

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

2(162)
2010

ТЕОРЕТИЧЕСКИЙ И ПРИКЛАДНОЙ НАУЧНО-ТЕХНИЧЕСКИЙ ЖУРНАЛ

Издается с ноября 1995 г.

УЧРЕДИТЕЛЬ
Издательство "Новые технологии"

СОДЕРЖАНИЕ

ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ

- Балакирев Н. Е., Добрышина Е. В. Концептуальная модель и структура системы обработки текстовой информации 2
Голомазов Д. Д. Выделение терминов из коллекции текстов с заданным тематическим делением 8
Шапкин П. А. Модели и методы разработки веб-приложений на основе онтологии предметной области 13

МОДЕЛИРОВАНИЕ И ОПТИМИЗАЦИЯ

- Переварюха А. Ю. Новый метод компьютерного моделирования режима переходного хаоса 18
Карпенко А. П., Селиверстов Е. Ю. Глобальная оптимизация методом роя частиц. Обзор 25
Лиманова Н. И., Мамзин Е. А. Дискретная математическая модель детерминированного клеточного автомата и ее программная реализация 34
Ковель И. В. Вероятностный метод построения окрестности статистического глобального оптимума NPC-задач 38

СИСТЕМЫ АВТОМАТИЗИРОВАННОГО ПРОЕКТИРОВАНИЯ

- Бабак Л. И., Дорофеев С. Ю., Песков М. А., Черкашин М. В., Шеерман Ф. И., Барышников А. С., Кошевой С. Е. Интеллектуальная система автоматизированного проектирования СВЧ устройств INDESYS 42
Разумовский А. И., Сизова Л. Н. Проектирование и трассировка печатных плат с использованием программного комплекса "Графика—ТР" 48

КОДИРОВАНИЕ И ОБРАБОТКА СИГНАЛОВ

- Назаров Л. Е. Алгоритмы посимвольного приема сигналов 53
Дворников С. В., Дворников С. С. Обнаружение сигналов с высоким различием динамики их амплитуд 56
Савченко В. В., Акатьев Д. Ю. Обучение звуковому строю языка глухонемых и слабослышащих на основе информационной теории восприятия речи 60

КОНТРОЛЬ И ДИАГНОСТИКА СИСТЕМ

- Дюндиков Е. Т., Качкин А. А. Технология динамической интеграции и представления разнородных данных для анализа и оценки состояния многопараметрических объектов 66

ПРИКЛАДНЫЕ ИНФОРМАЦИОННЫЕ СИСТЕМЫ

- Углова Е. С., Моисеева Л. С. Математическая модель и информационная система поддержки принятия решений в задачах коррозионной защиты 74
Contents 78
Приложение. Нариньяни А. С., Гофман И. Д., Липатов А. А., Инишев Д. А. Интеллектуальная технология недоопределенного календарно-ресурсного планирования и управления проектами Time-EXP®

Главный редактор
НОРЕНКОВ И. П.

Зам. гл. редактора
ФИЛИМОНОВ Н. Б.

Редакционная
коллегия:

АВДОШИН С. М.
АНТОНОВ Б. И.
БАТИЩЕВ Д. И.
БАРСКИЙ А. Б.
БОЖКО А. Н.
ВАСЕНИН В. А.
ГАЛУШКИН А. И.
ГЛОРИОЗОВ Е. Л.
ГОРБАТОВ В. А.
ДОМРАЧЕВ В. Г.
ЗАГИДУЛЛИН Р. Ш.
ЗАРУБИН В. С.
ИВАННИКОВ А. Д.
ИСАЕНКО Р. О.
КОЛИН К. К.
КУЛАГИН В. П.
КУРЕЙЧИК В. М.
ЛЬВОВИЧ Я. Е.
МАЛЬЦЕВ П. П.
МЕДВЕДЕВ Н. В.
МИХАЙЛОВ Б. М.
НАРИНЬЯНИ А. С.
НЕЧАЕВ В. В.
ПАВЛОВ В. В.
ПУЗАНКОВ Д. В.
РЯБОВ Г. Г.
СОКОЛОВ Б. В.
СТЕМПКОВСКИЙ А. Л.
УСКОВ В. Л.
ЧЕРМОШЕНЦЕВ С. Ф.
ШИЛОВ В. В.

Редакция:

БЕЗМЕНОВА М. Ю.
ГРИГОРИН-РЯБОВА Е. В.
ЛЫСЕНКО А. В.
ЧУГУНОВА А. В.

Информация о журнале доступна по сети Internet по адресу <http://www.informika.ru/text/magaz/it/> или <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 004.912

Н. Е. Балакирев, канд. техн. наук, проф.,
Е. В. Добрышина, аспирант,
"МАТИ" — РГТУ им. К. Э. Циолковского,
e-mail: balakirev1949@yandex.ru

Концептуальная модель и структура системы обработки текстовой информации

Проанализированы проблемы формализации и автоматизированной обработки текста, сформулирован подход к анализу естественно-языковых текстов, учитывающий наиболее существенные свойства языка, предложена концептуальная модель и структура системы обработки текстовой информации.

Ключевые слова: автоматизированный анализ текста, система анализа, обработка естественно-языкового текста, представление знаний, формализация естественного языка

Повсеместное использование вычислительной техники, особенно после создания персональных компьютеров и появления сети Интернет, привело к резкому росту объема информации, представленной в текстовой форме, и, как следствие, к необходимости создания средств для работы с их постоянно нарастающим потоком в целях решения задач поиска, классификации, кластеризации, извлечения и хранения информации.

Под обработкой текстов на естественном языке понимается совокупность процессов анализа текстов, их понимания и синтеза. В процессе анализа в наиболее развитых системах обработки проводятся морфологический, синтаксический и семантический анализ, в результате чего выявляется глубинная структура текста, которая переводится во внутреннее представление, используемое в базе знаний интеллектуальной системы. Соотнесение этой структуры с теми знаниями, которые хранятся в системе, позволяет понять смысл исходного текста. С обработкой естественно-языкового текста связано решение задач машинного перевода, автоматического реферирования, общения с пользователем на ограниченном профессиональном естественном языке и т. п. [1].

Качество любого программно-аппаратного комплекса во многом определяется адекватностью моделирования системы, с которой он рабо-

тает. При решении задач анализа и понимания текстовой информации в идеале необходимо заложить в компьютер возможности, соответствующие человеческому интеллекту. Но большинство подходов в создании систем ориентировано лишь на частный случай решения конкретной задачи, что вполне оправдано в связи со сложностью проблемы. К сожалению, решение проблемы автоматизированного понимания текста не может быть представлено совокупностью отдельных решенных задач, они могут лишь являться модулями единой системы, степень зависимости друг от друга которых характеризуется подчиненностью их работы общему замыслу. Создан ряд моделей естественного языка [2, 3], на основе которых если и разрабатываются компьютерные системы, то чаще всего они не используются на практике. Поэтому необходима разработка модели системы, позволяющей отражать специфику анализа текстов на естественном языке, учитывать особенности языка и проводить обработку больших объемов текстов за приемлемое время.

При создании системы необходимо выбрать базовую основу и принципы построения, которые определяют перспективу ее использования и дальнейшего развития. С одной стороны, система должна строиться на основе некоторых основных функций, с другой стороны — позволять комбинировать эти функции, учитывая цели решаемой задачи. Также структура системы должна давать возможность серьезного развития и перестройки системы.

В данной работе предлагается модель универсальной системы автоматизированной обработки текста на естественном языке, которая обладает необходимыми возможностями для извлечения, накопления и анализа информации, используемой для решения основных задач, требующих обработки текста.

Сложность формализации текста на естественном языке

Стремление разработчиков максимально улучшить качество результатов анализа текста и создать по-настоящему интеллектуальную систему приводит к необходимости понимания и моделирования процессов умственной деятельности человека [2]. Ее изучение интенсивно продвигается [4—7], однако сложность объекта изучения до сих

пор не позволяет полностью понять мыслительный процесс и создать его полноценную модель.

Основным способом отображения и хранения результатов интеллектуальной деятельности человека является запись текстов на естественных языках. Естественные языки складывались эволюционно под воздействием множества факторов [8, 9], следствием чего стали следующие их особенности, отражающиеся в текстах:

- *многозначность*, порождающая неопределенность;
- *разнообразие форм* выражения мысли;
- *противоречивость* трактовки и отношения к описываемой ситуации, событию, явлению и т. д.;
- *многофакторность*, *параллельность* (наличие скрытого смыслового контекста);
- *сложность и многоплановость* представления, мыслей, наблюдений и фактов;
- *многоязыковость*;
- *недосказанность, недоопределенность*;
- *контекстная привязанность*;
- *изменчивость смысла слова* с течением времени.

Все эти особенности являются препятствиями на пути формализации естественного языка и причиной ограничений на решение проблемы "глубинного" анализа текста на основе доступных в настоящее время знаний.

Гигантское количество накопленной текстовой информации приводит к необходимости создания высокопродуктивных систем анализа текста и определяет следующие требования к таким системам:

- возможность обработки больших объемов информации за приемлемое время;
- возможность накопления больших объемов обработанной информации;
- возможность совместного использования и пополнения средств анализа;
- возможность оперирования сложными конструкциями представления данных.

Проблемы компьютерного анализа текстов

Компьютерная обработка текста на естественном языке имеет ряд специфических особенностей, сильно затрудняющих ее практическое и полное решение. Главная особенность — это сложность формализации естественного языка и, как следствие, трудность разработки его полной лингвистической модели, а затем и реализации этой модели в виде программы, правильно и в разумное время обрабатывающей любые тексты на естественном языке [2].

Понимание того, какую информацию потребуется извлечь из текста, в определенной степени обуславливает методы и алгоритмы решения задачи извлечения знаний. Сложность задачи формализации естественного языка такова, что, не-

смотря на создание множества моделей [2, 10], к настоящему моменту нет какой-либо единой системы представления и накопления знаний [10], да и созданные модели не слишком пригодны для практической реализации и использования.

Громадные объемы данных, которые требуют обработки, сильно затрудняют решение задачи общего анализа информации, принадлежащей достаточно большой предметной области, не говоря уже об окружающем мире в целом. Наряду с этим существуют технические ограничения, связанные со скоростью обработки и объемом предоставляемой памяти.

К сожалению, существующие системы автоматизированного анализа решают лишь конкретные задачи, связанные с обработкой текста, так же как и множество экспериментальных систем, которые, безусловно, интересны моделями представления знаний, принципами логических выводов [2, 11—14].

В зависимости от целей анализа текста и информации, которую необходимо извлечь для их достижения, в системах используются лингвистические и (или) статистические методы. Первые базируются на схемах анализа текста, ставших общепринятыми в компьютерной лингвистике [11, 12], статистические же методы позволяют получать необходимую информацию на основании количественных характеристик текстов. В последнее время наблюдается тенденция комбинирования статистических и лингвистических методов анализа [15, 16].

Все изложенное привело к мысли поставить вопрос об анализе текста в более широком аспекте и, исходя из этого, предложить свой подход к построению универсальной системы автоматизированного анализа текста. В качестве ключевой идеи решения проблемы технических ограничений предлагается следующее: анализировать отдельные блоки данных (тексты) и впоследствии объединять результаты обработки и переходить к анализу всей области уже на основе обобщенной информации.

Концептуальный подход к построению системы обработки информации

Рост объема обрабатываемой информации неминуемо ведет к нелинейному увеличению времени обработки. Решение задач, связанных с углублением (например, от синтаксиса к семантике) анализа текста, даже в простейшем варианте также приводит к нелинейному росту времени обработки, что делает практическое использование таких систем невозможным [2, 17]. Поэтому необходим как тщательно продуманный общий алгоритм анализа текстов и накопления результатов обработки, так и максимальная оптимизация ка-

ждого отдельно взятого алгоритма анализа конкретного текста.

Кроме этого, важно не потерять и правильно понять каждый элемент текста. Особенно это важно в случаях, когда невозможно получение однозначного ответа на конкретном этапе анализа. Необходимо допустить фиксацию неопределенных ответов, что позволит сохранить всю информацию и нормально завершать работу отдельно взятого блока системы.

Исходя из этого был сформулирован концептуальный подход, определяющий следующие требования к системе обработки текстовой информации.

1. *Допустимость третьего логического состояния — неопределенности.* Позволяет учитывать особенности естественного языка и отражать **неоднозначности языковых конструкций**.

2. *Открытость, повышающая объективность анализа результатов.* Предполагает возможность коллективных работ, обмена опытом, на базе открытости для пользователя структуры системы, форматов данных и всех тех элементов, которые напрямую определяют алгоритмы анализа текста.

3. *Параллельность обработки для увеличения скорости при наличии свободных ресурсов.* Основой параллельности является унификация всех используемых структур. Кроме этого параллельность в конечном итоге предполагает сведение в единую структуру полученных результатов за счет использования разрабатываемых **операций анализа и обобщения результатов**.

4. *Масштабируемость для расширения возможностей системы.* Расширение и замена модулей не должны затрагивать базовой основы системы.

5. *Многоуровневость и многоэтапность как средство поддержки разбиения обработки текста на отдельные шаги и уровни.* Все это должно облегчить локализацию обработки ошибок и дать возможность оформления самостоятельных ветвей для частных задач.

Кроме этого, предлагается трехмерная модель системы анализа текста (рис. 1), в которой имеются три направления обработки: лингвистическое (лингвистические методы обработки); статистическое (статистические методы обработки); аналитическое (операции над результатами). Такая трехмерная схема обработки возможна как со стороны средств системы, так и со стороны пользователя, который может воспользоваться как базовыми инструментами, так и своими собственными либо их комбинацией. Средой объединения является система хранения, которая предполагает хранение промежуточных и окончательных данных и инструментов, предназначенных для анализа текста. Инструменты могут представлять собой наборы операций, отдельных процедур и трансляторов [18], специализирующихся на соз-

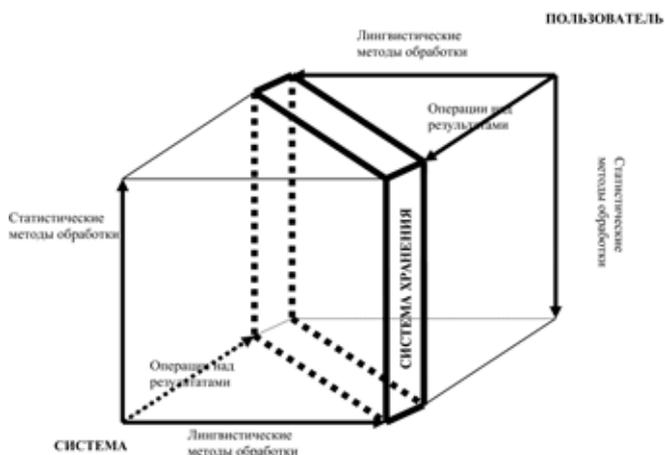


Рис. 1. Трехмерная модель системы анализа текста

дании алгоритмов текстовой обработки. Новым направлением анализа текста является введение **операций анализа и обобщения результатов** обработки, которые позволяют обеспечить параллельность обработки и сбора получаемых результатов в один пул информации и являются основой анализа полученных лингвистическо-статистических материалов под выбранным углом зрения.

Предлагаемая структура системы обработки информации

Описанный концептуальный подход требует соответствующей структуры системы (рис. 2), приведенной ниже.

1. *Система базовой обработки* предоставляет возможность обработки произвольного текста, управления процессом обработки, настройки его параметров.

2. *Система анализа результатов* включает в себя операции над структурами данных, полученными после базовой обработки. Они позволяют проводить дальнейший анализ и дают возможность написания алгоритмов для решения конкретной задачи.



Рис. 2. Схема системы

3. *Аналитическая система накопления поступающей информации* — набор операций над результатами обработки, который автоматически запускается при обработке каждого текста и позволяет постоянно обновлять информационную базу системы.

4. *Система хранения* включает в себя базу данных и совокупность файлов, создаваемых программой при обработке или после ее завершения по желанию пользователя.

5. *Пользовательская система* — среда взаимодействия пользователя с программой, обеспечивающая возможность доступа к системе хранения и имеющемуся набору инструментов.

Блоки взаимодействуют между собой с помощью данных разных типов и структур: словники, списки предложений, списки связей, результаты статистической обработки, синтаксические и семантические представления и т. д.

Требование многоуровневости и многоэтапности обработки делает необходимым введение понятий глубины и ширины обработки. Под *глубиной* будем понимать уровень обработки с лингвистической точки зрения, под *шириной* — совокупность действий, определяющих операции по анализу и обобщению получаемых данных после прохождения определенного этапа обработки. Движение в ширину — это перемещение между отдельными частями системы, в глубину — перемещение по уровням обработки и анализа внутри отдельного блока системы.

Система базовой обработки

Весь процесс обработки текста целесообразно представлять в виде отдельных последовательно выполняемых этапов. В первую очередь, текст необходимо подготовить к загрузке в систему, что осуществляется на этапе предварительной обработки. На этом этапе может выполняться удаление из текста лишней информации, приведение к нужному формату, разделение на части и т. д.

Основу разбиения процесса обработки текста на блоки должна определять последовательность этапов анализа, принятая в компьютерной лингвистике. На этапе *графематического анализа* текст рассматривается как последовательность символов. При этом выделяются отдельные слова, предложения и разделители; могут выделяться абзацы, пункты и другие структурные единицы текста в зависимости от решаемой задачи. Кроме этого может проводиться расшифровка сокращений, выделение цифровых комплексов, дат, фамилий с инициалами, электронных адресов.

Графематический анализ проводится первым и не менее важен, чем предварительная обработка, поскольку дает возможность привести обрабаты-

ваемый текст к виду, более пригодному для дальнейшей обработки. На этапе графематического анализа возможно также получение различных частотных характеристик текста. Статистические данные полезны при решении задач классификации текстов и иногда дают интересные результаты при анализе информации. Статистические данные на уровне слов важны для дальнейшей обработки текста, поскольку частоты использования слов влияют как на значимость того или иного понятия, так и на весомость связи пары слов. При этом подобная информация впоследствии дает возможность оптимизации поиска, что очень важно при обработке больших объемов текстов. Уже на этапе деления текста на слова и предложения возникают неопределенности [11], например, при употреблении дефисного написания слов, использовании многоточий, сокращений и т. д.

Если графематический анализ проводится на уровне отдельных символов текста, то *морфологический анализ* — на уровне отдельных слов. На этом этапе определяются грамматические признаки и начальная форма для каждой словоформы. Таким образом, для морфологического анализа необходим словарь основ и словоформ с их грамматическими признаками.

Синтаксический анализ определяет синтаксическую структуру входного предложения. Одним из вариантов представления синтаксической структуры текста является дерево подчинения, т. е. дерево, узлы которого — слова (члены предложения), а корнем является один из главных членов предложения, как правило, сказуемое. Подчиненными узлами являются слова, зависящие от данного. При этом каждая связь относится к какому-либо типу, который описывается соответствующей меткой. Другой вариант представления результатов синтаксического анализа — выделение групп в предложении, в частности, группы подлежащего и группы сказуемого.

Синтаксический анализ, во-первых, часто дает возможность получения ответов на вопросы, для ответа на которые не было достаточно информации ранних этапов, например разрешение некоторых видов омонимии. Во-вторых, синтаксические связи между словами в определенной степени определяют и семантические. Таким образом, морфологические характеристики слов и синтаксические отношения между ними во многом предопределяют последующие смысловые связи и их значимость.

Семантический анализ определяет смысловые связи между понятиями в тексте. Его результатом является некоторое формальное представление текста, которое может строиться на основе различных моделей [11, 12]. Наиболее адаптирована к представлению текстовой информации форма

семантической сети, поэтому алгоритмы перехода от текста к такой структуре понятны, модифицируемы и удобны для программной реализации, тогда как механизмы перевода в другие формы представления знаний четко не сформулированы. При этом сетевая форма в качестве базовой дает возможность дальнейшего усовершенствования модели представления знаний.

На каждом этапе анализа получаемая информация сохраняется, что позволяет осуществлять процесс обработки поэтапно, возобновлять ранее начатый анализ текста, а также реализовать механизм возврата на ранние этапы анализа после получения дополнительной информации о неопределенном ранее понятии или конструкции.

Результаты блока обработки являются исходными данными блока анализа, в котором имеются соответствующие операции для каждого типа результатов.

Система анализа результатов

Система анализа результатов предоставляет набор операций над результатами обработки, т. е. структурами данных, полученными после использования хотя бы части инструментов системы базовой обработки. Введение таких операций дает возможность получать необходимую для решения конкретной задачи информацию на основе общей, извлеченной непосредственно из текста.

Кроме того, проблема уменьшения времени обработки может быть решена за счет разбиения исходного текста на "разумные" блоки и последующего применения операций над структурами, полученными после базовой обработки текста, в первую очередь, операции объединения. В том числе этот способ дает возможность осуществления параллельной обработки частей текста. Последующее применение операций позволяет получить единый результат обработки всего текста.

Список операций над результатами может уточняться по мере необходимости. По аналогии с операциями в теории множеств в качестве базовых предлагается принять *объединение, пересечение, вычитание, объединение с отсечением, отношение* полученных результатов обработки текста. При этом действие каждой операции определяется видом результата, к которому она применяется, и параметров, которые при этом учитываются (например, частей речи, частоты, веса и т. д.).

Представляется целесообразным введение следующих операций:

1. *Объединение.* Дает возможность решать задачу обобщения получаемых результатов в виде объединенных словарей, объединенных структур, получаемых в ходе обработки текстов, относящихся

к одному классу, автору, теме и т. д. Является одной из ключевых операций в системе.

2. *Пересечение.* Позволяет автоматизировать решение задач сравнения текстов по определенному признаку, определения авторства, классификации текстов и т. д.

3. *Вычитание.* Операция вычитания может использоваться для сравнения текстов, определения авторства, решения задачи классификации текстов. Также используется в качестве подоперации для нахождения отношения типа 2.

4. *Объединение с отсечением.* Операция объединения с отсечением необходима для уменьшения объемов накапливаемой информации путем удаления наименее существенных результатов после объединения.

5. *Отношение.* Операция отношения может использоваться для сравнения текстов, для определения авторства, решения задачи классификации текстов, оценки словарного запаса человека или группы людей, принимая первый текст за базовый (например, словарь) и находя отношения к нему второго текста.

Введение операций позволяет создавать и сохранять для последующего использования алгоритмы в виде последовательности их применения для решения задач обработки текста. Это делает возможным обобщение последовательностей операций в виде самостоятельных функций в системе.

Аналитическая система накопления поступающей информации

Учитывая, что одна из целей создания системы — накопление и анализ больших объемов извлеченной из текстов информации, необходим алгоритм добавления новых получаемых результатов анализа к имеющимся в системе хранения.

Введение операций над результатами обработки существенно упрощает создание такого алгоритма. Он может быть реализован с помощью этих операций и автоматически применен к вновь получаемым результатам.

То есть алгоритм накопления использует средства системы анализа результатов и автоматически включается при обработке каждого текста. Это позволяет постоянно обновлять информацию в системе хранения и проводить анализ следующих текстов с учетом новой информации.

Система хранения

Система хранения информации включает в себя базу данных и все файлы, создаваемые при обработке автоматически или по указанию пользователя. Она должна иметь, во-первых, возможность хранения очень больших объемов данных,

во-вторых, максимально высокую скорость обращения к данным, выполнения запросов.

Структура системы хранения позволяет иметь широкий спектр данных об обработанных текстах, получаемых результатах анализа, информацию о пользователях и обработанных ими текстах, в ней хранятся наборы последовательностей операций над результатами обработки и результаты суммирования всех обработанных текстов в виде структур различных типов (списки слов, предложений, связей и т. д.).

Таким образом, база данных является хранилищем:

- извлекаемой из текстов информации;
- общей накопленной информации;
- служебной информации.

Под *извлекаемой* понимается информация, получаемая непосредственно при обработке текстов, а также на ее основе путем использования операций над результатами. *Общая накопленная информация* — результат работы аналитической системы накопления поступающей информации, совокупность всей полученной информации по каждому типу результатов. *Служебной информацией* в системе являются данные, необходимые для основных этапов обработки текста (например, морфологический словарь), информация о пользователях и обработанных ими текстах, а также создаваемые пользователями или разработчиками наборы операций для решения конкретных задач.

Пользовательская система

Взаимодействие пользователя с системой осуществляется через систему хранения и предоставляемый набор средств работы с текстами и промежуточными данными, полученными в ходе обработки. Возможно как использование имеющихся инструментов, так и создание собственных. Это могут быть самостоятельные программы, использующие результаты обработки, или части самой системы, алгоритмы применения предоставляемых операций.

Средства для решения конкретных задач могут реализовываться в виде отдельных программных модулей системы, а могут представлять собой набор операций над получаемыми результатами обработки. Такие наборы операций могут сохраняться и в дальнейшем представлять собой самостоятельный модуль для решения некоторой задачи.

Заключение

В статье предлагается модель системы анализа текста, ориентированной на обработку произвольных массивов текстовой информации, на решение как конкретных задач с использованием результатов автоматизированного анализа текста,

так и на накопление общей информации, извлекаемой из текстов.

Учет и хранение результатов, получаемых на всех этапах обработки, позволяет решать широкий круг отдельных практических задач путем введения дополнительных программных модулей, базирующихся на имеющихся алгоритмах обработки текста, использовании средств анализа на разных уровнях, применении операций над результатами обработки.

Основные принципы, на которых базируется система, отражают специфику работы с естественно-языковым текстом и позволяют учитывать его особенности.

Список литературы

1. **Аверкин А. Н., Гаазе-Рапорт М. Г., Поспелов Д. А.** Толковый словарь по искусственному интеллекту. [Электронный ресурс] URL: <http://www.raai.org/library/tolk/aivoc.html> (29.01.2009)
2. **Попов Э. В.** Общение с ЭВМ на естественном языке. Изд. 2-е, стер. М.: Эдиториал УРСС/2004. (Науки об искусственном).
3. **Налимов В. В.** Вероятностная модель языка. О соотношении естественных и искусственных языков. 2 изд., перераб. и доп. М.: Наука, 1979.
4. **Фрумкина Р. М.** Психоллингвистика. М.: Академия, 2003.
5. **Выготский Л. С.** Мышление и речь. Изд. 5, испр. М.: Лабиринт, 1999. 352 с.
6. **Саймон Г.** Науки об искусственном: Пер. с англ. Изд. 2-е. М.: Эдиториал УРСС, 2004. 144 с. (Науки об искусственном).
7. **Хоккинс Дж., Блейкли С.** Об интеллекте: Пер. с англ. М.: Вильямс, 2007. 240 с.
8. **Звегинцев В. А.** Очерки по общему языкознанию. М.: Изд-во МГУ, 1962.
9. **Серебренников Б. А.** Общее языкознание. М.: Наука, 1970.
10. **Тарасов В. Б.** Логико-лингвистические модели в искусственном интеллекте: прошлое, настоящее, будущее // Политехнические чтения: Сб. тр. Вып. 7. Искусственный интеллект — проблемы и перспективы. М.: Политехн. музей, 2006. 194 с.
11. **Леонтьева Н. Н.** Автоматическое понимание текстов: системы, модели, ресурсы: учеб. пособие для студ. лингв. фак. вузов. М.: Академия, 2006.
12. **Евдокимова И. С.** Естественно-языковые системы: курс лекций. Улан-Удэ: Изд-во ВСГТУ, 2006.
13. **Нечипоренко А., Русин А.** Система автоматизированного извлечения знаний из текстов на естественном языке // Труды Международной научно-технической конференции "Информационные системы и технологии — 2003". Новосибирск: НГТУ, 2003.
14. **Баталина А. М., Епифанов М. Е., Кобзарева Т. Ю.** и др. Опыт экспериментальной реализации алгоритмов поверхностно-синтаксического анализа // Компьютерная лингвистика и интеллектуальные технологии: Труды Международной конференции Диалог'2006. М.: Наука, 2006.
15. **Компьютерная лингвистика и интеллектуальные технологии:** Труды Международной конференции Диалог'2005. М.: Наука, 2005.
16. **Компьютерная лингвистика и интеллектуальные технологии:** Труды Международной конференции Диалог'2006. М.: Наука, 2006.
17. **Добрышина Е. В.** Основные проблемы и принципы автоматизированного определения смысла текста на естественном языке // XXXIII Гагаринские чтения. Научные труды Международной молодежной научной конференции в 8 т. Т. 6. М.: "МАТИ"-РГТУ, 2007. С. 221—222.
18. **Балакирев Н. Е., Гирин Б. Б.** Основные концепции языка описания алгоритмов анализа естественно-языкового текста // Информационные технологии. 2008. № 9. С. 42—47.

Д. Д. Голомазов, аспирант,
МГУ им. М. В. Ломоносова,
мл. науч. сотр. НИИ механики
МГУ им. М. В. Ломоносова,
e-mail: denis.golomazov@gmail.com

Выделение терминов из коллекции текстов с заданным тематическим делением

*Предлагается разработанный и программно реализованный автором статистический алгоритм *Brainstern* решения задачи выделения терминов из текстовых документов, разбитых на рубрики. Термином считается пара слов, характеризующая документ с позиции его принадлежности к одной или нескольким рубрикам. В работе приводятся результаты сравнения эффективности упомянутого алгоритма с аналогичными по назначению алгоритмами *TF-IDF* и *LSI* на задаче классификации. Для алгоритма *Brainstern* были получены высокие показатели точности классификации (до 85%), сравнимые с соответствующими показателями для алгоритма *LSI* и превосходящие результаты, полученные с использованием алгоритма *TF-IDF*. Высокая эффективность и скорость работы подтверждают практическую значимость предложенного алгоритма.*

Ключевые слова: выделение терминов, классификация текстов, кластеризация текстов, векторная модель, рубрикация текстов, навигация по документам, коллекции документов, автоматическая обработка, значимость

Введение

Задача выделения терминов является важной на направлении исследований, связанных с извлечением информации из текстовых документов. Актуальность этих исследований обусловлена быстро увеличивающимся объемом подобной информации, например, в крупных корпоративных информационных системах, в сети Интернет.

Под терминами понимаются слова или словосочетания, несущие информацию о тематике документа. Термины, или ключевые слова, представляют собой краткое *описание документа*, которое может быть использовано как непосредственно для выдачи пользователю, например в результатах поиска, так и для дальнейшей обработки, например в задачах классификации и кластеризации информации.

Задача выделения терминов привлекает внимание многих исследователей. Существует большое число алгоритмов, выделяющих термины-одиночные слова (их обзор представлен в работе [1]).

В настоящей работе предлагается новый алгоритм выделения терминов, получивший название *Brainstern*. С помощью этого алгоритма решается задача выделения терминов, состоящих из пар слов, из текстовых документов, разделенных на рубрики. Выбор в качестве терминов пар слов обусловлен тем обстоятельством, что пары слов могут описывать документ точнее, чем одиночные слова. Например, по отдельности слова "граф" и "теория" могут характеризовать различные тематические направления. Однако употребляемые вместе, они с большой вероятностью свидетельствуют о том, что данный документ посвящен математике.

Важной особенностью рассматриваемой задачи является требование разделения подлежащих анализу документов на рубрики. Выделенные термины зависят от исходных рубрик. Такое требование накладывает дополнительное условие на систему, использующую алгоритм *Brainstern*. Его суть в том, что документы, из которых нужно выделить термины, должны быть предварительно классифицированы. Для этого можно использовать обучающую выборку, составленную экспертом. Термины, извлеченные из документов, входящих в обучающую выборку, можно применять для классификации и кластеризации новых документов, а также для других задач. Требование рубрицируемости документов дает важное преимущество. Благодаря "относительному" характеру терминов, которые зависят от исходных рубрик, достигается их высокая информативность в рамках конкретной задачи. Это обстоятельство позволяет применять алгоритм *Brainstern* в узкоспециализированных областях, в которых применение алгоритмов выделения "абсолютных" (не зависящих от рубрик) терминов не позволяет получить должных результатов.

Термин — пара слов, характеризующая документ, в котором она встречается, с точки зрения его принадлежности к одной или нескольким рубрикам.

Заметим, что свойство "быть термином" зависит от контекста имеющихся рубрик. Например, пусть даны рубрики "наука", "искусство", "спорт". Тогда пара слов "доказательство теоремы" будет идеальным термином, так как документ, в котором эта пара встречается, почти со стопроцентной вероятностью принадлежит рубрике "наука". Рассмотрим пример других рубрик, а именно — "алгебра", "геометрия", "топология". В контексте данных рубрик пара "доказательство теоремы" уже не будет термином, так как такая пара характерна для всех рубрик, и судить о тематике документа по наличию в нем этой пары нельзя.

Необходимо отметить, что, несмотря на большое число работ, посвященных рассматриваемо-

му классу задач, а также проделанный автором обширный анализ источников, результатов исследований, содержащих решение задачи, которая поставлена в данной работе, обнаружить не удалось. Большинство существующих методов направлены на выделение терминов из документов, не разделенных на рубрики (разделение, если оно есть, не используется), либо позволяют выделять только одиночные слова. В связи с отмеченным обстоятельством провести прямое сравнение алгоритма *Brainsterm* с полностью аналогичными методами не представляется возможным. В рамках данного исследования было проведено тестирование, формально оценивающее качество терминов с точки зрения задачи классификации документов по тематике. Заметим, что эта задача играет ключевую роль во многих современных системах обработки данных. Проведено сравнение эффективности алгоритма *Brainsterm* с двумя широко распространенными в области обработки данных алгоритмами TF-IDF [2] и LSI [3], которые, и это следует отметить, не учитывают разделения исходных документов на рубрики.

Для тестирования использовалась векторная модель представления документов. Результат работы каждого из перечисленных алгоритмов, включая разработанный автором, можно рассматривать как базис векторного пространства, состоящий из терминов. Качество этого базиса формально оценивалось с помощью алгоритма классификации. Результаты тестирования показывают, что эффективность алгоритма *Brainsterm* сравнима с эффективностью алгоритма LSI и превышает показатели TF-IDF. Методика тестирования описана в разделе 3.1, результаты тестирования приведены в разделе 3.2. Подробное описание алгоритма *Brainsterm* приведено в разделе 2.2.

1. Существующие подходы к решению задачи

В контексте рассматриваемой в настоящей работе задачи можно выделить два популярных подхода к ее решению, которые принято именовать лингвистическим и статистическим.

