k$ ;  $n_\varphi, n_\lambda, n_R$  — мак-

симальные значения степеней соответствующих полиномов;  $a_i^\varphi(k), a_i^\lambda(k), a_i^R(k)$  — коэффициенты полиномов, отождествляемые со скоростями изменения соответствующих координат и функциями от их более старших производных;  $T = t_{k+1} - t_k, t_k \in [t_1, t_m]$ .

Информационная ситуация, обеспечиваемая сетью из  $L$  радаров, описывается моделью вида

$$z_k^{(j)} = \begin{bmatrix} r^{(j)}(k) \\ \psi^{(j)}(k) \end{bmatrix} + \begin{bmatrix} \xi_r^{(j)}(k) \\ \xi_\psi^{(j)}(k) \end{bmatrix}, \quad (2)$$

где  $z_k^{(j)}$  — вектор  $k$ -го измерения  $j$ -й станцией;  $r^{(j)}(k)$  — дальность от объекта до  $j$ -й станции в момент времени  $t_k^{(j)}$  (время  $k$ -го измерения  $j$ -й станцией);  $\psi^{(j)}(k)$  — азимут объекта по отношению к  $j$ -й станции в момент времени  $t_k^{(j)}$ ;  $t_{k+1}^{(j)} - t_k^{(j)} = T^{(j)}$ ;  $T^{(j)}$  — период вращения  $j$ -й станции;  $\xi_r^{(j)}(k), \xi_\psi^{(j)}(k)$  — инструментальные измерительные погрешности, причем  $M[\xi_r^{(j)}(k)] = 0, M[\xi_r^{(j)}(k), \xi_r^{(i)}(m)] = D_r^{(j)} \delta_{ji} \delta_{km}, M[\xi_\psi^{(j)}(k)] = 0, M[\xi_\psi^{(j)}(k), \xi_\psi^{(i)}(m)] = D_\psi^{(j)} \delta_{ji} \delta_{km}; j = \overline{1, L}; M[*]$  — оператор математического ожидания,  $\delta_{ij}$  — символ Кронекера.

В свете указанных модельных представлений может быть поставлена обратная траекторная задача, описываемая уравнениями (1) и (2), целью решения которой является определение  $u$ -мерного вектора  $s_k = (\varphi_k, a_1^\varphi(k), \dots, a_{n_\varphi}^\varphi(k), \lambda_k, a_1^\lambda(k), \dots, a_{n_\lambda}^\lambda(k), R_k, a_1^R(k), \dots, a_{n_R}^R(k))^T$  по измерениям  $z_k^{(j)}, j = \overline{1, L}, u = \dim s_k; ^T$  — символ транспонирования.

### Метод решения задачи

Общим методом решения таких обратных задач является их линеаризация около некоторого опорного решения, характеризующего априорные представления о движении объекта. Допуская наличие опорного решения, будем говорить о сведении исходной задачи к задаче "в малом" с искомым вектором

$$\begin{aligned} \delta s_k &= (\delta \varphi_k, \delta a_1^\varphi(k), \dots, \delta a_{n_\varphi}^\varphi(k), \delta \lambda_k, \delta a_1^\lambda(k), \dots, \\ &\delta a_{n_\lambda}^\lambda(k), \delta R_k, \delta a_1^R(k), \dots, \delta a_{n_R}^R(k))^T, \end{aligned}$$

где  $\delta s_k$  — вектор погрешностей априорных представлений. Линеаризация исходной задачи (1), (2) приводит ее к следующему виду "состояние—измерение":

$$\begin{aligned} \delta s_{k+1} &= A_k \delta s_k + q_k; \\ \delta z_k^{(j)} &= H_k \delta s_k + \xi_k^{(j)}, \\ j &= \overline{1, L}, \end{aligned} \quad (3)$$

где  $q_k$  — вектор немоделируемых параметров движения;  $A, H$  — матричные коэффициенты (матрицы частных производных) с размерностью соответственно  $(u \times u)$  и  $(2 \times u)$ . Преобразование уравнений (3) к конечномерному виду, характерному для задач метода наименьших квадратов, приводит исходную задачу к модели

$$\delta Z = \tilde{H} \delta s_i + \tilde{q}, \quad (4)$$

где  $\delta Z$  — полный вектор измерений на интервале наблюдения;  $\delta s_i$  — вектор погрешностей априорных представлений в момент времени  $t_i$ ;  $\tilde{q}$  — вектор

приведенных погрешностей измерений;  $\tilde{H}$  — матричный коэффициент размерности  $N \times \dim s_i$ , являющийся композицией матриц  $A$  и  $H$ ;  $N$  — общее число обрабатываемых измерений (от всех станций).

Несмотря на то, что при  $a_1^\phi(i)$  и  $a_1^\lambda(k)$ , не равных одновременно нулю, система (4) не вырождена уже для одного радара ( $L = 1$ ), а при наличии в системе нескольких радаров ( $L > 1$ ) задача в принципе разрешима при любых возможных траекториях движения наблюдаемого объекта [10], для обеспечения практической разрешимости задачи необ-

ходимо ограничить размерность задачи (1), (2) так, чтобы движение объекта описывалось полиномами первой степени для угловых компонент и нулевой степенью для радиальной (т. е.  $n_\phi = 1, n_\lambda = 1, n_R = 0$ ,  $s_i = (\phi_i, a_1^\phi(i), \lambda_i, a_1^\lambda(i), R_i)^T$ ). Это соответствует движению объектов на постоянной высоте без маневрирования на интервале наблюдения.

Характерным свойством рассматриваемой задачи (1), (2) является нерегулярность оценок радиальной координаты (т. е. высоты) маловысотных удаленных объектов, что связано с плохой обусловленностью системы (4), исходной нелинейностью задачи и конечной точностью измерений [10, 12, 13]. Эта особенность задачи продемонстрирована на рис. 1, на котором приведена оценка высоты надводного объекта (рис. 1, а) и воздушных объектов, движущихся на высоте 100 м (рис. 1, б) и 200 м (рис. 1, в) для случая двух РЛС, измеряющих дальность с погрешностью  $\pm 5$  м и азимут с погрешностью  $\pm 0,1^\circ$ . Видно, что начиная с некоторого расстояния от системы радаров, воздушный объект становится (по оценке высоты) неотличимым от морского: в данном случае это 5 км для объекта с высотой 100 м и 9 км для объекта с высотой 200 м. Сами оценки высоты носят "изрезанный" характер со случайными выбросами. Такая картина является побудительным мотивом, наряду с оцениванием собственно высоты объекта, определять дополнительно еще и "высотный класс" объекта, т. е. диапазон высот, которому принадлежит траектория объекта. В рамках данной работы возможные диапазоны высот ограничены понятиями "морской" и "воздушный". При таком взгляде на проблему оказываются продуктивными идеи обучения, отождествляемые в настоящее время с понятием искусственных нейронных сетей.

Пусть  $\hat{h}_i = \hat{R}_i - R_3$  — оценка высоты объекта

над уровнем моря ( $\hat{R}_i$  — оценка радиальной компоненты вектора  $s_i$ ;  $R_3$  — радиус Земли на уровне моря). С учетом особенности задачи будем считать, что основными информативными признаками, дающими представление о "высотном классе" объекта, являются оценка его высоты и сравнительный характер (степень "изрезанности", "нерегулярности") оценок высоты в различные моменты времени  $t_i$ . Введем лингвистическую переменную  $P_h$  "оценка высоты объекта" с термами "большая" и "малая" и функциями принадлежности типа "дополнение":

$$\mu_{\text{малая}}(h) = 1 - \frac{1}{1 + \exp(-a_h(h - c_h))};$$

$$\mu_{\text{большая}}(h) = \frac{1}{1 + \exp(-a_h(h - c_h))}.$$

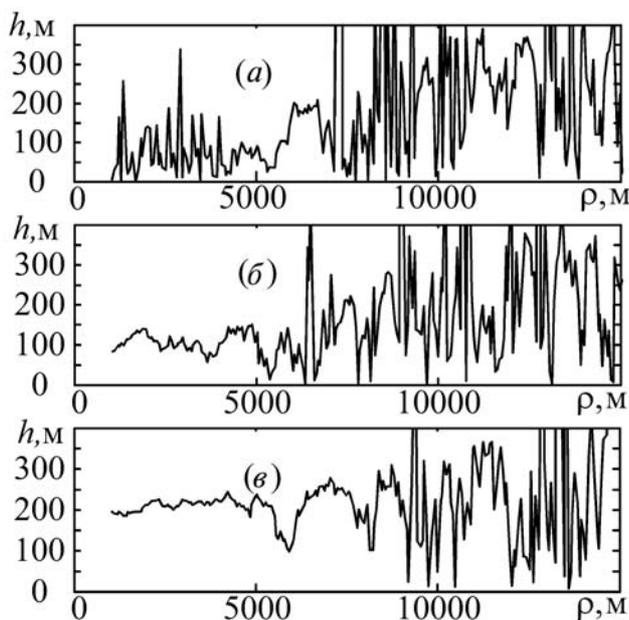


Рис. 1. Оценка высоты объекта по мере удаления от радаров. Здесь  $\rho$  — расстояние от объекта до системы радаров