Лингвистический подход означает выделение ключевых слов и фраз на основе морфологии и семантики текста. В рамках этого подхода широко используются базы семантических связей между словами (например, база WordNet¹), словари синонимов, средства морфологического разбора и анализа. Самые известные разработки на основе лингвистического подхода — это системы TERMS [4] и LEXTER [5]. Основным проблемным вопросом при реализации лингвистических методов является "шум", а именно — большое число слово-

сочетаний, характерных для естественного языка, но не являющихся терминами.

Статистический подход подразумевает оценивание важности слова на основе информации о числе его вхождений в документы. Для этого чаще всего используют модели на основе TF-IDF — нормализованного веса слова. В рамках статистического подхода широко применяют средства, основанные на вероятностных моделях [6, 7]. С одной стороны, это позволяет использовать мощный математический аппарат и получать теоретические доказательства эффективности алгоритмов. С другой стороны, большинство таких подходов имеют один существенный недостаток. Его суть в том, что они опираются на предположение о независимости появления слов в предложении. Для естественного языка такая гипотеза является слишком сильным условием. Например, слова "теорема" и "доказывать" являются гораздо более зависимыми, чем слова "теорема" и "петь".

Другим методом в рамках статистического подхода является формулирование "базовых требуемых свойств", на основе которых строится критерий отбора терминов. Именно этот метод используется в алгоритме *Brainsterm*.

В отличие от лингвистического подхода, основным проблемным вопросом статистического подхода является "тишина" — ситуация, при которой термины, состоящие из одного слова, остаются незамеченными системой.

В исследованиях последних лет статистические и лингвистические методы часто комбинируются. Здесь можно выделить систему ACABIT [8], а также TRUCKS [9].

2. Алгоритм *Brainsterm*

В настоящем разделе вводится формальная модель, в рамках которой определяется алгоритм *Brainsterm*, а также приводится описание алгоритма *Brainsterm*, который условно можно разделить на четыре этапа.

2.1. Описание модели

Пусть W — множество всех слов, встречающихся во всех документах заданной коллекции, включая ε — пустое слово, а PW — множество всех пар слов, т. е. $PW = W \times W$. Определим документ d как отображение $d: \mathbb{N} \rightarrow W$, ставящее в соответствие каждому натуральному числу n слово, стоящее на n -й позиции в данном документе коллекции. Номера позиций, на которых нет слов (после конца документа), отображаются в пустое слово. Обозначим множество документов коллекции через D .

Аналогично определим абзац p как отображение $p: \mathbb{N} \rightarrow W$, ставящее в соответствие каждому натуральному числу n слово, стоящее на n -й по-

¹ <http://wordnet.princeton.edu>

зиции в данном абзаце. Номера позиций, на которых нет слов, отображаются в пустое слово. Обозначим множество всех абзацев в коллекции через P .

Определим рубрику r как произвольное подмножество множества документов, а именно — $r \in 2^D$. Мощность рубрики, как число документов в ней, будем обозначать через $|r|$. Обозначим также множество всех заданных рубрик через R .

Определим еще несколько вспомогательных отображений:

- $\tau_1: PW \rightarrow W, \tau_2: PW \rightarrow W$ — проекции пары на множество слов, ставящие в соответствие паре первое (соответственно, второе) слово пары;
- $Freq: PW \times D \rightarrow \mathbb{N} \cup \{0\}$ — функция, определяющая число вхождений пары $pw \in PW$ в документ $d \in D$;
- $Freq: W \times D \rightarrow \mathbb{N} \cup \{0\}$ — функция, определяющая число вхождений слова $w \in W$ в документ $d \in D$;
- $L(d) = |\{n \in \mathbb{N} | d(n) \neq \varepsilon\}|$ — длина документа d ;
- $id(a) = a$ — тождественное отображение;
- $Av(f, A) = \frac{\sum_{a \in A} f(a)}{|A|}$ — среднее значение функции f на конечном множестве A .

Например, $Av(|\cdot|, R)$ — среднее число документов в рубрике; $Av(L, D)$ — средняя длина документа; $Av(id, A)$ — среднее арифметическое чисел из множества $A = \{a_1, \dots, a_k\}$.

2.2. Описание алгоритма

Алгоритм *Brainsterm* включает четыре этапа. На каждом из них с помощью некоторого правила выбирается подмножество M_i множества M_{i-1} , полученного на предыдущем шаге. На первом этапе выбор осуществляется из множества PW (всех пар слов), т. е. $M_0 = PW$. Множество M_4 и есть термины — пары, удовлетворяющие всем четырем критериям.

Пространственный критерий. Первоначальный выбор пар для последующей обработки основывается на предположении о том, что слова, образующие термин, расположены в тексте достаточно близко (хотя и не обязательно рядом):

$$M_1 = \{pw \in M_0 | \exists p \in P: |p^{-1}(\tau_1(pw)) - p^{-1}(\tau_2(pw))| \leq \text{MAX_DIST}\}.$$

Таким образом, пару образуют два слова, находящиеся в одном абзаце, между которыми в тексте стоят не более $\text{MAX_DIST}-1$ других слов.

Критерий частотности. Как отмечалось ранее, выделенные в результате работы алгоритма пары слов можно рассматривать как базис векторного пространства текстов. Для того, чтобы такой базис был информативным, в нем должны быть сло-

ва, достаточно часто встречающиеся в документах. С целью обеспечить это требование используется критерий частотности, исключающий из множества M_1 пары, которые встретились во всей коллекции меньше, чем MIN_FREQ раз:

$$M_2 = \{pw \in M_1 | \sum_{r \in R} \sum_{d \in r} \text{Freq}(pw, d) \geq \text{MIN_FREQ}\}.$$

Критерий характерности. Критерий характерности — основной критерий алгоритма. Его суть заключается в определении термина: "пара должна быть характерной для некоторых рубрик".

Вес пары в рубрике. Каждой паре соответствует набор чисел — весов пары в каждой из рубрик. Вес пары pw в рубрике r вычисляется по следующей формуле:

$$\text{Weight}(pw, r) = \frac{\sqrt{\sum_{d \in r} \ln \left(\frac{\sqrt{\text{Freq}(pw, d)}}{\ln \left(\frac{L(d)}{Av(L, D)} + 1 \right) + 1 \right)}}{\ln \left(\frac{|r|}{Av(|\cdot|, R)} + 1 \right)}.$$

Аналогично определяется вес слова в рубрике $\text{Weight}(w, r)$: в приведенной выше формуле $\text{Freq}(pw, d)$ заменяется на $\text{Freq}(w, d)$.

Функция характерности пары. К вычисленному набору весов пары в рубриках применяется функция характерности *Discr*, которая ставит в соответствие набору неотрицательных действительных чисел $A = \{a_1, \dots, a_n\}$, не равных одновременно нулю, число — "показатель характерности набора":

$$\text{Discr}(a_1, a_2, \dots, a_n) = 1 - \frac{Av(id, A)}{\max_{a \in A} a(1 + \ln(1 + \max_{a \in A} a))}.$$

Эта функция принимает значения из отрезка $[0, 1]$, и чем более характерным является набор с точки зрения определения термина, тем ближе значение этой функции к 1. На рис. 1 показаны примеры параметров и значений функции *Discr* на них.

Пусть $R = \{r_1, \dots, r_n\}$, тогда $M_3 = \{pw \in M_2 |$

$$\text{Discr}(\text{Weight}(pw, r_1), \dots, \text{Weight}(pw, r_n)) \geq \text{MIN_DISCR}\}.$$

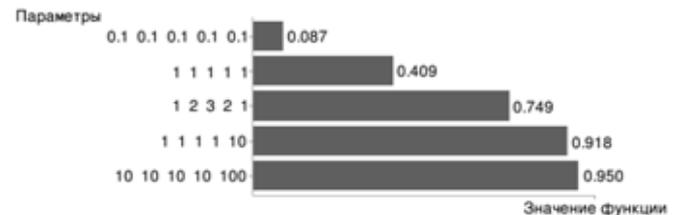


Рис. 1. Функция характерности *Discr*

Таким образом, отбираются пары, значение функции *Discr* на которых не меньше константы MIN_DISCR, принадлежащей отрезку [0, 1].

Критерий значимых рубрик. Идея критерия значимых рубрик состоит в том, что слова, образующие термин, должны встречаться достаточно часто в паре и сравнительно редко по отдельности. В частности, ее учет позволяет исключить следующую возможную ошибку. Пусть в документах рубрики некоторое число раз встретилось слово, уникальное для данной рубрики (например, "гидрофосфат" в рубрике "химия"). Тогда все слова, находящиеся на небольшом расстоянии от этого слова, скорее всего будут удовлетворять критерию характерности, так как в других рубриках эти пары не встречаются вообще. Вместе с тем, многие из них, очевидно, не являются терминами (например, "гидрофосфат считается" или "имея гидрофосфат").

Определение. Рубрика *r* называется **значимой** для пары *pw*, если

$$Weight(pw, r) \geq Av(Weight(pw, \cdot), R).$$

Пусть $2^R = \{A \mid A \subseteq R\}$. Определим отображение $imp: M_3 \rightarrow 2^R$:

$$imp(pw) = \{r \in R \mid Weight(pw, r) \geq Av(Weight(pw, \cdot), R)\}$$

(каждой паре соответствует множество значимых для нее рубрик).

Тогда $M_4 = \{pw \in M_3 \mid$

$$\min_{r \in imp(pw)} \left(\frac{Weight(\tau_1(pw), r)}{Weight(pw, r)} + \frac{Weight(\tau_2(pw), r)}{Weight(pw, r)} \right) \leq MAX_FREQ_RATIO \}.$$

Таким образом, отбираются пары, у которых среди значимых для них рубрик найдется рубрика, в которой сумма весов каждого из слов пары превышает вес пары не более, чем в MAX_FREQ_RATIO раз.

3. Практическая апробация алгоритма

Алгоритм *Brainsterm* реализован на языке C++. Программа составляет более 1200 строк. На языке C++ также написана программа анализа эффективности алгоритмов *Brainsterm*, LSI и TF-IDF. Она включает в себя реализацию алгоритмов LSI и TF-IDF и составляет около 1700 строк, не считая используемой сторонней библиотеки *svd²* (около 7500 строк), реализующей сингулярное разложение матрицы, необходимое для алгоритма LSI.

² <http://www.alglib.net/matrixops/general/svd.php>

3.1. Методика тестирования

Процедура анализа эффективности алгоритмов состоит в следующем. Рассматриваются две выборки документов, разделенных на рубрики, одна из которых обучающая, другая тестовая. Выборки содержат приблизительно равное число документов, рубрики в обеих выборках одни и те же. Затем с помощью обучающей выборки проводится обучение системы, в ходе которого выделяются термины, образующие базис евклидова пространства наперед заданной размерности. На следующем шаге документам обучающей и тестовой выборок ставятся в соответствие точки в пространстве на основе отмеченного базиса, у которых *i*-я координата равна числу вхождений пары-термина с номером *i* в документ. Таким образом, из *N* документов получаются *N* точек, которые разделены, с одной стороны, на два множества — обучающую и тестовую выборки (каждое множество образовано точками-образами документов из соответствующей выборки), а с другой стороны, на кластеры — образы исходных рубрик. Кластер, соответствующий некоторой рубрике, образован точками-образами документов этой рубрики. Анализ эффективности алгоритма заключается в тестировании конфигурации кластеров с помощью метода классификации.

С помощью алгоритмов LSI и TF-IDF выполняется аналогичная операция: документы отображаются в точки, а затем получившаяся конфигурация кластеров формально оценивается с помощью алгоритма классификации.

Оценка с помощью алгоритма классификации.

Проводится классификация точек тестовой выборки с помощью метода "К ближайших соседей". Для ее реализации используется разбиение точек обучающей выборки на кластеры. В результате классификации каждая точка тестовой выборки попадает в некоторый кластер, который является образом одной из исходных рубрик. При этом точка представляет документ, который сам изначально относится к одной из рубрик. Таким образом, в результате классификации точек документы тестовой выборки еще раз распределяются по рубрикам. Далее проверяется, насколько это распределение совпадает с исходным распределением документов по рубрикам. Для этого вычисляется процентное отношение числа документов тестовой выборки, попавших после классификации в "свою" рубрику (то есть документов, отнесенных в результате классификации к той же рубрике, к которой они принадлежали изначально). Чем больше полученное число (которое находится в промежутке от 0 до 100 %), тем эффективнее рассматриваемый алгоритм.

3.2. Результаты тестирования

В целях сравнения алгоритмов *Brainsterm*, LSI и TF-IDF была использована выборка, содержащая около 1,4 млн слов в 7 тыс. документов. Обучающая выборка содержала 3591 документ, остальные документы составили тестовую выборку. Заметим, что алгоритм LSI не допускает размерности выше, чем число документов в обучающей выборке. Именно поэтому на иллюстрациях графики модификаций алгоритмов LSI обрываются на точке 3591.

Производительность. Общее представление о сравнении производительности рассматриваемых алгоритмов можно получить по времени их работы в ходе тестовых испытаний. Параметры вычислительной установки: CPU AMD Opteron 2 Ghz, 8 Gb RAM.

На рис. 2 представлены показатели времени работы алгоритмов *Brainsterm*, LSI, TF-IDF. Каждый алгоритм был запущен с параметром размерности целевого пространства, имеющим оптимальное значение, при котором данный алгоритм достигает максимальных результатов точности классификации. Для алгоритма *Brainsterm* это значение равно 10000, для алгоритма LSI—250. Алгоритм TF-IDF не позволяет выбрать размерность целевого пространства. Фактически она равна общему числу уникальных слов в коллекции, в данном случае — около 21000.

Как следует из графика, алгоритмы *Brainsterm* и TF-IDF работают гораздо быстрее алгоритма LSI. Отметим, что при тестировании на имеющейся выборке алгоритму *Brainsterm* потребовалось около 1 Гбайт оперативной памяти, алгоритму TF-IDF — около 550 Мбайт, а алгоритму LSI — около 2,2 Гбайт.

Точность классификации. На рис. 3 приведены результаты сравнения точности классификации. Их анализ позволяет сделать следующие выводы.

1. Максимальный результат алгоритма *Brainsterm* (85,8 %) превосходит результаты всех остальных алгоритмов. Следует отметить, что этот результат был достигнут на достаточно высокой размерности — 10000.

2. На размерностях ниже 2000 алгоритм *Brainsterm* показывает достаточно низкие результаты, уступающие показателю TF-IDF (73,5 %).

3. Алгоритмы LSI с матрицей частот и булевой матрицей эффективны на размерностях 50...3000, при дальнейшем увеличении размерности их эффективность падает.

4. Точность классификации, полученная при использовании алгоритма LSI с матрицей из TF-IDF, гораздо ниже показателей, полученных с помощью остальных алгоритмов.



Рис. 2. Сравнение времени работы алгоритмов

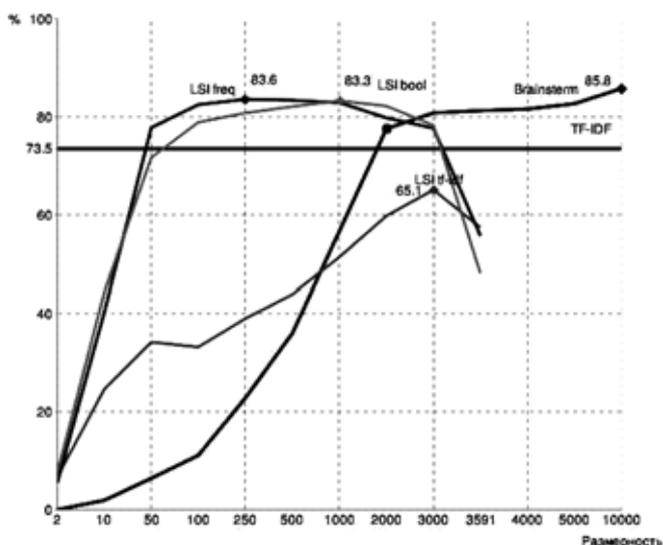


Рис. 3. Точность классификации

5. В целом точность классификации для алгоритма *Brainsterm* на размерностях от 2000 сопоставима с максимальными показателями алгоритмов LSI.

Заключение

В настоящей работе предложен статистический алгоритм *Brainsterm* решения задачи выделения терминов из текстовых документов, предварительно разделенных на рубрики. Разработана методика анализа, с применением которой проведено исследование эффективности алгоритма путем сравнения его с другими методами, широко используемыми в области обработки текстовых данных. Как показал анализ, с позиций эффективности результаты работы алгоритма *Brainsterm* сравнимы с показателями алгоритма LSI и превосходят аналогичные показатели алгоритма TF-IDF. Изложенные факты подтверждают практическую применимость алгоритма *Brainsterm* и указывают на возможность его использования в составе систем классификации и поиска по текстовым документам.

Работа поддержана грантом РФФИ № 09-07-00366-а.

Список литературы

1. Kageura K., Umino B. Methods of automatic term recognition: A review. // Terminology. 1996. Vol. 3, N 2. P. 259—289.

2. **Salton G., Fox E. A., Wu H.** Extended boolean information retrieval // Commun. ACM. 1983. Vol. 26. N 11.
3. **Deerwester S., Dumais S. T., Furnas G. W., Landauer T. K., Harshman R.** Indexing by latent semantic analysis // Journal of the American society for information science. 1990. N 41. P. 391—407.
4. **Justeson J. S., Katz S. M.** Technical terminology: Some linguistic properties and an algorithm for identification in text // Natural Language Engineering. 1995. Vol. 1.
5. **Bourigault D.** Lexter: A terminology extraction software for knowledge acquisition from texts // Proc. 9th Banff Knowledge Acquisition for Knowledge Based Systems Workshop (Banff, Canada). 1995. Vol. 1.

6. **Lin D.** Extracting Collocations from Text Corpora // First Workshop on Computational Terminology. 1998. P. 57—63.
7. **Amati G., Carpineto C., Romano G., Roma F. U. B.** Fub at tree-10 web track: A probabilistic framework for topic relevance term weighting // Proc. of the 10th Text Retrieval Conference TREC 2001 (Gaithersburg, Maryland). NIST Special Publication. 2001. P. 182—191.
8. **Daille B.** Repérage et extraction de terminologie par une approche mixte statistique et linguistique // Traitement Automatique des Langues (T. A. L.). 1995. Vol. 36. N 1—2.
9. **Maynard D.** Term Recognition Using Combined Knowledge Sources // PhD thesis, Department of Computing and Mathematics, Manchester Metropolitan University, Manchester (UK). 1999.

УДК 004.4'244, 004.436.4.

П. А. Шапкин, аспирант, ассистент,
Национальный исследовательский
ядерный университет "МИФИ",
e-mail: p.shapkin@gmail.com

Модели и методы разработки веб-приложений на основе онтологии предметной области

Описываются модели и методы, применимые для обработки информации, представленной в виде RDF-графов и оснащенной онтологией предметной области. Модели строятся таким образом, чтобы приблизить структуру системы к механизмам, используемым при обработке данных в формате XML, а также к принципам объектно-ориентированного программирования. Рассматривается архитектура информационной системы, использующей описываемые подходы.

Ключевые слова: онтологии, веб-программирование, дескрипционная логика

Введение

Развитие Интернет-технологий привело к появлению новых форматов, направленных на представление смысла, или семантики данных. Данные форматы составляют основу инициативы "семантического Интернета".

Для представления данных с учетом их семантики используется язык RDF [1], позволяющий описывать данные в виде семантических сетей, строящихся из троек *<объект, атрибут, значение>*. При этом одна и та же сеть может быть по-разному записана на языке RDF, но одинаково воспринята при интерпретации этих записей. Эта особенность делает RDF более подходящим средством для описания данных в среде Веб, чем XML [2]: в XML также существуют различные способы описания семантически эквивалентных сущностей, однако не существует способов установления данной эквивалентности при интерпретации этих описаний.

При использовании RDF для представления данных схема данных может быть определена в виде онтологии на языке OWL [3]. Онтология представляет собой формальное описание предметной области в виде определений используемых понятий (концептов) и их свойств (ролей).

Концепция семантического Интернета получила большое развитие, но все еще не нашла широкого применения. Во многом это связано с дефицитом инструментария, необходимого для разработки приложений, полностью ориентированных на семантическое представление данных. В данной статье описываются модели и методы, с помощью которых проводится разработка веб-ориентированной информационной системы, на всех этапах использующей формат RDF для представления данных. Настройка системы на предметную область осуществляется с помощью онтологий.

Статья организована следующим образом: в разделе 1 описаны принцип построения модели и метамодели предметной области на основе онтологии; в разделе 2 приведены модели и методы, используемые при обработке онтологии и RDF-данных; в разделе 3 представлена структура системы, основанной на принципах, описываемых в данной работе.

1. Онтологическое представление модели предметной области

1.1. Онтологии и их формальные основы

Каждая информационная система позволяет выполнять операции над некоторым множеством объектов (сущностей) Δ , используемых для представления информации. Онтология представляет собой формальную модель предметной области, в которой типы сущностей описываются в виде *концептов* (понятий). Концепты определяются с помощью *дескрипций* (описаний). Каждому концепту ставится в соответствие некоторый предикат на множестве Δ , а следовательно, некоторое мно-

жество значений, удовлетворяющих этому предикату. Функция, реализующая такое соответствие, называется интерпретацией и обозначается как $(\cdot)^I$. В общем случае множество Δ зависит от выбранной интерпретации, что обозначается как Δ^I .

В основе языка описания онтологий OWL лежит формальная система, называемая дескрипционной логикой (ДЛ) [4]. Дескрипционная логика определяет правила построения дескрипций концептов. Помимо концептов в рассмотрение вводятся роли и индивиды. Роли понимаются как типы связей между объектами предметной области, интерпретация ставит в соответствие каждой роли множество упорядоченных пар значений из Δ . Индивиды трактуются как объекты предметной области; интерпретацией индивида является отдельное значение из Δ . При этом делается соглашение о том, что различные индивиды имеют различные интерпретации.

Дескрипции в ДЛ строятся по следующей схеме:

- базовыми строительными блоками являются универсальный и пустой концепты, атомарные концепты и роли, а также индивиды;
- более сложные концепты и роли строятся из атомарных с помощью конструкторов концептов и ролей.

Универсальный и пустой концепты, обозначаемые \top и \perp , описывают соответственно любую сущность и пустую. Атомарные концепты — первичные, неделимые концепты, специфичные для каждой предметной области.

Существуют различные варианты дескрипционных логик, отличающиеся предоставляемым набором конструкторов. Язык OWL DL основан на разновидности под названием $\mathcal{SHOIN}(\mathcal{D})$. Списки основных конструкторов концептов, определяемых для $\mathcal{SHOIN}(\mathcal{D})$, представлены в таблице. Помимо конструкторов, перечисленных в табли-

це, $\mathcal{SHOIN}(\mathcal{D})$ включает в себя конструкторы, реализующие работу с простыми ("конкретными") типами данных, такими как строки, числа и т. п. Для каждого конкретного типа может быть задан набор дополнительных предикатов, например, предикаты сравнения чисел и т. п.

На концептах задаются отношения эквивалентности и вложенности. Два концепта C и D считаются эквивалентными, что записывается как $C \equiv D$, тогда и только тогда, когда $C^I = D^I$ для всех I . Концепт C называется вложенным в D , когда $C^I \subseteq D^I$ для всех I . Вложенность концептов реализует родовидовые отношения: видовые концепты "наследуют" все свойства родовых, т. е. являются вложенными в них.

Наибольший интерес представляют те разновидности ДЛ, в которых отношение вложенности является вычислимым, т. е. существует процедура, проверяющая истинность либо ложность формулы $C \subseteq D$ для любых допустимых дескрипций концептов C и D в любой интерпретации. Особенность $\mathcal{SHOIN}(\mathcal{D})$ состоит в том, что при высокой выразительности она сохраняет вычислимость вложенности [5]. Из вычислимости вложенности концептов следует разрешимость других связанных задач, таких как проверка противоречивости концепта ($C \subseteq \perp$) и проверка несовместимости пары концептов ($C \cap D = \perp$).

Тогда как вложенность концептов схожа с наследованием классов в объектно-ориентированном программировании (ООП), вычислимость данного отношения включает в себе основное отличие онтологий от ООП: иерархия наследования не обязательно должна быть задана разработчиками, она может быть вычислена в любой момент.

1.2. Составление онтологии предметной области

Представление модели предметной области в виде онтологии дает ряд преимуществ. Во-первых, онтологии предоставляют большие возможности для описания метаданных. Системы, широко использующие метаданные модели предметной области, являются легко настраиваемыми. Для внесения изменений в такие системы во многих случаях требуются лишь небольшие изменения модели предметной области.

Во-вторых, онтологии поддерживают расширение и объединение. При составлении модели можно либо основываться на уже существующих онтологиях, либо разрабатывать онтологию для конкретной предметной области с нуля. Но и в последнем случае возможно позднее объединение разработанной онтологии со схожими онтологиями. Для этого требуется составить определения эквивалентных концептов среди пере-

Список конструкторов концептов в ДЛ $\mathcal{SHOIN}(\mathcal{D})$

Название	Обозначение	Интерпретация
Атомарное понятие	C, D, \dots	Фиксированное подмножество Δ
Универсальное понятие	\top	Δ^I
Пустое понятие	\perp	\emptyset
Отрицание	$\neg A$	$D^I \setminus C^I$
Пересечение	$C \cap D$	$C^I \cap D^I$
Объединение	$C \cup D$	$C^I \cup D^I$
Ограничение существования	$\exists R.C$	$\{a \in \Delta^I \mid \exists b.(a, b) \in R^I \wedge b \in C^I\}$
Ограничение диапазона	$\forall R.C$	$\{a \in \Delta^I \mid \forall b.(a, b) \in R^I \rightarrow b \in C^I\}$
Ограничение численности	$\geq nR$ и $\leq nR$	$\{a \in \Delta^I \mid \{b \mid (a, b) \in R^I\} \geq n\}$ и $\{a \in \Delta^I \mid \{b \mid (a, b) \in R^I\} \leq n\}$, где $ M $ — мощность множества M

секающихся онтологий и указать их эквивалентность в явном виде.

При разработке онтологии предметной области каждый тип сущностей описывается в виде концепта, атрибуты и связи сущностей представляются ролями. Для построения системы, основанной на онтологии предметной области, в данную модель необходимо внести определенный набор метаданных. Основная задача — выделение особых сущностей, которые будем называть *ресурсами*. Они отличаются тем, что при каждом обращении к системе пользователь работает с каким-либо определенным ресурсом или набором однотипных ресурсов, т. е. это те классы объектов, к которым система предоставляет непосредственный доступ для просмотра и редактирования. Ресурсы могут разделяться на непересекающиеся типы. Типы ресурсов могут различаться методами хранения: одни ресурсы могут храниться в базе данных (БД), другие — в файловой системе, третьи могут быть доступны через внешние службы. Кроме того, система может предоставлять пользователям различные наборы действий для работы с ресурсами различных типов.

Метаданные указанного вида вносятся в модель предметной области путем наследования концептов различных сущностей от концептов соответствующих типов ресурсов. А в зависимости от того, от каких концептов наследуется концепт ресурса, система может обрабатывать его тем или иным образом. Для связывания операций, требуемых для обработки объектов с концептами, к которым они относятся, предлагается применять *системы шаблонов*.

2. Принцип обработки RDF-данных на основе шаблонов

Обработка данных подразумевает либо их преобразование для получения производных данных или представления исходных данных в другом формате, либо выполнение действий, управляемых обрабатываемыми данными. Обе эти задачи можно свести к вычислению некоторой функции, зависящей от природы обрабатываемых данных: при преобразовании результатом вычисления будут являться новые данные, во втором случае — выполнение некоторых действий. Для построения метода обработки данных в формате RDF будем отталкиваться от задачи преобразования, которая потом будет обобщена и на задачу выполнения действий. Разрабатываемая модель схожа с методом преобразования данных в формате XML с использованием шаблонов XSL [6], поэтому будем называть ее моделью обработки RDF-данных на основе шаблонов.

Преобразование RDF-данных может быть построено по тому же принципу, что и язык XSL. Любое преобразование может быть определено в виде набора шаблонов, представляющих собой кортежи $\langle c, e \rangle$, где c — концепт, определяющий класс объектов, к которым применим данный шаблон, e — функция, проводящая преобразование, соответствующее данному шаблону. Для удобства будем пользоваться функцией *concept*, возвращающей концепт, связанный с заданным шаблоном.

В общем случае система шаблонов $TS \in TemplateSystem(InputType, OutputType)$, где *InputType* — тип входных данных, *OutputType* — тип результата обработки, представляет собой кортеж $\langle ts, e \rangle$, в котором ts — множество шаблонов, а $e \in InputType \rightarrow OutputType$ — функция, проводящая преобразование входного объекта в выходной. Зададим проекцию *templates*, возвращающую список шаблонов заданной системы шаблонов.

Для целей создания информационной системы, обрабатывающей семантическую информацию, требуется наличие шаблонных систем двух типов: первый тип систем (*концептуальные системы*) принимает на вход концепты, второй (*индивидуальные системы*) — индивиды. При этом на вход как тех, так и других систем может подаваться набор дополнительных параметров.

Следует заметить, что описанный принцип обработки информации схож с подходом ООП: шаблоны связываются с концептами так же, как методы связываются с классами, они представляют собой операции, выполняемые одинаковым образом для некоторого множества объектов предметной области.

При применении системы шаблонов к входному объекту должен выполняться шаблон, концепт которого наиболее близок к входному объекту. В случае концептуальной системы шаблонов входной концепт должен быть вложен в концепт выбранного шаблона, в случае индивидуальной системы входной индивид должен принадлежать концепту выбранного шаблона. При этом встает вопрос разрешения конфликтов, т. е. выбора наиболее подходящего шаблона в том случае, когда для обработки входного объекта подходят сразу несколько шаблонов.

2.1. Алгоритм выбора шаблона

Требуется описать функцию *choose(i, ts)*, осуществляющую выбор шаблона, соответствующего индивиду i из системы шаблонов ts . Прежде всего требуется наложить некоторые ограничения на допустимые системы шаблонов.

Определение. Система шаблонов T называется *правильно сформированной*, если выполняется следующее условие:

$$\forall t_1, t_2 \in \text{templates}(T). \text{concept}(t_1) \subseteq \text{concept}(t_2) \vee \\ \vee \text{concept}(t_1) \supseteq \text{concept}(t_2) \vee \text{concept}(t_1) \cap \\ \cap \text{concept}(t_2) = \emptyset.$$

Данное ограничение говорит о том, что в правильно сформированной системе не должно быть шаблонов, концепты которых пересекаются между собой: для любых двух шаблонов связанные с ними концепты должны быть либо вложенными один в другой, либо быть несовместимыми.

Дальнейшие рассуждения будем строить для индивидуальной системы шаблонов. Для применения их к случаю концептуальной системы шаблонов везде, где используется операция проверки принадлежности индивида концепту \in , должна применяться операция проверки вложенности концептов \subseteq .

Определение. Глубиной шаблона τ в системе шаблонов T будем называть значение функции $\text{depth}(\tau, T): \text{Template} \times \text{TemplateSystem} \rightarrow \mathbb{N}_0$, где $\mathbb{N}_0 = \{0, 1, 2, \dots\}$ — множество неотрицательных целых чисел, вычисляемой следующим образом:

1) если $\nexists t \in \text{templates}(T). \text{concept}(\tau) \subset \text{concept}(t)$, то $\text{depth}(\tau, T) = 0$;

2) если $\exists t \in \text{templates}(T). \text{concept}(\tau) \subset \text{concept}(t)$, то $\text{depth}(\tau, T) = \text{depth}(t, T) + 1$.

Теорема. В правильно сформированной системе шаблонов, если $\tau = \text{choose}(i, T) = \max_{\lambda, t. \text{depth}(t, T)} \{t \mid i \in \text{concept}(t) \& t \in \text{templates}(T)\}$, то

1) данный шаблон единственен;

2) данный шаблон является наиболее близким к индивиду i , т. е. не существует шаблона, концепт которого более специфичен, чем у τ , и при этом содержит i ; т. е. $\forall t \in \text{templates}(T). i \in \text{concept}(t) \Rightarrow \tau \subset \text{concept}(t) \vee t = \tau$.

Доказательство. Докажем первое утверждение: шаблон, выбираемый с помощью описанной функции choose , единственен. Если бы это было не так, то существовало бы несколько шаблонов с равной глубиной, концептам которых принадлежит i . Рассмотрим любые два из них. По условию правильно сформированной системы шаблонов концепты этих шаблонов могут либо включаться один в другой, либо быть несовместимыми. Так как i входит в оба концепта, несовместимыми они быть не могут. Следовательно, один из них должен быть включен в другой. Но в таком случае они имеют разную глубину, что противоречит предположению о том, что их глубины равны другу. Утверждение доказано.

Докажем от противного второе утверждение. Пусть $\exists t \in \text{templates}(T). i \in \text{concept}(t) \& t \neq \tau \& t \subset \tau$. Так как $t \subset \tau$, то $\text{depth}(T, t) > \text{depth}(T, \tau)$; но по ус-

ловию теоремы $\text{depth}(T, \tau)$ максимальна — противоречие. ■

Доказанная теорема позволяет составить алгоритм выбора наиболее подходящего шаблона для обработки конкретного индивида или концепта. Основная задача состоит в том, чтобы просматривать шаблоны в порядке убывания глубины. Первый из просматриваемых шаблонов, концепт которого включает в себя объект обработки, и является наиболее подходящим. Это говорит об адекватности данного механизма поставленной задаче.

2.2. Проверка непротиворечивости систем шаблонов

Рассмотрим некоторую систему шаблонов S . При разработке информационной системы, работающей на системе S , важным свойством является наличие гарантии того, что при функционировании не возникнет ошибок. Некоторые виды ошибок могут быть выявлены до выполнения системы, на основе статического анализа структуры шаблонов. Под проверкой системы на непротиворечивость будем подразумевать следующие операции:

- проверка существования шаблонов для визуализации всех ресурсов;
- проверка существования шаблонов для обработки всех действий с БД для всех ресурсов.

Указанные проверки позволяют получить гарантию, что система не выдаст ошибок, если пользователем введен корректный запрос. Эти проверки используют структуру онтологии предметной области и выполняются путем сопоставления с ней используемых систем шаблонов.

Кроме того, возможно осуществлять проверку индивидуальных шаблонов на предмет использования только тех ролей индивидов, которые связаны с концептом шаблона. В противном случае может произойти ошибка времени выполнения, так как обрабатываемый индивид может не иметь требуемой роли.