Пусть  $\Delta_i = 2|\hat{h}_i - \hat{h}_{i-1}| / |\hat{h}_i + \hat{h}_{i-1}|$  — относительная разность между соседними оценками высоты. Введем лингвистическую переменную  $P_\Delta$  "разность соседних оценок высоты объекта" с термами "большая" и "малая" и функциями принадлежности термов типа "дополнение":

$$\lambda_{\text{малая}}(\Delta) = 1 - \frac{1}{1 + \exp(-a_\Delta(\Delta - c_\Delta))};$$

$$\lambda_{\text{большая}}(\Delta) = \frac{1}{1 + \exp(-a_\Delta(\Delta - c_\Delta))}.$$

Величины  $\hat{h}_i$  и  $\Delta_i$  (вход) обрабатываются нейронечеткой сетью, показанной на рис. 2, на выходе которой формируется числовое значение  $u_i$  — степень принадлежности наблюдаемого объекта к диапазону высот "воздушный" в момент времени  $t_i$  (считается, что  $u_i = 0$  для надводных объектов и  $u_i = 1$  для воздушных объектов). Сеть состоит из пяти слоев [14].

В узлах первого слоя  $\mu_1, \mu_2, \lambda_1, \lambda_2$  вычисляются значения функций принадлежности  $\mu_{\text{малая}}, \mu_{\text{большая}}, \lambda_{\text{малая}}, \lambda_{\text{большая}}$  соответственно.

Узлы  $\Pi$  второго слоя (четыре узла) соответствуют посылкам четырех возможных нечетких правил:

- $P_h = \text{"малая"} \text{ И } P_\Delta = \text{"малая"};$
- $P_h = \text{"малая"} \text{ И } P_\Delta = \text{"большая"};$
- $P_h = \text{"большая"} \text{ И } P_\Delta = \text{"малая"};$
- $P_h = \text{"большая"} \text{ И } P_\Delta = \text{"большая"}.$

Каждый узел второго слоя соединен с теми узлами первого слоя, которые формируют посылки соответствующего правила. Выходом каждого узла второго слоя является степень выполнения  $j$ -го правила  $\tau_j$ , которая рассчитывается как произведение входных сигналов.

Узлы  $N$  третьего слоя (четыре узла) рассчитывают относительную степень выполнения каждого нечеткого правила по формуле

$$\tau_j^* = \frac{\tau_j}{\sum_{k=1}^4 \tau_k}.$$

Узлы  $b_1, b_2, b_3, b_4$  четвертого слоя формируют заключения нечетких правил. Каждый узел соединен с одним узлом третьего слоя и рассчитывает вклад одного нечеткого правила в выход сети по формуле  $y_j = b_j \tau_j^*$ .

Единственный узел пятого слоя агрегирует результат, полученный по разным правилам, суммируя вклады всех правил  $u_i = \sum_{k=1}^4 y_j$ .

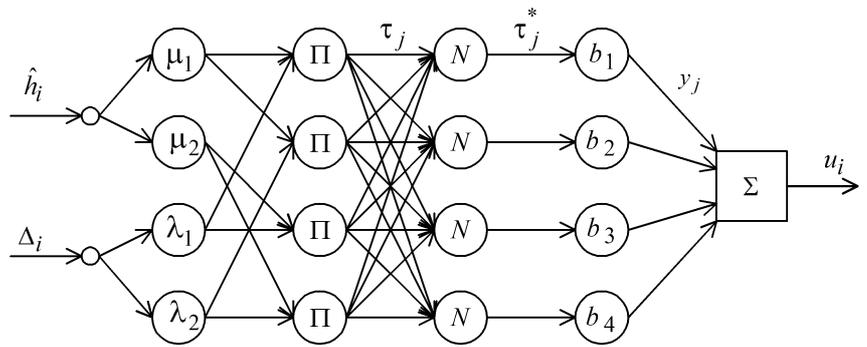


Рис. 2. Схема нейронечеткой сети, распознающей воздушные объекты

## Обучение системы

Обучение нейронечеткой сети (см. рис. 2) состоит в настройке коэффициентов функций принадлежности  $a_h, c_h, a_\Delta, c_\Delta$  и коэффициентов  $b_1, b_2, b_3, b_4$  узлов четвертого слоя. Обучение может быть проведено с применением трех различных стратегий.

**Стратегия 1.** Обучение полностью экспертным способом. В этом случае все коэффициенты назначаются экспертом.

**Стратегия 2.** Обучение на обучающей выборке с экспертным формированием заключений нечетких правил. В этом случае коэффициенты  $b_1, b_2, b_3, b_4$  назначаются экспертом, а коэффициенты  $a_h, c_h, a_\Delta, c_\Delta$  определяются настройкой системы на обучающей выборке.

**Стратегия 3.** Обучение полностью на обучающей выборке. В этом случае все коэффициенты системы определяются настройкой на обучающей выборке.

Обучающая выборка формируется следующим образом. Моделируется решение задачи (1), (2) при

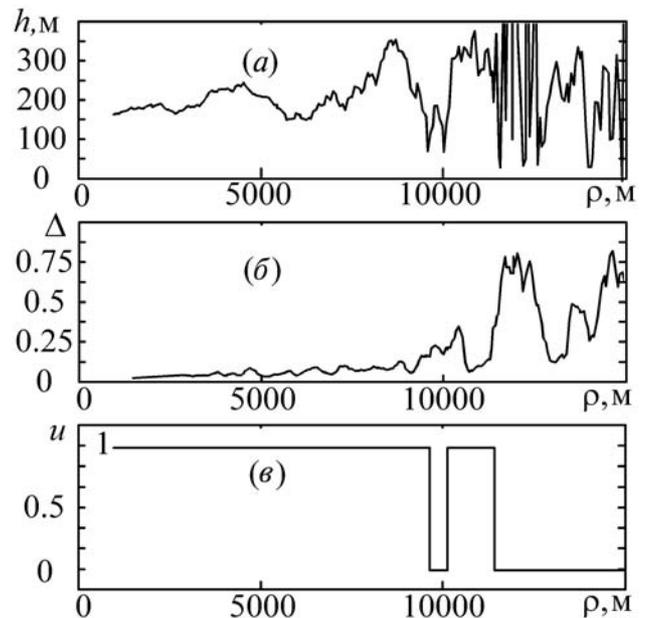


Рис. 3. Пример формирования обучающей выборки из оценок высоты, относительной разности соседних оценок высоты (вход) и степени принадлежности объекта к диапазону высот "воздушный" (выход)

движении объекта на различных высотах, в том числе и при движении надводного объекта. В результате получаются оценки высоты объекта типа изображенных на рис. 1, формирующие входные данные обучающей выборки. Соответствующие им выходные данные обучающей выборки формируются экспертом: если характер оценки высоты объекта дает возможность отличить его от надводного, считается, что система выдает значение  $u_i = 1$ , значение  $u_i = 0$  в противном случае. На рис. 3 показан пример формирования фрагмента такой обучающей выборки.

На рис. 3, *a* — оценка высоты наблюдаемого объекта по мере его удаления от системы радаров (движение объекта моделируется на высоте 200 м); на рис. 3, *b* — относительная разность соседних оценок высоты; 3, *в* — решение эксперта о возможности выделить объект как воздушный: оценки высоты позволяют устойчиво сделать это до дальности приблизительно 10 км ( $u_i = 1$ ), далее следует короткая зона неопределенности ( $u_i = 0$ ), после чего до дальности приблизительно 11,5 км объект снова может быть отнесен к воздушным ( $u_i = 1$ ), после чего следует сплошная зона неопределенности ( $u_i = 0$ ).

Накапливая данные для различных высот движения объекта и множества возможных траекторий, формируют общую обучающую выборку, на базе которой и обучают нейронечеткую сеть (см. рис. 2) в рамках стратегии 2 или стратегии 3, пользуясь известными методами обучения сетей такого типа [14, 15].