Подобная проверка может быть осуществлена с помощью тестового применения каждого шаблона к *индивиду-заглушке* для его концепта. При попытке обращения к роли R для заглушки m возвращается некоторое тестовое значение, если $m \in \exists R.T$, в противном случае генерируется ошибка. Если при применении шаблона к заглушке ошибок не произошло, его можно считать непротиворечивым.

3. Реализация информационной системы на основе шаблонного подхода

3.1. Структура системы

За основу для построения системы взят подход Модель-Вид-Контроллер (Model-View-Control-



Алгоритм обработки запроса диспетчером запросов

ler, MVC) [7], используемый во многих веб-каркасах, таких как Apache Struts, Spring MVC, Ruby on Rails и т. п. Система взаимодействует с окружающим миром через отдельный объект, называемый *диспетчером запросов*. Диаграмма алгоритма обработки запроса диспетчером приведена на рисунке, он включает в себя следующие действия:

- анализ запроса;
- выполнение операций с БД;
- визуализация результата;
- возвращение результата пользователю.

Непосредственно самим диспетчером выполняются только анализ запроса и возвращение результата пользователю, для выполнения остальных операций управление передается другим объектам: *исполнителю операций с БД* и *визуализатору*.

При анализе запроса рассматриваются запрошенные пользователем URL и метод HTTP и определяются ресурс и действие, которые над ним необходимо выполнить, а также формат, в котором пользователь ожидает получить ответ. Если разобрать запрос не удалось, то пользователю возвращается сообщение об ошибке.

В случае удачного разбора запроса результат передается исполнителю операций с БД, результатом работы которого является некоторый информационный объект, представленный индивидом в RDF-графе. Данный индивид передается одному из визуализаторов — тому, который соответствует запрошенному пользователем формату представления данных. Обработанный визуализатором результат возвращается пользователю в качестве ответа.

Исполнители операций с БД и визуализаторы представляются в виде шаблонных систем, соответственно концептуальных и индивидуальных. Исполнитель операций с БД является концептуальной системой шаблонов, так как после анализа запросов известен только тип ресурсов, к которому осуществляется запрос, т. е. на вход системы подается концепт данного типа ресурсов. Визуализатор же обрабатывает индивид, полученный в результате работы исполнителя операций с БД.

3.2. Пример: система доступа к взаимосвязанным объектам научной информации

В качестве примера рассмотрим реализацию портала, предоставляющего доступ к базе данных научных публикаций. Ограничим функционал системы только просмотром данных. Система включает в себя информацию о публикациях, авторах, научных мероприятиях, организациях. Между указанными сущностями существует множество связей. Кроме того, необходимо реализовать использование внешних источников данных. Внешней по отношению к реализуемой системе информацией являются классификационные схемы, с помощью которых выполняется тематическая разметка публикаций. Доступ к классификационным схемам может быть осуществлен через веб-службу, имеющую RDF-интерфейс.

Модель данной предметной области можно представить в виде онтологии, где сущностям соответствуют концепты, связям — роли. При этом возможно применение стандартных онтологий, таких как Dublin Core [8]. Концепты, описывающие модель классификационных схем, могут быть взяты из онтологии, сопровождающей службу классификационных схем.

В системе выделяются два вида ресурсов: внутренние ресурсы — публикации, авторы, мероприятия, организации, и внешние ресурсы — классификационные схемы и их рубрики. Данные ресурсы различаются по способу доступа к данным.

Для каждого типа ресурсов система должна предоставлять следующие функции:

- просмотр информации о каком-либо объекте;
- просмотр списка объектов какого-либо типа;
- поиск объектов.

Данные действия имеют различные наборы параметров. Для просмотра информации об объекте необходимо идентифицировать его в системе. Для совместимости с реляционными БД идентификацию объектов удобно проводить путем выявления наборов идентифицирующих свойств. Комбинация значений для данных свойств должна однозначно идентифицировать объект внутри заданного ресурса. Такой механизм наиболее близок к механизму первичных ключей, используемому в реляционных БД. Таким образом, для просмотра информации о конкретном объекте необходимо задать его ресурс и набор значений идентифицирующих свойств.

Для просмотра списка объектов необходимо указать ресурс, объекты которого будут просматриваться. Для просмотра объектов, связанных с каким-либо определенным объектом, необходимо идентифицировать данный объект: указать его ресурс и значения идентифицирующих свойств, а также тип связи.

Поиск объектов можно считать надстройкой над операцией просмотра списка. Для осуществления поиска необходимы те же параметры, что и для просмотра соответствующего списка объектов, плюс поисковое выражение.

Каждому из указанных действий должна соответствовать система шаблонов, загружающих необходимую информацию из источников данных. Конкретная система выбирается при обработке запроса диспетчером и затем используется для получения необходимых данных. Шаблоны визуализации зависят от требуемого формата. Они принимают на вход индивид и модель, загруженную на первом этапе, и возвращают отформатированное представление данных. В системе поддерживается два формата представления данных: HTML — для просмотра данных пользователями-людьми, RDF — для программного доступа к данным.

Заключение

В данной статье рассмотрены модели и методы, позволяющие работать с данными в формате RDF-способами, аналогичными используемым в подходе ООП. Основу подхода составляют системы шаблонов, схожие с методами классов в ООП.

Использование формальных моделей позволяет логически доказывать адекватность используемых алгоритмов и непротиворечивость разрабатываемых систем шаблонов.

Список литературы

1. **Lassila O., Swick R. R.** Resource description framework (RDF) model and syntax specification // W3C Recommendation 22 February 1999. <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222>.
2. **Bray T., Paoli J., Sperberg-McQueen C. M.** et al. Extensible markup language (XML) 1.0 // W3C Recommendation 6 October 2000. <http://www.w3.org/TR/2000/REC-xml-20001006>.
3. **Смит М. К., Велти К., МакГиннес Д. Л.** OWL, язык веб-онтологий. Руководство // Рекомендация W3C 10 февраля 2004. http://sherdim.rsu.ru/pts/semanticweb/REC-owl-guide-20040210_ru.html
4. **Baader F.** The Description Logic Handbook: Theory, Implementation, and Applications. Cambridge University Press, 2007. 622 p.
5. **Borgida A.** On the relative expressiveness of description logics and predicate logics // Artificial Intelligence. 1996. Vol. 82. № 1–2. P. 353–367.
6. **Clark J.** XSL transformations (XSLT) version 1.0 // W3C Recommendation 16 November 1999. <http://www.w3.org/TR/1999/REC-xslt-19991116>.
7. **Гамма Э., Хелм Р., Джонсон Р.** и др. Приемы объектно-ориентированного проектирования. Паттерны проектирования. СПб.: Питер, 2007. 366 с.
8. **Weibel S., Kunze J., Lagoze C.** et al. Dublin core metadata for resource discovery // Internet Engineering Task Force RFC. 1998. Vol. 2413.

МОДЕЛИРОВАНИЕ И ОПТИМИЗАЦИЯ

УДК 629.075

А. Ю. Переварюха, канд. техн. наук,

Санкт-Петербургский институт информатики и автоматизации РАН, e-mail: temp_elf@mail.ru

Новый метод компьютерного моделирования режима переходного хаоса

Рассматриваются проблемы перехода к хаосу и кризиса странных аттракторов в моделях биологических систем. Предлагается новый метод компьютерного моделирования динамики популяций, основанный на применении формализма гибридных автоматов. Разработанные модели реализованы и исследованы с применением современных вычислительных сред, что дало возможность описывать нелинейные эффекты в динамике биологических популяций. Динамическая система позволяет моделировать аperiodические колебания популяции в определенной области хаотического репеллера и переход к хаотическому движению происходит без реализации каскада бифуркаций удвоения периода.

Ключевые слова: моделирование динамики популяций; нелинейные динамические системы; переходный хаотический режим

Введение

После работ Лесли, Хатчинсона, Пирла, Никольсона, Гурни и Вольтерра может сложиться неверное представление, что такое современное научное направление, как компьютерное моделирование экологических процессов, основывается на достаточно сформированном и изученном математическом аппарате. Соответственно, задача мате-

матической биологии заключается не только в корректной идентификации выбранной известной модели по длинному ряду данных наблюдений за популяциями в естественной среде обитания, которую можно провести с применением ЭВМ, используя стандартный набор методов и алгоритмов.

Классические модели математической биологии разрабатывались еще до появления эффек-

тивных вычислительных средств, и это особенно отражается на существующей практике применения систем дискретных разностных уравнений. Преимущество методов вычислительной математики стало очевидным после знаменитого открытия Эдвардом Лоренцом чувствительной зависимости от начальных условий ("эффекта бабочки") и работы О. Реслера [1]. Возможность такой чувствительности предсказывал еще Анри Пуанкаре, но он никак не смог бы обнаружить объект, названный в 1971 г. Д. Рюэлем и Ф. Такенсом в статье "О природе турбулентности" [2] термином странный аттрактор, прочно вошедшим в научный оборот. Наличие современных информационных технологий позволило исследовать различные уравнения, а не ограничиваться теми, которые можно интегрировать аналитически, например исследовать уравнение Кортевега — де Фриза.

Развитие методов нелинейной динамики показало, что задача определения значений параметров модели сталкивается с более существенными проблемами, чем предполагали их разработчики, и не является тривиальной и рутинной задачей. Дополнительно стало очевидно, что некоторые из классических моделей имеют существенно более сложную динамику с наличием качественно различных режимов поведения траектории, что плохо согласуется с данными наблюдений и представлениями предметной области, в рамках которой они разрабатывались. В действительности, как показывает изучение большого числа работ по моделированию биологических систем, эти две проблемы тесно связаны.

В работе [3] мы показали, как нелинейная зависимость смертности от плотности приводит к стремительной деградации популяции при небольшом превышении допустимой доли изъятия. В настоящей статье рассматривается моделирование перехода популяции в режим хаотических колебаний, происходящего при чрезмерном промысле.

Бифуркации и развитие хаоса в биологических моделях

Хорошим примером иллюстрации описанных проблем может являться нелинейный анализ применяемых в экологических исследованиях математических моделей.

Многие применяемые в экологии модели описывают унимодальные кривые, имеющие единственный экстремум — максимум. Например, функции Шепарда и Рикера, являющиеся моделями вида запас—пополнение и решающие задачу формализации причинно-следственных связей, действующих на определенном интервале жизненного цикла поколения. Модели данного вида применяют для моделирования динамики попу-

ляций, которые характеризуются очень высокими показателями смертности на ранних стадиях развития. Показатели смертности большинства видов рыб на ранних этапах онтогенеза намного превышают естественную смертность взрослой части популяции, что оказывается решающим фактором в динамике развития всей популяции.

Модели Дж. Шепарда и У. Рикера предложены в разное время и основываются на частично сходных представлениях о действии факторов смертности, регулирующих численность особей при формировании поколений популяции.

Модель Шепарда [4], названная им самим универсальной, определяет численность пополнения R в зависимости от численности запаса S и некоей критической биомассы K :

$$R = \frac{aS}{1 + (S/K)^\beta}, \quad (1)$$

где a — константа, имеющая размерность — пополнение на единицу биомассы; K — интерпретируется как биомасса, при превышении которой начинают действовать зависящие от плотности факторы смертности; β — параметр, определяющий интенсивность воздействия компенсационной смертности, увеличивающейся от возрастания плотности особей. Однако модель (1), названная универсальной, не нашла широкого применения во многом из-за отсутствия методик определения нетривиальных для ихтиологической практики коэффициентов. Наиболее широко используемой при изучении зависимости запаса и пополнения остается модель Рикера [5], которая может быть выведена из дифференциального уравнения, включающего среди параметров, определяющих скорость убыли численности поколения $N(t)$, величину начальной численности поколения $N(0)$:

$$\frac{dN}{dt} = -(\alpha N(0) + \beta)N(t), \quad t \in [0, \tau].$$

После перехода к другим модельным константам $a = \lambda \exp(-\beta\tau)$, $b = \alpha\tau$, λ — средняя плодовитость особей популяции, влияющая на определение начальных условий $N(0) = \lambda S$, интегрируя дифференциальное уравнение, получаем функцию унимодальной кривой:

$$R = aS \exp(-bS), \quad (2)$$

где b — коэффициент, отражающий величину, обратную S , при которой число выжившей молоди максимально. Соответственно, параметр должен быть меньше единицы; a — безразмерный параметр, связанный с репродуктивным потенциалом, но биологическая интерпретация a трактуется неоднозначно в различных работах. Графиком функции (2) является унимодальная кривая, имеющая

одну точку перегиба и единственное пересечение с биссектрисой координатного угла $R = S$.

Но более интересны другие свойства. Данные модели относятся к семейству унимодальных отображений с отрицательным дифференциальным инвариантом Шварца:

$$S_\psi = \frac{\psi'''(R)}{\psi'(R)} - \frac{3}{2} \left(\frac{\psi''(R)}{\psi'(R)} \right)^2.$$

Величина S_ψ , называемая обычно шварцианом, обладает таким важным свойством, как сохранение знака при итерациях. Условием для реализации сценария перехода к хаосу через бесконечный каскад бифуркаций удвоения периода в унимодальных отображениях является отрицательное значение дифференциального инварианта Шварца $S_\psi < 0$ [6]. Как следует из теоремы Гробмана—Хартмана, устойчивость неподвижной точки $R^* = \psi(R^*)$ отображения $R_{n+1} = \psi(R_n)$ определяется значением тангенса угла наклона касательной, вычисленного в неподвижной точке. Условием для первой бифуркации удвоения периода является равенство $\psi'(R^*) = -1$. При этом у отображения появляются две новые циклические точки $\psi^n(R^*) = \psi^{n+2}(R^*)$, являющиеся неподвижными точками второй итерации (рис. 1, см. третью сторону обложки) $\psi^2(R)$, аттрактором становится устойчивый цикл. Далее при увеличении параметра a и выполнении условия $\psi^{2n'}(R^*) = -1$ происходят следующие бифуркации удвоения периода, появляются циклы периода 2^n . После достижения некоторого значения параметра период становится бесконечным. До развития методов исследования нелинейных динамических систем предполагалось, что сложные флуктуации, наблюдаемые в физических системах, соответствуют движению по поверхности бесконечномерного тора (согласно теории развития турбулентности Ландау—Хопфа) в фазовом пространстве. Но в данном случае возникает аттрактор, отличный от конечного объединения гладких многообразий. В работе М. Фейгенбаума [7] была установлена константа, характеризующая скорость перехода к хаотическому движению, заключающегося в сходимости отношений разницы значений параметра, при которых происходят последовательные бифуркации удвоения, к некоторому постоянному числу. Однако развитием хаоса по известному сценарию Фейгенбаума интересные свойства несложных на первый взгляд моделей (1) и особенно (2) далеко не ограничиваются.

Кризис странного аттрактора и окна периодичности

В предыдущей работе мы рассмотрели эффекты, связанные с возможностью мягкой потери устой-

чивости стационарного состояния одномерного дифференцируемого отображения, возникающей в определенной точке пространства параметров. При бифуркации удвоения периода качественные изменения в поведении траектории происходят только в окрестности цикла. Однако для данных динамических систем возможны и более резкие метаморфозы фазового портрета. Оказалось, что после образования критического аттрактора и перехода к хаосу по сценарию Фейгенбаума в модели Рикера может происходить ряд интересных явлений. Модель Рикера отличается от квадратичного отображения наличием точки перегиба, где вторая производная меняет свой знак. Митчелл Фейгенбаум в своей знаменитой работе рассматривал только четные итерации квадратичного отображения и циклы периодов 2^n , но ранее Т. Ли и Дж. Йорк опубликовали работу под названием "Период три означает хаос" [8].

Цикл периода 3 действительно возникает в модели (2) при $a_t = 22,253$ в виде окна периодичности, прерывающего большой диапазон значений параметра, для которого поведение траектории динамической системы является хаотическим. Это важнейшее свойство для моделей динамики биологических популяций, так как из теорем, доказанных Ли, Йорком и ранее А. Н. Шарковским, следует, что наличие цикла периода 3 означает и наличие для динамической системы циклов других всевозможных периодов.

Окна периодичности различных периодов возникают после тангенсальной (*tangent*) бифуркации. Необходимо отметить особенности используемой терминологии, в русском переводе книги Ф. Муна "Хаотические колебания" [6] допущена ошибка — тангенсальная бифуркация, о которой идет речь, названа касательной, но *tangent* и *tangency bifurcation* это две разные бифуркации, возникающие в различных типах дискретных отображений.

В момент тангенсальной бифуркации производная в появляющихся неподвижных точках $\psi^n(R^*) = 1$, и далее следует жесткая потеря устойчивости, так как эта бифуркация изменяет не только тип аттрактора динамической системы, но и структуру его области притяжения. Появление окна периодичности с циклом периода 3 происходит, когда у третьей итерации $\psi^3(R)$ функции (2) появляются шесть новых (помимо нетривиальной неподвижной точки $R^* = \psi(R^*)$) неподвижных точек — пересечений с биссектрисой $R_n = R_{n+1}$ (рис. 2, см. третью сторону обложки), три из которых образуют устойчивый цикл, а три другие — неустойчивый.

Далее необходимо рассматривать особенности $R_{n+1} = \psi^3(R_n)$. В диапазоне значений $\Delta a = a_c - a_t$ существования периодического окна, который будем называть длиной окна, отображение имеет

три аттрактора, и границы областей притяжения этих аттракторов становятся фрактальными согласно определению Ц. Гребоджи, Дж. Йорка и авторов [9].

При увеличении параметра каждая устойчивая точка $\psi^3(R)$ претерпевает каскад бифуркаций удвоения периода с переходом к хаосу. Однако область хаотического режима, возникшего после удвоений цикла периода 3, представляет собой три несвязные полосы в фазовом пространстве. Поведение траектории интересно тем, что она попадает в

полосы строго периодически, возвращая каждую третью итерацию в одну и ту же хаотическую полосу. Далее периодическое окно "закрывается" при значении параметра a_c , когда три хаотические полосы объединяются в момент, совпадающий с пересечением нестабильного цикла, возникшего при тангенсальной бифуркации, и хаотических подмножеств. Когда неустойчивые неподвижные точки $\psi^3(R)$ оказываются внутри хаотических полос, размеры составного хаотического аттрактора резко увеличиваются, так как в пустой области между хаотическими полосами моментально появляются непериодические точки траектории и вновь возникает единый аттрактор.

Подобное явление, согласно современным представлениям, определяется как внутренний кризис странного аттрактора, при котором его размеры резко увеличиваются. В семействе квадратичных отображений существует и другая разновидность кризиса — граничный кризис (*boundary crises*) по терминологии мэрилендской школы [10], когда увеличивающийся аттрактор соприкасается с границей области притяжения и исчезает. Предлагавшиеся сценарии развития хаоса подразумевают усложнение характера регулярного нехаотического аттрактора и превращение его в странный хаотический. При граничном кризисе хаотический аттрактор возникнет моментально, если значение управляющего параметра будет изменяться в противоположном направлении.

Значение изменения структуры области притяжения в окне периодичности выражается в том, что, перед тем как динамическая система переходит в устойчивый циклический режим, движение траектории долгое время является хаотическим. Подобный переходный хаотический режим (называемый *chaotic transient* [11]) очень интересен с точки зрения моделирования динамики популяции, но естественно не в рамках узкого интервала значений периодических окон в динамике отображений (1) и (2). Проведенные вычислительные

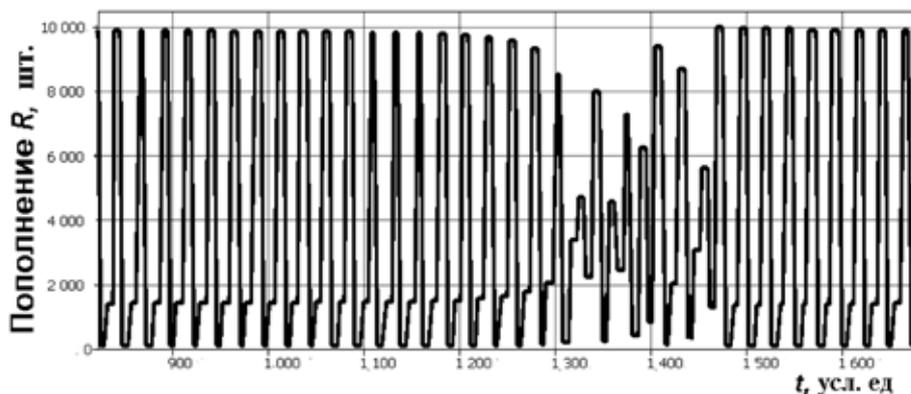


Рис. 3. Режим перемежаемости

эксперименты показывают, что длительность переходного хаотического движения обратно пропорциональна значению того диапазона значений управляющего параметра Δa , при котором существует периодическое окно.

С тангенсальной бифуркацией и фрактальностью границ областей притяжения связано появление еще одного важного для теории динамического хаоса явления перемежаемости (*intermittency* [12]). До того как происходит данная бифуркация, при значениях параметра $a \approx a_t$ длительные периоды почти периодического движения прерываются нерегулярно возникающими хаотическими всплесками (рис. 3, $a = 22,237$). Реализуется еще одна из возможных разновидностей хаотического движения траектории (перемежаемость типа I), появление такого эффекта было описано в первых опытах по изучению гидродинамической турбулентности. Подобное поведение траектории связано с тем, что когда кривая $\psi^3(R)$ приближается к биссектрисе координатного угла, движение точек третьей итерации в этой области становится очень "медленным", но выход из этой области сопровождается значительными непериодическими колебаниями.

Таким образом, в исследованных моделях, применявшихся в математической биологии, установлено наличие хаотической динамики трех различных типов и явление внутреннего кризиса, связанного с резкими изменениями строения странного аттрактора.

Новый метод моделирования хаоса в гибридной модели динамики популяции

Для решения некоторых актуальных задач математической биологии, в частности исследования динамики некоторых промысловых популяций анадромных видов рыб, необходимо разработать модель, позволяющую проводить компьютерное моделирование режима аперриодических флуктуаций численности популяций.

С точки зрения биологической обоснованности использование для моделирования реальной популяции свойств странного аттрактора представляется корректным только приблизительно. Такой аттрактор является объектом фрактальной топологической размерности [13], и траектории всюду настолько плотны в аттракторе, сколь угодно близко от апериодической точки траектории может быть другая точка. Но популяция — не числовая прямая, и представляет собой счетное множество, ее численность измеряется дискретными единицами. Возникновение странного аттрактора невозможно на счетном множестве. Существуют несколько возможностей возникновения хаоса в динамических системах, и хаотическая динамика может быть не только следствием каскада бифуркаций, наиболее распространенного и известного сценария перехода к хаосу с образованием аттрактора.

В целях создания модели в рамках теории зависимости запаса и пополнения популяций, обладающей необходимыми свойствами, мы решили расширить представления, лежащие в основе теории, за счет идей другой биологической концепции.

Изучение биологических особенностей некоторых видов рыб, обитающих в море и совершающих миграции в реки для размножения (называемых анадромными), и анализ данных наблюдений за численностью молоди волжской севрюги в различные годы привели к выводу о существенном влиянии важных и быстро происходящих изменений в развитии особей на характер действия факторов компенсационной смертности.

Закономерность возникновения таких пороговых эффектов имеет биологическое обоснование с точки зрения теории этапности развития организмов, разрабатывавшейся В. В. Васнецовым и в последующих работах О. И. Шмальгаузен и А. И. Смирнова [14].

В предыдущей работе [3] автором была предложена новая модель, описывающая динамику убыли первоначальной численности поколения на определенном интервале времени в виде системы дифференциальных уравнений, учитывающих влияние на смертность роста организмов и действие эффекта Олли при малой численности запаса:

$$\begin{cases} \frac{dN}{dt} = -(\alpha w(t)N(t) + \theta(S)\beta)N(t), \\ \frac{dw}{dt} = \frac{g}{N^k(t) + \zeta}, \quad k < 1, \quad \theta(S) = \frac{1}{1 - \exp(-cS)}, \\ S = N(T), \quad w(0) = w_0, \quad N(0) = \lambda S. \end{cases} \quad (3)$$

где S — численность нерестового запаса; $w(t)$ — отражает уровень размерного развития поколе-

ния, влияющий на увеличение пищевых потребностей; g — параметр, учитывающий ограниченность числа доступных для молоди кормовых объектов; убывающая функция $\theta(S) \rightarrow 1$ отражает действие эффекта Олли (*W. Allee*), выражающегося в том, что при существенном снижении запаса S резко сокращается эффективность воспроизводства, что было установлено для популяций волжской севрюги и русского осетра; ζ — параметр, учитывающий ограничение темпов развития, не зависящее от численности; λ — средняя плодовитость особей; α — коэффициент компенсационной смертности; β — коэффициент декомпенсационной смертности; $t \in [0, T]$ — интервал уязвимости; c — параметр, характеризующий степень действия эффекта Олли, который не проявляется, если численность взрослой популяции достаточно велика.

Графиком исследованной с применением численного решения в инструментальной среде моделирования системы уравнений (3) для определения численности выживших особей поколения к моменту времени T новой модели является унимодальная кривая с уменьшающимся наклоном ниспадающей правой ветви. Кривая имеет ненулевую горизонтальную асимптоту, две нетривиальные точки пересечения с биссектрисой координатного угла $R = S$ и характеризуется знакопостоянным дифференциальным инвариантом Шварца. Такие свойства соответствуют результатам некоторых наблюдений, в частности, за воспроизводством лабораторных популяций. Фазовое пространство в случае динамической системы (3) разделяется на две области притяжения: Ω_1 и Ω_2 . Границей между ними является неустойчивая особая точка первого пересечения кривой с биссектрисой координатного угла. Анализ устойчивости неподвижных точек динамической системы, реализованной в инструментальной среде моделирования, можно проводить с использованием свойства второй итерации $\psi^2(S)$. Необходимым и достаточным условием устойчивости неподвижной точки x^* одномерного отображения является неравенство второй итерации $\psi^2(x) > x$ при $x < x^*$ и $\psi^2(x) < x$ при $x > x^*$.

Для динамической системы (3) невозможен сценарий перехода к хаосу через каскад бифуркаций удвоения периода. Условием для реализации сценария М. Фейгенбаума в унимодальном отображении является отрицательное значение дифференциального инварианта Шварца, но знакопостоянство S_ψ невозможно, если точек перегиба две.

Концепция этапности развития основывается на том, что для последующего развития на каждой стадии организм требует иных условий. При наличии необходимых условий организм переходит на следующую стадию, причем переход происхо-

дит быстро, скачкообразно. Как установлено экспериментально, резкие изменения строения совершаются в очень короткий срок, в некоторых случаях меньше 3 ч; при этом изменяются все системы органов одновременно. Изменяется характер движения, трансформируется способ захвата пищи в связи с изменением ее состава. Рыбы при этом меняют места своего обитания, переходя в новую по своему характеру стадию.

Два важнейших скачкообразных изменения в раннем онтогенезе анадромных рыб происходят при переходе на активное питание и при прекращении тактильного контакта с дном, после которого молодь начинает миграцию в потоке воды по направлению к морю.

Действие резких изменений в развитии на факторы смертности приводит к тому, что зависимость запаса и пополнения анадромных рыб перестает адекватно описываться унимодальной кривой. Наблюдается появление дополнительного второго максимума.

С учетом представлений о важных качественных изменениях коэффициент α нельзя считать ни постоянным, ни монотонно изменяющей значение функцией $\alpha_w(t) = \alpha w(t)$, и правая часть первого уравнения (3) уже не может рассматриваться как гладкая.

Основным математическим аппаратом, используемым для моделирования систем, обладающих способностью к резким структурным изменениям, являются дифференциальные уравнения с разрывными правыми частями и переменной структурой.

Первое уравнения системы (3) для описания убыли численности поколения N заменяется уравнением с переменной структурой (изменяющейся правой частью) и отклоняющимся аргументом при переходе к третьей стадии развития:

$$\begin{cases} \frac{dN}{dt} = \begin{cases} -(\alpha w(t)N(t) + \theta(S)\beta)N(t), & 0 < t \leq \tau \\ -(\alpha_1 N(\tau)/w(\tau) + \beta)N(t), & t > \tau, w(t) < w_{k1} \\ -(\alpha_2 w(t)N(t))N(t - \zeta), & w_{k1} < w(t) < w_k \end{cases} \\ \frac{dw}{dt} = \frac{g}{N^k(t) + \zeta}, \end{cases} \quad (4)$$

где τ — длительность самого раннего этапа развития, характеризующегося тем, что питание происходит за счет желточного мешка; продолжительность этого периода зависит от биологических особенностей конкретного вида; w_k — интерпретируется как уровень развития, достижение которого изменяет характер действия факторов смертности; ζ — небольшое по сравнению с интервалом уязвимости $t \in [0, T]$ запаздывание.

Модель можно реализовать алгоритмически в виде некоторого набора правил, согласно кото-

рому вычислительная среда будет определять, какое дифференциальное уравнение с непрерывными правыми частями следует интегрировать в данный момент времени, и как необходимо провести расчет начальных условий.

Исследование нелинейных эффектов модели (4) на основе представлений о скачкообразных изменениях в развитии и их влиянии на скорость убыли численности поколения требует концептуального изменения в представлении модельного времени, и это возможно с применением современных вычислительных технологий.

В разработанной модели (4) рассматриваются непрерывные процессы, описываемые дифференциальными уравнениями, и события, меняющие характер протекания процессов. Подобные модели относятся к классу гибридных (в некоторых работах использовался термин "событийно управляемые"). Основной особенностью данной структуры является непрерывно-дискретное модельное время, не применявшееся ранее в предметной области.

Если возникает необходимость приостанавливать расчет уравнений непрерывного процесса, выделять события и выполнять определенные дискретные действия, то время в таких моделях представляется множеством последовательно расположенных отрезков, где между концом текущего интервала и началом следующего можно изменять переменные состояния. Следовательно, для реализации модели (4) будет использовано следующее представление модельного времени:

$$\varphi = \{ \{ Gap_pre_1, [0, T_1], Gap_post_1 \}_1, \dots, \{ Gap_pre_n, [T_{n-1}, T_n], Gap_post_n \}_n \},$$

где Gap_pre — временная щель для вычисления согласованных начальных условий и проверки предиката на левом конце промежутка очередного длительного поведения; Gap_post — аналогичная щель для вычисления новых начальных условий на правом конце текущего промежутка φ_i ; T_i — время срабатывания перехода или точка, в которой становится истинным предикат события, приводящего к смене поведения. Гибридное время состоит из пронумерованной и упорядоченной последовательности кадров, что позволяет непрерывной составляющей времени сменяться дискретными отсчетами.

Гибридное поведение может являться следствием совместного функционирования непрерывных объектов и дискретных регуляторов или наличия качественных изменений в непрерывном объекте. При достижении особых состояний в пространстве переменных состояний (событий) могут изменяться значения параметров в правых частях, форма правой части или число уравнений. События описываются предикатами, выделяю-

щими из всех состояний системы события, приводящие к смене поведения. Каждому состоянию в модели (4) соответствует такой период развития модели, в течение которого происходит только рост, но не совершаются принципиальные изменения в поведении рыбы.

Таким образом, модель (4) определяется конечным множеством режимов изменения состояния, с каждым из которых связана правая часть системы дифференциальных уравнений первого порядка, и множеством переходов между режимами. Переходам поставлены в соответствие условие завершения активности и функция инициализации новых начальных условий. Возможность алгоритмической реализации модельной структуры в виде гибридного автомата играет особую роль, так как наглядно вводит дискретное время. Гибридный автомат по своей сути есть дальнейшее расширение и усовершенствование концепции дискретных карт состояний, где узлам графической формы соответствуют описываемые дифференциальными уравнениями непрерывные процессы.

Исследование динамики разработанной гибридной модели

Как было рассмотрено выше, существуют различные типы хаотической динамики для отображений, в том числе перемежаемость и переходный хаос. Разработанная модель (4), определяющая одномерное неунимодальное отображение с незнакопостоянным шварцианом (рис. 4, см. третью сторону обложки), является примером возникновения в фазовом пространстве динамической системы хаотического подмножества, не являющегося аттрактором.

Кривая запас—пополнение модели (4) 4 раза пересекает прямую $R = S$, соответственно, нетривиальными неподвижными особыми точками динамической системы будут R_1^* , R_2^* , R_3^* , R_4^* . Если отображение взаимно однозначно, то устойчивые и неустойчивые особые точки должны чередоваться. В случае с разработанной гибридной моделью (4), как показали вычислительные эксперименты и анализ второй итерации, в области между неустойчивыми первой и третьей особыми точками не существует устойчивой точки. Аттракторов, как и для динамической системы на основе (3), по-прежнему два, один из которых — тривиальное положение равновесия R_0^* .

В диапазоне, ограниченном первой и третьей неустойчивыми точками, появляется последовательность значений точек траектории, которые никогда точно не повторяются, и все приближенные повторения имеют конечную продолжительность. Реализуется хаотический режим, являющийся переходным (*chaotic transient*). Траекто-

рия постепенно заполняет эту область фазового пространства, но не остается навсегда в этой области, как в странном аттракторе, а с равной единицей вероятностью покидает ее по направлению к одному из двух возможных аттракторов (рис. 5, см. третью сторону обложки). Данное явление является следствием того, что Ω_1 и Ω_2 не образуют непрерывных подпространств в фазовом пространстве и изменение характера границы между областями притяжения аттракторов влияет на режим поведения траектории системы, так же как и изменение типа аттрактора при бифуркации.

В случае гладких границ, как для динамической системы (3), при определении начальных условий траектория определенно соответствовала Ω_1 или Ω_2 и малое изменение начальных условий не приводило к выходу фазовой траектории на альтернативный аттрактор. В случае (4) области Ω_1 и Ω_2 изрезаны и имеют сложную структуру. Область притяжения аттрактора R_4^* прерывается отрезками, принадлежащими области притяжения R_0^* , и границы областей притяжения локально несвязны (*locally disconnected basin boundaries*) по разработанной мерилэндской группой классификации фрактальных границ [15]. Таким образом, движение траектории происходит не по гладкому множеству, а по бесконечно разорванному, впервые особенности подобного поведения отметил С. Смейл на примере геометрического отображения, известного под названием "подкова Смейла".

Наличие подобного типа границ приводит к невозможности предсказания относительно того, в непрерывную часть области притяжения какого именно из возможных аттракторов попадет фазовая траектория. Динамические системы, имеющие непритягивающее хаотическое подмножество в фазовом пространстве (хаотический репеллер), характеризуются чувствительностью к начальным условиям, но это другой вид чувствительности, чем в хаотических аттракторах, определяемый обычно термином "*final state sensitivity*" [16], т. е. чувствительность траектории по отношению попадания в определенное асимптотическое состояние из нескольких возможных.