### Результаты численного моделирования

При моделировании задачи было принято, что информационной базой СУДС являются два радара кругового обзора (например, типа Raytheon), находящихся на расстоянии 5 км друг от друга, с периодом обращения 3 с и погрешностями измерений угла и дальности соответственно  $\xi_{\psi}^{(j)}(k) \in [-0,1^\circ, 0,1^\circ]$ ,  $\xi_r^{(j)}(k) \in [-5 \text{ м}, 5 \text{ м}]$ . Число измерений  $m$  от каждой

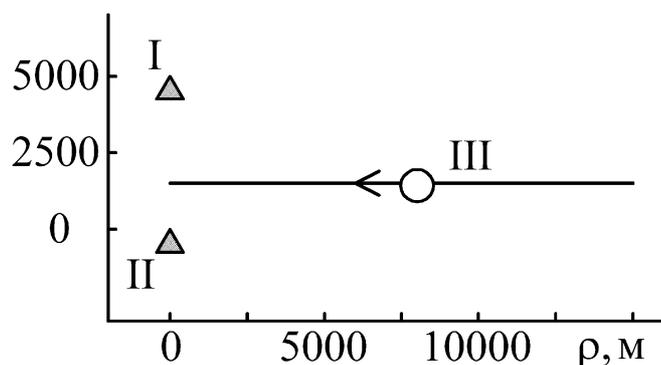


Рис. 4. Моделируемая конфигурация системы двух радаров и траектория движения объекта

станции было принято равным  $m = 10$  и  $m = 20$  (т. е. измерения набираются в течение 30 с и 1 мин).

Обучение системы проводилось в рамках стратегии 3, объем обучающей выборки составил около 10 000 значений "вход—выход", полученных при моделировании движения объекта по различным траекториям. При этом параметры функций принадлежности приняты следующие значения:  $a_h = 0,11$ ,  $c_h = 20,21$ ,  $a_{\Delta} = 2,51$ ,  $c_{\Delta} = 0,43$ ; коэффициенты узлов четвертого слоя приняли следующие значения:  $b_1 = 0,62$ ,  $b_2 = -0,05$ ,  $b_3 = 1,02$ ,  $b_4 = 0,16$ .

На рис. 4 показана траектория движения воздушного объекта, моделируемая для демонстрации решения задачи распознавания воздушных объектов с помощью предварительно обученной нейронечеткой сети (см. рис. 2). Здесь I и II — радиолокационные станции, III — траектория объекта. Объект движется издалека по прямой со скоростью 20 м/с, приближаясь к РЛС. Здесь  $\rho$  — расстояние от объекта до линии, соединяющей радиолокационные станции.

На рис. 5 показаны результаты решения задачи оценки высоты объекта (левая колонка рисунков) и оценки его высотного диапазона нечеткой системой (правая колонка рисунков). Здесь  $h$  — высота объекта,  $u$  — степень принадлежности объекта к диапазону высот "воздушный". Задача моделирова-

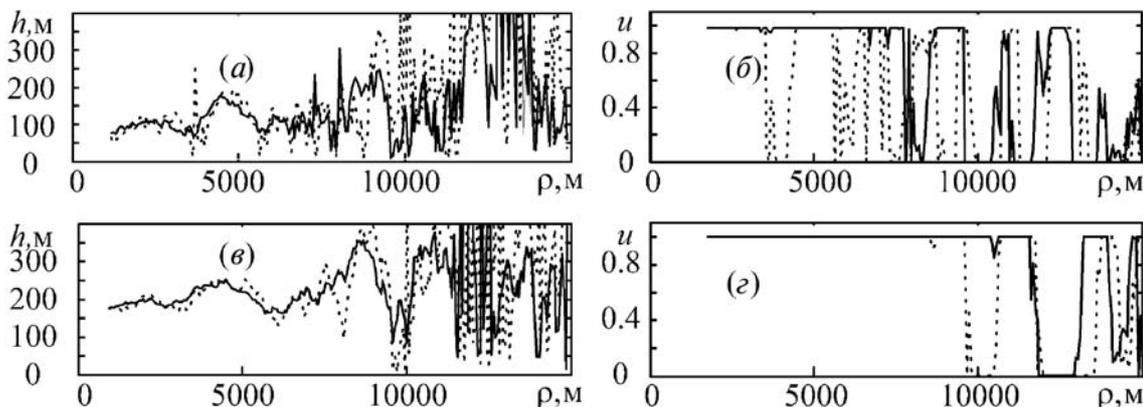


Рис. 5. Результат решения задачи

лась для объектов, движущихся на высоте 100 м (рис. 5, а и б) и 200 м (рис. 5, в и г). Сплошные графики соответствуют числу измерений  $m = 20$ , точки — числу измерений  $m = 10$ . Из рисунка видно, что, например, уверенное выделение воздушного объекта, движущегося на высоте 100 м возможно до дальности  $\approx 3000$  м при  $m = 10$  и до дальности  $\approx 7000$  м при  $m = 20$  (рис. 5, б). Для объекта, движущегося на высоте 200 м, выделение его как воздушного возможно до дальности  $\approx 9000$  м при  $m = 10$  и до дальности  $\approx 12\,000$  м при  $m = 20$  (рис. 5, г). Такие дальности (по сути — границы применимости метода) вполне соответствуют размерам зон ответственности в акваториях морских портов, что позволяет говорить о пригодности предлагаемого метода селекции воздушных объектов для судоводительской практики.

### Заключение

Перечислим основные результаты работы. В статье обозначена проблема генерации ложных тревог при управлении коллективным движением судов, связанная с присутствием над акваторией воздушных объектов (вертолетов). Для корректной обработки СУДС таких объектов необходимо их идентифицировать. Алгоритм идентификации основан на вычислении высоты объекта по результатам измерений его дальности и азимута системой двухкоординатных РЛС и обработке полученных данных нейронечеткой сетью. Предлагаемый алгоритм позволяет принять решение о степени принадлежности объекта к классу воздушных. В статье продемонстрированы границы применимости предлагаемой методики. В целом на основании анализа представленных данных можно сделать вывод о конструктивной разрешимости рассматриваемой задачи распознавания воздушных объектов. Результаты

работы ориентированы на расширение навигационных функций современных систем управления движением судов.

### Список литературы

1. **Huges T.** When is a VTS is not a VTS // The J. of Navigation. 2009. V. 62, N 3. P. 439—442.
2. **ОАО Норфес.** [Электронный ресурс]. URL: <http://www.norfes.ru/>
3. **Tam Ch. K., Bucknall R., Greig A.** Review of Collision Avoidance and Path Planning Methods for Ships in Close Range Encounters // The J. of Navigation. 2009. V. 62, N 2. P. 455—476.
4. **Астреин В. В.** Системы предупреждения столкновения судов, тенденции развития (к 40-летию МППСС-72) // Вестник Астраханского государственного технического университета. Сер. "Морская техника и технология". 2012. № 1. С. 7—17.
5. **Berle F. J.** Multy radar tracking and multy sensor trackig in air defebse systems // Electronic Technologies. 1984. V. 28, N 4.
6. **Hudel P.** Dreidimensional arbeitendes Radarsystem: Patent DE 4123898 A1 (Патент DE 4123898 A1, 18.07.91 г. Трехкоординатная радиолокационная система, ИСМ 85-08—94 г.).
7. **Nabaa N., Bishop R. H.** Estimate Fusion for 2D Search Sensors // AIAA Guidance, Navigation and Control Proceedings, 1995. V. 1. P. 677—684.
8. **Кузьмин С. З.** Основы теории цифровой обработки радиолокационной информации. М.: Сов. радио, 1974. 431 с.
9. **Фарина А., Студер Ф.** Цифровая обработка радиолокационной информации. Сопровождение целей. М.: Радио и связь, 1993. 320 с.
10. **Гриняк В. М.** Исследование пространственной задачи навигации в условиях неполной измерительной информации // Дальневосточный математический журнал. 2000. Т. 1, № 1. С. 93—101.
11. **Девятисильный А. С., Дорожко В. М., Гриняк В. М.** Способ распознавания удаленных воздушных объектов. Патент № 2206104 // Б.И. 2003. № 16.
12. **Девятисильный А. С., Дорожко В. М., Гриняк В. М.** Идентификация воздушных объектов двухкоординатными измерителями // Измерительная техника. 2004. № 11. С. 19—21.
13. **Девятисильный А. С., Дорожко В. М., Гриняк В. М.** Нейроподобные алгоритмы высотной классификации воздушных объектов // Информационные технологии. 2001. № 12. С. 45—51.
14. **Штовба С. Д.** Проектирование нечетких систем средствами MatLab. М.: Горячая линия телеком. 2007. 288 с.
15. **Nauk D., Klawonn F., Kruse R.** Foundations of Neuro-Fuzzy Systems. New York: John Wiley & Sons. 1997. 305 p.

## ИНФОРМАЦИЯ

Теоретический и прикладной научно-технический журнал

# ПРОГРАММНАЯ ИНЖЕНЕРИЯ

ISSN 2220-3397

В журнале освещаются состояние и тенденции развития основных направлений индустрии программного обеспечения, связанных с проектированием, конструированием, архитектурой, обеспечением качества и сопровождением жизненного цикла программного обеспечения, а также рассматриваются достижения в области создания и эксплуатации прикладных программно-информационных систем во всех областях человеческой деятельности.

Журнал распространяется только по подписке.

Журнал входит в перечень периодических изданий, рекомендованных ВАК РФ для публикации основных результатов диссертаций на соискание степеней кандидата и доктора наук, а также в базу данных Российского индекса научного цитирования (РИНЦ).

Оформить подписку можно через подписные Агентства или непосредственно в редакции журнала.

Подписные индексы по каталогам:

"Роспечать" — 22765; "Пресса России" — 39795

107076, Москва, Стромьинский пер., 4

Тел./факс: (499) 269-55-10 e-mail: [prin@novtex.ru](mailto:prin@novtex.ru) <http://novtex.ru/pi.html>

**С. А. Горбатков**, д-р техн. наук, проф.,  
e-mail: sgorbatkov@mail.ru,

**И. И. Белолипец**, преподаватель, e-mail:  
red7315@gmail.com,

**А. М. Солнцев**, преподаватель,  
e-mail: i0280@r02.nalog.ru,

Уфимский филиал Финуниверситета

## О состоятельности процедур регуляризации нейросетевых моделей налогового и финансового контроля

*Предложен способ повышения однородности и информативности данных для улучшения качества обучения и прогностических свойств нейросетевых моделей в задачах с сильным зашумлением данных. Вычислительным экспериментом доказана необходимость и состоятельность предлагаемых процедур предобработки данных.*

**Ключевые слова:** нейросеть, регуляризация обучения, противоречивые данные, устойчивость, искаженные данные

### Введение

Предметом исследования в работе является построение нейросетевых моделей, которые могут быть использованы для решения широкого круга прикладных экономических задач. Примером такого типа задач являются задачи налогового и финансового администрирования. Целью данной работы является повышение качества и эффективности нейросетевых моделей налогового и финансового контроля в сложных условиях моделирования, характерных для ряда экономических объектов. Под эффективностью модели в данном случае понимается повышение достоверности выводов, полученных на основе моделирования, что, в свою очередь, приводит к росту экономической эффективности от принятия научно обоснованных и достаточно достоверных управленческих решений. Для достижения поставленной цели необходимо обеспечить хорошие прогностические свойства и добиться устойчивости модели к возмущению входных данных. Для задач налогового и финансового контроля характерны достаточно сложные условия моделирования. Отметим основные трудности, характерные для задач подобного класса:

- данные бухгалтерской отчетности могут быть намеренно искажены в целях уменьшения налогооблагаемой базы либо в целях маскировки истинного финансового состояния хозяйствующего субъекта;

- данные могут быть достаточно сильно зашумлены, при этом закон распределения шумов неизвестен и может отличаться от нормального; как показали ранее проведенные исследования, около 35...40 % данных являются искаженными, при этом искажаются не все входные факторы, а только некоторые из них [1];

- как правило, входные факторы связаны между собой и с моделируемой величиной существенно нелинейными зависимостями.

В связи с этим изучение круга проблем, связанных с построением подобных моделей, в указанных сложных условиях моделирования является актуальной задачей. Сформулируем задачи, решаемые в рамках данной работы:

- повышение однородности и информативности данных для улучшения качества обучения и прогностических свойств нейросетевых моделей;
- исследование условий устойчивости нейросетевого отображения в сложных условиях моделирования.

Вопрос об обеспечении устойчивости нейросетевых моделей исследовал С. А. Шумский [2], который сформулировал основные положения регуляризации нейросетевых моделей на основе байесовского подхода. Однако в работе С. А. Шумского предполагается, что закон распределения шумов (искажений) в данных является нормальным, что не всегда справедливо для рассматриваемого класса задач. Это требует разработки новых подходов к обеспечению устойчивости нейросетевых моделей, которые были бы нечувствительны к виду закона распределения шумов.

Ранее авторами был разработан и опубликован [1, 3, 4] метод построения нейросетевых моделей налогового контроля, получивший название метода вложенных математических моделей (МВММ-метод), включающий в себя ряд оригинальных процедур и алгоритмов.

В данной статье подробно рассматриваются две процедуры, входящие в МВММ: процедуры удаления противоречивых и аномальных наблюдений. В отличие от классических подходов к обеспечению однородности данных предлагается совершенно иная идеология: процедуры предобработки данных должны проводиться не изолированно, а взаимосвязанно с обучением нейросетей. Процесс повышения однородности и информативности данных является итерационным и продолжается до тех пор, пока не будут достигнуты приемлемые значения критериев качества обучения.

### Процедура удаления противоречивых данных

Предлагаемая процедура позволяет выявить самые грубые искажения данных до непосредственного обучения нейросети. Суть "противоречивости данных" в следующем: двум близким по некоторой

числовой мере вектор-строкам  $\mathbf{x}_\alpha$  и  $\mathbf{x}_\beta$  могут соответствовать существенно отличающиеся значения выходной величины  $y_\alpha$  и  $y_\beta$ . То есть пара вектор-строк  $\mathbf{x}_\alpha$  и  $\mathbf{x}_\beta$  "растягивают" будущее нейросетевое отображение модели, что ухудшает качество обучения сети и негативно влияет на устойчивость модели к изменению входных факторов. В качестве числовой меры для выявления локальных неоднородностей предлагается использовать аналог константы Липшица:

$$L_{\alpha,\beta} = \|y_\alpha - y_\beta\|_{E_n} / \|\mathbf{x}_\alpha - \mathbf{x}_\beta\|_{E_n}, \quad (1)$$

где  $\mathbf{x}_\alpha, \mathbf{x}_\beta$  — близкие по евклидовой норме вектор-строки входных факторов;  $y_\alpha, y_\beta$  — соответствующие им значения выходной величины (здесь и далее большими латинскими буквами  $X$  и  $Y$  обозначаются случайные входные и выходные величины модели, а малыми буквами  $x$  и  $y$  — их конкретные реализации в расчете или эксперименте);  $E_n$  —  $n$ -мерное евклидово пространство;  $\|\bullet\|$  — норма в  $E_n$ . Все записи исходной базы данных предварительно сортируются в порядке возрастания  $\|\mathbf{x}_j\|$ . Тогда, если  $L_{\alpha,\beta} > 1$ , то это значит, что двум близким по норме векторам  $\mathbf{x}_\alpha$  и  $\mathbf{x}_\beta$  соответствуют существенно различающиеся значения  $y_\alpha$  и  $y_\beta$ , и это может говорить о наличии локальной неоднородности в данных. Большие значения величины  $L_{\alpha,\beta}$  могут быть вызваны двумя причинами:

1) случай двух близких по норме векторов  $\mathbf{x}_\alpha$  и  $\mathbf{x}_\beta$  и большой нормы  $\|y_\alpha - y_\beta\|$  говорит либо о грубых ошибках измерения, либо о сознательном искажении данных;

2) норма  $\|\mathbf{x}_\alpha - \mathbf{x}_\beta\|$  близка к нулю, и даже при малых, но конечных значениях  $\|y_\alpha - y_\beta\|$ , получаются  $L_{\alpha,\beta} \gg 1$ .

Введем дополнительное условие: для пары близких по норме векторов  $\mathbf{x}_\alpha, \mathbf{x}_\beta$  должно выполняться соотношение

$$\|\mathbf{x}_\alpha - \mathbf{x}_\beta\| / \|\mathbf{x}_\alpha\| \geq \zeta, \quad \zeta > 0, \quad (2)$$

где  $\zeta$  — коэффициент запаса, малое число (например,  $10^{-3}$ ). Это условие исключает большой рост  $\|y_\alpha - y_\beta\|$  за счет очень малой разницы между сравниваемыми точками  $\mathbf{x}_\alpha, \mathbf{x}_\beta$  в факторном пространстве. При обнаружении пары вектор-строк, для которых  $L_{\alpha,\beta} > 1$ , без дополнительного исследования неясно, какая из них будет "растягивать" нейросетевое отображение при обучении сети. Если удалить из базы данных обе строки, то вместе с противоречивыми данными могут быть удалены и "хорошие" информативные наблюдения. Для рассматриваемого класса задач часто характерен дефицит данных, поэтому на этапе предобработки желательно сохранить как можно больше непротиворечивых наблюдений. Для решения этой задачи была разработана модификация предлагавшегося ранее алгоритма удаления противоречивых данных. Изложим ее подробно.

1. Для всех строк исходной базы данных  $\langle \mathbf{x}_i, \mathbf{y}_i \rangle$ ,  $i = \overline{1, N}$ , вычисляются евклидовы нормы векторов  $\mathbf{x}_i$

$$\|\mathbf{x}_i\| = \sqrt{\sum_{j=1}^n x_{ij}^2}, \quad i = \overline{1, N},$$

где  $j$  — номер компоненты вектора  $\mathbf{x}_i$ ;  $i$  — номер наблюдения.