Заключение

Результаты проведенных исследований с точки зрения теории динамических систем показывают ограниченность возможностей применения обычных для биологов математических методов, основным из которых остается регрессионный анализ. Это объясняет и то, что известно множество примеров неудачных попыток аппроксимации, связанных с моделью Рикера, по их результатам даже ставились под сомнение фундаментальные положения, на которых основывается теория

формирования пополнения. Показано, что в биологической интерпретации переход к хаосу в модели Шепарда осуществляется при усилении лимитирующих факторов, а в модели Рикера — при увеличении репродуктивного потенциала. Таким образом, исследование динамики этих двух известных в экологии моделей привело к выводу, что они взаимно исключают друг друга, несмотря на то, что были разработаны в рамках одной предметной области.

В разработанной новой гибридной модели вида запас—пополнение применяются дифференциальные уравнения с разрывными правыми частями для описания динамики численности поколения на разных этапах развития. Результаты исследования динамических свойств (4) согласуются с данными экспериментальных наблюдений ихтиологов и данными о промысле анадромных рыб. При превышении допустимой доли изъятия популяция может переходить в режим аperiodических колебаний численности, и для такой популяции существует вероятность полного исчезновения, что предсказывает существование у модели (4) второго аттрактора — точки (0, 0) на фазовой плоскости.

Список литературы

1. **Rossler O. E.** An equation for continuous chaos // *Physical Letters A*. 1976. Vol. 57. P. 397—399.

2. **Ruelle D., Takens F.** On the nature of turbulence // *Comm. Math. Phys.* 1971. Vol. 20. P. 167—192.

3. **Переварюха А. Ю.** Нелинейные модели и особенности оптимизации в задаче системного анализа динамики популяций // *Информационные технологии*. 2009. № 1. С. 77—81.

4. **Shepherd J. G.** A versatile new stock and recruitment relationship for fisheries and the construction of sustainable yield curve // *J. Cons. Int. Explor. Mer.* 1982. Vol. 40. N 1. P. 67—75.

5. **Ricker W.** Stock and recruitment // *J. Fish. Res. Board. Can.* 1954. Vol. 11. N 5. P. 559—623.

6. **Moon F.** Chaotic vibration. NY: Wiley Interscience, 1987. 312 p.

7. **Li T., Yorke J.** Period three implies chaos // *Amer. Math. Monthly*. 1975. Vol. 82. N 10. P. 985—990.

8. **Feigenbaum M. J.** Universal behavior in nonlinear systems // *Physica D*. 1983. Vol. 7. N 1—3. P. 16—39.

9. **Park B., Grebogi C., Ott E., Yorke J.** Scaling of fractal basin boundaries near intermittency transitions to chaos // *Physical Review A*. 1989. Vol. 40. N 3. P. 1576—1581.

10. **Grebogi C., Ott E., Yorke J.** Chaotic attractors in crisis // *Physical Review Letters*. 1982. Vol. 48. N 22. P. 1507—1510.

11. **Blackburn J., Jensen N., Smith H.** Stochastic noise and chaotic transients // *Physical Review Letters*. 1995. Vol. 74. N 6. P. 908—911.

12. **Hirsch E., Huberman B., Scalapino J.** Theory of intermittency // *Physical Review A*. Vol. 25. N 1. P. 519—532.

13. **Farmer J., Ott E., Yorke J. A.** The dimension of chaotic attractors // *Physica D*. 1983. Vol. 7. P. 153—170.

14. **Smirnov S. A., Makeyeva A. P., Smirnov A. I.** Development of ecomorphology of fishes in Russia // *Environmental Biology of Fishes*. 1995. Vol. 44. N 1. P. 23—33.

15. **MacDonald S., Grebogi C., Ott E., Yorke J.** Fractal basin boundaries // *Physica D*. 1985. Vol. 17. N 2. P. 125—153.

16. **Grebogi C., MacDonald S., Ott E., Yorke J.** Final state sensitivity: an obstruction to predictability // *Physical Letters A*. 1983. Vol. 99. P. 415—418.

УДК 519.6

А. П. Карпенко, д-р. физ.-мат. наук, проф.,
Е. Ю. Селиверстов, студент,
МГТУ им. Н. Э. Баумана,
e-mail: karpenko@nog.ru

Глобальная оптимизация методом роя частиц. Обзор

Рассматриваются метод роя частиц для задачи глобальной многопараметрической оптимизации, а также известные последовательные и параллельные модификации этого метода.

Ключевые слова: глобальная оптимизация, метод роя частиц, метод PSO

Введение

Среди задач непрерывной конечномерной оптимизации самым важным с практической точки зрения и одновременно самым сложным является

класс задач глобальной условной оптимизации. Методы решения задач этого класса можно разделить на две большие группы:

- методы сведения задачи глобальной условной оптимизации к задаче глобальной безусловной оптимизации с помощью штрафных или барьерных функций;
- методы, специально сконструированные для решения задачи глобальной условной оптимизации.

Рассматриваемый в работе метод роя частиц относится к методам первой группы и, таким образом, предназначен для решения задач глобальной безусловной оптимизации.

Методы решения задачи глобальной безусловной оптимизации делятся на детерминированные, стохастические и эвристические методы [1].

Эвристические методы являются относительно новым быстро развивающимся классом методов глобальной безусловной оптимизации. Среди этих методов выделяются эволюционные и поведенческие (имитационные) методы.

Поведенческие методы решения задачи глобальной безусловной оптимизации основаны на моделировании коллективного поведения самоорганизующихся живых или неживых систем. Взаимодействующие элементы этих систем в общем случае называются агентами. Ключевыми идеями поведенческих методов являются децентрализованность, взаимодействие агентов, простота поведения агентов.

Известность получили следующие поведенческие методы решения задачи глобальной безусловной оптимизации: метод поведения пчел; метод колонии муравьев; метод роя частиц, рассматриваемый в данной работе.

В методе оптимизации роем частиц (*particle swarm optimization — PSO*) агентами являются частицы в пространстве параметров задачи оптимизации. В каждый момент времени (на каждой итерации) частицы имеют в этом пространстве некоторые положения и векторы скоростей. Для каждого положения частицы вычисляется соответствующее значение целевой функции, и на этой основе по определенным правилам частица меняет свое положение и скорость в пространстве поиска.

В основу метода PSO положена социально-психологическая поведенческая модель толпы. Существует несколько разновидностей метода. Например, в каноническом методе роя частиц, предложенном в 1995 г. в работе Kennedy, Eberhart [2], на каждой итерации при определении следующего положения частицы учитывается информация о наилучшей частице из числа "соседей" данной частицы, а также информация о данной частице на той итерации, когда этой частице соответствовало наилучшее значение целевой функции. Модификация канонической модели FIPS учитывает значения целевой функции, соответствующие всем частицам роя; в некоторых моделях частицы группируются в несколько роев и т. д.

Большинство известных методов роя частиц являются последовательными. Параллельных методов известно немного, и все они появились после 2004 г. В работе рассмотрены как последовательные, так и параллельные методы PSO.

Основным стимулом для подготовки данной работы явилось почти полное отсутствие русскоязычных исследований и публикаций по методу роя частиц.

1. Канонический метод роя частиц

Рассмотрим задачу глобальной безусловной минимизации целевой функции $\Phi(X)$ в n -мерном арифметическом пространстве R^n :

$$\min_{X \in R^n} \Phi(X) = \Phi(X^*). \quad (1)$$

Множество частиц обозначим $S = \{P_i, i \in [1:N]\}$, где N — число частиц в рое S (размер популяции). В момент времени $t = 0, 1, 2 \dots$ координаты частицы P_i определяются вектором $X_{i,t} = (x_{i,t,1}, x_{i,t,2}, \dots, x_{i,t,n})$, а ее скорость — вектором $V_{i,t} = (v_{i,t,1}, v_{i,t,2}, \dots, v_{i,t,n})$. Начальные координаты и скорости частицы P_i равны $X_{i,0} = X_i^0, V_{i,0} = V_i^0$ соответственно.

Итерации в каноническом методе PSO выполняются по следующей схеме:

$$X_{i,t+1} = X_{i,t} + V_{i,t+1}; \quad (2)$$

$$V_{i,t+1} = \alpha V_{i,t} + U[0, \beta] \otimes (X_{i,t}^b - X_{i,t}) + U[0, \gamma] \otimes (X_{g,t} - X_{i,t}). \quad (3)$$

Здесь $U[a, b]$ представляет собой n -мерный вектор псевдослучайных чисел, равномерно распределенных в интервале $[a, b]$; \otimes — символ покомпонентного умножения векторов; $X_{i,t}^b$ — вектор координат частицы P_i с наилучшим (в смысле (1)) значением целевой функции $\Phi(X)$ за все время поиска; $X_{g,t}$ — вектор координат соседней с данной частицы с наилучшим за время поиска значением целевой функции $\Phi(X)$; α, β, γ — свободные параметры алгоритма.

Пересчет координат частиц по формулам (2), (3) может происходить по синхронной схеме (обновление координат частиц выполняется только после определения текущих скоростей всех N частиц) или по асинхронной схеме (расчет новых координат частиц проводится до завершения указанных вычислений).

В процессе итераций вектор $X_{i,t}^b$ образует так называемый собственный путь (*private guide*) частицы P_i , а вектор $X_{g,t}$ — локальный путь (*local guide*) этой частицы.

Свободный параметр α определяет "инерционные" свойства частиц (при $\alpha < 1$ движение частиц, очевидно, замедляется). Рекомендуемое значение параметра α равно 0,7298. В процессе оптимизации может быть эффективным постепенное уменьшение коэффициента α от 0,9 до 0,4. При этом большие значения параметра обеспечивают широкий обзор пространства поиска, а малые — точную локализацию минимума целевой функции. Рекомендуемые значения свободных параметров β, γ равны 1,49618.

Важнейшим понятием в методе PSO является понятие соседства частиц, которое определяется соответствующей топологией соседства. В силу важности этого понятия, оно рассмотрено в самостоятельном разделе (см. раздел 4).

Второй компонент в формуле (3) называется "когнитивным" компонентом (по социальной аналогии) и формализует тенденцию частицы

вернуться в положение с минимальным значением целевой функции. Третий компонент в формуле (3) называется "социальным" компонентом. Компонент отражает влияние на данную частицу ее соседей.

Вместо формулы (3) часто используют ее вариант

$$V_{i,t+1} = \alpha(V_{i,t} + \varphi \oplus (Y_{i,t} - X_{i,t})). \quad (4)$$

Коэффициент α в этом случае называется ограничивающим фактором (*restriction factor*);

$$\varphi = U[0, \beta] + U[0, \gamma],$$

$$Y_{i,t} = \frac{U[0, \beta] \otimes X_{i,t}^b + U[0, \gamma] \otimes X_{g,t}}{\varphi}.$$

Рекомендуемые значения параметров β, γ в этом случае равны 2,05, а их оптимальные значения определяются классом целевой функции $\Phi(X)$ [3].

2. Метод роя частиц с полной информацией (FIPS)

Метод роя частиц с полной информацией FIPS (*fully informed particle swarm*) предложен в работах [4, 5] и представляет собой обобщение канонического метода PSO.

В основе метода FIPS лежит наблюдение, что поведение особи определяется не только наиболее успешным соседом, но и всей совокупностью соседей. Таким образом, основное отличие метода FIPS от канонического метода PSO состоит в том, что в методе FIPS учитывается влияние на данную частицу всех соседних частиц. Ясно, что при этом различные топологии соседства частиц существенно влияют на процесс оптимизации. В формуле (3) в данном случае исчезает необходимость в слагаемом, содержащем вектор $X_{g,t}$, т. е. "локального пути" частицы в данном случае не существует.

В итерационной схеме метода FIPS координаты частиц определяются по формуле (2), а их скорости — по формуле (4), в которой следующим образом изменяется схема вычисления члена $Y_{i,t}$:

$$Y_{i,t} = \frac{\sum_k (\varphi_k \otimes X_{k,t}^b)}{\sum_k \varphi_k}, \quad k \in \mathbf{N}_i. \quad (5)$$

Здесь \mathbf{N}_i — множество номеров частиц, являющихся соседями частицы P_i (в смысле используемой топологии соседства); $\varphi_k = U[0, \beta/N_i]$; N_i — число элементов во множестве \mathbf{N}_i .

Метод обычно показывает лучшую сходимость, чем канонический метод PSO. Основной проблемой при использовании метода FIPS является сильное влияние на его эффективность топологии пространства соседства. Особенностью

метода является тенденция его сходимости к центру начального положения частиц $X_i^0, i \in [1 : N]$. В этой связи рекомендуется использовать асимметричное начальное расположение частиц.

Метод FIPS с весами. Метод FIPS с весами (*weighted FIPS*) является модификацией метода FIPS и заключается в добавлении весов в последнее слагаемое уравнения (4). При соответствующем выборе весов такая модификация позволяет усилить влияние тех соседних частиц, которые имеют лучшие значения целевой функции. Итерационная формула (4) для метода FIPS с весами имеет вид

$$V_{i,t+1} = \alpha \left(V_{i,t} + \frac{\sum_k w(k,t) U[0, \gamma] \otimes (X_{k,t}^b - X_{i,t})}{\sum_k w(k,t)} \right),$$

$$k \in \mathbf{N}_i, \quad (6)$$

где $w(k,t)$ — скалярный весовой множитель, в общем случае меняющийся со временем, который определяет влияние на данную частицу k -й соседней частицы.

В качестве веса $w(k,t)$ чаще всего используют величину, обратную целевой функции $\Phi(X_{k,t})$ (напомним, что рассматривается задача минимизации этой функции):

$$w(k,t) = \frac{1}{\Phi(X_{k,t})}. \quad (7)$$

В роли весов $w(k,t)$ могут использоваться и другие величины, например, расстояние между частицами.

Ранжированный метод FIPS. Как отмечалось выше, имеет место нестабильность метода FIPS, обусловленная сильным влиянием на рассматриваемую частицу топологии соседства частиц. Кроме того, при использовании в качестве весов величин (8) может проявляться следующий эффект. На поздних итерациях, когда основные улучшения целевой функции проведены и все частицы находятся в окрестности локального минимума, имеют место близкие значения целевой функции, соответствующие всем частицам. Поэтому и значения весов $w(k,t)$ для всех $k \in \mathbf{N}_i$ оказываются близкими. В результате поиск замедляется или даже останавливается, так что применение в качестве весов значений (8) теряет смысл. Ранжированный метод FIPS (*ranked FIPS*) [6] позволяет преодолеть данный недостаток.

Идея ранжированного метода FIPS состоит в отказе от использования весов, построенных на основе значений целевой функции. Вместо этого веса $w(k,t), k \in \mathbf{N}_i$ заменяются рангами r_k , которые определяются по следующим правилам:

$$r_k = 2r_{k+1}; \quad \sum_k r_k = 1; \quad \Phi(X_k) < \Phi(X_{k+1}). \quad (8)$$

Отметим, что в соответствии с этими правилами вес наилучшей частицы в рое равен 0,5.

Таким образом, в ранжированном методе FIPS формула (3) принимает вид

$$V_{i,t+1} = \alpha(V_{i,t} + U[0, \beta] \otimes (X_{i,t}^b - X_{i,t})) + \sum_k r_k U[0, \gamma] \otimes (X_{g,t} - X_{i,t}), \quad k \in \mathbf{N}_j \quad (9)$$

Ранжированный метод FIPS обеспечивает также меньшее влияние на данную частицу топологии соседства частиц, так что увеличение числа частиц во множестве соседних частиц не оказывает отрицательного влияния на эффективность поиска.

3. Метод роя частиц, основанный на отношении "значение — расстояние"

Метод роя частиц, основанный на отношении "значение — расстояние" FDR PSO (*fitness-distance ratio based PSO*), предложен в работе [7]. Метод отличается от канонического метода PSO учетом влияния нескольких окружающих частиц. Заметим, что метод FIPS имеет такое же принципиальное отличие от канонического метода PSO. Однако в методе FDR PSO учет влияния соседних частиц реализуется иначе, чем в методе FIPS.

В методе FDR PSO используется модифицированный канонический итерационный процесс (2), (3). Суть модификации заключается в добавлении в формулу (3) еще одного слагаемого

$$k_3 U[0, \delta] \otimes (X_{q,t} - X_{i,t}), \quad (10)$$

где δ — свободный параметр метода; $X_{q,t}$ — вектор координат некоторой частицы $P_q \in \mathbf{S}$, определяемой по следующим правилам:

- частица P_q является соседней с частицей P_i ;
- частице P_q соответствует лучшее значение целевой функции $\Phi(X)$ по сравнению со всеми соседними частицами.

Здесь в том и другом случае соседство понимается в смысле используемой топологии соседства частиц.

Таким образом, в методе FDR PSO скорости частиц определяются по формуле

$$V_{i,t+1} = \alpha V_{i,t} + k_1 U[0, \beta] \otimes (X_{i,t}^b - X_{i,t}) + k_2 U[0, \gamma] \otimes (X_{g,t} - X_{i,t}) + k_3 U[0, \delta] \otimes (X_{q,t} - X_{i,t}), \quad (11)$$

где k_1, k_2, k_3 — свободные параметры.

Введем в рассмотрение функцию $FDR(X_j, X_i)$ положений частиц P_j, P_i из популяции \mathbf{S} :

$$FDR(X_j, X_i) = \frac{\Phi(X_j) - \Phi(X_i)}{\|X_j - X_i\|}, \quad (12)$$

где $\|*\|$ — некоторая векторная норма.

Тогда частица, наилучшим образом удовлетворяющая сформулированным правилам, доставляет минимум функции FDR :

$$FDR(X_q, X_i) = \min_j FDR(X_j, X_i), \quad j \in \mathbf{N}_i.$$

Рекомендованные в работе [7] значения параметров k_1, k_2, k_3 равны 1, 1, 2 соответственно. Экспериментально показано, что метод малочувствителен к значениям этих параметров — даже их обнуление не приводит к существенному ухудшению скорости сходимости метода. Основной вклад в скорость сходимости метода обеспечивает компонент (10).

Отметим, что в том виде, в котором метод FDR PSO рассматривается в работе [7], случайные коэффициенты $U[a, b]$ не используются, что эквивалентно принятию $U[a, b] = \text{const} = 1$. Однако в той же работе указывается, что в качестве этих коэффициентов имеет смысл использовать случайные величины, равномерно распределенные в диапазоне $[0; 2,05]$.

4. Топологии соседства частиц

Эффективность метода PSO в значительной мере зависит от топологии соседства частиц (*population topology, neighbourhood topology, swarm topology, sociometry*). Топология соседства определяется неориентированным графом, вершины которого соответствуют частицам роя, а ребра связывают непосредственных соседей.

В реализации метода PSO чаще всего используются следующие топологии соседства частиц [8]:

- клика (gbest-топология — глобально оптимальная топология);
- кольцо (lbest-топология — локально оптимальная топология);
- двумерный тор (топология фон Неймана);
- кластерная топология.

В топологии "клика" (в полносвязном графе) соседями каждой из частиц $P_i, i \in [1, N]$ являются остальные $(N - 1)$ частицы. Диаметр полносвязного графа равен единице [9].

В топологии "кольцо" соседями каждой из частиц P_i являются две частицы. Например, в графе, приведенном на рис. 1, соседями частицы P_3 являются частицы P_2, P_4 . Диаметр соответствующего графа равен $\sim N/2$.

В топологии "двумерный тор" соседями каждой из частиц $P_i, i \in [1, N]$ являются четыре частицы. Например, в графе, приведенном на рис. 2, соседями частицы P_3 являются частицы P_2, P_4, P_8, P_{18} . Если граф типа "двумерный тор" построен на основе $(m \times m)$ -решетки, то диаметр этого графа равен $2\lceil m/2 \rceil$, где $\lceil * \rceil$ — символ ближайшего целого меньшего; $N = m^2$.

В кластерной топологии граф имеет в качестве узлов клики из C_{clique} узлов. Клики числом N_{clique} объединяются в полносвязный граф, ребра которого связаны с теми узлами в кликах, степени которых равны ($C_{clique} - 1$). Очевидно, что диаметр такого графа равен 3. Соседями каждой из частиц $P_i, i \in [1, N]$, являются C_{clique} частиц. К примеру, в графе, приведенном на рис. 3, $N_{clique} = 4, C_{clique} = 5$, диаметр равен 3, соседями частицы P_3 являются пять следующих частиц: $P_1, P_2, P_4, P_5, P_{15}$.

Диаметр графа, соответствующего используемой рою топологии соседства частиц, определяет скорость распространения информации в рою. Поэтому в рою с топологией соседства "клика" лучшее значение целевой функции, достигнутое той или иной частицей, сразу становится известным всем остальным частицам рою. Из числа рассмотренных топологий самую низкую скорость распространения информации обеспечивает топология "кольцо". Промежуточные значения скорости распространения информации достигаются в роях, обладающих топологией соседства частиц "двумерный тор" и кластерной топологией.

Известно значительное число исследований эффективности и сходимости метода PSO и его модификаций при различных топологиях соседства частиц [4, 5, 8, 10].

Результаты исследований показывают, что топология типа "клика" применима для поиска минимума функции $\Phi(X)$ с малыми градиентами. Топология обеспечивает высокую эффективность, но может привести к преждевременной сходимости метода к некоторому локальному минимуму функции $\Phi(X)$.

Напротив, топология типа "кольцо" эффективна при оптимизации сложных многоэкстремальных целевых функций. Это обстоятельство обусловлено тем, что частицы рою, принадлежащие "далеким" частям кольца, слабо связаны друг с другом и могут эффективно исследовать различные области пространства поиска. Топология обеспечивает слабое "притяжение" частиц локальными минимумами целевой функции и позволяет избежать преждевременной сходимости.

Топология "двумерный тор" и кластерная топология обладают промежуточными свойствами. Они обеспечивают высокую вероятность локализации глобального минимума целевой функции и достаточно высокую эффективность поиска.

Эффективность метода PSO в значительной степени зависит от размера популяции N . Эта зависимость может иметь различный характер для различных топологий соседства частиц.

Для топологии "клика" и большого размера популяции канонический метод PSO может привести к ранней сходимости. В тех же условиях в методе FIPS большая популяция может охватывать

несколько локальных минимумов целевой функции, и значительное число частиц может колебаться между этими минимумами. Такие же колебания частиц возможны и в каноническом методе PSO.

Топология "кольцо" более устойчива к большому размеру популяции, так как среди рассмотренных топологий является наименее связанной и обес-

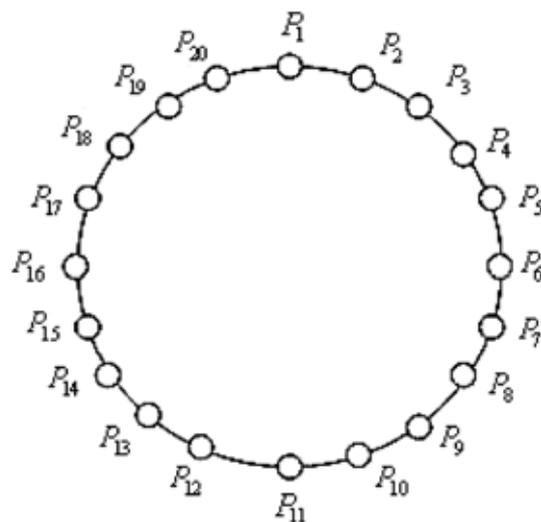


Рис. 1. Топологии соседства частиц "кольцо"; $N = 20$

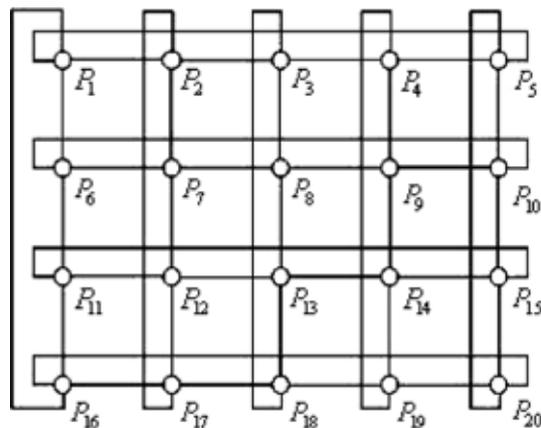


Рис. 2. Топологии соседства частиц "двумерный тор"; $N = 20$

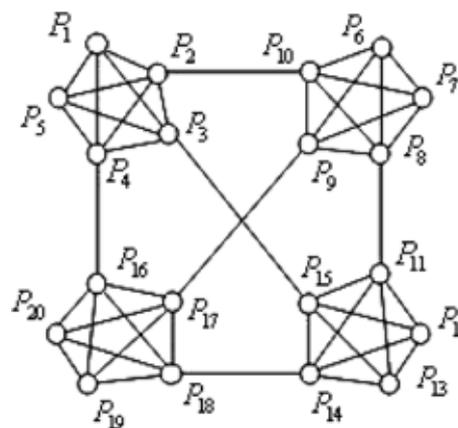


Рис. 3. Кластерная топология соседства частиц; $N = 20$

печивает самую низкую скорость передачи информации.

С точки зрения эффективности метода PSO при больших популяциях в качестве универсальной топологии рекомендуется использовать топологию "двумерный тор", которая обеспечивает достаточно высокую эффективность любого из рассмотренных методов.

5. Методы, использующие динамические топологии соседства частиц

Рассмотренные методы PSO используют статические топологии соседства частиц, т. е. соседство частиц для заданного роя задается до начала итерации и во время итерационного процесса не меняется. Наряду с этим известно значительное число методов PSO, использующих топологии соседства частиц, которые меняются в процессе итераций — динамические топологии соседства.

Метод роя частиц с организацией в клубы. Метод роя частиц с организацией в клубы СВ PSO (*clubs-based PSO*) предложен в работе [11]. Суть метода состоит в следующем. Частицы роя образуют динамические группы, называемые клубами. Каждая частица может входить в несколько клубов. В процессе итераций частица может мигрировать между клубами. Соседними частицами считаются частицы из одного клуба. Скорости частиц определяются по формулам (2), (3) канонического метода PSO.

Число клубов M_i , в которые входит частица P_i , называется уровнем вхождения частицы в клубы (*membership level*). В начальный момент времени каждая из частиц случайным образом распределяется по N_{cl} клубам, где N_{cl} — так называемый стандартный уровень вхождения в клубы. При росте величины N_{cl} скорость сходимости метода СВ PSO возрастает. Обычно ограничиваются выбором уровня вхождения в пределах 4—15.

В методе СВ PSO на каждой итерации оптимизационного процесса находят частицы P_{best} и P_{worse} с наилучшим и наихудшим значениями целевой функции соответственно. Частица P_{best} исключается из произвольного клуба, к которому она принадлежит, а частица P_{worse} включается в произвольный клуб. Каждые k итераций для частиц, не являющихся наилучшими или наихудшими в клубе, проводится изменение их уровней вхождения в клубы в направлении стандартного уровня вхождения. Такой прием улучшает стабильность метода и равномерность распределения частиц по клубам.

На параметр M_i накладывается ограничение $M_i \in [M^{\min}, M^{\max}]$, так что каждая из частиц P_i , $i \in [1 : N]$, во всех случаях входит не менее чем в M^{\min} и не более чем в M^{\max} клубов. Значение M^{\max}

не должно быть меньше N , чтобы частица могла состоять в разных клубах. В работе [11] для роя из 20 частиц используются значения $M^{\min} = 5$, $M^{\max} = 33$.

Исключение частицы P_{best} из клуба имеет целью предотвратить преждевременную сходимость метода, так как эта частица оказывает сильное влияние на соседние частицы в клубе. Включение частицы P_{worse} в некоторый другой клуб позволяет увеличить число соседних с нею частиц. Так как частицы в новых клубах в общем случае не хуже частиц, находящихся в исходном клубе, то это дает возможность частице P_{worse} улучшить значение целевой функции.

Как отмечалось выше, в методах PSO, основанных на статической топологии соседства частиц, ранняя сходимость является типичной проблемой. Причина этого эффекта заключается в том, что частица, нашедшая какой-либо локальный минимум целевой функции, оказывает сильное и постоянное влияние на соседние частицы, стремящиеся локализовать эти частицы в окрестности того же локального минимума. В результате уменьшается вероятность того, что эти частицы могут исследовать другие возможные локальные минимумы. В методе СВ PSO эта проблема решается путем исключения из числа соседей наиболее успешных частиц. При этом на ранних стадиях поиска частицы имеют довольно мало соседей, и нахождение отдельных локальных минимумов слабо влияет на остальные частицы.

В работе [11] показано, что метод СВ PSO по сравнению с каноническим методом PSO дает лучшие результаты с точки зрения сходимости и выхода из локальных минимумов целевой функции как для топологии соседства "клика", так и для топологии "кольцо".

Метод роя частиц со стереотипированием. Метод роя частиц со стереотипированием (*PSO with stereotyping*) предложен в работе [12]. Основная идея метода состоит в статическом разделении частиц по фиксированному числу кластеров. Критерием разделения является геометрическая близость частиц в пространстве параметров R^n . В качестве центра кластера используется геометрический центр координат частиц, образующих этот кластер. В формулах вычисления скорости частицы P_i , $i \in [1 : N]$, ее собственные координаты заменяются координатами центра кластера, которому принадлежит эта частица. Например, в формуле вычисления скоростей частиц для канонического метода PSO (3) указанным образом заменяется компонента $X_{i,r}$

Метод может быть скомбинирован как с каноническим методом PSO, так и с его модификацией FIPS.

Метод роя частиц с дополнением графа соседства частиц. Метод PSO с дополнением графа соседства частиц DS PSO (*dynamic sociometry PSO*) предложен и исследован в работах [13—15]. Метод реализует технику так называемого "перехода от обзора к поиску" (*switch from exploration to exploitation*).

Как отмечалось выше, разреженные графы соседства частиц обеспечивают высокую эффективность при обзоре, а плотные графы — при поиске. Идея метода DS PSO состоит в следующем. На начальных итерациях оптимизации используется сильно разреженный граф соседства частиц. Через фиксированное число итераций k в этот граф регулярным или случайным образом добавляется новое ребро. Например, в топологии соседства частиц "кольцо" после каждых k итераций в текущий граф соседства частиц добавляется ребро (P_j, P_{j+i}) . В результате, в конце концов, при достаточном числе итераций получается граф соседства типа "клика".

На ранних стадиях поиска метод DS PSO реализует стратегию глобального исследования пространства параметров (обзор), а на поздних стадиях — проводит уточнение полученных результатов (поиск).

6. Многороевые методы

Все рассмотренные методы PSO, как со статической, так и с динамической топологией соседства частиц, использовали один рой. Известны также методы, использующие несколько роев и миграцию частиц между ними.

Многороевый метод с миграцией частиц. Многороевый метод с миграцией частиц MPSO (*PSO with migration*) предложен в работе [6]. Идея метода состоит в формировании нескольких роев частиц (*subswarms*) и организации миграции частиц между роями. При этом каждый из роев использует, вообще говоря, свои значения свободных параметров и свою топологию соседства частиц. При переходе частицы из одного роя в другой топология соседства взаимодействующих роев не меняется. Метод является адаптивным и может применяться при неполных знаниях о минимизируемой функции $\Phi(X)$. Напомним, что для метода FIPS, например, топология соседства частиц должна выбираться в зависимости от класса функции в соответствии с эмпирически полученными рекомендациями.

В простейшем случае миграция частиц между роями осуществляется через фиксированное число итераций k . Для каждого из роев S_i на основе текущих значений целевой функции $\Phi(X)$ вычисляется значение функции оптимальности этого роя $\Psi(S_i)$, равное значению функции $\Phi(X)$ для лучшего собственного пути в рое S_i . Таким образом, функция оптимальности роя отражает наилучшие достижения частиц в рое.

Каждые k итераций осуществляется переход одной частицы из роя с минимальным значением функции Ψ в рой с максимальным значением этой функции. Здесь k — свободный параметр метода. В работе [6] использовалось значение $k = 60$.

Модифицированный многороевый метод с миграцией частиц. Схема модифицированного метода MPSO имеет следующий вид [6].

1. Для каждой пары роев S_i, S_j через каждые k итераций вычисляется отношение значений их модифицированных функций оптимальности $\Psi^*(S_i), \Psi^*(S_j)$

$$R(S_i, S_j) = R_{i,j} = \frac{\Psi^*(S_i)}{\Psi^*(S_j)}. \quad (13)$$

2. Находится множество роев $\{S_k\}$, для которых справедливо неравенство $R_{i,j} > R_{lim}$.

3. Каждому из роев $\{S_k\}$ назначается штрафной балл B_k (*penalty*).

4. Из числа роев $\{S_k\}$ определяется новое множество роев $\{S^*\}$ таких, что для каждого из них имеет место неравенство $B_k > B_{lim}$.

5. Рои множества $\{S^*\}$ сортируются по возрастанию их штрафных баллов.

6. Из каждого из роев отсортированного множества $\{S^*\}$ (кроме последнего роя) мигрирует произвольно выбранная частица в следующий по порядку рой.

Назначение функции $\Psi^*(S_i)$ состоит в классификации множества роев на "успешные" и "неуспешные". В качестве этой функции может использоваться относительное улучшение функции оптимальности за некоторое число итераций d :

$$\Psi^*(S_i) = \frac{\Psi_i(S_i)}{\Psi_{t-d}(S_i)}. \quad (14)$$

В модифицированном методе MPSO параметры R_{lim}, B_{lim}, d являются свободными. Для тестирования метода авторы применяли $R_{lim} = 2, B_{lim} = 60, d = 200$. Возможна оптимизация этих параметров для конкретных классов целевых функций.

7. Параллельные методы роя частиц

Идеи, лежащие в основе метода PSO, дают основания надеяться на существование эффективных параллельных аналогов метода.

Действительно, известно значительное число хорошо масштабируемых параллельных методов PSO, в своей массе алгоритмически отличающихся от последовательных аналогов.

Для генетических методов, гораздо более исследованных по сравнению с методами PSO, выделяются три следующих класса параллельных методов [16]:

- методы, основанные на глобальной модели параллелизма;
- миграционные методы, основанные на островной модели параллелизма;
- методы, основанные на диффузной модели параллелизма.