2. Все вектор-строки базы данных сортируются в порядке возрастания  $\|\mathbf{x}_i\|$ :

$$\|\mathbf{x}_1\|, \|\mathbf{x}_2\|, \dots, \|\mathbf{x}_s\|, \dots, \|\mathbf{x}_{N-1}\|, \|\mathbf{x}_N\|, \quad s = \overline{1, N},$$

где  $s$  — номер члена проранжированного ряда.

3. Вычисляются нормы разностей

$$\|\mathbf{x}_1 - \mathbf{x}_2\|, \|\mathbf{x}_2 - \mathbf{x}_3\|, \dots, \|\mathbf{x}_s - \mathbf{x}_{s+1}\|, \dots, \|\mathbf{x}_{N-1} - \mathbf{x}_N\|.$$

4. Для каждой пары векторов  $\|\mathbf{x}_s\|$  и  $\|\mathbf{x}_{s+1}\|$  вычисляется величина  $L_{s,s+1}$  по критерию (1).

5. Проводится итерационный процесс поиска и удаления противоречивых наблюдений. Определяются все пары вектор-строк, для которых  $L_{s,s+1} > 1$ ,  $s = \overline{1, N_k - 1}$ , где  $N_k$  — число наблюдений в базе на  $k$ -й итерации.

6. Организуется вложенный цикл: для каждой пары вектор-строк с  $L_{s,s+1} > 1$  проверяется условие (2). Если оно не выполняется, переходим к следующей паре вектор-строк с  $L_{p,p+1} > 1$ . Если же условие (2) выполняется, вычисляется  $L_{s-1,s+1}$ . Если  $L_{s-1,s+1} < 1$ , то это означает, что для пары векторов с номерами  $s$  и  $s+1$  именно строка с номером  $s$  приводит к тому, что  $L_{s,s+1} > 1$ . Вектор-строка  $\mathbf{x}_s$  признается противоречивой, удаляется из базы данных и не участвует в дальнейшем обучении. И наоборот, если  $L_{s-1,s+1} > 1$ , то противоречивой, "растягивающей" является строка с номером  $s+1$ .

7. В конце каждой  $k$ -й итерации проверяется условие репрезентативности данных по правилу

$$N_k > \zeta n, \quad \zeta \approx (5 \dots 10), \quad (3)$$

где  $N_k$  — число оставшихся наблюдений на  $k$ -й итерации;  $\zeta$  — коэффициент запаса, зависящий от требуемого качества (ошибки обобщения) нейросети;  $n$  — число входных факторов. Если условие (3) не выполняется, фиксируется состояние, достигнутое на предыдущей  $(k-1)$ -й итерации, процесс удаления противоречивых точек останавливается.

8. Шаги 3...7 повторяются до тех пор, пока в базе данных не останется пар строк, для которых  $L_{\alpha,\beta} > 1$ , либо пока процедура не будет прервана по правилу (3).

В результате выполнения предложенной процедуры из базы данных удаляются сознательно искаженные наблюдения, которые могут негативно повлиять на качество обучения нейросети. Сформированная таким образом база данных  $\langle \mathbf{x}_i, \mathbf{y}_i \rangle$ ,  $i = \overline{1, N_0}$ , будет более однородной и обученная на ней нейросеть будет менее чувствительна к случайным изменениям входных факторов.

## Удаление аномальных наблюдений

В результате удаления противоречивых наблюдений из обучающей выборки исключаются самые явные и грубые локальные неоднородности. Но в процессе исследований выяснилось, что этого недостаточно для обеспечения необходимой ошибки обобщения и устойчивости модели. Поэтому на этапе обучения нейросетей проводится более тщательная "чистка" базы данных от наблюдений, препятствующих хорошему обучению. Такие наблюдения были названы авторами "аномальными". Вектор-строку  $\mathbf{x}_i$  будем считать аномальной, если значение  $y_i$  существенно отклоняется от аппроксимируемого сетью значения  $\hat{y}_i$ . Удаление аномальных наблюдений проводится итерационно по следующему правилу:

$$i^{*(k)}: \delta_{i^{*(k)}} = \left| \frac{y_{i^{*(k)}} - \hat{y}_{i^{*(k)}}}{y_{i^{*(k)}}} \right| \cdot 100 \% > \xi^{(k)}, \xi^{(k)} > 0, \\ k = 1, 2, \dots, \quad (4)$$

где  $k$  — номер итерации;  $i^*$  — номер удаляемой из данных вектор-строки;  $\xi^{(k)}$  — задаваемый уровень отбраковки; крышечкой сверху здесь и далее отмечены рассчитанные в нейросетевых моделях (НСМ) (прогнозные) значения величин. Параметр  $\xi^{(k)}$  на первой итерации составляет порядка сотен процентов, на второй итерации — порядка десятков процентов и т. д. Цель данной процедуры — сглаживание восстанавливаемой нейросетью поверхности функции  $\hat{y}_i = F(\mathbf{X}_i, W)$ . Смысл правила (4) состоит в том, что разные примеры в обучающем множестве имеют существенно различную информационную ценность вклада в модификацию синаптических весов сети, связанную с разной степенью зашумленности. Поэтому и удаляются аномальные, малоинформативные точки. Более того, эти точки искажают вклад в модификацию синаптических весов других (информативных) точек.

Качество обучения сети на каждой итерации оценивается по мультипликативному обобщенному критерию  $J$ , представляющему собой произведение частных критериев:

$$J = ES; \quad (5)$$

$$E = \|\hat{Y}_i - Y_i\| / \|Y_i\|; S = \frac{1}{N_{(k)}^{\Omega_{test}}} \sum_{i=1}^{N_{(k)}^{\Omega_{test}}} \frac{|\hat{y}_i - \hat{y}_{i+1}|}{\|\mathbf{x}_i - \mathbf{x}_{i+1}\|}; \\ i = \overline{1, N}; N \in \Omega_{test}. \quad (6)$$

Величина  $E$  в (5), (6) является ошибкой обобщения, рассчитанной на тестовом множестве  $\Omega_{test}$  наблюдений, не участвовавших в обучении НСМ, и характеризует точность и прогностические свойства сети. Величина  $S$  также вычисляется на тестовом множестве  $\Omega_{test}$  и характеризует устойчивость НСМ к вариации независимых переменных: чем меньше  $S$ , тем меньше разброс значений  $\hat{y}(\mathbf{x})$  на

новых наблюдениях после обучения нейросети. По мере удаления аномальных по (4) наблюдений улучшается и точность  $E$ , и устойчивость  $S$  модели. После каждой  $k$ -й итерации проверяется условие

$$J^k < J^{k-1} \quad (7)$$

Если это условие нарушается, то итерация останавливается.

Итерационный процесс (4)—(7), как правило, быстро сходится: за несколько итераций происходит уменьшение критерия  $J$  более чем на порядок. Процедура очистки данных от аномальных наблюдений останавливается в случае достижения желаемого значения  $J$  (порядка сотых долей) либо при нарушении условия репрезентативности данных [1].

## Результаты вычислительных экспериментов

Введем величины  $p_i \in [p_0; p_N]$ , где  $p_0 = 1, p_N = 7,8$ ,  $p_i = p_0 + i(p_N - p_0)/N$ , и  $q_i \in [q_0; q_N]$ , где  $q_0 = 4,6$ ,  $q_N = 1, q_i = q_0 + i(q_N - q_0)/N$ ,  $i = \overline{1, N}$ ,  $N = 200$ . Сконструируем 10 факторов:

$$X_1 = p + q; X_2 = pq; X_3 = 1/p - 1/q; X_4 = p/q; \\ X_5 = p - q; X_6 = (\sqrt{p} - \sqrt{q})^2; X_7 = \sqrt{p^2 + q^2}; \\ X_8 = (p^2 + q)/(p + q^2); X_9 = (p^3 + q^3)/(p^2 - pq + q^2); \\ X_{10} = (p^2 + q^2)/(\sqrt{p^3 - q^3}). \quad (8)$$

Зададим моделируемую детерминированную зависимость:

$$Y = \sum_{j=1}^m \text{th}(X_j), m = 10. \quad (10)$$

Вычислительный эксперимент проводился на модельных данных для того, чтобы знать неискаженные значения входных факторов (8)—(9) и выходной величины (10). Составим матрицу наблюдений  $A = \{a_{ij}\}$ . Здесь  $a_{ij} = x_{ij}$ ,  $i = \overline{1, N}$ ,  $j = \overline{1, m}$ ,  $a_{i, m+1} = y_i$ . В получившейся матрице строки перемешаем случайным образом. Первые 100 строк будут обучающим множеством, другие 100 — тестовым. Тестовое множество обозначим  $T = \{t_{ij}\}$ .

Здесь  $i = \overline{1, N_{test}}$ ,  $j = \overline{1, m+1}$ ,  $N_{test} = 100$ . Детерминированная зависимость (10) аппроксимировалась с помощью нейросети типа многослойный перцептрон с двумя скрытыми слоями и активационной функцией гиперболического тангенса. Сначала сеть была обучена на незашумленных данных, показатели качества обучения, вычисленные по формулам (4)—(5) следующие:  $E = 0,0013$ ;  $S = 0,01078$ ;  $J = 0,000014$ . Далее обученной на "эталонных" данных сети предъявлялись зашумленные данные для того, чтобы определить, как будут меняться показатели качества при разном уровне шума.

Подвергнем зашумлению тестовое множество исходной базы данных, состоящее из 100 вектор-строк. Вводить возмущение (шум) в данные будем, учитывая перечисленные во введении особенности, характерные для задач налогового и финансового контроля. Зашумлению будем подвергать не все столбцы (факторы), а только  $X_1, X_3, X_5, X_7, X_9, Y$ . Столбцы  $X_2, X_4, X_6, X_8, X_{10}$  останутся не зашумленными. Выбор зашумленных и незашумленных факторов случаен. При этом для факторов  $X_1, X_5, X_9$  и моделируемой величины  $Y$  был выбран нормальный закон распределения шумов, а для факторов  $X_3, X_7$  — равномерный. Рассмотрим случайные величины  $\xi_j^{(k)}, \eta_j^{(k)}$ , распределенные по нормальному закону

$$\xi_j^{(k)} \sim N(k_{x_j} \cdot \bar{x}_j; k_{x_j} \cdot S_{x_j}^2); \quad (11)$$

$$\eta_j^{(k)} \sim N(k_Y \cdot \bar{y}; k_Y \cdot S_Y^2), \quad (12)$$

где  $\bar{x}_j, \bar{y}$  — средние арифметические вдоль столбцов;  $S_{x_j}^2, S_Y^2$  — выборочные дисперсии величин  $X_j, Y$ ;  $k_{x_j}, k_Y$  — параметры, характеризующие уровень шума в столбцах  $X_j$  и  $Y$  соответственно. Рассмотрим также случайные величины  $\lambda_j^{(k)}$  и  $\tau_j^{(k)}$ , равномерно распределенные на интервалах  $(k_{x_j} \cdot \bar{x}_j - k_{x_j} \cdot S_{x_j}; k_{x_j} \cdot \bar{x}_j + k_{x_j} \cdot S_{x_j})$  и  $(k_Y \cdot \bar{y} - k_Y \cdot S_Y; k_Y \cdot \bar{y} + k_Y \cdot S_Y)$ .

Сгенерируем по 100 значений случайных величин  $\xi_j^{(k)}, \eta_j^{(k)}, \lambda_j^{(k)}$  и  $\tau_j^{(k)}$ . Добавив шумовую составляющую к исходным данным, получим возмущенное тестовое множество  $\tilde{T} = \{\tilde{t}_{ij}\}$ , где

$$\tilde{t}_{ij}^{(k)} = t_{ij} + \xi_{ij}^{(k)}, \quad \tilde{t}_{i,m+1}^{(k)} = t_{i,m+1} + \eta_j, \quad (13)$$

$$i = \overline{1, N_{test}}; j = 1, 5, 9, m + 1;$$

$$\tilde{t}_{ij}^{(k)} = t_{ij} + \lambda_j^{(k)}, \quad \tilde{t}_{i,m+1}^{(k)} = t_{i,m+1} + \tau_j, \quad (14)$$

$$i = \overline{1, N_{test}}; j = 3, 7.$$

Параметр  $K$ , характеризующий интенсивность шума, выбирался произвольным образом для каждого фактора  $X_j$  и моделируемой величины  $Y$ .

Величину, характеризующую меру интенсивности возмущения тестового множества, определим как

$$\mu = \|\tilde{T} - T\|/\|T\|, \quad (15)$$

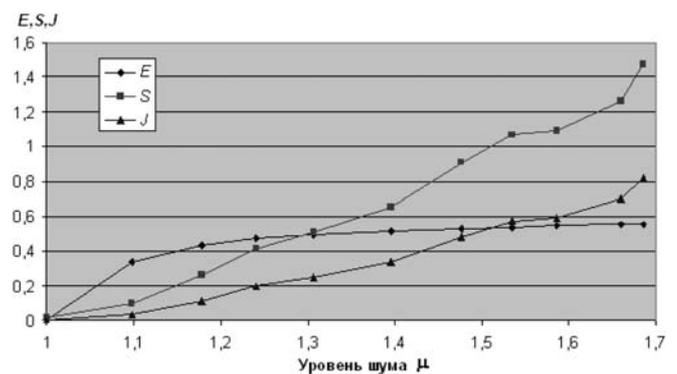
$$\text{где } \|T\| = \sqrt{\sum_{i=1}^{N_{test}} \sum_{j=1}^{m+1} t_{ij}^2}, \quad \|\tilde{T}\| = \sqrt{\sum_{i=1}^{N_{test}} \sum_{j=1}^{m+1} \tilde{t}_{ij}^2} \quad (16)$$

— евклидовы нормы исходной и зашумленной матриц, составляющих тестовое множество.

Рассмотрим зависимость показателей качества обучения сети (5)—(6) от интенсивности возмущения данных  $\mu$ . Для этого будем последовательно увеличивать число зашумленных строк и предъявлять зашумленные данные нейросети, обученной на "эталонных" данных. Долю зашумленных строк матрицы  $\tilde{T} = \{\tilde{t}_{ij}\}$  обозначим  $\omega$ . В табл. 1 и на рисунке отражены результаты тестирования нейросети.

При появлении шума в данных все показатели качества сети резко ухудшаются. Например, уже при  $\omega = 0,1$  (зашумлено 10 % строк тестового множества) ошибка обобщения составляет 33,9 %, что делает сеть непригодной для практического использования. Следует отметить, что при работе с реальными (не модельными) данными доля зашумленных данных может быть значительно больше 10 %.

При уровне шума  $\mu = 1,535$  (соответствует  $\omega = 0,7$ ) критерий устойчивости  $S > 1$ , т. е. нейросетевое отображение становится "растягивающим". Можно сделать вывод, что при некотором критическом уровне зашумленности данных НСМ становится неустойчивой. При проведении вычислительных экспериментов на реальных данных [4] выяснилось, что сеть становится неустойчивой при гораздо меньшем значении  $\omega$ . Объясняется это более



Зависимость  $E, S, J$  от интенсивности шума  $\mu$

Таблица 1

Показатели качества обучения при разном уровне шума

$\omega$	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
$\mu$	1,000	1,098	1,179	1,241	1,307	1,395	1,477	1,535	1,586	1,660	1,685
$E$	0,00130	0,339	0,435	0,477	0,496	0,515	0,528	0,536	0,545	0,550	0,553
$S$	0,01078	0,098	0,266	0,417	0,506	0,654	0,906	1,068	1,091	1,267	1,473
$J$	0,000014	0,033	0,115	0,199	0,251	0,337	0,479	0,572	0,593	0,697	0,814

сложной и недетерминированной зависимостью факторов и моделируемой величины. Относительно небольшие возмущения входных данных могут приводить к значительному искажению нейросетевого отклика  $\hat{Y}$ .

Полученные на модельном примере результаты подтверждают идею о необходимости и полезности предложенных выше процедур. Можно сделать следующий вывод: для построения качественной НСМ необходима предварительная подготовка (предобработка) данных, направленная на повышение их однородности. Предложенные авторами и описанные выше процедуры удаления противоречивых и аномальных наблюдений позволяют ослабить нежелательные последствия шума и добиться нужных прогностических свойств и точности модели. Интересно, что зависимость ошибки обобщения от интенсивности шума носит экспоненциальный характер.

Для того чтобы подтвердить эффективность предлагаемых процедур удаления противоречивых и аномальных наблюдений проведем еще один вычислительный эксперимент. Суть его в следующем: к исходной базе данных  $A = \{a_{ij}\}$ , состоящей из 200 вектор-строк, добавляется шумовая составляющая согласно правилам (13)—(14), т. е. формируется зашумленная база данных  $\tilde{A} = \{\tilde{a}_{ij}\}$ . Пусть  $\omega = 0,2$ , это значит, что 40 из 200 строк базы данных зашумлены. Зашумленные данные были промаркированы для того, чтобы оценить эффективность предлагаемых процедур регуляризации.