По аналогии с этой классификацией, для PSO выделяется три рассматриваемых ниже группы параллельных методов [17].

Методы, основанные на глобальной модели параллелизма. Параллельные методы PSO, относящиеся к данной группе методов, представляют собой параллельные аналоги рассмотренных выше последовательных методов. Эти методы используют параллелизм по данным и ориентированы на организацию параллельных вычислений по типу "*master — slave*". Мастер-процесс выполняет работу по поиску глобального минимума целевой функции $F(X)$, а подчиненные процессы проводят вычисления значений целевой функции, а также скоростей и координат частиц.

Рабочие процессы после каждой итерации отсылают мастер-процессу значения полученных координат частиц и их скоростей, а также соответствующие значения целевой функции. Перед каждой итерацией рабочие процессы получают от мастер-процесса координаты "лучших" частиц. Мастер-процесс на основе данных, полученных от рабочих процессов, вычисляет координаты "лучших" частиц и посылает их рабочим процессам.

Параллельные вычисления в методах данного класса могут быть как синхронными, так и асинхронными.

Достоинство методов, основанных на глобальной модели параллелизма, состоит в использовании информации о глобально "лучших" частицах. Результат оптимизации при использовании данных методов получается точно таким же, как при аналогичных последовательных вычислениях. Недостатком методов являются большие накладные расходы на коммуникации (при большом диаметре коммуникационной сети).

Методы, основанные на островной модели параллелизма. Методы данного класса используют параллелизм по данным и ориентированы на реализацию на кластерных вычислительных системах. Весь рой из N частиц делится на m островов (по числу процессоров в системе), и частицы, принадлежащие каждому из островов, обрабатываются на своем процессоре. После каждых k независимых итераций острова обмениваются между собой лучшими частицами (в соответствии с используемой топологией соседства частиц).

При большом k метод имеет малые коммуникационные расходы, но недостаточно эффективно использует информацию о лучших частицах из

других островов. Существуют различные стратегии замещения частиц в островах [18].

Методы, основанные на диффузной модели параллелизма. Методы этого класса можно считать частным случаем методов, основанных на островной модели параллелизма (когда число частиц N равно числу островов m и, тем самым, числу используемых процессоров вычислительной системы).

Основная идея диффузных методов состоит в том, что "лучшие" частицы определяются параллельно и только среди соседних частиц (в смысле используемой топологии соседства частиц). Для этого на каждой итерации процессор, на котором обрабатывается частица P_i , получает от всех тех процессоров, на которых обрабатываются частицы, соседние с данной, их координаты, скорости и соответствующие значения целевой функции $F(X)$. Такие коммуникации выполняются параллельно всеми N процессорами.

Коммуникационные расходы в диффузной модели напрямую определяются используемой топологией соседства частиц и являются высокими в случае, если соответствующий граф имеет высокую связность. Методы данного класса обеспечивают высокую производительность только при невысокой связности указанного графа и/или высокой вычислительной сложности целевой функции.

В целом, метод ориентирован на массивно-параллельные вычислительные системы с сильно связными (например, гиперкуб или двумерный тор) и быстрыми коммуникационными сетями.

Параллельный синхронный метод. Параллельный синхронный метод предложен в работе [19] (в этой работе метод называется *parallel PSO*). Метод реализует глобальную модель параллелизма, при которой параллельно вычисляются лишь значения целевой функции.

На каждой итерации мастер-процессор рассылает по рабочим процессорам текущие координаты частиц и ожидает возвращения соответствующих значений целевой функции $F(X)$. Текущая итерация завершается только после того, как мастер-процессором будут получены значения целевой функции от всех рабочих процессоров (т. е. используется глобальная барьерная синхронизация). Обновление состояния роя (скоростей и координат частиц) происходит только после получения мастер-процессором указанной информации обо всех частицах роя. В методе используется статическая балансировка загрузки процессоров.

Условиями высокой производительности параллельного синхронного метода являются однородность вычислительной системы, одинаковое время вычисления значений целевой функции в любой точке пространства параметров, а также равномерное распределение вычислений по процессорам системы. Обеспечить выполнение всех

этих условий обычно не удается. Так, часто приходится использовать гетерогенные вычислительные системы, в реальных оптимизационных задачах вычислительная сложность целевой функции $F(X)$ зависит от значений компонент вектора варьируемых параметров X .

Параллельный асинхронный метод. Параллельный асинхронный метод PAPSO (*parallel asynchronous PSO*) предложен в работах [20, 21]. Метод также реализует глобальную модель параллелизма.

В отличие от синхронного метода, в этом случае мастер-процессор получает данные от рабочих процессоров не после глобальной синхронизации, а в любой момент времени по мере готовности этих данных. На основе полученной информации мастер-процессор обновляет скорости и координаты соответствующих частиц роя и немедленно возвращает их свободным рабочим процессорам для продолжения итераций.

Метод реализует динамическую балансировку загрузки процессоров. Номера необработанных частиц хранятся в очереди FIFO. Мастер-процессор обновляет координаты и скорость первой частицы из этой очереди и посылает их первому свободному процессору.

Сравнение эффективности синхронного и асинхронного методов. Применение асинхронного метода оправдано, когда вычисление целевой функции $F(X)$ при разных значениях вектора X происходит за различные интервалы времени, т. е. когда вычислительная сложность этой функции переменна.

В работе [20] исследовано влияние дисперсии вычислительной сложности целевой функции и гетерогенности вычислительной системы на ускорение и эффективность синхронного и асинхронного методов. Вычислительная сложность функции $F(X)$ полагалась случайной величиной с заданным законом распределения и моделировалась с помощью соответствующих временных задержек.

Исследование показало, что при увеличении дисперсии вычислительной сложности функции $F(X)$ от нуля (что эквивалентно постоянной вычислительной сложности) эффективность синхронного метода линейно падает. Асинхронный метод на малом числе процессоров (менее 15) показывает в этих условиях низкую эффективность, но при дальнейшем увеличении числа процессоров эффективность метода асимптотически возрастает до 93 %. Такая зависимость эффективности от числа процессоров не является типичной для параллельных вычислений и объясняется малой долей коммуникаций в процессе вычислений, а также распределением их по времени (в отличие от единовременных коммуникаций в синхронном методе). Вариации вычислительной сложности целевой функции практически не оказывают

влияния на эффективность асинхронного метода. Указанные достоинства асинхронного метода обуславливают его высокую эффективность при использовании сильно гетерогенных кластерных вычислительных систем.

Таким образом, метод PAPSO неэффективен при малом числе процессоров, но обеспечивает высокую эффективность при большом числе процессоров (вплоть до N). Для обеспечения эффективной работы асинхронного метода при большем N числе процессоров необходимо распараллеливать вычисления значений целевой функции $F(X)$.

Важно отметить, что при параллельной реализации одного и того же метода PSO в синхронном и асинхронном вариантах, при прочих равных условиях, результаты оптимизации могут быть различными.

Заключение

Известно большое число модификаций метода PSO, направленных на повышение эффективности поиска. Эти модификации могут содержать и, как правило, содержат значительное число свободных параметров, оптимальные значения которых зависят от класса оптимизируемых функций. Отыскание оптимальных значений указанных параметров для каждого из классов функций требует очень высоких вычислительных затрат. Поэтому рекомендации по априорному выбору значений этих параметров очень скупы. Перспективной представляется разработка адаптивного программного обеспечения, реализующего метод PSO и его модификации, в котором поиск оптимальных значений свободных параметров метода осуществляется в процессе нормальной эксплуатации этого программного обеспечения.

Поскольку практически значимые оптимизационные задачи имеют, как правило, высокую вычислительную сложность целевой функции $F(X)$, важным положительным свойством метода PSO и его модификаций является возможность построения хорошо масштабируемых параллельных вариантов метода, ориентированных на вычислительные системы, как с общей, так и с распределенной памятью. Параллельный метод PSO и его модификации имеют еще большее число свободных параметров. Поэтому указанная выше проблема поиска их значений, оптимальных для классов целевых функций, является в данном случае еще более актуальной.

Для практически значимых задач оптимизации типичной является ситуация, когда вычислительная сложность целевой функции зависит от вектора варьируемых параметров X . В связи с этим другой актуальной проблемой при параллельной реализации метода PSO и его модификаций является проблема разработки и реализации соответ-

ствующих методов балансировки загрузки вычислительной системы.

В настоящее время активно разрабатываются различные параллельные алгоритмы и соответствующее программное обеспечение, ориентированные на вычисления на графических картах. Представляется актуальной разработка такого же программного обеспечения для метода PSO.

Список литературы

1. **Weise T.** Global Optimization Algorithms Theory and Application: Ph. D. Thesis. University of Kassel, 2008.
2. **Kennedy J., Eberhart R.** Particle swarm optimization // Proceedings of IEEE International conference on Neural Networks. 1995. P. 1942–1948.
3. **Li X., Engelbrecht A. P.** Particle swarm optimization: an introduction and its recent developments // Proceedings of the 2007 GECCO conference companion on Genetic and evolutionary computation. London: ACM. 2007. P. 3391–3414.
4. **Mendes R., Kennedy J., Neves J.** The fully informed particle swarm: Simpler, maybe better // IEEE Transactions on Evolutionary Computation. 2004. V. 8. P. 204–210.
5. **Mendes R., Kennedy J., Neves J.** Watch thy neighbor or how the swarm can learn from its environment // Proceedings of the Swarm Intelligence Symposium 2003. IEEE. 2003. P. 88–94.
6. **Jordan J., Helwig S., Wanka R.** Social interaction in particle swarm optimization, the ranked FTPS, and adaptive multi-swarms // Proceedings of the 10th annual conference on Genetic and evolutionary computation. Atlanta, USA: ACM. 2008. P. 49–56.
7. **Veeramachaneni K., Peram Th., Mohan Ch., Osadciw L. A.** Optimization Using Particle Swarm with Near Neighbor Interactions // Lecture Notes Computer Science. Springer Verlag, 2003.
8. **Kennedy J., Mendes R.** Population structure and particle swarm performance // Proceedings of the 2002 Evolutionary Computation Congress. Washington: IEEE Computer Society. 2002. P. 1671–1676.
9. **Воеводин В. В., Воеводин Вл. В.** Параллельные вычисления. СПб.: БХВ-Петербург, 2002.
10. **Kennedy J., Mendes R.** Neighborhood topologies in fully informed and best-of-neighborhood particle swarms // Systems, Man, and Cybernetics. 2006. V. 36. P. 515–519.
11. **Elshamy W., Emarah H. M., Bahgat A.** Clubs-based Particle Swarm Optimization // Swarm Intelligence Symposium. 2007. P. 289–296.
12. **Kennedy J.** Stereotyping: improving particle swarm performance with cluster analysis // Proceedings of the 2000 Congress on Evolutionary Computation. 2002. V. 2. P. 1507–1512.
13. **Suganthan P. N.** Particle swarm optimizer with neighborhood operator // Proceedings of the 1999 Congress on Evolutionary Computation. 1999. V. 3. P. 1958–1962.
14. **Richards M., Ventura D.** Dynamic Sociometry in Particle Swarm Optimization // Proceedings of the Joint Conference on Information Sciences. 2003. P. 1557–1560.
15. **Mohais A. S., Mendes R., Ward Ch., Posthoff Ch.** Neighborhood Re-structuring in Particle Swarm Optimization // Proceedings of the Australian Conference on Artificial Intelligence. 2005. P. 776–785.
16. **Dorigo M., Maniezzo V.** Parallel Genetic Algorithms: Introduction and Overview of Current Research // Parallel Genetic Algorithms: Theory and Applications / Ed. by J. Stenders. Amsterdam: IOS Press, 1992.
17. **Belal M., El-Ghazawi T.** Parallel Models for Particle Swarm Optimizers // International Journal of Intelligent Computing and Information Sciences. 2004. V. 1. P. 100–111.
18. **Shu-Chuan Chu, Jeng-Shyang Pan.** Intelligent Parallel Particle Swarm Optimization Algorithms // Parallel Evolutionary Computations. Springer Berlin, Heidelberg. 2006. V. 22. P. 159–175.
19. **Schutte J. F., Reinbolt J. A., Fregly B. J., Haftka R. T., George A. D.** Parallel global optimization with particle swarm algorithm // Journal of Numerical Methods in Engineering. 2004. V. 61. P. 2296–2315.
20. **Koh B., George A. D., Haftka R. T., Fregly B. J.** Parallel asynchronous particle swarm optimization // International Journal for Numerical Methods in Engineering. 2006. V. 67 (4). P. 578–595.
21. **Venter G., Sobieszcanski-sobieski J.** A parallel particle swarm optimization algorithm accelerated by asynchronous evaluations // Journal of Aerospace Computing, Information, and Communication. 2006. V. 3. P. 123–137.

УДК 621.3

Н. И. Лиманова, д-р техн. наук, проф., **Е. А. Мамзин**, аспирант,
Тольяттинский государственный университет, e-mail: N.Limanova@tltsu.ru

Дискретная математическая модель детерминированного клеточного автомата и ее программная реализация

Предложены математическая модель детерминированного клеточного автомата на шестигранной пространственной сетке и ее программная реализация. Особенностью разработанного автомата является пошаговая смена конфигураций элементов, соседних с текущим, по состояниям которых определяется состояние текущего элемента. Программирование выполнено на языке Си. Визуализация результатов вычислений проводилась с помощью графической библиотеки OpenGL. Использование методов параллельного программирования и пошаговой смены конфигураций элементов позволили увеличить скорость вычислений более чем в 2 раза по сравнению с известными аналогами.

Ключевые слова: клеточный автомат, математическая модель, пространственная сетка, конфигурация соседних элементов, скорость вычислений

Введение

Моделирование различных физических процессов и явлений в большинстве случаев основано на решении дифференциальных уравнений, используемых для описания этих процессов. Од-

нако аналитическое решение таких уравнений относится к довольно трудоемким задачам, к тому же его можно найти не всегда. Этим объясняется широта использования различных численных методов, одним из которых является метод клеточ-

ных автоматов [1]. В соответствии с [1] под клеточными автоматами понимаются сети элементов, изменяющих свое состояние в последовательные дискретные моменты времени по определенному закону в зависимости от того, каким было состояние самого элемента и его соседей в предыдущий дискретный момент времени. В данной статье рассматривается математическая модель детерминированного клеточного автомата, для которого определен закон перехода из предыдущего состояния в последующее. В основу реализации подавляющего большинства моделей клеточных автоматов положен шаблон, имеющий вид четырехгранной сетки. Однако в ряде случаев такой автомат не является оптимальным. Если рассматривать его с точки зрения взаимодействия соседних элементов сетки, то можно заметить, что диагональные элементы дальше удалены от центрального и должны оказывать меньшее воздействие на него, что необходимо учитывать при написании программного кода. Таким образом, получается, что для вычисления состояния текущего элемента клеточного автомата необходимо учитывать состояния восьми соседних, причем четыре диагональных клетки влияют на центральную слабее, чем клетки, соприкасающиеся с ней общими гранями. Целью данной статьи являлась разработка клеточного автомата на основе шестигранной сетки, не обладающего недостатками известных аналогов.

Разработка дискретной математической модели детерминированного клеточного автомата

Рассмотрим задачу моделирования волновых колебаний поверхности вязкой жидкости без учета близости дна. Поставленная задача была решена на основе разработанного авторами шестигранного клеточного автомата [2] с правилами переходов, основанными на методе конечных разностей. Число ближайших соседей для любого элемента шаблона в данном случае равно шести, причем все клетки соприкасаются друг с другом только общими гранями. Данная модель позволяет получать достоверные картины волновых колебаний поверхности жидкости для различного типа волн, произвольного расположения источников колебаний, препятствий и барьеров без решения уравнения колебаний для каждого случая.

Для моделирования волновых колебаний детерминированным клеточным автоматом необходимо получить разностную схему решения широко известного уравнения колебаний [3]

$$\Delta u = \frac{1}{a^2} \frac{\partial^2 u}{\partial t^2}, \quad (1)$$

где Δ — оператор Лапласа; $u = u(t, r)$ — неизвестная функция; $r \in R^n$ — пространственная переменная; $t \in R$ — время; $\frac{1}{a^2} = v^2$ — фазовая скорость.

Рассмотрим уравнение колебаний (1) для двумерного случая [3]:

$$\frac{\partial^2 u}{\partial t^2} = a^2 \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right). \quad (2)$$

Составим для данного уравнения разностную схему, проведя следующие замены [4]:

$$\frac{\partial^2 u_x}{\partial x^2} = \frac{u_{x+1} - 2u_x + u_{x-1}}{h^2} \quad (3)$$

и

$$\frac{\partial^2 u}{\partial t^2} = \frac{u^{t+1} - 2u^t + u^{t-1}}{\tau^2}, \quad (3^*)$$

где $u_{i,j}^t$ — значение неизвестной функции в момент времени t в точке с координатами (i, j) . Здесь индексы вида $(i+1, j)$ показывают, что рассматривается следующая точка по координате i относительно текущей.

Подставив замены (3) и (3*) в волновое уравнение (2), получим:

$$\frac{u_{i,j}^{t+1} - 2u_{i,j}^t + u_{i,j}^{t-1}}{\tau^2} = a^2 \left(\frac{u_{i+1,j}^t - 2u_{i,j}^t + u_{i-1,j}^t}{h^2} + \frac{u_{i,j+1}^t - 2u_{i,j}^t + u_{i,j-1}^t}{h^2} \right), \quad (4)$$

где h — шаг пространственной сетки, τ — шаг по времени.

Выразив из (4) следующий по времени шаг для функции $u = u_{i,j}^{t+1}$, получим разностную схему для решения волнового уравнения (2):

$$u_{i,j}^{t+1} = \left(\frac{a^2 \tau^2}{h^2} \right) (u_{i-1,j}^t + u_{i+1,j}^t + u_{i,j-1}^t + u_{i,j+1}^t - 4u_{i,j}^t) + 2u_{i,j}^t - u_{i,j}^{t-1}, \quad (5)$$

где $a^2 = \frac{1}{v^2}$ — величина, обратная фазовой скорости.

Уравнение (5) представляет собой математическую модель детерминированного клеточного автомата в дискретные моменты времени. Из (5) видно, что множитель $u_{i-1,j}^t + u_{i+1,j}^t + u_{i,j-1}^t + u_{i,j+1}^t - 4u_{i,j}^t$ определяет среднее значение амплитуды колебаний центральной клетки на основании известных значений амплитуд соседних, из которого вычитается текущая амплитуда центральной ячейки. Основываясь на этом, можно перейти к шестигранной сетке, переписав данное выражение в виде $\left(\sum_{k=1}^6 u_k^t - 6u_{i,j}^t \right)$, где $\sum_{k=1}^6 u_k^t$ — сумма значений

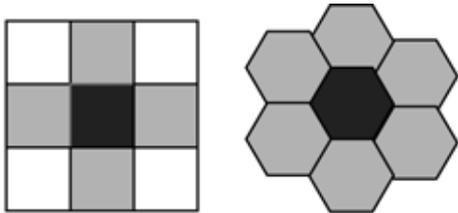


Рис. 1. Шаблоны для моделирования на основе стандартной четырехгранной сетки (слева) и предлагаемой шестигранной (справа)

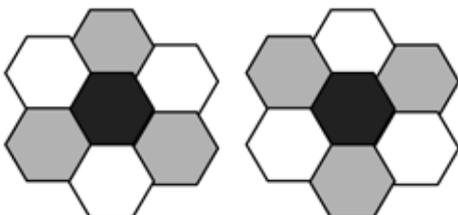


Рис. 2. Шаблон для моделирования на основе шестигранной сетки с пошаговой сменой конфигурации учитываемых соседних элементов

амплитуд колебаний по всем соседям центральной ячейки на шестигранной сетке.

На рис. 1 представлены рассматриваемые шаблоны для моделирования. Черным цветом показана центральная ячейка, расчет состояния которой проводится в текущий момент времени, серым цветом — соседние ячейки, по которым определяется состояние центральной.

Переход к шестигранной сетке позволяет выровнять свойства дискретного пространства по всем направлениям, но дает ускорение только в том случае, когда для расчета требуется брать еще и диагональные элементы. Необходимо отметить, что диагональные элементы не учитываются только в случае очень грубых, приблизительных расчетов. В целях повышения точности вычислений, а также для получения наглядной картины протекания процессов эти элементы необходимо учитывать. В этом случае скорость вычислений, которую дает шестигранный клеточный автомат, увеличивается на 30—40 % по сравнению с известными аналогами на четырехгранных сетках. Для дальнейшего увеличения скорости работы программ на основе рассматриваемого клеточного автомата следует учитывать одновременно на каждом шаге вычислений параметры не шести соседних элементов, а трех:

$$u_{i,j}^{t+1} = \left(\frac{a^2\tau^2}{h^2}\right) \left(\sum_{k=1}^3 u_k^t - 3u_{i,j}^t\right) + 2u_{i,j}^t - u_{i,j}^{t-1}. \quad (6)$$

Такой подход позволяет сократить число вычислений на одном шаге, что приводит к ускоре-

нию расчета смены состояний. Для повышения стабильности способа расчета (6) авторы предлагают использовать пошаговую смену конфигурации учитываемых в расчетах соседних элементов, т. е. на каждом шаге вычисления менять соседние ячейки так, как это показано на рис. 2.

Программа, использующая пошаговую реконфигурацию шаблона, показала увеличение скорости вычислений на 50 % по сравнению с программным кодом без ее использования.

Применение параллельных методов вычислений в программной реализации модели

Эффективным способом продвижения в область исследования сложных математических моделей является применение параллельных методов программирования, которые позволяют разбить процесс вычисления на потоки, выполняющиеся одновременно на базе высокопроизводительных вычислительных комплексов или кластеров.

К неоспоримым достоинствам модели разработанного клеточного автомата можно отнести возможность распараллеливания процесса вычислений [5], что позволяет значительно увеличить скорость расчетов. Для данного случая наиболее оптимальным методом распараллеливания является метод декомпозиции данных. На рис. 3 представлена наглядная схема такой декомпозиции.

При таком способе распараллеливания весь массив, представляющий моделируемую плоскость, разбивается на равные элементы, которые рассылаются с сервера на прочие компьютеры в сети для обработки. После получения результатов обработки компьютеры отправляют их обратно серверу, который, в свою очередь, соединяет элементы и отображает на экране полученную кар-

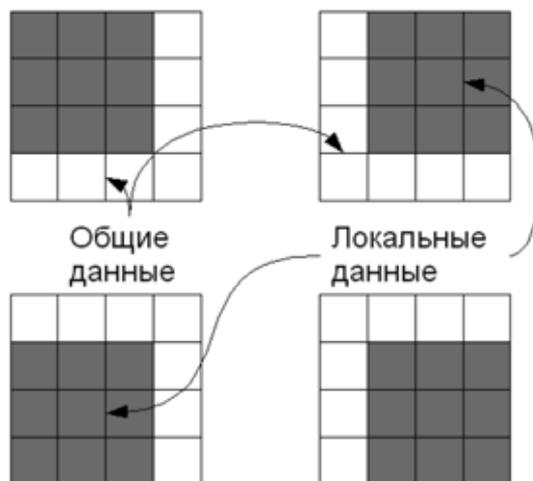


Рис. 3. Принцип декомпозиции данных

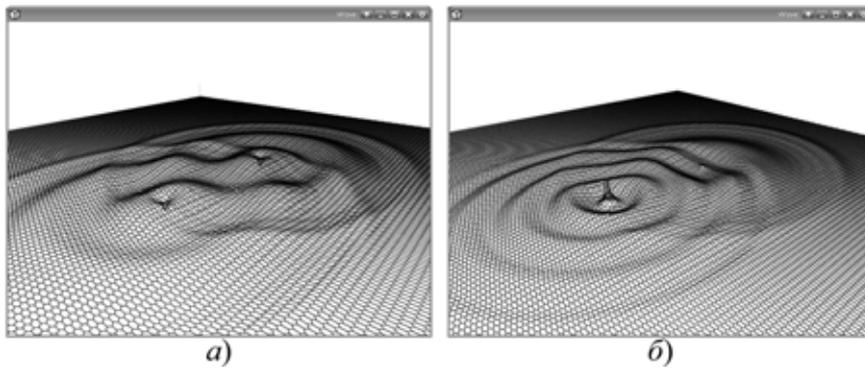


Рис. 4. Распространение волн от двух когерентных (а) и некогерентных (б) источников

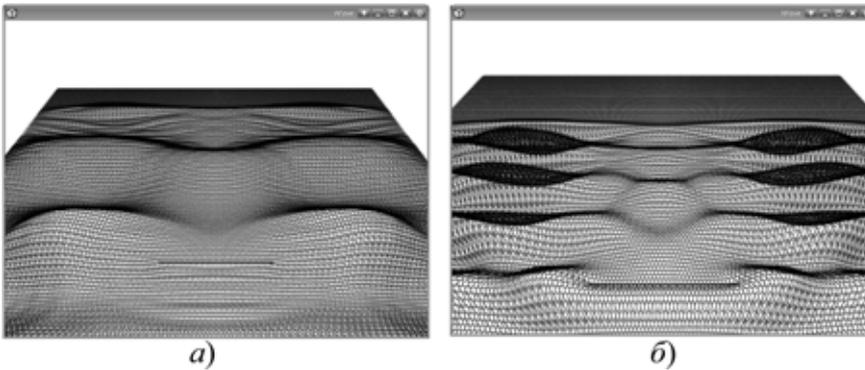


Рис. 5. Огибание препятствия плоской волной в случаях, когда длина волны больше линейных размеров препятствия (а) и когда меньше (б)

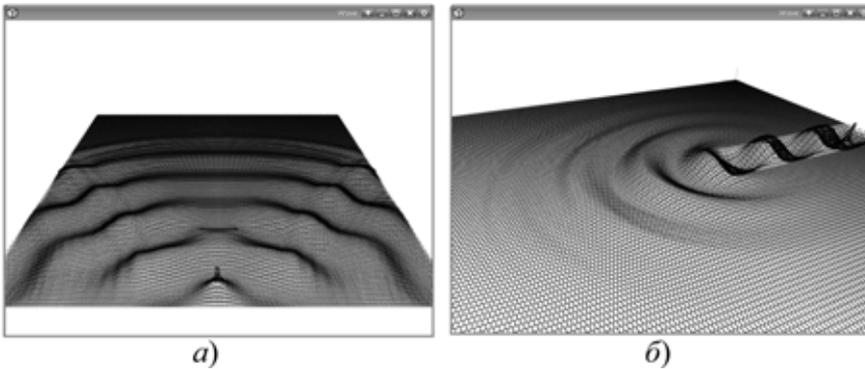


Рис. 6. Огибание сферической волной малого препятствия (а). Источник колебаний располагается в пространстве, ограниченном с трех сторон барьерами, на выходе из которых образуется сферическая волна (б)

Результаты моделирования

На основе разработанной модели детерминированного шестигранного клеточного автомата и библиотеки распараллеливания MPI [6] была написана программа, наглядно демонстрирующая динамику распространения волн на поверхности вязкой жидкости без учета близости дна. Программирование выполнено на языке Си. Визуализация результатов вычисления осуществлялась с помощью графической библиотеки OpenGL.

На рис. 4, 5 и 6 представлены результаты работы программы. Рис. 4 наглядно демонстрирует интерференцию волн, распространяющихся от двух когерентных (а) и некогерентных (б) источников. На рис. 5 представлена дифракция плоской волны для двух случаев: длина волны больше линейных размеров препятствия (а) и длина волны меньше линейных размеров препятствия (б).

Разработанный детерминированный шестигранный клеточный автомат позволяет рассматривать и такие сложные для моделирования случаи, как взаимодействие сферической волны с препятствием (рис. 6, а) и распространение волны в пространстве, ограниченном с трех сторон барьерами, на выходе из которых образуется сферическая волна (рис. 6, б).

Заключение

Переход к автомату на шестигранной сетке позволяет увеличить скорость вычислений на 30—40 % по сравнению с известными аналогами на четырехгранных сетках.

Эффективность работы рассматриваемого клеточного автомата оценивалась по соотношению характерного времени работы его программной реализации и времени работы программы, реализующей метод конечных разностей, на 10 000 итераций.

Дальнейшему ускорению работы программы способствовало использование распараллеливания программного кода и пошаговой смены конфигураций, учитываемых в расчетах соседних элементов, что в конечном итоге дало возможность наблюдать моделируемые процессы в динамике в реальном масштабе времени.

тину. Число и размер элементов, на которые разбивается исходный массив данных, зависит от размеров обрабатываемого массива. В процессе разработки программы учитывалось также, что пересылка данных занимает сравнительно большее время.

Распараллеливание вычислений выполнялось на компьютерном кластере Тольяттинского государственного университета. Кластер объединяет 168 компьютеров. Как показало тестирование, быстродействие системы на 60 узлах составило 30 Gflops.

Рассмотренный пример практического использования предложенной модели показывает, что детерминированный шестигранный клеточный автомат прекрасно описывает такие волновые явления, как дифракция и интерференция, позволяет в интерактивном режиме изменять параметры моделируемого пространства, т. е. задавать граничные и начальные условия, устанавливать центры источников колебания, преграды и т. д. Были реализованы вращение моделируемой области, приближение и возможности установки паузы для детального рассмотрения участка сетки. Необходимо отметить, что область практического применения предложенного автомата не ограничивается моделированием только волновых колебаний, а позволяет рассматривать широкий круг явлений и процессов в динамике. Программная реализация разработанного клеточного автомата

обладает скоростью вычислений, более чем в 2 раза превышающей известные аналоги.

Список литературы

1. **Тоффоли Т., Марголюс Н.** Машины клеточных автоматов. М.: Мир, 1991. 280 с.
2. **Лиманова Н. И., Мамзин Е. А.** Моделирование физических явлений на основе использования клеточных автоматов // V Всесибирский конгресс женщин-математиков: Матер. конф. 15—18 января 2008 г. Красноярск: РИО СФУ, 2008. С. 262—266.
3. **Тихонов А. Н., Самарский А. А.** Уравнения математической физики. М.: Наука, 1977. 404 с.
4. **Гулин А. В., Самарский А. А.** Численные методы. М.: Наука, 1989. 432 с.
5. **Лиманова Н. И.** Распределенные параллельные вычисления в задачах математического моделирования // Матер. Международной науч. конф. "Космос, астрономия и программирование (Лавровские чтения)" 20—22 мая 2008 г. СПб.: СПбГУ, 2008. С. 211—215.
6. **Богачев К. Ю.** Основы параллельного программирования. М.: БИНОМ. Лаборатория знаний, 2003. 342 с.

УДК 004.3.07

И. В. Ковель, аспирант,
Кубанский государственный
технологический университет,
e-mail: vinitar@yandex.ru

Вероятностный метод построения окрестности статистического глобального оптимума NPC-задач

Рассматривается вероятностный метод получения ранжированных субоптимальных решений NPC-задач в окрестности статистического глобального оптимума. Суть метода заключается в построении решений с использованием генераторов псевдослучайных чисел, работающих по статистически обоснованной геометрической функции вероятности рангов локальных решений.

Ключевые слова: NPC-задача, статистический глобальный оптимум, окрестность, ранг, локальный оптимум, задача коммивояжера, оптимизация раскроя-упаковки

Постановка задачи

Круг рассматриваемых недетерминированных полиномиальных полных задач (NPC-задач) ограничен задачами потребления невозобновляемых ресурсов произвольной природы. К таким задачам можно отнести задачу коммивояжера, задачу оптимизации раскроя-упаковки и нестинга, зада-

чу оптимизации размещения обслуживающих центров и т. п.

Исходные данные задачи описываются в виде графов без петель и кратных ребер $G = (V, E, d)$, где V — множество вершин-ресурсов v_i произвольной природы, $|V| = n$, каждая вершина v_i характеризуется вектором параметров ресурсов $P_i = (p_1, p_2, \dots, p_d)$, определенных на множестве \mathbb{R}^+ , $p_i \in \mathbb{R}^+$; d — размерность ресурсов, определенная на множестве \mathbb{Z}^+ ; E — множество ребер e_{ij} , задающих отношение смежности произвольной природы между вершинами v_i и v_j , $e_{ij} = e_{ji}$, $i \neq j$, $|E| \leq \frac{n^2 - n}{2}$.

Потребление ресурсов заключается в пошаговом исключении из графа G выбираемых по определенному рангу r_i ресурсов v_i и формировании ранжированной цепи $\pi(n, k) = [r_1, r_2, \dots, r_{k-1}]$, k — длина цепи, $k \leq n + 1$ [1]. Ранжированное дерево решений задается множеством ранжированных цепей $\Pi_{n,k} = \{\pi(n, k) \mid 0 < k \leq n + 1\}$.

Пусть вес вершины v графа G задается функцией $q(v) = \prod_{i=1}^d p_i$, а вес цепи $h(\pi) = \sum_{i=1}^k q(v_i)$. Задача состоит в нахождении подмножества субоптимальных решений $\Pi_{n,k}^* \subset \Pi_{n,k}$, которое содержит только те цепи $\pi^* \in \Pi_{n,k}^*$, у которых наи-

большая вероятность того, что их вес $h(\pi^*)$ минимален во всем множестве $\Pi_{n, k}$:

$$\Pi_{n, k}^* = \{\pi^* | P\{h(\pi^*) = \min_{\Pi_{n, k}} (h(\pi_i))\} > P\{h(\pi') = \min_{\Pi_{n, k}} (h(\pi_i))\}\}, \quad (1)$$

где $\pi' \in (\Pi_{n, k} \setminus \Pi_{n, k}^*)$.

Генератор рангов разрядов субоптимальных ранжированных цепей

Указанная в [2] статистическая закономерность распределения значений рангов r в i -х разрядах цепи $\pi(n, k) = [r_1, r_2, \dots, r_{k-1}]$ позволяет задавать закон генерации величин рангов посредством датчиков псевдослучайных чисел с геометрической функцией вероятности

$$P(r_i) = \vartheta_i(1 - \vartheta_i)^{r_i}. \quad (2)$$

Для поиска субоптимальных цепей $\pi(n, k)$ предлагается реализовать генератор рангов в виде k параллельно работающих генераторов псевдослучайных чисел (ГПСЧ). На рис. 1 представлена структурная схема генератора рангов. На данном рисунке величины рангов r_i каждого разряда цепи выставляются на выходах ГПСЧ. На входы ГПСЧ подается тактирующий сигнал "Разрешение", а также параметры цепи n и k .