К полученной базе данных  $\tilde{A} = \{\tilde{a}_{ij}\}$  применим описанные выше процедуру удаления противоречивых наблюдений и итерационную процедуру удаления аномальных наблюдений. После добавления шума в базу данных  $\tilde{A} = \{\tilde{a}_{ij}\}$  появилось 19 противоречивых пар вектор-строк. После их удаления в базе данных осталось 188 строк, 28 из которых содержат шум. Отметим, что в результате выполнения этой процедуры не все зашумленные строки были выявлены и удалены из базы данных. Можно дать следующее объяснение: процедуру удаления противоречивых наблюдений можно отнести к процедурам предрегуляризации модели, т. е. на этом этапе из зашумленной базы данных удаляются самые грубые неоднородности, дальнейшая чистка базы данных происходит на этапе обучения сети.

Таблица 2

Показатели качества обучения сети на каждой итерации при очистке базы данных от аномальных наблюдений

$k$	1	2	3	4	5	Обучение без шума
$N$	188	180	172	163	158	200
$A$	8	8	9	5	—	—
$\xi^{(k)}$	10 %	4 %	3 %	2 %	—	—
$E$	0,5280	0,0741	0,0509	0,0325	0,0079	0,00130
$S$	0,9064	0,0913	0,0939	0,0675	0,0978	0,01078
$J$	0,4786	0,0067	0,0047	0,0021	0,00077	0,000014

После этого была проведена итерационная процедура обучения сети с удалением на каждой итерации аномальных наблюдений. На каждой  $k$ -й итерации фиксировались показатели качества обучения (5)—(6), число аномальных наблюдений  $A$  и уровень отбраковки  $\xi^{(k)}$  (табл. 2).

Как видно из табл. 2, после 4-й итерации в базе данных осталось 158 наблюдений. После обучения и тестирования сети на оставшихся данных показатели качества обучения достигли желаемых значений, и процесс удаления аномальных наблюдений был остановлен. Из 158 наблюдений, оставшихся в базе данных, только три строки содержат шум (они были промаркированы и отслеживались). Таким образом, 37 из 40 зашумленных строк были обнаружены и удалены из данных. При этом показатели качества обучения приблизились к значениям, наблюдавшимся при обучении сети на эталонных (незашумленных) данных (для удобства сравнения они приведены в последнем столбце табл. 2).

### Заключение

Проведенные вычислительные эксперименты подтвердили необходимость предобработки данных в целях повышения однородности, информативности и повышения качества обучения нейросетей в задачах, для которых характерны зашумление или сознательное искажение первичных данных.

На модельном примере показана эффективность и состоятельность процедур удаления противоречивых и аномальных наблюдений. Особое значение имеет то, что процесс поиска и удаления аномальных наблюдений проводится итерационно, взаимосвязанно с обучением нейросети. При этом реализуется сформулированная авторами концепция "взаимосвязанного управления качеством данных и их аппроксимации нейросетью" [1].

### Список литературы

1. Горбатков С. А., Полупанов Д. В., Макеева Е. Ю., Бируков А. Н. Методологические основы разработки нейросетевых моделей экономических объектов в условиях неопределенности / Под ред. С. А. Горбаткова. М.: Издат. дом "Экономическая газета", 2012. 494 с.
2. Шумский С. А. Байесова регуляризация обучения // Научная сессия МИФИ 2002. IV Научно-техническая конференция "Нейроинформатика—2002": Лекции по нейроинформатике. Ч. 2. М.: МИФИ, 2002. С. 30—93.
3. Нейросетевое математическое моделирование в задачах ранжирования и кластеризации в бюджетно-налоговой системе регионального и муниципального уровней / С. А. Горбатков, Д. В. Полупанов, А. М. Солнцев, И. И. Белолипцев, М. В. Коротнева, С. А. Фархиева, О. Б. Рашитова. Уфа: РИЦ БашГУ, 2011. 224 с.
4. Горбатков С. А., Белолипцев И. И., Фархиева С. А. Приближенный метод байесовской регуляризации и двухступенчатая оценка адекватности гибридной нейросетевой модели // Научная сессия НИЯУ МИФИ—2011. XIII Всероссийская научно-техн. конф. "Нейроинформатика—2011": Сб. науч. тр. В 3-х частях. Ч. 2. М.: НИЯУ МИФИ, 2011. С. 144—154.
5. Ежов А. А., Шумский С. А. Нейрокомпьютинг и его применение в экономике и бизнесе: учебник / Под ред. В. В. Харитонов. М.: Изд. Московского инженерно-физического ин-

Институт системного программирования РАН

**В. В. Липаев**, д-р техн. наук, проф.

## **Программная инженерия: электронный комплекс учебников и монографий**

**Программная инженерия** — это область компьютерной науки и технологии, одной из главных задач которой является исследование методов и средств построения и сопровождения больших и сложных программных продуктов, для работы с которыми необходимо участие слаженных команд квалифицированных исполнителей различных специальностей и квалификаций. Методология программной инженерии состоит в применении систематизированного, научно обоснованного процесса **проектирования, производства и сопровождения программных комплексов** реального времени, включающих:

- постановку задач;
- исследования и подготовку проектных решений;
- планирование процесса производства;
- организацию работ, контроль за выполнением работ и использованием ресурсов.

Основные концептуальные положения программной инженерии сконцентрированы в **целостном комплексе систематизированных международных стандартов**, регламентирующих практически все процессы жизненного цикла сложных программных продуктов. Практическое применение стандартов, сосредоточивших мировой опыт создания различных типов крупных комплексов программ, способствует значительному **повышению производительности труда специалистов и качества** создаваемых программных продуктов. Во многих случаях в качестве базового рекомендуется использовать стандарт **ISO 12207:2008** — процессы жизненного цикла программных средств.

Объектами профессиональной деятельности специалистов являются:

- программные проекты;
- программные продукты;
- методы и инструментальные средства разработки программных продуктов;
- коллективы специалистов, участвующие в процессах жизненного цикла сложных комплексов программ.

Они должны **быть компетентными** в основных видах профессиональной деятельности.

Резко повысилась **ответственность специалистов за качество** результатов их труда и создаваемых программных продуктов. Это требует непрерывного совершенствования, обучения и повышения квалификации заказчиков, разработчиков и пользователей в области программной инженерии, освоения ими современных методов обеспечения верификации, тестирования и международных стандартов, а также **высокой корпоративной культуры коллективов специалистов**, обеспечивающих жизненный цикл критических программных продуктов.

Для управления затратами ресурсов на производство программных продуктов рекомендуется владеть знаниями по **экономике жизненного цикла программных комплексов**, промышленного производства средств и систем на базе вычислительной техники. Объективно исследования на этом направлении осложняются трудностью измерения и прогнозирования экономических характеристик производства программного продукта и выполняемых ими функций.

Резко возросла необходимость **гарантировать высокое качество программных продуктов**, а также регламентировать и корректно формировать требования к характеристикам реальных комплексов программ. Рассматриваются системные основы разработки требований к сложным комплексам программ, эталоны при **их проектировании и производстве**. Представлены методы тестирования потоков управления и потоков

данных программных модулей. Изложены программа и методики **верификации и тестирования компонентов и сложных комплексов программ**, организация и процессы их испытаний.

По мере расширения применения и увеличения сложности систем выделились области, в которых дефекты, недостаточное качество комплексов программ или данных могут наносить значительный ущерб, намного превышающий положительный эффект от их использования. В таких **критических системах** (например, управления космической и авиационной техникой, атомными электростанциями, крупными банками или системами вооружения) требуется **сертификация программных продуктов** и недопустимы проявления высоких рисков, нарушений принятых режимов функционирования программных продуктов при любых искажениях исходных данных, сбоях, частичных отказах аппаратуры, ошибках пользователей и других нештатных ситуациях. Значительное внимание уделено **надежности и функциональной безопасности сложных программных продуктов**, программе и методикам испытаний компонентов и комплексов программ, динамическим характеристикам программных продуктов.

Для **обучения**, повышения **квалификации и компетенции специалистов**, обеспечивающих все этапы жизненного цикла крупномасштабных программных средств, создающих и применяющих системы и комплексы программ обработки информации, автором подготовлен комплекс учебников и монографий, содержащий полные электронные копии перечисленных ниже книг, которые в свободном доступе представлены заинтересованным предприятиям и специалистам на сайте **Института системного программирования РАН** (<http://www.ispras.ru>).