Экспериментально полученные для $n \leq 15$ вероятности $P(r_i)$ однозначно определяют значения коэффициента ϑ_i при $r_i = 0$. Поэтому для нахождения ϑ_i в зависимости от n, i достаточно найти функциональную зависимость $P(0_i) = \varphi(n, i)$.

Нахождение функции $\varphi(n, i)$

Рассмотрим алгоритм определения функции $\varphi(n, i)$ для случая нахождения цепи $\pi(n, n)$ минимального веса на неориентированном полном графе, расположенном на произвольной криволинейной поверхности, для которого не выполняется неравенство треугольника, т. е. сумма расстояний между произвольными вершинами A, B и B, C может быть меньше расстояния между вершинами A, C .

1. Представим функцию $\varphi(n, i)$ в виде

$$\varphi(n, i) = a(i)e^{-b(i)n} + c(i) \quad (3)$$

с условиями $a(i) > 0, b(i) > 0, c(i) > 0$. В настоящем исследовании определено, что только экспоненциальная аппроксимация имеет наименьшее значение среднего квадратичного отклонения среди прочих видов функций. Значения $a(i), b(i), c(i)$ функции (3) для $i < 10$ приведены в таблице, которая дополнительно содержит значения σ_i — средние квадратичные отклонения функции (3) от

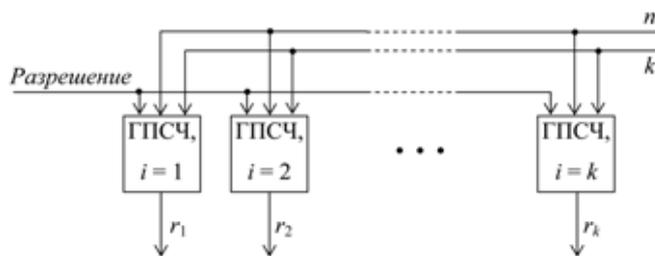


Рис. 1. Структурная схема генератора рангов

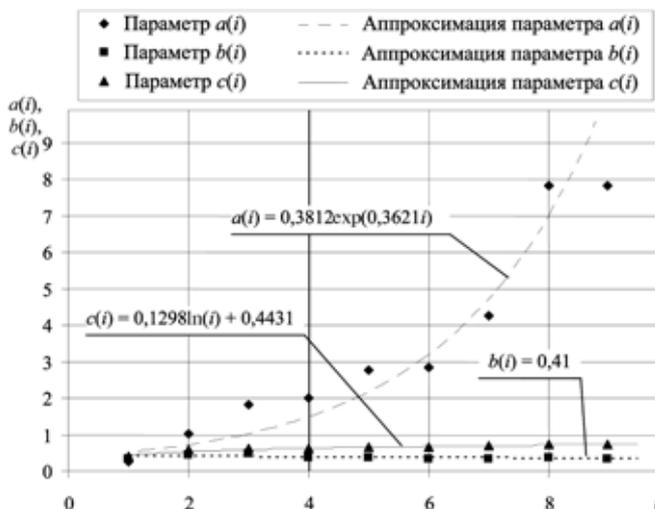


Рис. 2. Экстраполяция параметров функции (3)

Параметры аппроксимирующих экспонент

i	$a(i)$	$b(i)$	$c(i)$	σ_i
1	0,266	0,338	0,384	0,001
2	1,004	0,437	0,595	0,004
3	1,837	0,457	0,627	0,003
4	2,008	0,377	0,633	0,004
5	2,759	0,361	0,643	0,003
6	2,838	0,319	0,653	0,003
7	4,270	0,335	0,680	0,002
8	7,834	0,372	0,711	0,002
9	7,834	0,339	0,721	0,0006

экспериментальных значений коэффициента $P(0_i)$ из (2).

2. Рассматривая столбцы таблицы $a(i), b(i), c(i)$ в качестве исходных данных, можно легко описать закономерности изменения значений в указанных столбцах для $i > 15$, которые имеют наименьшие средние квадратичные отклонения среди прочих функций:

$$\begin{cases} a(i) = 0,3812e^{0,3621i}; \\ b(i) = 0,41; \\ c(i) = 0,1298\ln(i) + 0,4431. \end{cases} \quad (4)$$

На рис. 2 представлены значения коэффициентов $a(i), b(i), c(i)$, а также приведены функции, описывающие изменение этих коэффициентов.

3. Подставив (4) в (3), получим формулу для расчета $\varphi(n, i)$:

$$\varphi(n, i) = 0,3812e^{0,3621i - 0,41n} + 0,1298\ln(i) + 0,4431.$$

Определение коэффициента ϑ_i геометрического распределения для цепи $\pi(n, k)$ с произвольным параметром k основано на анализе априорно полученных экспериментальных данных о распределении рангов всех разрядов цепей $\pi(n, k)$ при $5 \leq n \leq 15$, а также на допущении о том, что выявленные закономерности сохраняются и при $n > 15$.

Построение окрестности глобального оптимума

Глобальным оптимумом (ГО) в рассматриваемых NPC-задачах является решение, получаемое по принципу Беллмана. Такое решение описывается ранжированной цепью $\pi(n, k) = [0_1, 0_2, \dots, 0_{k-1}]$, т. е. такой цепью, в которой все ранги имеют нулевое значение, определяя тем самым на каждом шаге решения выбор локального оптимума. Множество (1), определенное в постановке задачи, назовем окрестностью ГО. Для построения окрестности ГО предлагается использовать генератор цепей, структурная схема которого представлена на рис. 3.

Конвертер преобразует ранжированную цепь $\pi(n, k) = [r_1, r_2, \dots, r_{k-1}]$ в цепь вершин $\pi(n, k) = [v_1, v_2, \dots, v_k]$ в целях последующего вычисления ее веса $h(\pi)$ как суммы весов вершин и/или ребер. Селектор субоптимальной цепи сохраняет в своем регистре наилучшую текущую цепь, т. е. цепь с минимальным весом среди прочих построенных, и выдает сигнал на разрешение генерации следующей ранжированной цепи.

Оценка эффективности

В настоящее время существует ряд вероятностных методов, использующих различные универсальные и специальные эвристики локализации субоптимальных решений в пространстве поиска [3]. Данные эвристики основаны на здравом смысле и логике. К универсальным относятся эвристики, используемые в методе имитации отжига, муравьином алгоритме, генетическом алгоритме и др. [4]. Специальные эвристики используются для решения конкретных задач. Например, для задач раскроя-упаковки к таким эвристикам можно отнести вероятностную схему Усмановой А. Р. [5], рандомизированный метод динамического перебора с усечением Валеевой А. Ф., Гареева И. Р. и Мухачевой Э. А. [6], ряд вероятностных методов локального поиска для задач дискретной оптимизации, описанный Кочетовым Ю. А. [7], вероятностный поиск с запретами для задачи упаковки в контейнеры Кочетова Ю. А. и Усмановой А. Р. [8], рандомизированную эвристику на базе двойст-

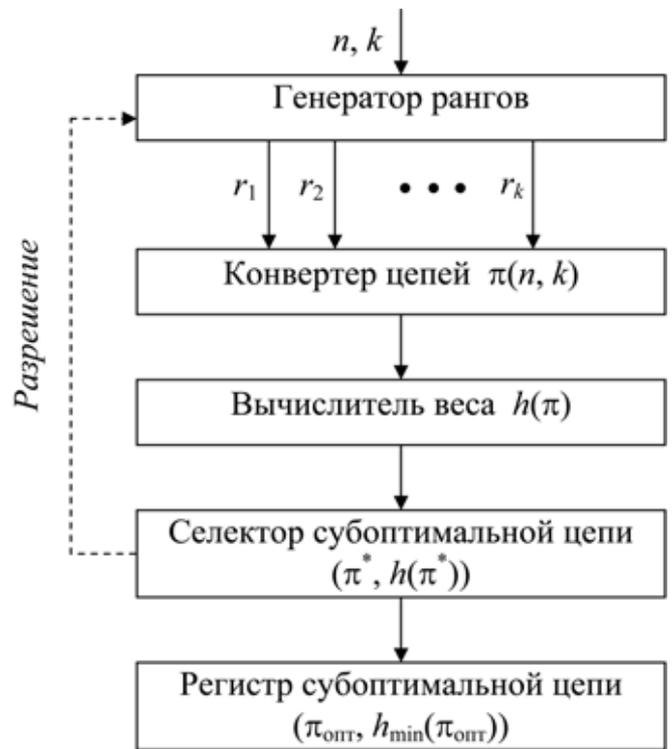


Рис. 3. Структурная схема генератора цепей в окрестности ГО

венной схемы локального поиска оптимума Мухачевой Э. А. и Ширгазина Р. Р. для задач упаковки прямоугольников [9].

Отличием предлагаемого вероятностного метода от перечисленных является использование статистически обоснованных эвристик распределения рангов выбираемых локальных решений.

Оценка эффективности предложенного вероятностного метода проведена путем сравнения решений задачи коммивояжера для $20 \leq n \leq 120$ с решениями, полученными методом "имитации отжига" [3], который использует экспоненциальный закон распределения вероятностей $P(n, i) = A(i)e^{-B(i)n}$ локальных решений. В качестве показателя оценки эффективности выбрано среднее за N опытов отношение теоретической оценки веса наилегчайшей цепи h_{inf} каждого графа к весу h_{min} цепи, определяемой каждым методом:

$$M(i) = \frac{1}{N} \sum_{i=1}^N \frac{h_{inf,i}}{h_{min,i}}.$$

Теоретическая нижняя оценка веса h_{inf} вычислялась как половинная сумма весов двух наилегчайших ребер, т. е. ребер с рангами $r = 0, 1$, инцидентных каждой вершине v_i исходного графа:

$$h_{inf} = \frac{1}{2} \sum_{i=1}^n \sum_{r=0}^1 w_r(e_i),$$

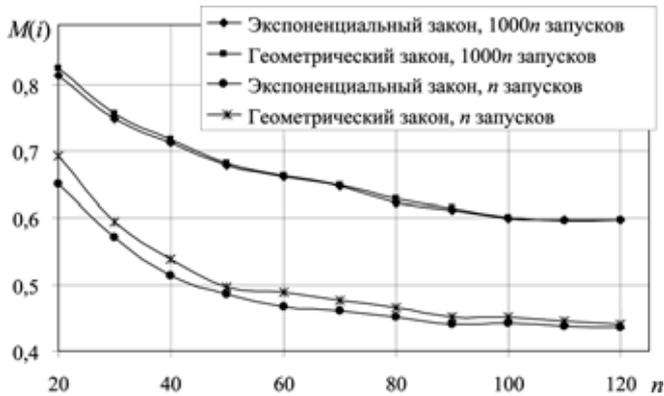


Рис. 4. Экспериментальное сравнение эффективности вероятностных методов, использующих геометрический и экспоненциальный законы распределения вероятностей

где $w_r(e_i)$ — вес ребра ранга r , инцидентного вершине v_i . Веса ребер полного графа для каждого опыта генерировались случайной функцией с равномерным распределением.

На рис. 4 приведены результаты сравнения, которые позволяют сделать вывод о том, что предлагаемый вероятностный метод быстро получает одно из субоптимальных решений, но при последующих запусках достаточно медленно его улучшает, что объясняется повтором наиболее вероятных ранее проверенных решений. Одновременно с этим алгоритм "имитации отжига" быстрее улучшает текущее субоптимальное решение, постепенно компенсируя проигрыш на первых запусках.

С ростом размерности задачи n различия методов уменьшаются. Это объясняется малым значением разности между вероятностями рангов первых и последующих разрядов субоптимальной цепи, что значительно увеличивает окрестность СГО (1).

Выводы

Предлагаемый метод имеет достоинства и недостатки, присущие численным вероятностным методам.

Достоинства предлагаемого метода:

- быстрое получение субоптимального решения;
- отсутствие притяжения к локальным оптимумам;
- наиболее часто проверяются субоптимальные по вероятности решения;

- возможность останова либо по времени, либо по заданному объему поиска, либо при достижении заданного порога веса решения;
 - возможность повторного запуска алгоритма.
- Недостатки предлагаемого метода:
- повторное вычисление решений из окрестности СГО;
 - наличие дополнительных вычислительных затрат, обусловленных генерацией ранжированных цепей и их последующего преобразования в цепи вершин графа исходных данных;
 - эффективность процесса решения пропорциональна времени работы алгоритма;
 - медленное улучшение первоначально полученного субоптимального решения.

Таким образом, предлагаемый метод выгодно использовать для быстрого получения субоптимального решения, которое в дальнейшем может служить "затравкой" для универсальных или специальных алгоритмов решения оптимизационных NPC-задач.

Список литературы

1. Касьянов В. Н., Евстигнеев В. А. Графы в программировании: обработка, визуализация и применение. СПб.: БХВ-Петербург, 2003. 1104 с.
2. Марков В. Н. Ранжирование решений NP-полных задач методом обратных приращений // Исследование, разработка и применение высоких технологий в промышленности: Сб. тр. III Международ. НПК. Санкт-Петербург. 2007. Т. 5. С. 249—250.
3. Макконелл Дж. Основы современных алгоритмов. Изд. 2-е, допол.: Пер. с англ. М.: Техносфера, 2004. 368 с.
4. Джонс М. Т. Программирование искусственного интеллекта в приложениях: Пер. с англ. М.: ДМК Пресс, 2004. 312 с.
5. Усманова А. Р. Вероятностные жадные эвристики для задачи упаковки в контейнеры. СПб.: ОПТИМ, 2001. С. 141—146.
6. Валеева А. Ф., Гареев И. Р., Мухачева Э. А. Задача одномерной упаковки: рандомизированный метод динамического перебора с усечением // Информационные технологии. 2003. № 10. Приложение. 24 с.
7. Кочетов Ю. А. Вероятностные методы локального поиска для задач дискретной оптимизации // Дискретная математика и ее приложения. Сб. лекций молодежных и научных школ по дискретной математике и ее приложениям. М.: Изд-во центра прикладных исследований при механико-математическом факультете МГУ, 2000. С. 87—117.
8. Кочетов Ю., Усманова А. Вероятностный поиск с запретами для задачи упаковки в контейнеры // XII Байкальская международ. конф. Иркутск. 2001. С. 22—27.
9. Мухачева А. С., Ширгазин Р. Р. Задачи упаковки прямоугольников: рандомизированная эвристика на базе двойственной схемы локального поиска оптимума // Информационные технологии. 2003. № 5. С. 18—23.

УДК 658.512.011.56:004.42

Л. И. Бабак, канд. техн. наук, доц.,
С. Ю. Дорофеев, аспирант,
М. А. Песков, аспирант,
М. В. Черкашин, канд. техн. наук, доц.,
Ф. И. Шеерман, канд. техн. наук, доц.,
А. С. Барышников, аспирант,
С. Е. Кошевой, аспирант,
ООО "Элликс", Томский государственный
университет систем управления
и радиоэлектроники,
e-mail: babak@tomica.ru

Интеллектуальная система автоматизированного проектирования СВЧ устройств INDESYS

Описывается интеллектуальная система автоматизированного проектирования (синтеза) СВЧ устройств INDESYS, позволяющая генерировать принципиальную схему и топологию устройства по поставленным требованиям при использовании точных моделей элементов. Кратко рассматриваются структура системы, а также подходы, методы и алгоритмы, положенные в ее основу.

Ключевые слова: СВЧ, радиоэлектронные устройства, синтез, INDESYS, САПР, визуальные вычисления, генетические алгоритмы, декомпозиционный подход, монолитные интегральные схемы, UML, API, Locust-MMIC, Microwave Office, морфологическое дерево

Особенностью современного этапа развития радиоэлектроники является все более широкое освоение СВЧ диапазона, ведущее к появлению качественно новых радиоэлектронных и телекоммуникационных систем с улучшенными характеристиками. Однако существующие САПР, предназначенные для проектирования СВЧ устройств (такие, как ADS, Microwave Office, Serenade, Genesys и др.), обычно решают только задачу моделирования, т. е. расчета характеристик по уже заданной принципиальной схеме и топологии устройства. Выбор последних осуществляется на основе эвристического подхода с использованием опыта разработчика, многократного моделирования, оптимизации различных вариантов и т. д. Это делает процесс проектирования СВЧ уст-

ройств длительным и трудоемким и не гарантирует получения наилучших результатов.

Для повышения эффективности проектирования требуется программное обеспечение, позволяющее решать обратную задачу — задачу синтеза, т. е. определения принципиальной схемы и топологии СВЧ устройства по заданным требованиям. Однако в настоящее время на коммерческой основе предлагаются в основном лишь программы синтеза ограниченного класса пассивных СВЧ устройств — фильтров и согласующих цепей (СЦ), в частности, модуль синтеза фильтров в составе Microwave Office (MWO), модули синтеза СЦ в составе комплексов ADS и Genesys и др. Что касается активных СВЧ полупроводниковых устройств (ППУ), то существует лишь единственная коммерческая программа MultiMatch фирмы Ampsa (США) для автоматизированного синтеза СВЧ транзисторных усилителей, которая также имеет различные ограничения.

В настоящее время наибольший интерес представляют СВЧ устройства, выполненные по *монолитной интегральной технологии*. Для их проектирования требуются программы автоматизированного синтеза СВЧ монолитных интегральных устройств (МИУ), использующие точные модели монолитных элементов. Существующие программы позволяют синтезировать лишь пассивные и активные СВЧ устройства, содержащие идеальные пассивные элементы. Поэтому они плохо пригодны для проектирования СВЧ МИУ. Кроме того, в указанных программах не контролируются или недостаточно контролируются структура и значения номиналов элементов синтезируемых цепей, что может привести к получению практически неосуществимых решений.

В настоящей статье описывается разрабатываемая интеллектуальная система автоматизированного проектирования (синтеза) СВЧ устройств INDESYS (Intelligent Design System). Она позволит повысить эффективность проектирования СВЧ устройств (в том числе МИУ), т. е. упростить и ускорить процесс проектирования, улучшить качественные показатели устройств, снизить требования к квалификации разработчиков.

Следует отметить, что исследования в области теории, методов и алгоритмов структурного синтеза СВЧ цепей выполняются авторами в течение ряда лет [1—10]. В результате был создан комплекс экспериментальных программ, предназна-

ченных для синтеза определенных типов СВЧ устройств [6, 8—13]. Однако эти программные продукты обладают жесткой внутренней структурой и используются практически отдельно друг от друга. В связи с этим их применение не очень удобно, а поддержка и дальнейшее развитие значительно осложнены. Одной из целей разработки системы INDESYS является преодоление указанных недостатков.

Основные характеристики системы

Система INDESYS предназначена для проектирования различных типов активных СВЧ устройств, в том числе транзисторных усилителей (линейных, малошумящих и мощных), активных фильтров, конвертеров иммитанса, а также пассивных согласующих и корректирующих (КЦ) цепей и фильтров. При развитии в нее будут включены также СВЧ устройства других типов, в частности, автогенераторы, полупроводниковые управляющие устройства (аттенюаторы, фазовращатели, амплитудные и фазовые манипуляторы) и т. д. При этом система создается не только как средство синтеза указанных классов СВЧ устройств, но благодаря открытой архитектуре может легко дополняться новыми модулями. Данные модули могут реализовывать новые методы синтеза и быть адаптированными под другие типы решаемых задач и т. д.

Для повышения эффективности в системе INDESYS одновременно применяются два различных принципа "интеллектуализации" программ автоматизированного проектирования:

- использование интерактивных человеко-машинных процедур, объединяющих возможности современных ЭВМ по обработке и отображению информации с интеллектуальными способностями человека по анализу визуальной информации и принятию решений;
- использование элементов искусственного интеллекта и, в частности, генетических алгоритмов (ГА), основанных на имитации моделей природной эволюции.

В основу системы положен ряд разработанных авторами новых подходов, в том числе:

- декомпозиционный метод синтеза активных СВЧ ППУ [1, 2];
- процедура решения систем нелинейных неравенств на основе построения проекций многомерных областей [14, 15];
- оригинальные интерактивные процедуры "визуального" проектирования пассивных цепей [8, 9] на базе технологии "визуальных вычислений" [15];
- подход к проектированию СВЧ МИУ на основе преобразования моделей элементов [16];

- автоматический синтез пассивных и активных СВЧ цепей на базе ГА [10, 13].

В отличие от существующих программных продуктов разрабатываемая система позволяет, во-первых, генерировать принципиальную схему и топологию устройства по поставленным требованиям и, во-вторых, точно учитывать на этапе синтеза потери и паразитные параметры монолитных интегральных элементов. Преимущества разрабатываемой интеллектуальной системы автоматизированного проектирования СВЧ устройств по сравнению с известными программами состоят в следующем:

- благодаря использованию декомпозиционного подхода она позволит решить на единой основе задачи схемотехнического синтеза для широкого класса СВЧ ППУ;
- синтез разных типов СВЧ устройств выполняется при учете полного комплекса характеристик;
- при проектировании полностью контролируются структура, значения номиналов элементов и топология синтезируемых устройств, что позволяет получить практически реализуемые решения;
- сочетание автоматической генерации схем и интерактивного "визуального" проектирования обеспечивает гибкость и расширяет возможности системы для пользователей различного уровня.

В настоящее время в России и за рубежом отсутствуют программные продукты, обладающие подобными возможностями.

Используемые методы и алгоритмы

Опишем кратко основные методы и алгоритмы, используемые в системе INDESYS при решении задач проектирования СВЧ устройств.

Декомпозиционный метод синтеза (ДМС) активных СВЧ ППУ. ДМС [1, 2] — это систематический подход к проектированию линейных и нелинейных СВЧ ППУ, которые представляются в виде соединения полупроводниковых приборов и пассивных цепей (например, корректирующих, согласующих и стабилизирующих цепей, цепей обратной связи и т. д.). При этом КЦ и СЦ рассматриваются как управляемая часть цепи, выбором которой добиваются нужных характеристик устройства в целом.

Для упрощения задачи синтеза и получения практических решений проводится более детальная декомпозиция активной цепи, в результате которой определяется ее структура в виде блок-схемы (структурной схемы). В этой схеме конкретизируются типы блоков (например, активные элементы, двухполюсные и четырехполюсные, реактивные и диссипативные пассивные цепи и т. д.) и задаются связи между ними. Параметры

части блоков заданы (полупроводниковые приборы, цепи с известной структурой и элементами). Остальные блоки (пассивные КЦ и СЦ) являются "черными ящиками", структуру которых предстоит раскрыть в процессе синтеза.

После выбора структурной схемы проектирование активных устройств выполняется в два основных этапа:

- определение на ряде фиксированных частот рабочего диапазона областей допустимых значений (ОДЗ) входного иммитанса КЦ или СЦ исходя из совокупности требований к характеристикам устройства;
- проектирование (синтез) пассивных КЦ или СЦ по ОДЗ иммитанса.

Представление требований к КЦ и СЦ в виде ОДЗ иммитанса позволяет учесть произвольные ограничения на иммитанс цепей и дает ряд других важных преимуществ [1, 2].

Синтез СВЧ МИУ на основе преобразования моделей элементов. Метод преобразования моделей пассивных элементов [16] позволяет, используя существующие процедуры синтеза цепей на идеальных элементах, осуществить синтез устройств на монолитных интегральных (МИ) элементах. При этом могут быть применены любые типы моделей, разрешающие с необходимой точностью описать монолитные элементы, например, модели в виде эквивалентных схем, многомерных полиномов, нейросетей и др., модель должна лишь отвечать требованию достаточного быстрого действия.

Процедура структурного синтеза СВЧ МИУ состоит в следующем. Для выбранного или сгенерированного начального приближения — идеальной цепи-прототипа (цепь на идеальных элементах) — автоматически строится соответствующая МИ-цепь [16, 17]. В ней каждый идеальный пассивный элемент (сопротивление, индуктивность, емкость, идеальная линия передачи) заменяется моделью соответствующего МИ-элемента (например, в виде эквивалентной схемы). По известным электрическим параметрам идеальных элементов осуществляется автоматический расчет геометрических размеров МИ-элементов, эти размеры служат параметрами моделей МИ-элементов. Таким образом, параметры моделей элементов МИ-цепи оказываются связанными со значениями номиналов элементов идеальной цепи.

Кроме моделей МИ-элементов МИ-цепь включает также модели соединительных линий (микроразрывные линии, тройники), заземляющих отверстий в плате и т. д. Активные элементы могут быть представлены параметрами рассеяния, эквивалентными схемами либо другими моделями.

После преобразования выполняется моделирование МИ-цепи. Если ее характеристики не удовлетворяют предъявляемым требованиям, то далее генерируются новая структура идеальной цепи-прототипа и значения номиналов ее элементов, и все повторяется. Важно отметить, что целенаправленное изменение элементов и структуры осуществляется на основе формируемой целевой функции, учитывающей характеристики *реальной* МИ-цепи и требования к ней, это позволяет при синтезе использовать точные модели МИ-элементов.

Используемые в системе INDESYS алгоритмы моделирования СВЧ устройств, а также построения ОДЗ параметров КЦ и СЦ описаны в [3—6, 8, 14, 15]. После определения ОДЗ иммитанса КЦ и СЦ они отображаются в графическом виде на экране дисплея.

Интерактивные процедуры "визуального" проектирования пассивных и активных СВЧ устройств

Под "визуальным" проектированием [8, 15] понимается процесс, при котором пользователь с помощью графических средств интерфейса непосредственно управляет ходом проектирования, активно вмешиваясь в него (например, изменяя проектные параметры и т. д.), и одновременно наблюдает за результатами своих действий. При выполнении некоторой проектной процедуры цель и результат проектирования одновременно отображаются (визуализируются) на экране монитора. Разработчик с помощью интерфейсных средств воздействует на результат проектирования (точнее говоря, на его графический образ), пытаясь добиться соответствия цели.

Характерной чертой "визуального" проектирования является то, что пользователь отвлекается от реального содержания задачи проектирования и стремится достичь целей проектирования только средствами визуализации (подобно компьютерной игре). Указанный подход позволяет привлечь интеллектуальные способности человека по восприятию визуальной информации и принятию решений, это приводит к повышению эффективности проектирования.

"Визуальное" проектирование КЦ и СЦ на идеальных элементах. Интерактивная "визуальная" процедура проектирования КЦ и СЦ на идеальных элементах [4, 8, 9] включает два шага:

- выбор структуры КЦ (СЦ) исходя из расположения ОДЗ на плоскости импеданса. При этом пользователь на экране компьютера одновременно наблюдает сами ОДЗ, а также библиотеку типовых структур цепей и их годографов импеданса или проводимости. В результате пользователь выбирает несколько структур це-

пей, годограф которых может попасть в заданные области;

- расчет элементов КЦ (СЦ) с выбранной структурой. Пользователь может управлять формой годографа, перемещая с помощью "мыши" опорную точку годографа на одной из частот и изменяя с помощью тюнера значения номиналов некоторых (управляемых) элементов цепи. Таким образом, он добивается попадания точек годографа на других частотах в соответствующие ОДЗ. Если этого удастся добиться, то задача проектирования СЦ выполнена, в противном случае необходимо выбрать другую структуру цепи.

"Визуальное" проектирование КЦ и СЦ на МИ-элементах. Аналогичная процедура для визуального проектирования монолитных КЦ и СЦ [17] основана на методе преобразования моделей пассивных элементов [16] (см. выше). Пусть для выбранной КЦ на идеальных элементах построена соответствующая МИ-цепь. На экране компьютера вместе с ОДЗ отображаются одновременно два взаимосвязанных годографа — идеальной цепи и соответствующей ей МИ-цепи. Если изменять форму годографа идеальной цепи, варьируя ее элементы, то при этом будет изменяться и форма годографа МИ-цепи.

Первоначально пользователь по ОДЗ иммитанса выбирает структуру идеальной КЦ (СЦ), тем самым задается и структура ассоциированной МИ-цепи. Процесс визуального определения элементов МИ-цепи заключается в том, что пользователь, управляя формой годографа *идеальной* цепи, добивается попадания в ОДЗ точек годографа *монолитной* цепи. При этом он может быстро оценить, возможно ли с помощью МИ-цепи выбранной структуры удовлетворить указанному требованию, и при необходимости взять другую структуру цепи. Таким образом, влияние потерь и паразитных параметров в МИ-элементах учитывается при выборе как структуры, так и значения номиналов элементов монолитной КЦ (СЦ).

При интеграции программы "визуального" проектирования монолитных КЦ и СЦ со средой MWO можно не только использовать встроенные в MWO модели МИ-элементов и алгоритмы моделирования, но также отображать в режиме реального времени изменение топологии и характеристик монолитной КЦ (СЦ) при варьировании элементов идеальной цепи-прототипа. Это позволяет быстро сравнить характеристики монолитных КЦ и СЦ с разными структурами и (или) топологиями и выбрать лучший вариант.

"Визуальное" проектирование СВЧ транзисторных усилителей [4, 6, 7, 12]. Вначале по требованиям к характеристикам усилителя на заданных частотах определяются и отображаются на

дисплее ОДЗ иммитанса первой из имеющихся КЦ или СЦ. Далее с использованием описанных выше процедур осуществляется "визуальное" проектирование указанной цепи, построенной на базе идеальных или монолитных элементов. После этого аналогичным образом проектируется следующая КЦ (СЦ) и т. д.

Проектирование СВЧ устройств на основе генетических алгоритмов

Генетические алгоритмы (ГА) — это класс алгоритмов, имитирующих эволюционные процессы в природе на основе механизмов генетического наследования и естественного отбора [18]. Одной из особенностей ГА по сравнению с классическими (детерминированными) методами оптимизации является то, что они оперируют не с единственным решением, а с целой совокупностью (популяцией) решений. Второй особенностью является использование одновременно стратегий детерминированного и случайного поиска. Это делает ГА более эффективными при поиске глобально-оптимальных решений в задачах высокой размерности, а также менее подверженными сходимости к локальным оптимумам, чем традиционные оптимизационные методы. Кроме того, ГА обычно работают в пространстве кодов, отображающих решения (часто бинарных), поэтому они особенно пригодны для решения задач структурного синтеза технических устройств.

Для осуществления структурного синтеза СВЧ устройства на основе ГА [10, 13] информация о его структуре и параметрах элементов представляется в виде числового кода — хромосомы. После генерации начального множества числовых кодов (решений) в процессе работы ГА популяция хромосом на каждом шаге обновляется путем моделирования механизмов наследственной генетики — скрещивания хромосом и мутации. Моделирование механизма естественного отбора на основе вычисления целевой функции приводит к постепенному улучшению решений в процессе эволюции.

Представление структуры СВЧ устройств с помощью шаблонов. В задачах синтеза цепей с использованием ГА все более частое применение находят так называемые морфологические методы описания структуры синтезируемых устройств [19]. Суть этих методов заключается в определении пространства поиска, которое называется *морфологическим множеством* (множеством альтернатив). Такое множество должно содержать все структурные решения устройств рассматриваемого класса, как реально существующие, так и потенциально возможные. Морфологическое множество можно представить в виде многоуровневых морфологических таблиц или морфологи-

ческих "и/или" деревьев, на которых будет проводиться поиск рациональных технических решений [19, 20].

Рассмотрим морфологический метод описания структуры СВЧ устройства, реализованный в системе INDESYS [21]. Для представления радиоэлектронных цепей применяется иерархический (многоуровневый) шаблонный подход, реализуемый паттерном "компоновщик". Использование такого подхода позволяет строить сложные схемы из более простых компонентов. Проектировщик может сгруппировать (соединить) простые компоненты для формирования более сложных, которые, в свою очередь, могут стать основой для создания еще более сложных. При этом он работает как с простыми, так и со сложными (составными) компонентами единообразно.

При построении компонентов на любом уровне используются шаблоны. Для каждого составного компонента они определяют (ограничивают) возможные способы его образования из компонентов низшего уровня, а также типы и значения параметров этих компонентов. В результате многоуровневая система шаблонов определяет множество возможных структур всего устройства.

В качестве примера можно рассмотреть типичную схему транзисторного СВЧ усилителя с СЦ. Шаблон "усилитель" (рис. 1, а) состоит из трех шаблонов более низкой иерархии: входная СЦ, выходная СЦ и активный элемент (транзистор с подключенными КЦ). В свою очередь, шаблон "активный элемент" (рис. 1, б) состоит из семи шаблонов, один из которых является самым низкоуровневым шаблоном — базовым элементом (транзистор), а шесть других (КЦ) также могут включать в себя другие шаблоны (составляющие компоненты или базовые элементы КЦ).

Рассмотренный иерархический шаблонный подход представления структуры СВЧ устройств имеет ряд достоинств.

1. При составлении шаблонов возможно учесть накопленный в технике СВЧ многолетний опыт построения эффективных и практически реализуемых устройств: каскадный (блочный) принцип построения устройств, типовые структурные схемы устройств и блоков, согласование отдельных блоков со стандартным волновым трактом, типовые схемы включения транзисторов, сравнительно небольшое число проверенных способов построения СВЧ транзисторных каскадов (например, с реактивными СЦ на входе и выходе, определенными типами обратной связи), наиболее удобные в реализации и малочувствительные

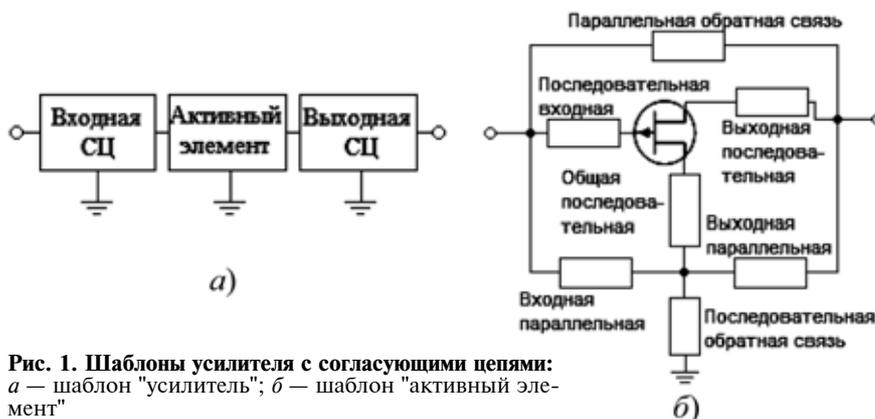


Рис. 1. Шаблоны усилителя с согласующими цепями: а — шаблон "усилитель"; б — шаблон "активный элемент"

структуры пассивных СЦ и КЦ (например, лестничного типа) и т. д.

2. Многоуровневая система шаблонов позволяет описать широкий класс структур СВЧ устройств; осуществить полный контроль над структурой устройства, вплоть до типов и значений отдельных элементов; гибко управлять множеством получаемых при синтезе структур; заранее отбросить большое число схем, заведомо неэффективных, с нереализуемыми на СВЧ топологиями или значениями номиналов элементов. Благодаря этому подход разрешает повысить шансы получения при синтезе удобных в исполнении и практически реализуемых устройств.

3. Такой подход позволяет также обеспечить высокую скорость моделирования СВЧ устройств. Это достигается за счет многоуровневой организации процесса моделирования, при которой вначале объединяются компоненты низшего уровня, затем компоненты более высокого уровня и т. д. При этом на каждом шаге для вычисления параметров составных компонентов используются сравнительно несложные аналитические формулы.

4. Применение паттерна "компоновщик" позволяет существенно упростить разработку алгоритмов анализа и синтеза, так как не требуется проводить различие между простыми и составными объектами.

В системе INDESYS морфологический метод описания структуры СВЧ устройства на основе шаблонов реализуется с использованием "и/или" дерева [19, 20].

Структура и реализация системы

Чтобы обеспечить возможность использования системы INDESYS в качестве основы для реализации широкого спектра задач в области проектирования (синтеза) и характеристики (измерения параметров) различных СВЧ устройств, было решено разделить ее на два уровня.

Первый (базовый) уровень Indesys Framework содержит общие механизмы, необходимые для реализации специфики данной предметной области:

комплексную арифметику, логическое и геометрическое представления радиоэлектронных цепей, алгоритмы моделирования и расчета характеристик цепей, алгоритмы оптимизации, задание ограничений, расчет целевых функций, управление измерительной аппаратурой через стандартные протоколы VISA и др. Кроме того, этот уровень реализует общий пользовательский интерфейс, построенный на библиотеке пользовательских элементов управления DXperience фирмы DevExpress, а именно: главное окно программы, дерево проекта, средства отображения данных (графические диаграммы — прямоугольная, полярная и Вольперта—Смита, таблицы), общие диалоговые формы — задания рабочих частот, требований к характеристикам, единиц измерения физических величин и т. д.

Доступ к этому уровню обеспечивается через набор внешних программных интерфейсов Indesys Framework API. Для обеспечения возможности автоматизации системы конечным пользователем в нее встроены скриптовый редактор Indesys ScriptEditor, позволяющий разрабатывать программы на языках C# и Visual Basic .NET. С помощью этого редактора пользователь может разработать программу, которая дает возможность задать последовательность определенных действий в системе через использование программных интерфейсов уровня Indesys Framework. Это является особенно удобным при работе с измерительной аппаратурой — например, можно разработать программный скрипт, который снимает данные с прибора, обрабатывает их и выводит результаты в виде графиков. В Indesys ScriptEditor пользователь даже может реализовать собственный алгоритм синтеза устройства заданного класса. Кроме того, Indesys Framework API доступен для внешних программ, что позволяет обеспечить интеграцию с ними, а также разработку программных расширений для системы INDESYS.

Второй уровень системы INDESYS является надстройкой над уровнем Indesys Framework и содержит методы синтеза различных типов СВЧ устройств. Методы реализованы в виде динамических подключаемых модулей (плагинов), которые могут содержать как конкретные алгоритмы методов синтеза, так и собственный пользовательский интерфейс. Кроме того, плагин может встраивать в систему свои панели инструментов, меню и MDI-окно, с помощью которых пользователь управляет процессом синтеза. Управление системой происходит через Indesys Framework API.

Благодаря тому, что в Indesys Framework содержится большое число общих компонентов, используемых при проектировании СВЧ устройств различных классов, на этой платформе в короткие сроки можно реализовать программы синтеза конкретных типов СВЧ устройств и осуществить их удобное взаимодействие.

В качестве основы для построения системы был использован шаблон проектирования программных систем (паттерн) Model—View—Controller ("Модель—Вид—Контроллер"), активно использовались и другие паттерны. Разработка велась с использованием языка описания моделей UML (Unified Model Language) в программе Enterprise Architect. Всего в ходе работы над системой было разработано более 70 UML-диаграмм классов, 12 UML-диаграмм вариантов использования программ, 20 UML-диаграмм деятельности. По этим диаграммам был сгенерирован каркасный код на языке C# платформы .NET Framework. Данный код в дальнейшем наполнялся функциональным кодом. После реализации каждого класса и модуля разрабатывался модульный тест с использованием инструмента NUnit. Такой подход к организации процесса разработки и тщательное проектирование структуры программы позволили разработать сложную программную систему, содержащую на данный момент около 1000 различных классов и интерфейсов (без учета используемых сторонних компонентов).

В настоящий момент на основе платформы Indesys Framework реализованы и реализуются ряд программ синтеза различных типов пассивных и активных СВЧ устройств, в частности:

- программа "визуального" проектирования КЦ и СЦ на идеальных элементах INDESYS-L;
- программа "визуального" проектирования монолитных КЦ и СЦ INDESYS-LM (интегрирована со средой проектирования СВЧ устройств MWO);
- программы автоматического синтеза на основе ГА фильтров и СЦ (INDESYS-G) и транзисторных усилителей (INDESYS-GA);
- программа управления измерительными СВЧ приборами INDESYS-MS.

Опыт использования программ подтвердил эффективность предложенных подходов и алгоритмов при решении практических задач проектирования СВЧ устройств. Реализация программ на базе платформы Indesys Framework позволила значительно уменьшить трудоемкость и сократить сроки их разработки.

Работа поддержана грантами международной организации INTAS (№ 06-100024-9199), Российского фонда фундаментальных исследований (РФФИ № 06-07-96916, № 08-07-99034 и № 09-07-99020), "Бизнес-старта с Microsoft", "У.М.Н.И.К." и "Ползуновские гранты".

Список литературы

1. **Бабак Л. И.** Структурный синтез СВЧ полупроводниковых устройств на основе декомпозиционного подхода // Труды ТПУ. 2006. Вып. 8. Томск. С. 160—165.
2. **Babak L. I.** Decomposition synthesis approach to design of RF and microwave active circuits // IEEE MTT-S Int. Microwave Symp. Digest. Phoenix. AZ. 2001. P. 1167—1170.

3. **Бабак Л. И.** Автоматизированный синтез двухполюсных цепей коррекции полупроводниковых устройств ВЧ и СВЧ // Известия вузов. Радиоэлектроника. Томск. 1993. Т. 36. № 10. С. 35—44; № 11. С. 3—11.

4. **Бабак Л. И.** Проектирование транзисторных широкополосных СВЧ усилителей с двухполюсными цепями коррекции и обратной связи // Электронная техника. Серия 1. СВЧ техника. Фрязино. 1994. № 2. С. 16—19. № 3. С. 9—16.

5. **Бабак Л. И.** Синтез согласующих цепей и цепей связи транзисторных широкополосных усилителей по областям иммитанса // Радиотехника и электроника. 1995. Т. 40. Вып. 10. № 8. С. 1550—1560.

6. **Бабак Л. И., Поляков А. Ю.** Автоматизированное проектирование малошумящих транзисторных СВЧ усилителей с реактивными согласующими цепями // Доклады ТУСУР. Томск: Изд-во ТУСУР. 1997. Т. 1. Вып. 1. С. 94—108.

7. **Babak L. I., Cherkashin M. V., Pokrovsky M. Yu.** Computer-aided design of ultrawide-band transistor amplifiers using decomposition synthesis method // Proc. 32th European Microwave Conf. Milan, Italy. 2002. P. 143—146.

8. **Бабак Л. И., Черкашин М. В., Зайцев Д. А.** "Визуальное" проектирование корректирующих и согласующих цепей полупроводниковых СВЧ устройств // Доклады ТУСУР. Томск. 2006. Вып. 6 (14). С. 11—23.

9. **Babak L. I., Cherkashin M. V.** Interactive "visual" design of matching and compensation networks for microwave active circuits // IEEE MTT-S Int. Microwave Symp. Digest. Phoenix. AZ. 2001. P. 2095—2098.

10. **Бабак Л. И., Вьюшков В. А.** Автоматизированный синтез согласующих цепей на основе генетического алгоритма // Сб. докладов. Межд. науч.-практ. конф. "Электронные средства и системы управления". Томск: Изд-во Института оптики и атмосферы СО РАН. 2005. Ч. 2. С. 102—105.

11. **Бабак Л. И., Черкашин М. В., Зайцев Д. А.** Программа "визуального" проектирования корректирующих и согласующих цепей СВЧ устройств // Матер. 15-й Междунар. Крымской конф. "СВЧ-техника и телекоммуникационные технологии" (КрыМиКо'2005): В 2 т. Севастополь: Вебер, 2005. Т. 2. С. 423—424.

12. **Бабак Л. И., Черкашин М. В., Поляков А. Ю., Бодунов К. С., Дягилев А. В.** Программы "визуального" проектирования тран-

зисторных СВЧ усилителей // Матер. 15-й Междунар. Крымской конф. "СВЧ-техника и телекоммуникационные технологии" (КрыМиКо'2005): В 2 т. Севастополь: Вебер, 2005. Т. 2. С. 425—426.

13. **Дорофеев С. Ю., Бабак Л. И.** Синтез согласующих цепей на сосредоточенных и распределенных элементах с использованием генетического алгоритма // Матер. 18-й Междунар. Крымской конф. "СВЧ-техника и телекоммуникационные технологии" (КрыМиКо'2008). Севастополь, 8—12 сентября 2008 г. Севастополь: Вебер, 2008. Т. 1. С. 133—134.

14. **Бабак Л. И., Касьмова Г. К., Поляков А. Ю., Черкашин М. В.** Решение задачи стабилизации систем управления на основе построения проекций области устойчивости // Вычислительные технологии. Новосибирск. 2003. Т. 8. Спец. выпуск. С. 103—113.

15. **Бабак Л. И.** "Визуальные" вычисления: решение систем нелинейных неравенств и многокритериальных проблем // Вестник Томского государственного педагогического университета. 2005. Вып. 7 (51). С. 21—29.

16. **Бабак Л. И.** Синтез СВЧ монолитных интегральных устройств на основе преобразования моделей пассивных элементов // Матер. 18-й Междунар. Крымской конф. "СВЧ-техника и телекоммуникационные технологии" (КрыМиКо'2008). Севастополь, 8—12 сентября 2008. Севастополь: Вебер, 2008. Т. 1. С. 129—130.

17. **Шеерман Ф. И., Бабак Л. И., Зайцев Д. А.** Интегрированная среда "визуального" проектирования корректирующих и согласующих цепей монолитных СВЧ устройств // Труды ТПУ. 2006. Вып. 8. С. 166—171.

18. **Панченко Т. В.** Генетические алгоритмы / Под ред. Ю. Ю. Тарасевича. Астрахань: Изд. Астраханского ун-та, 2007. 87 с.

19. **Акимов С. В.** Исследование и разработка методов структурно-параметрического синтеза линейных транзисторных усилителей СВЧ: Автореф. дис. канд. техн. наук / СПбГУТ, 2002.

20. **Норенков И. П., Маничев В. Б.** Основы теории и проектирования САПР. М.: Высшая школа, 1990. 336 с.

21. **Дорофеев С. Ю., Бабак Л. И.** Организация универсальной программной системы для синтеза радиоэлектронных устройств на основе генетического алгоритма // Доклады ТУСУР. 2007. № 2 (16). С. 151—156.

УДК 658.512.011.56

А. И. Разумовский, канд. техн. наук, ст. науч. сотр., e-mail: alr@ipu.rssi.ru,
Л. Н. Сизова, науч. сотр.,
 Институт проблем управления им. В. А. Трапезникова РАН

Проектирование и трассировка печатных плат с использованием программного комплекса "ГРАФИКА — ТР"

Описаны возможности интерактивного программного комплекса "Графика — ТР" как инструментального средства, предназначенного для проектирования электронных устройств средств связи, создания схемной документации в соответствии с ЕСКД, разработки чертежей, черчения схем алгоритмов, таблиц, поиска и визуализации кратчайшего пути между объектами.

Ключевые слова: интерактивный программный комплекс, проектирование электронных устройств, схемная документация, чертеж, принципиальная электрическая схема, монтажная схема, печатная плата, автоматическая трассировка соединений, контакты элементов, перечень элементов

Введение

Проектирование принципиальных электрических схем и разработка печатных плат требуют от конструктора знания и учета множества разнообразных требований, что в целом делает этот про-

цесс трудоемким и длительным и в настоящее время практически невозможным без применения средств автоматизации (САПР).

Для проектирования печатных плат на многих предприятиях радиоэлектронной промышленности используются ограниченное множество сис-

точных точек перегиба трасс таким образом, чтобы ни одна точка трассы не проходила через ранее занятую ячейку модели трассируемого поля.

8. Вывод результатов.

Свойства интерактивной среды проектирования программного комплекса "Графика — ТР"

Разработка печатных плат ведется в сквозном цикле, начиная с разработки библиотек элементов и принципиальных электрических схем и заканчивая созданием печатных плат с выпуском конструкторской документации в соответствии с ЕСКД и другой справочной информацией, необходимой для технологического производства изготовления плат.

Программный комплекс функционирует в двух интерактивных режимах:

- "графика-2D";
- "проектирование принципиальных электрических схем, монтажных схем и создание печатных плат".

В режиме "графика-2D" пользователю предоставляется возможность создавать векторные образы графических элементов схемы, редактировать их как на уровне отдельных примитивов (точка, ломаная, окружность, дуга окружности, эллипс, дуга эллипса, символ), так и составных узлов, создавать библиотеки элементов. Точность построения примитивов достигается масштабом визуализации и режимом привязки к узлам размерной сетки и к характерным точкам примитивов. Погрешность в системе составляет 0,025 мм.

Редакторские функции системы достаточно обширны: смещение, масштабирование, поворот, зеркальное отображение, копирование, копирование с зеркальным отображением, изменение типа, цвета и толщины линии, смещение конца отрезка, преобразование отрезка в ломаную линию, удаление и отмена текущей операции.

Режим "проектирование принципиальных электрических схем, монтажных схем и создание печатных плат" включает проведение ручного размещения элементов на схеме, автоматическую или полуавтоматическую трассировку соединений между элементами, простановку позиционных обозначений элементов на схеме в соответствии с требованиями ГОСТ 2.710—81, интерактивный ввод характеристик элементов и создание таблиц, содержащих их перечень.

Подготовка исходных данных для формирования геометриче-

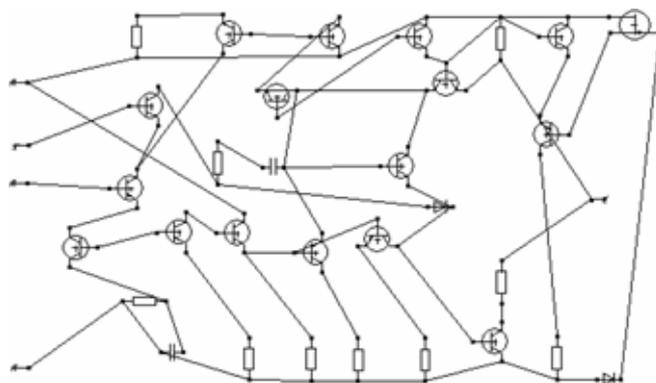


Рис. 1. Модель принципиальной схемы

ских моделей схем осуществляется в режиме "графика-2D". Исходные данные представляют собой описание типовых графических изображений элементов радиоэлектронной аппаратуры и создание библиотеки элементов. Библиотека элементов содержит два уровня описания: условные графические обозначения элементов на принципиальной схеме и графические изображения элементов на монтажной схеме. Система "Графика — ТР" позволяет сохранять экранные образы в векторном и растровом видах. В дальнейшем, просматривая растровые картинки внутри системы, можно быстро выбирать нужное условное графическое обозначение элемента, при этом обязательно должно существовать условное графическое обозначение в векторной форме, сохраненное ранее в виде блока в библиотеке. В программном комплексе также существует система внутреннего поиска необходимого компонента по его векторному изображению.

Рис. 1 иллюстрирует процесс создания принципиальной схемы. Из предварительно созданной библиотеки элементов выбраны графические об-

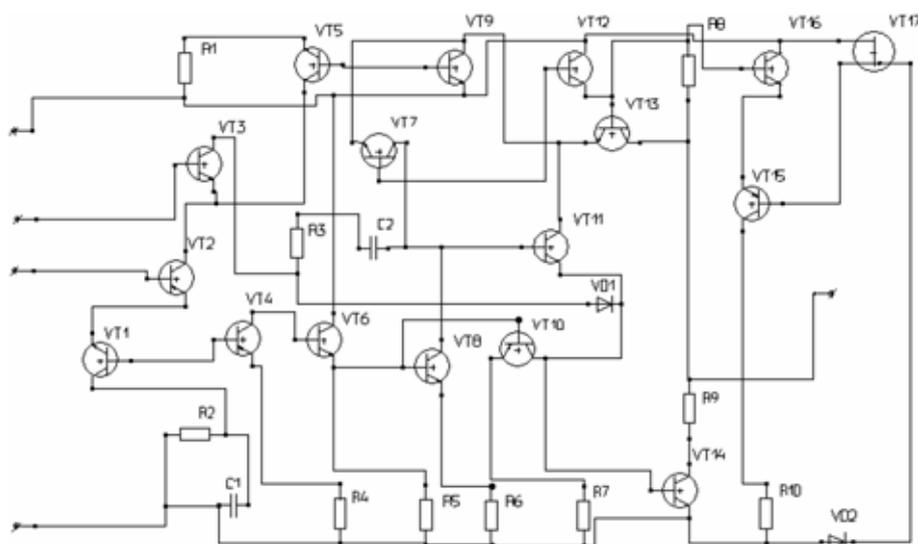


Рис. 2. Принципиальная схема

разы необходимых элементов и установлены связи между ними. Для проведения связей достаточно указать контакты элементов, входящих в текущую связь, с помощью нажатия клавиши мыши.

В результате последовательной обработки всех данных модулем автоматической трассировки соединений на экране отобразится принципиальная схема, показанная на рис. 2. На схеме можно видеть также результат автоматической простановки позиционных обозначений элементов. Эта операция активируется на любом этапе редактирования модели. Причем, в соответствии с действующими стандартами ЕСКД, предусмотрена возможность ручной доводки.

Порядок применения и результаты трассировки соединений

Автоматическая трассировка соединений реализует алгоритм определения кратчайшего пути на графе с пересечениями соединений и без пересечений. Задавая различные параметры трассировки, можно менять расстояние между трассами, число изгибов трасс, отдавать предпочтение вертикальным или горизонтальным трассам. В системе "Графика — ТР" имеется команда, по которой строится функция числа пересечений между связями в проекции на оси OX и OY чертежа схемы. Полученные результаты позволяют оценить качество размещения элементов схемы. Настройка параметров трассировки, режимы автоматической сортировки цепей и размещение элементов на схеме дают возможность получить максимальное число разведенных трасс.

На основе модели принципиальной электрической схемы автоматически реализуется модель печатной платы — графическое изображение элементов и связей между ними (рис. 3). Если же в наличии имеется модель печатной платы, то легко осуществить обратное преобразование в модель принципиальной электрической схемы. Индексация компонентов проставляется автоматически на основе позиционных обозначений элементов на принципиальной схеме. Разводка трасс на печатной плате предусматривает два способа: под углом 90° и 45° , причем если система не сможет развести трассы на одном слое, то автоматически создается второй слой. В этом случае переходы осуществляются через контактные площадки. Проведя необходимую компоновку элементов и применив команду трассировки к модели печатной платы, можно получить готовую разведенную плату. На рис. 4 и 5 показана однослойная разводка.

На каждой стадии проектирования в системе предусмотрено редактирование полученных результатов. При редактировании модели в любой момент можно добавить, удалить или изменить ме-

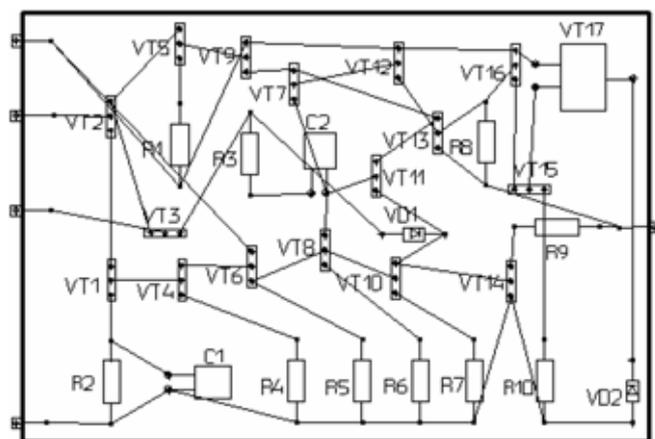


Рис. 3. Модель печатной платы

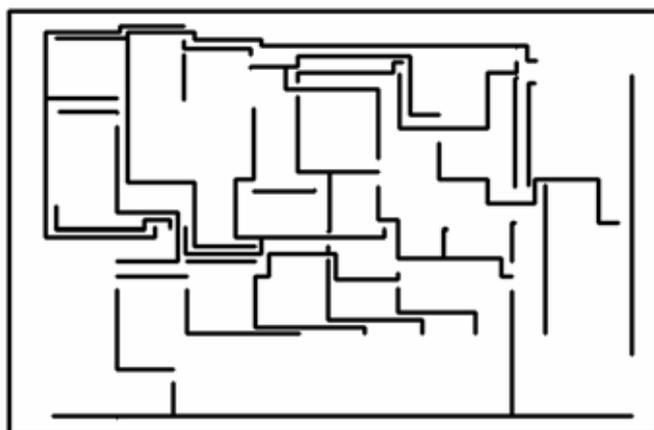


Рис. 4. Слой проводников печатной платы

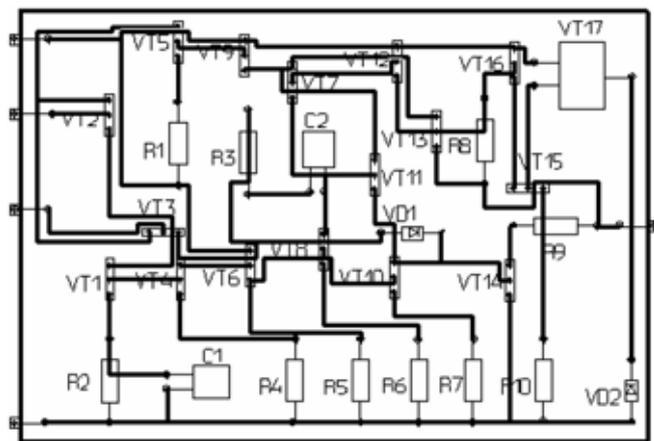


Рис. 5. Монтажная схема

сто положения элемента, при этом автоматически будут видоизменяться линии связей между элементами. Имеется возможность также для редактирования линий связей, добавления или удаления контакта элемента из соединения. При проектировании в модель схемы можно включать различные геометрические примитивы, непосредственно соз-

Рис. 6. Подготовка перечня элементов

Поз. обозн.	Наименование	Кол.	Примечание
	Конденсаторы		
C1, C2	KM - 4 - 47 МкФ ± 5 % ДЖД460043ТУ	2	
	Резисторы ГОСТ 7113-66		
R1, R5...R8	M/IT - 0,125 - 56 ± 5 %	5	
R2...R4	M/IT - 0,125 - 100 ± 5 %	3	
R9, R10	M/IT - 0,125 - 15 к ± 10 %	2	
	Диоды ГОСТ 14342-69		
V01, V02	KD522A	2	

Рис. 7. Фрагмент таблицы перечня элементов

давая их в момент работы с моделью или добавляя их из ранее сохраненного файла.

В программном комплексе имеется инструмент для создания и внесения изменений в перечень элементов (рис. 6).

После автоматической обработки данных списка элементов схемы по запросу пользователя на экран выводится перечень элементов в табличной форме (рис. 7). Число элементов с одинаковыми параметрами вычисляется автоматически.

Всю графическую информацию, отображенную на экране дисплея, в любой момент можно вывести на принтер, графопостроитель, координатограф или фотоплоттер. Взаимодействие с другими графическими системами осуществляется через файлы типа HPGL, DXF, GERBER.

Сферы применения программного комплекса "Графика — ТР"

Возможности системы "Графика — ТР" позволяют использовать ее также как инструмент для

создания чертежей в машиностроении, схемной документации радиоэлектронной аппаратуры, для черчения схем алгоритмов программ и таблиц, а также в картографии.

Специальной задачей картографии, решаемой на базе системы "Графика — ТР", является нахождение оптимального маршрута при централизованном обслуживании объектов социально-образовательной сферы: развоз завтраков, школьной мебели, инспектирование школ и т. д. Поскольку задача поиска оптимального маршрута подобна задаче трассировки соединений между объектами, то формулировка ее условий выглядит следующим образом: по входному описанию, содержащему n координат точек начала и конца трасс, определить дополнительные m координат промежуточных точек перегиба трасс таким образом, чтобы каждая точка встречалась только один раз.

Заключение

Программный комплекс "Графика — ТР" — это инструментальное средство по разработке печатных плат в сквозном цикле, начиная с разработки библиотек элементов и принципиальных электрических схем и заканчивая разработкой печатных плат и выпуском конструкторской документации в соответствии с ЕСКД. Кроме того, система может быть использована как инструмент для создания чертежей в машиностроении и картографии для определения кратчайшего пути между несколькими объектами, для черчения и графической обработки схем алгоритмов, таблиц. На основе комплекса организован специальный курс по исследованию общих принципов автоматизированного проектирования схемной документации, организации интерфейсов пользователей на основе средств операционной системы Windows, организации структур данных, а также управления вводом и выводом информации.

Список литературы

1. Уваров А. С. P-CAD. Проектирование и конструирование электронных устройств. М.: Горячая линия — Телеком, 2004. 760 с.
2. Артамонов Е. И., Ромакин В. А., Сизова Л. Н., Тенякшев А. М. Автоматизированное проектирование и выпуск схемной документации на аппаратуру средств связи: Методическое пособие. М.: МТУСИ, 2006. 26 с.
3. Сизова Л. Н. Автоматизация проектно-конструкторских работ в системе "Графика — ТР" // Проблемы геометрического моделирования в автоматизированном проектировании и производстве. Сб. матер. 1-й Междунар. научн. конф. / Под ред. В. И. Якунина. М.: МГИУ, 2008.
4. Артамонов Е. И., Хачумов В. М. Синтез структур специализированных средств машинной графики. М.: Институт проблем управления. 1991.

УДК 621.391.96

Л. Е. Назаров, д-р физ.-мат. наук, вед. науч. сотр.,
Институт радиотехники и электроники
им. В. А. Котельникова РАН, г. Фрязино,
e-mail: nazarov@ire.rssi.ru

Алгоритмы посимвольного приема сигналов

Приведены описания алгоритмов посимвольного приема ансамблей сигналов, соответствующих низкоскоростным линейным блоковым кодам. Основу разработанных алгоритмов составляет алгоритм быстрого спектрального преобразования в базисе Уолша, а также его модификация.

Ключевые слова: сигналы, посимвольный прием, алгоритм быстрого преобразования Уолша - Адамара

Введение

В работе [1] приведены описания методов приема сигналов на основе блоковых кодов, реализующих правило максимального правдоподобия. Применение данного правила обеспечивает минимум вероятности ошибки на дискретное сообщение. Основу разработанных методов составляет алгоритм быстрого спектрального преобразования в базисе Уолша—Адамара (БПУ) с размерностью, совпадающей с размерностью порождающих матриц соответствующих блоковых кодов [2]. Алгоритм БПУ имеет параллельную структуру, что позволяет при реализации разработанных алгоритмов оптимального приема применять программно-аппаратные средства параллельных вычислений. Это дает возможность существенно упростить реализацию алгоритмов приема и повысить их производительность. Аппарат быстрых спектральных преобразований в дискретных базисах составляет основу информационных технологий цифровой обработки сигналов в телекоммуникационных системах связи, в радио- и гидролокационных системах, при обработке изображений [2].

В работе [3, 4] приведены описания методов приема сигналов на основе блоковых кодов, реализующих правило посимвольного приема. Применение данного правила обеспечивает минимум вероятности ошибки на символ дискретных сооб-

щений. Основу этих методов составляет также алгоритм БПУ, размерность базиса Уолша определяется размерностью проверочных матриц кодов. Для класса высокоскоростных кодов, где размерность порождающих матриц существенно больше размерности проверочных матриц, разработанные методы посимвольного приема составляют альтернативу методам оптимального приема, реализующим правило максимального правдоподобия относительно сложности их реализации.

Методы посимвольного приема составляют основу процедур приема сигнально-кодовых конструкций под общим названием "турбокоды", рассматриваемых как наиболее перспективных для использования в телекоммуникационных системах широкого назначения. Это определяет актуальность проблемы разработки эффективных методов посимвольного приема сигналов.

В данной работе приведены описания разработанных методов посимвольного приема сигналов, формируемых на основе класса низкоскоростных блоковых кодов.

Общая постановка задачи

Пусть $A = (a_i; 0 \leq i < k)$ — последовательность информационных символов кодового слова $B = (b_l; 0 \leq l < n)$ двоичного блокового кода (n, k) с порождающей матрицей $G = (g_{ij}; 0 \leq i < k, 0 \leq l < n)$, где n — длительность кодовых слов; k — размерность кода [5]. Обозначим $Y = (y_l; 0 \leq l < n)$ дискретную реализацию на входе приемного устройства, соответствующую сигналу с двоичной фазовой модуляцией на основе кодового слова

B_m . Здесь $y_l = s_{ml}d + n_l$, $s_{ml} = (-1)^{b_{ml}}$, $d = \sqrt{2 \frac{kE_6}{nN_0}}$;

E_6 — энергия сигналов на информационный бит; N_0 — односторонняя плотность помехи в канале в виде аддитивного белого гауссовского шума; n_l — помеховые составляющие. Функцию правдоподобия $p\{Y|B\}$ полагаем известной. Рассматривается канал без памяти, последовательности A равновероятны.

Правило посимвольного приема заключается в вычислении функции отношения правдоподобия $L(b_l|Y)$ для апостериорных вероятностей [6]:

$$L(b_l|Y) = \ln \left(\frac{\sum_{B: b_l=0} p(Y|B)}{\sum_{B: b_l=1} p(Y|B)} \right). \quad (1)$$

Принимается решение $b_l = 0$, если $L(b_l = 0|Y) > 0$, в противном случае $b_l = 1$.

Поиск производительного алгоритма, реализующего приведенное правило посимвольного приема (1), составляет суть рассматриваемой проблемы.

Оптимальный посимвольный прием сигналов

Основу метода оптимального посимвольного приема сигналов при вычислении соотношения (1) при наличии в канале аддитивного белого гауссовского шума составляет спектральное преобразование в базисе Уолша—Адамара. В этом случае относительно $L(b_l|Y)$ верно соотношение

$$L(b_l|Y) = \ln \left(\frac{\sum_{B_m: b_{ml}=0} \exp \left(\sum_{i=0}^{n-1} dy_i s_{mi} \right)}{\sum_{B_m: b_{ml}=1} \exp \left(\sum_{i=0}^{n-1} dy_i s_{mi} \right)} \right), \quad l = 0, 1, \dots, n-1. \quad (2)$$

В работах [3, 4] показано, что при вычислении внутренних сумм выражения (2) можно применить алгоритм БПУ размерностью 2^k над анализируемой функцией $V(v)$, $v = 0, 1, \dots, 2^k - 1$:

$$V(v) = \begin{cases} y_b, & v = \sum_{i=0}^{k-1} g_{li} \cdot 2^i; \\ 0, & v \neq \sum_{i=0}^{k-1} g_{li} \cdot 2^i. \end{cases} \quad (3)$$

Алгоритм БПУ основан на представлении матрицы Уолша—Адамара размерностью 2^k в виде произведения k матриц размерностью 2^k , имеющих лишь 2^{k+1} ненулевых компонент [2]. Ориентированный граф БПУ имеет параллельную структуру, содержит k слоев, каждый слой содержит 2^{k-1} основных элементов — "бабочек" — с базовыми операциями "сложение—вычитание". Параллельная структура графа БПУ упрощает его исполнение средствами цифровой вычислительной техники.

Составляющие внешние суммы в выражении (2)

$z_{l\xi} = \sum_{B_m: b_{ml}=0} \exp \left(\sum_{i=0}^{n-1} dy_i s_{mi} \right)$ ($\xi = 0, 1$) могут быть записаны в виде

$$z_{l\xi} = \frac{1}{2} \sum_{B_m} \exp \left(\sum_{i=0}^{n-1} dy_i s_{mi} \right) + \frac{(-1)^\xi}{2} \sum_{B_m} (-1)^{b_{ml}} \exp \left(\sum_{i=0}^{n-1} dy_i s_{mi} \right). \quad (4)$$

Второе слагаемое в выражении (4) определяет вычисление спектральных компонент в базисе Уолша—Адамара над анализируемой функцией $U(v)$, $v = 0, 1, \dots, 2^k - 1$:

$$U(v) = \begin{cases} \exp \left(\sum_{i=0}^{n-1} dy_i s_{mi} \right), & v = \sum_{i=0}^{k-1} g_{li} \cdot 2^i, \\ 0, & v \neq \sum_{i=0}^{k-1} g_{li} \cdot 2^i. \end{cases} \quad (5)$$

Первое слагаемое в выражении (4) соответствует 0-й спектральной компоненте. Для низкоскоростных кодов, параметры которых удовлетворяют условию $n \approx 2^k$, например, для кодов максимальной длительности (2^k , k), соответствующих ансамблям ортогональных сигналов, требуется вычисление полного спектрального множества, в этом случае необходимо применять алгоритм БПУ размерностью 2^k .

При реализации метода оптимального посимвольного приема (2), (3), (4) требуется оценка энергетического параметра d , а также вычисление нелинейной функции экспоненциального типа. Это усложняет техническое исполнение соответствующей вычислительной процедуры. Ниже приведено описание метода подоптимального посимвольного приема, при его реализации не требуется оценки энергетического параметра и требуются лишь линейные операции "сложение—вычитание—сравнение—пересылки".