- **Проектирование и производство сложных заказных программных продуктов.** — М.: СИНТЕГ, 2011. — 408 с. — **25245 Кбайт.**
  - **Человеческие факторы в программной инженерии: рекомендации и требования к профессиональной квалификации специалистов.** Учебник. — М.: СИНТЕГ, 2009. — 328 с. — **25314 Кбайт.**
  - **Экономика производства программных продуктов.** Издание второе. — М.: СИНТЕГ, 2011. — 358 с. — **81135 Кбайт.**
  - **Тестирование компонентов и комплексов программ.** Учебник. — М.: СИНТЕГ, 2010. — 400 с. — **50233 Кбайт.**
  - **Сопровождение и управление конфигурацией сложных программных средств.** — М.: СИНТЕГ, 2006. — 372 с. — **61972 Кбайт.**
  - **Сертификация программных средств.** Учебник. — М.: СИНТЕГ, 2010. — 348 с. — **59739 Кбайт.**
  - **Надежность и функциональная безопасность комплексов программ реального времени.** — М.: 2013. — 200 с. — **1527 Кбайт.**
- Общий объем семи книг около **2400 страниц**, объем занимаемой памяти — **315 Мбайт.**  
**Справки:** e-mail: [vlip28@mail.ru](mailto:vlip28@mail.ru) телефоны: 8-916-192-48-32; 8-499-196-63-65.

# CONTENTS

**Maiorov A. A., Tsvetkov V. Ya.** *Geoinformatics as an Important Direction of Informatics* . . . . . 2

The article describes the evolution of computer science and geoinformatics. The features of the evolution of informatics in the form of "informatics 1" (information sciences), "informatics 2" (computer science), "informatics 3" (integrated informatics). Showing the current problems of the informatics science. We consider the origins of the origin and characteristics of the development of geoinformatics. The similarity and difference between computer science and geoinformatics. It is shown that in the computer science subject area of specialization is based on the methods of treatment to the subject area of expertise and data on this basis. It is shown that geoinformatics based on data integration and their application in various fields. The article describes the problems of informatics and geoinformatics. The necessity of integrating these areas. The article shows that the geo-information integration is the basis for the "informatics 1", "informatics 2", "informatics 3".

**Keywords:** geoinformatica, computer science, geographic data, GIS-based approach, GIS technology

**Filimonov A. B., Kromin O. A.** *Jade Semantic Agent Communication Mechanism* . . . . . 8

The structure of communication acts on the agents software platform JADE are considered. The singularities of semantic language of communication of agents FIPA-SL are stated. The possibilities of realization of semantic interaction of agents using framework of programming of multi-agent systems JSA are discussed.

**Keywords:** multi-agent systems, the language of communications of agents, semantics of communication acts, framework of programming of multi-agent systems JSA

**Vasylyev A. V.** *Numerically-Analytical Investigations of the Mathematical Model of Fractional Oscillator by S-Transform* . . . . . 15

The S-transform is applied to the numerical and analytical study of the fractional order differential equation — a mathematical model of a fractional oscillator. Illustrative examples show how to obtain solutions for different orders of fractional differential operator, initial conditions and external influences. A local versions of shifted Legendre polynomials are used as basis systems functions for S-transform. Computational experiments were carried out in the software environment of Mathematica®.

**Keywords:** fractional calculus, fractional oscillator, Caputo's fractional derivative, S-transform method, Legendre polynomial

**Kazakov P. V.** *The Model of Dynamic Varying Population Size for Multi-Objective Genetic Algorithms* . . . . . 20

The new approach for a population size control in the genetic algorithms for multi-objective optimization is considered. This approach is universal and may be implemented for any multi-objective evolutionary algorithms. The comparative performance analysis of using dynamic population size for solving different multi-objective problems are cited.

**Keywords:** multi-objective optimization, Pareto's principles, Pareto front, multi-objective genetic algorithms, dynamic varying population size

**Fedosov V. V., Fedosova A. V.** *The Stochastic Search of Protective Regulations Objects (Zones) Located on the Same Property with a Group of Sources of Industrial Emissions* . . . . . 27

A model of the algorithm and the numerical solution of the problem of finding safety regulations objects (zones) located on the same property with a group of emission sources. Objects (areas) are arbitrary limits and number. Emission sources are different in power, the arrangement and composition of emissions. The algorithm excludes from consideration the industrial zone of emission sources.

In MATLAB developed software system that provides mapping the total and component-contamination and points defining the standards of OZ in their borders. Allows you to simulate the design OZ accommodation in industrial areas.

**Keywords:** emissions, components of the emission, standards of pollution loads, zone area, stochastic algorithm

**Maksymenko-Sheyko K. V., Tolok A. V., Sheyko T. I.** *Connected Problem of the Convective Heat Transfer in the Fuel Cartridge of the Power Rods* . . . . . 32

The connected boundary-value problems of the heat exchange were considered for cases when viscous incompressible liquid goes on channels of complex cross-section, flowing about the bundle of rods. The effect of the packing type on distribution of velocity and temperature was investigated. The R-functions theory in a combination with variational Ritz method was used for the solution. The various packings power rods were surveyed. The each packing contains 91 rod, and the corresponding equations built with the help of new constructive means of the R-functions theory.

**Keywords:** R-functions theory, Ritz method, connected problem, heat exchange, fuel cartridge, triangular packing, cyclic symmetry

**Vostrikov A. A.** *On Hadamard—Mersenne Matrices and Image Masking* . . . . . 37

The issue of significant extending of orthogonal matrices basis is considered in the paper. Images masking procedure, based on M-matrices as orthogonal basis, is proposed and described. It is noted that orders of matrices in this case are of Mersenne prime.

**Keywords:** orthogonal matrices, M-matrices, Hadamard matrices, Hadamard—Mersenne matrices, Mersenne prime

**Belyaeva E. A.** *Comparative Assessment of Hardware—Software Units of Trusted Boot Based on the Classification of their Functionality* . . . . . 40

This paper aims of identifying the main and additional features of hardware-software unites of trusted boot, while confirming the need to evaluate their impact on the protection efficiency against unauthorized access and offering ways of classifying them. Science-based approach to obtaining integrated assessment of features for hardware-software unites of trusted boot based on an algorithm of receiving a comprehensive measure to assess the effectiveness of protection against unauthorized access and comparative evaluation methodology by "cost-effectiveness" is presented by the auther or the article.

**Keywords:** feature, classification, criterion, unauthorized access, method, algorithm

**Eremenko A. V., Sulavko A. E.** *Analysis of Algorithms for Cryptography Key Generation Based on User-Specific Biometric Information of Computer Systems Users* . . . . . 47

The article is devoted to the problem of safety ensuring cryptography keys taking into account a human factor. Fuzzy extractors are offered to use for constructing keys from biometric data for protection against unauthorized access to confidential keys. A research of fuzzy logic algorithm for transformation of dynamic characteristics of hand-written passwords to authentically reproduced, randomized, uniformly distributed key material was carried out.

**Keywords:** hand written dynamics, error-correcting codes, error-correcting capability, access control, correlation of signals, identifiable characteristic, fuzzy data, key sequence

**Domunian A. A.** *Three Stages of Gesture Recognition* . . . . . 52

Considered problem of gesture recognition. For preprocessing scene, choose method of picture difference and shaping. Marked out 3 levels of the solve problem. Describe method of features extraction, describe functions of positional distance, suggest method of machine learning of the system.

**Keywords:** gesture recognition, feature extraction, position distance, systems with training

**Grinyak V. M., Devyatitsilny A. S.** *Adaptive Network Based Fuzzy Inference System of Air Target Identification* . . . . . 58

The problem of air target identification by two-coordinate radar is watched. Identification method is based on target height estimation from range and azimuth measurements. Target height estimations are processed by neuro-fuzzy system (ANFIS). System estimates level of belonging to "air target" class. ANFIS configuration and learning method are given in this paper. Computer modeling shows the usability of proposed approach for typical situations.

**Keywords:** vessel traffic control, air target, radar measurement, target height, neuro-fuzzy system, ANFIS

**Gorbatkov S. A., Beloliptsev I. I., Solntsev A. M.** *To the Question of the Viability of Regularization Procedures of Neural Network Model of Tax and Financial Control* . . . . . 64

A method of increasing the uniformity and informativeness of data is proposed to improve the quality of learning and prediction strength of neural network models in problems with highly skewed data. Computational experiments have proved the necessity and viability of the proposed data preprocessing procedures.

**Keywords:** neural network, regularization of the learning, conflicting data, stability, data corruption

---

---

**Адрес редакции:**

107076, Москва, Стромьинский пер., 4

Телефон редакции журнала (499) 269-5510

E-mail: it@novtex.ru

Дизайнер *Т.Н. Погорелова*. Технический редактор *Е.В. Конова*.

Корректор *Т.В. Пчелкина*.

Сдано в набор 02.09.2013. Подписано в печать 22.10.2013. Формат 60×88 1/8. Бумага офсетная.

Усл. печ. л. 8,86. Заказ IT1113. Цена договорная.

Журнал зарегистрирован в Министерстве Российской Федерации по делам печати, телерадиовещания и средств массовых коммуникаций.

Свидетельство о регистрации ПИ № 77-15565 от 02 июня 2003 г.

Оригинал-макет ООО "Авансед солюшнз". Отпечатано в ООО "Авансед солюшнз".

119071, г. Москва, Ленинский пр-т, д. 19, стр. 1.

---