Подоптимальный посимвольный прием

Суть метода подоптимального посимвольного приема заключается в применении приближения к исходному соотношению (2) в виде

$$L(b_l|Y) \cong \max_{B_m: b_{ml}=0} \left(\sum_{i=0}^{n-1} y_i s_{mi} \right) - \max_{B_m: b_{ml}=1} \left(\sum_{i=0}^{n-1} y_i s_{mi} \right). \quad (6)$$

Подобное приближение используется в алгоритмах итеративного посимвольного приема турбокодов. Исследования показывают, что при применении данного приближения энергетические

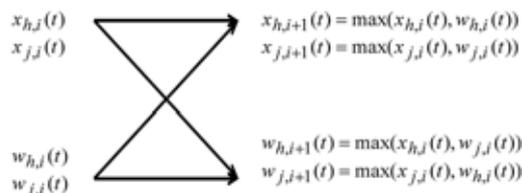


Рис. 1. Схематическое изображение элемента "бабочка" модифицированного алгоритма быстрого преобразования Уолша—Адамара с базовыми операциями "сравнение—пересылки"

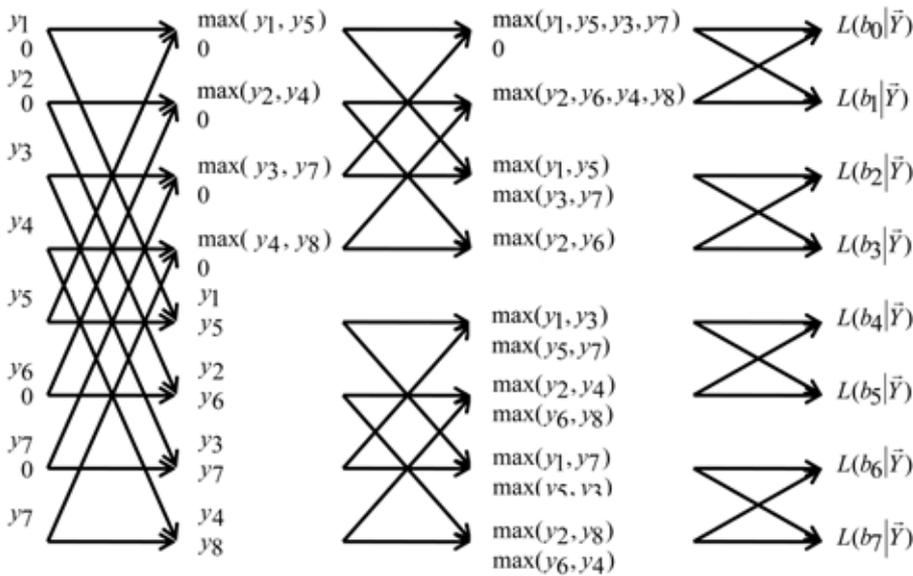


Рис. 2. Результирующий граф модифицированного алгоритма БПУ размерностью 8 с базовыми операциями "сравнение—пересылки"

$L(b_j|Y)$, задаваемые соотношением (6).

На рис. 2 в качестве примера приведен результирующий граф модифицированного алгоритма БПУ размерностью 8. Данный граф состоит из трех слоев, слои содержат по четыре "бабочки" с базовыми операциями "сравнение—пересылки".

Заключение

Приведены описания методов оптимального и подоптимального посимвольного приема сигналов, соответствующих двоичным блоковым кодам. Основу данных методов составляет производительный алгоритм быстрого

потери по отношению к процедуре итеративного приема с использованием оптимального посимвольного приема не превышают 0,3 дБ.

При вычислении соотношения (6) можно применить модифицированный алгоритм БПУ размерностью 2^k с базовыми операциями "сравнение—пересылки". На рис. 1 приведена схема основного элемента ("бабочки") i -го слоя модифицированного алгоритма БПУ. В этом случае выходные парные отсчеты $x_{h, i+1}(t)$, $x_{j, i+1}(t)$ и $w_{h, i+1}(t)$, $w_{j, i+1}(t)$, являющиеся входными для $(i+1)$ -го слоя, формируются по правилу $x_{h, i+1}(t) = \max(x_{h, i}(t), w_{h, i}(t))$; $x_{j, i+1}(t) = \max(x_{j, i}(t), w_{j, i}(t))$; $w_{h, i+1}(t) = \max(x_{h, i}(t), w_{j, i}(t))$; $w_{j, i+1}(t) = \max(x_{j, i}(t), w_{h, i}(t))$. Здесь $x_{h, i}(t)$, $x_{j, i}(t)$ и $w_{h, i}(t)$, $w_{j, i}(t)$ — парные отсчеты на входе i -го слоя, $t = 0, 1, \dots, 2^{k-1} - 1$. На первом слое ($i = 0$) парные отсчеты задаются в следующем виде: $x_{j, 0}(t) = w_{j, 0}(t) = 0$; $x_{h, 0}(t) =$

$$= \begin{cases} y_p, t = \sum_{i=0}^{k-1} g_{i,t} \cdot 2^i \\ 0, t \neq \sum_{i=0}^{k-1} g_{i,t} \cdot 2^i \end{cases}, w_{h, 0}(t) = \begin{cases} y_r, t = \sum_{i=0}^{k-1} g_{i, 2t+1} \cdot 2^i \\ 0, t \neq \sum_{i=0}^{k-1} g_{i, 2t+1} \cdot 2^i \end{cases}$$

На последнем слое определяются значения

спектрального преобразования в базисе Уолша—Адамара, а также его модификация с базовыми операциями "сравнение—пересылки".

При реализации подоптимального посимвольного приема не требуется оценки энергетического параметра, требуются лишь линейные операции типа "сложение—вычитание—сравнение—пересылки". Это существенно упрощает его исполнение по отношению к методу оптимального посимвольного приема.

Список литературы

1. Назаров Л. Е. Параллельные алгоритмы приема сигналов, соответствующих двоичным блоковым кодам // Информационные технологии. 2004. № 6. С. 31—36.
2. Трахтман А. М., Трахтман В. А. Основы теории дискретных сигналов на конечных интервалах. М.: Мир, 1976.
3. Смольянинов В. М. Дискретные мультипликативные групповые сигналы и их связь с групповыми кодами // Радиотехника и электроника. 1985. № 12. С. 2391—2394.
4. Назаров Л. Е. Параллельные алгоритмы посимвольного приема ансамблей сигналов, соответствующих двоичным кодам // Информационные технологии. 2005. № 1. С. 41—46.
5. Зюко А. Г., Фалько А. И., Панфилов И. П., Банкет В. Л., Иващенко П. В. Помехоустойчивость и эффективность систем передачи информации. М.: Радио и связь, 1985.
6. Hartmann C. R., Rudolph L. D. An optimum symbol-by-symbol decoding rule for linear codes // IEEE Transactions on Information Theory. 1978. V. 22. N 1. P. 514—517.

С. В. Дворников, д-р техн. наук, доц.,
Военная академия связи, г. Санкт-Петербург,
e-mail: practicsdv@yandex.ru,
С. С. Дворников, студент,
Санкт-Петербургский государственный
политехнический университет

Обнаружение сигналов с высоким различием динамики их амплитуд

Предлагается метод обнаружения радиосигналов с высоким различием динамики их амплитуд на основе многоэтапного алгоритма. Рассматривается один из подходов оптимизации процедур принятия решения по критерию максимального правдоподобия с автоматическим выбором порога без предварительного обучения системы. Приводятся результаты теоретических исследований и практических экспериментов.

Ключевые слова: обнаружение радиосигналов, многоэтапный алгоритм, различная динамика амплитуд, автоматический выбор порога без предварительного обучения системы

Вопросы автоматизации процедур, связанных с расчетом порога обнаружения в широкодиапазонных приемных системах, в настоящее время являются актуальными для радиомониторинга (РМ). Значимость их решения возрастает при обработке реализаций, содержащих смесь слабых сигналов и мощных излучений. Поскольку высокая динамика изменения сигнально-помеховой обстановки значительно затрудняет эффективное применение классических методов, базирующихся на статистической обработке, то необходим поиск новых подходов, позволяющих получить требуемый результат, в условиях ограничений на получение статистики.

В общем случае задача обнаружения состоит в определении алгоритма проверки гипотез о наличии сигнала $H_1: x(t) = s(t) \oplus \xi(t)$ в обрабатываемой реализации $x(t)$ или его отсутствии $H_0: x(t) = \xi(t)$. Здесь \oplus — произвольное взаимодействие шумов $\xi(t)$ с сигналом $s(t)$. Для широкодиапазонных систем $s(t)$ — совокупный сигнал.

Однако синтез оптимальных алгоритмов обнаружения даже для критерия максимального правдоподобия (МП) возможен лишь при априорных данных об условной плотности вероятности для $\omega_n(x|H_0)$ или $\omega_n(x|H_1)$ (здесь $x = \{x_1, \dots, x_n\}$ —

входная выборка длиной n) [1]. Поэтому решающая функция на выборочном пространстве X

$$\Phi(x) = \begin{cases} 1, & x \in X_1, \\ 0, & x \in X_0, \end{cases}$$

строится относительно G_p порога, разбивающего его на X_1 — область принятия гипотезы H_1 , и X_0 — область принятия гипотезы H_0 , как правило, на основе субъективных данных.

Учитывая ограничения, накладываемые на статистическую обработку, предлагается вместо условных плотностей вероятности $\omega_n(x|H_0)$ и $\omega_n(x|H_1)$ строить условные распределения амплитудных значений или условные распределения спектральных компонент принимаемых реализаций $w_n(y|H_1)$ и $w_n(y|H_0)$, где $y = (y_1, \dots, y_n)$ — входная выборка спектральных компонент длиной n . В общем случае число спектральных и временных отчетов обрабатываемой реализации может не совпадать.

Важным моментом остается вопрос выбора порогового значения для принятия решения о наличии или отсутствии полезного сигнала в обрабатываемой выборке. В работе [2] предложено решение об обнаружении сигналов принимать непосредственно по порогу G , рассчитываемому на основе статистических параметров реализации $x(t)$. Но такой подход предполагает значительные объемы выборок для получения требуемых статистик при том, что он не гарантирует обнаружение слабых сигналов на фоне мощных излучений, с точки зрения минимизации ложной тревоги α , при максимизации вероятности обнаружения $P = 1 - \beta$.

Одним из вариантов решения данной задачи в рассматриваемой ситуации является применение так называемых многоэтапных алгоритмов принятия решения, в которых асимптотически уточняется пороговое значение принятия гипотезы H_1 или H_0 .

В частности, в работе [3] предложен двухэтапный алгоритм, позволяющий за счет предварительного выделения из спектральной совокупности компонент обрабатываемой реализации сигнальных составляющих в ходе первого этапа уже на втором этапе принимать окончательное решение об обнаружении. В качестве порогового значения на каждом из этапов для указанного подхода обосновано и принято следующее:

$$G = 3m_1\{y(f)\}. \quad (1)$$

Здесь $m_1\{*\}$ — функция условных средних; $y(f)$ — спектральное преобразование обрабатываемой реализации $x(t)$.

В свою очередь пороговое значение G , определяемое формулой (1), выбрано на основании результатов эксперимента, представленных в работе [4].

Проведенный анализ подтверждает правомерность такого выбора, но только при условии, что динамика амплитуд обнаруживаемых сигналов не превышает 10...15 дБ [3].

Однако нередко в полосе анализа комплексов РМ оказываются источники, мощность радиоизлучения которых на 20 дБ и более превышает мощность контролируемых сигналов. В указанных ситуациях эффективность обнаружителей, реализующих двухэтапные алгоритмы принятия решения, резко падает [5].

В частности, на рис. 1 (см. вторую сторону обложки) представлен групповой спектр в полосе 50 кГц, содержащий два сигнала, разница динамики амплитуд которых составляет 28 дБ. Для рассматриваемых условий выбор порога, в соответствии с формулой (1), не позволит обнаружить контролируемый сигнал, поскольку его спектральные компоненты будут находиться ниже расчетного значения G .

С учетом рассмотренных обстоятельств предлагается следующий подход к решению задачи обнаружения сигналов с низкой энергетикой на фоне мощных радиоизлучений. В частности, необходимо исключить или хотя бы минимизировать влияние мощного радиоизлучения на рассчитываемое пороговое значение G . С этой целью рекомендуется определить уровень для спектральных компонент, и тогда для расчета G использовать только те компоненты, которые бы не превысили установленный уровень.

В результате искомый метод обнаружения сигналов с низкой энергетикой на фоне мощных радиоизлучений можно представить в виде следующих основных этапов.

1. По частотному представлению входной реализации $y(f)$ рассчитывается первый порог разбиения G_{p1} , который устанавливает уровень для спектральных компонент, используемых в последующих расчетах.

2. Для массива спектральных компонент, не превысивших G_{p1} , рассчитывается второй порог разбиения G_{p2} .

3. Формируются два независимых массива $^1y(f)$ и $^2y(f)$: первый $^1y(f)$ — для спектральных компонент, превысивших порог G_{p2} , второй $^2y(f)$ — для не превысивших его.

4. Для каждого из сформированных массивов на этом этапе строится распределение и в соответствии с выбранным критерием находится искомое значение порога принятия решения G об обнаружении сигнала в обрабатываемой реализации.

Если в качестве решающего правила на каждом из этапов определить критерий МП, то искомое

пороговое значение будет являться точкой пересечения массивов условных плотностей распределений спектральных компонент $w_n(y|H_1)$ и $w_n(y|H_0)$, соответственно превысивших пороговое значение G_p и не превысивших его.

В этом случае окончательное решение следует принимать по результатам статистической обработки сформированных массивов [4]:

$$G = \frac{m_1\{y_1, \dots, y_n|H_1\} + m_2\{y_1, \dots, y_n|H_0\}}{2}. \quad (2)$$

С этим обоснованием пороговых значений был проведен эксперимент, где в качестве обрабатываемой реализации рассматривалась аддитивная совокупность двух сигналов $s_0(t) = s_1(t) + s_2(t)$, в которой $s_1(t)$ представлял сигнал с низкой энергетикой, а $s_2(t)$ имитировал мощное радиоизлучение. В качестве расчетных значений рассматривались различные варианты комбинаций параметров условных средних $m_1\{*\}$ и среднего квадратичного отклонения $\sigma\{*\}$ спектральных параметров $y_0(f)$ обрабатываемой реализации $s_0(t)$.

Исследуемые варианты комбинаций были условно разделены на три группы. Отнесение к той или иной группе осуществлялось в зависимости от близости расчетных показателей.

В группу $G1_p$ вошли пороги, рассчитанные следующим образом: $m_1\{*\} + \sigma\{*\}$; $m_1\{*\} + 1,5\sigma\{*\}$; $2m_1\{*\} + \sigma\{*\}$; $4m_1\{*\} - \sigma\{*\}$. К группе $G2_p$ были отнесены пороги $2m_1\{*\} - \sigma\{*\}$; $3m_1\{*\} - \sigma\{*\}$. Группу $G3_p$ составили пороги, имеющие значения: $3m_1\{*\} + \sigma\{*\}$; $m_1\{*\} + 2\sigma\{*\}$; $3m_1\{*\}$; $2m_1\{*\} + 2\sigma\{*\}$; $3m_1\{*\} + 2\sigma\{*\}$; $4m_1\{*\} + \sigma\{*\}$; $2m_1\{*\} + 2\sigma\{*\}$.

Целью эксперимента являлось установление факта обнаружения сигнала с низкой энергетикой на фоне сильного радиоизлучения в зависимости от отношения их мощностей при использовании многоэтапного алгоритма принятия решения.

Данные эксперимента сведены в таблицу, где они сравнивались с результатами обнаружения

Результаты обнаружения сигнала с низкой энергетикой на фоне мощного радиоизлучения для различных пороговых значений

Соотношение амплитуд сигналов, дБ	Пороговые значения							
	G_{opt}		Группа $G1_p$		Группа $G2_p$		Группа $G3_p$	
40	-	-	-	+	-	+	-	-
39	-	+	-	+	-	+	-	-
38	-	+	-	+	-	+	-	+
37	-	+	-	+	-	+	-	+
36	-	+	+	+	+	+	-	+
35	-	+	+	+	+	+	-	+
34	-	+	###	+	+	+	-	+
30	+	+	###	+	###	+	-	+
26	+	+	###	###	###	+	+	+
20	+	+	###	###	###	###	+	+
14	+	+	###	###	###	###	+	###
6	+	###	###	###	###	###	+	###

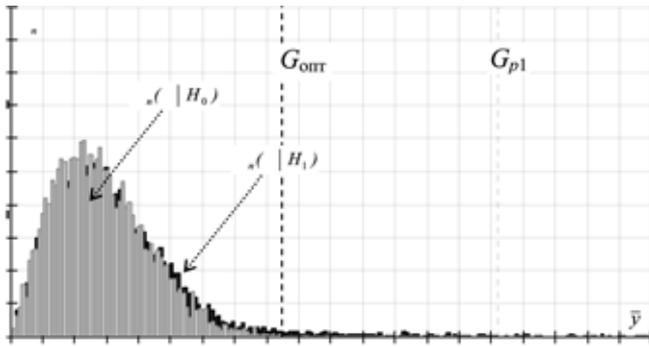


Рис. 2. Распределение спектральных компонент обрабатываемой реализации

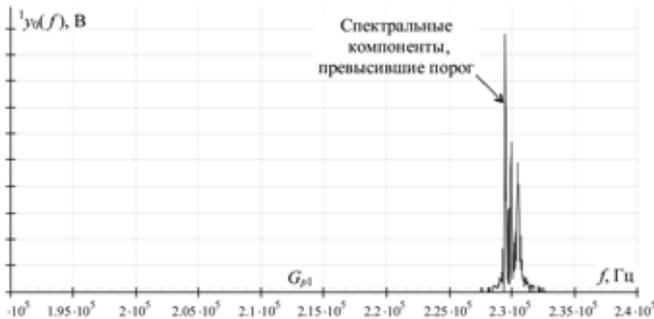


Рис. 3. Компоненты группового спектра двух сигналов с различной динамикой амплитуд, превысившие порог G_{p1}

относительно оптимального порога $G_{\text{опт}}$, полученного на основе полных априорных данных о параметрах шума и сигнального ансамбля. Каждая ячейка таблицы имеет два поля. В левом поле указаны результаты обнаружения при выполнении первого этапа, а в правом поле — второго этапа. Причем "+" указывает на положительное решение, "-" на отрицательное, а "###" — на переход порога в область шумов, обусловленную большой вероятностью ложной тревоги.

Анализ полученных данных таблицы показывает, что наиболее близкие значения к $G_{\text{опт}}$ по конечной результативности имеют пороги, принадлежащие третьей группе.

Дальнейшие исследования в данном направлении показали [4, 5] приоритетность выбора $G = 3m_1\{*\}$.

В целях получения аналитического значения для порога принятия решения по критерию МП были изучены виды распределений $y_0(f)$ при различных соотношениях мощностей $s_1(t)$ и $s_2(t)$. Выяснилось, что они асимптотически, с возрастанием числа учитываемых спектральных компонент, приближаются к релейским (рис. 2).

Так, на рис. 2 показано распределение спектральных компонент $y_0(f)$ при отношении мощностей $s_1(t)$ и $s_2(t)$ в 28 дБ (значения по оси абсцисс нормированы) и при условии $H_0: x(t) = \xi(t)$.

На рис. 2 также нанесен порог G_{p1} , определяемый формулой (1), и порог $G_{\text{опт}}$, обеспечивающий обнаружение сигнала с низкой энергетикой в условиях полной априорной информации о параметрах шума и каждого из сигналов.

Предлагаемый подход ориентирован на работу со спектральными компонентами, ограниченными порогом G_{p1} , поэтому вся совокупность $y_0(f)$ разбивается на две выборки ${}^1y_0(f)$ — превысивших G_{p1} (рис. 3) и ${}^2y_0(f)$ — не превысивших его (рис. 4, см. вторую сторону обложки).

Затем значения ${}^1y_0(f)$ исключаются из дальнейших расчетов. Этим обеспечивается снижение влияния мощного радиоизлучения на энергетически слабый сигнал, а компоненты ${}^2y_0(f)$ снова разбиваем порогом $G_{p2} = 3m_1\{y_{02}(f)\}$ на две выборки $z(f)$ и $u(f)$, для которых строим условные распределения $w_n({}^2y_{01}, \dots, {}^2y_{0n}|H_1)$ и $w_n({}^2y_{01}, \dots, {}^2y_{0n}|H_0)$, считая значения выборки $z(f)$, превысившие порог разбиения G_{p2} , сигнальными компонентами (рис. 5, см. вторую сторону обложки). Здесь графики построены на основе статистической обработки результатов экспериментов.

Результаты многочисленных экспериментов показали близость значений порогов $G_{\text{опт}}$ и G_{p2} , что позволяет сделать вывод о рациональности выбранного подхода к решению задачи обнаружения сигнала с низкой энергетикой на фоне мощного радиоизлучения.

Следует отметить, что для полученных распределений $w(z(f))$ и $w(u(f))$ оптимальным порогом принятия решения по критерию МП будет значение, соответствующее их номинальному равенству:

$$w(z(f)) = w(u(f)). \quad (3)$$

Для вычисления значения G представим распределения, определяющие условие (3), следующими аналитическими выражениями:

$$w_\xi(u(f)) = \frac{u(f)}{\sigma_\xi^2(u(f))} \exp\left(-\frac{u^2(f)}{2\sigma_\xi^2(u(f))}\right);$$

$$w_s(z(f)) = \frac{1}{\sigma_s(z(f))\sqrt{2\pi}} \exp\left(-\frac{(z(f) - m_s(z(f)))^2}{2\sigma_s^2(z(f))}\right).$$

Здесь $\sigma_\xi^2(u(f))$ и $\sigma_s^2(z(f))$ — дисперсии распределения соответственно для спектральных компонент шума и сигнальных компонент, превысивших порог G_{p2} ; $m_\xi(u(f))$ и $m_s(z(f))$ — функции условных средних распределения спектральных компонент шума и сигнальных компонент, превысивших порог G_{p2} . Условия выбора закона распределения для $w(z(f))$ и $w(u(f))$ рассмотрены в работах [3, 4]. Тогда для равенства (3) запишем:

$$\frac{1}{\sigma_s(z(f))\sqrt{2\pi}} \exp\left(-\frac{(z(f) - m_s(z(f)))^2}{2\sigma_s^2(z(f))}\right) = \frac{u(f)}{\sigma_\xi^2(u(f))} \exp\left(-\frac{u^2(f)}{2\sigma_\xi^2(u(f))}\right). \quad (4)$$

Решение уравнения (4) относительно аргумента и будет являться искомым аналитическим значением порога принятия решения.

Заметим, что выбор G как искомой величины позволит повысить достоверность обнаружения по сравнению с любым из двух порогов разбиения G_{p1} и G_{p2} , но вместе с тем возрастет и значение ложной тревоги α .

В целях проверки эффективности разработанного подхода были проведены дополнительные исследования по обнаружению в полосе 50 кГц реализации $s_0(t)$, состоящей из двух реальных сигналов частотной манипуляции (ЧМ). В качестве $s_1(t)$ рассматривался ЧМ2 450 Гц, 50 бод, а в качестве $s_2(t)$ сигнал ЧМ8 250 Гц, 125 бод (на рис. 1 в качестве примера рассматривались сигналы именно с такими параметрами). В ходе эксперимента изменялось соотношение амплитуд $D = s_1(t)/s_2(t)$ от 100/1 до 5/1. В качестве порогового значения рассматривалась величина, определяемая формулой (1). Оценивались вероятность обнаружения $P_{обн}$ ложной тревоги α и пропуска цели β . Решение принималось на основе двухэтапного подхода. Так, на рис. 6 показана зависимость вероятности правильного обнаружения от соотношения амплитуд сигналов $P(D)$, а на рис. 7 — зависимость $\alpha(D)$.

В эксперименте обнаружение сигнала с низкой энергией считалось достоверным, если превышение порога G_{p2} происходило в пределах не менее 50 Гц (ширина спектра по первой гармонике одной из поднесущих).

Анализ представленных на рис. 6 и 7 результатов позволяет предположить, что трехэтапный алгоритм принятия решения должен существенно повысить достоверность обнаружения энергетически слабых сигналов на фоне мощных радиопомех.

Так, на рис. 8 (см. вторую сторону обложки) представлены результаты обнаружения реальных сигналов ВЧ-диапазона. На рисунке изображен совокупный спектр реализаций $x_0(t) = s_0(t) + \xi(t)$ нескольких сигналов в полосе от 13 775 000 до 13 825 000 Гц. Источник радиопомех на частоте 13 812 000 Гц является мощным и не позволяет при одноэтапном алгоритме обнаружить остальные сигналы, и только переход к многоэтапному алгоритму позволил решить задачу обнаружения.

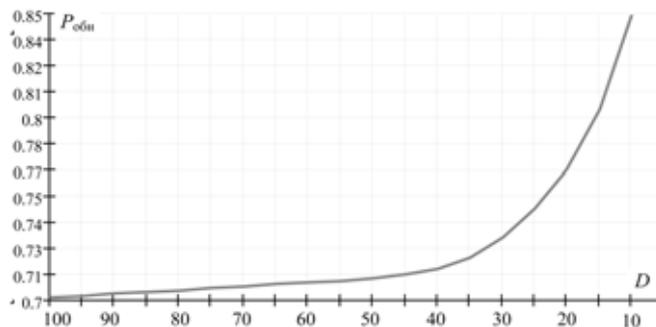


Рис. 6. Зависимость вероятности правильного обнаружения энергетически слабого сигнала от соотношения D

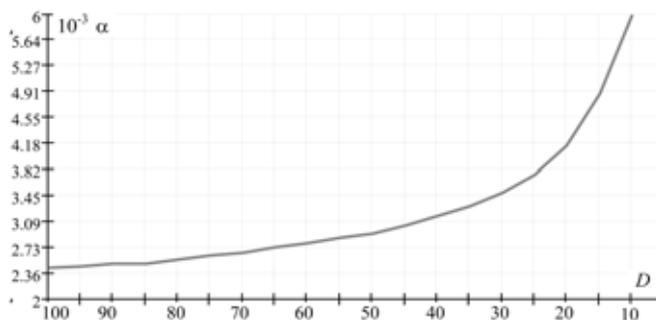


Рис. 7. Зависимость вероятности ложной тревоги при обнаружении сигнала с низкой энергией от соотношения D

В общем случае для обнаружения факта наличия сигнала в полосе анализа необязательно фиксировать все его спектральные составляющие. Вполне достаточно, чтобы уровень выбранного порога превысил лишь то их число, которое необходимо для принятия решения [5].

Предложенный подход к принятию решения при обнаружении гораздо проще в реализационном аспекте, чем трехэтапный алгоритм [5], и поэтому у него существенно расширяется область практического применения.

Список литературы

1. Левин Б. Р. Теоретические основы статистической радиотехники. М.: Сов. радио, 1968. 504 с.
2. Миддлтон Д. Введение в статистическую теорию связи / Пер. с англ. под ред. Б. Р. Левина. М.: Сов. радио, 1962. 782 с.
3. Дворников С. В. Метод обнаружения сигналов диапазона ВЧ на основе двухэтапного алгоритма принятия решения // Научное приборостроение. 2005. Т. 15, № 3. С. 114–119.
4. Дворников С. В. Метод обнаружения на основе посимвольного перемножения реализаций спектра наблюдаемого процесса с автоматическим расчетом порога принятия решения // Научное приборостроение. 2004. Т. 14, № 4. С. 92–97.
5. Дворников С. В., Клепиков В. П., Алексеева Т. Е. и др. Трехэтапный алгоритм принятия решения при обнаружении сигналов с различной динамикой амплитуд // Информация и Космос. 2008. № 4. С. 14–18.

В. В. Савченко, д-р техн. наук, проф., зав. каф.,
Д. Ю. Акатьев, канд. техн. наук, проф.,
 Нижегородский государственный
 лингвистический университет,
 e-mail: svv@lunn.ru; akatjev@lunn.ru

Обучение звуковому строю языка глухонемых и слабослышащих на основе информационной теории восприятия речи¹

Поставлена задача обучения разговорному языку на базовом фонетическом уровне людей с врожденными дефектами в слуховом механизме. При учете отличительной особенности данной категории обучаемых — их неспособности к самостоятельной, т. е. на слух, оценке качества своего произношения в процессе обучения — предложен новый подход к решению поставленной задачи на основе критерия минимума информационного рассогласования с визуализацией его решающей статистики.

Ключевые слова: речь, обучение речи, информационная теория восприятия речи, теоретико-информационный подход, визуализация информационного рассогласования

Введение

Стандартный процесс обучения языку разбивается на несколько уровней [1]. Среди них базовым, определяющим, является звуковой уровень. На нем решается задача настройки голосового механизма человека на звуковой, или фонетический, строй языка, т. е. задача постановки произношения [2]. При этом разным языкам соответствуют разные фонетические списки: и по составу, и по количеству R их элементов [3]. Вместе с тем, фонетический строй языка предъявляет определенные требования к его носителям, посредством которых (и только так) этот строй и реализуется в коммуникациях [4]. Сколько носителей — столько и разных реализаций фонетического списка национального языка. В этом проявляется краеугольная проблема вариативности устной речи. Более того, каждый отдельный носитель-диктор не произносит одинаково дважды даже одну и ту же фонему, причем изменения носят случайный, непредсказуемый характер. Нетрудно поэтому представить себе масштабы многообразия реализаций фонетического строя в пределах каждого

отдельного языка. Как выбрать среди указанного многообразия подходящий эталон для обучения конкретному языку? Будет ли он единственным? И по какому критерию оценивать степень близости тестируемого сигнала к эталонному произношению? Все это — принципиальные вопросы первого ряда при организации процесса обучения языкам.

Задача обучения языкам многократно усложняется для контингента с принципиально ограниченными возможностями их произношений — глухонемых и слабослышащих. Проблемы здесь сводятся главным образом к организации их самостоятельной работы. Налицо очевидное противоречие в самой постановке задачи: обучаемый с недостаточной на данный момент языковой подготовкой и ограниченными возможностями в перспективе должен приблизиться в процессе своего обучения по произношению к некоторому эталону, который он слабо себе представляет (плохо слышит). Указанное противоречие может быть преодолено, например, на основе визуализации произношения обучаемого по отношению к соответствующему речевому сигналу-эталону [5]. Например, в современной информационной системе IBM "Видимая речь-III" для визуализации данных используются временные диаграммы и/или частотные спектры речевых сигналов. Однако здесь возникает еще одна, может быть наиболее острая, проблема: многокритериальность качества речи. Спектры сравниваемых сигналов могут быть близкими друг другу на одних частотах и сильно различаться на других. Добиться близости речевых сигналов одновременно на всех участках (формантах) звукового диапазона частот (в полосе 4 кГц) в процессе постановки произношения — это практически (и теоретически тоже) неразрешимая проблема. Именно указанным обстоятельством объясняется тот общеизвестный факт, что современные IT-методики и технологии постановки речи и обучения языкам до настоящего времени не находят в мире действительно широкого применения несмотря на остроту проблемы и ее высокую гуманитарную значимость. В таком случае очевидным представляется интерес исследователей и пользователей обучающих систем к разработке эффективного математического аппарата для экономного (с визуализацией и сжатием данных) описания фонетического строя национального языка, причем в расчете как на отдельных его носителей с их индивидуальными особенностями речи, так и на произвольные группы разных дикторов, объединенных возрастом, полом, местом жительства и т. п. Общий подход к решению данной задачи в рамках информационной теории

¹ Работа выполнена при поддержке гранта РГНФ 09-06-12125в.

восприятия речи (ИТВР) [3] и составляет содержание настоящей статьи. В отличие от всех известных своих аналогов предлагаемый подход основывается на визуализации речевых сигналов в информационной метрике Кульбака—Лейблера, а процедура самообучения сводится к оптимизации произношения самими слабослышащими по единственному (!) критерию: минимума информационного рассогласования (МИР).

Постановка задачи

Работа большинства систем автоматической постановки произношения основывается на сопоставлении произнесенного слова или фонемы с заранее подготовленным эталоном. Чем меньше расстояние между эталоном и входным (тестируемым) сигналом, тем лучше считается произношение. Недостатком данного подхода является то, что диктор (обучаемый) может быть в силу ряда причин (например, из-за дефектов его органов слуха) просто не в состоянии произнести звук, в достаточной степени близкий (на слух) по произношению к эталону. При этом даже не ясно, что это значит: "в достаточной степени"? Выходом из данной ситуации может служить сравнение тестируемого сигнала x одновременно с несколькими эталонами $x_{r,j}$, $j = \overline{1, J_r}$, по каждой r -й фонеме, где $r = \overline{1, R}$. Диктору будет достаточно приблизить свое произношение к любому из них. Этим существенно ослабляется рассматриваемая проблема: каждый конкретный диктор в процессе своего обучения выбирает наиболее для себя удобный, достижимый вариант эталонного произношения каждой отдельной фонемы из множества имеющихся альтернатив. Одновременно становится понятным и собственно критерий "достаточной степени близости" тестируемого сигнала: он должен войти в границы множества эталонов $\{x_{r,j}\}$ как полноправный $(J_r + 1)$ -й его элемент. В таком случае задача переходит в сугубо предметную плоскость: сначала по каждой из R рассматриваемых фонем требуется сформировать множество ее эталонов $x_{r,j}$ и после этого проверить сигнал x в отношении степени его близости к каждому из J_r элементов этого множества, т. е. вычислить для него J_r -вектор "расстояний" $\rho_{r,j}(x)$, $r = \overline{1, R}$. В рамках ИТВР роль указанного "расстояния" выполняет информационное рассогласование (ИР) по Кульбаку—Лейблеру [6]

$$\rho_{r,j}(x) \triangleq \iint \ln \frac{dP(x)}{dP_{r,j}(x)} P(dx)$$

между выборочными распределениями (вероятностей) тестируемого сигнала $P(x)$ и j -го эталона

r -й фонемы $P_{r,j}(x)$, которая, как известно [7], легко и наглядно визуализируется на основе стандартных средств современной вычислительной техники.

Синтез алгоритма

Речевая функция является продуктом высшей нервной деятельности человеческого организма и служит неперенным условием абстрактного или образного мышления. Образ каждого явления в сознании человека сопровождается соответствующей речевой "меткой". Причем в восприятии разных людей один и тот же образ имеет разные на слух речевые метки. Сказанное в полной мере относится и к фонетическому строю языка. В этом проявляется острейшая проблема вариативности устной речи. В работе [3] предложен оригинальный подход к ее решению.

Несмотря на существующие различия в реализациях некоторой r -й фонемы все они воспринимаются человеком как нечто общее, иначе речь утратила бы свою информативность. Можно поэтому утверждать, что одноименные реализации $x_{r,j}$, $j = \overline{1, J_r}$, $J_r \gg 1$, в сознании человека группируются в соответствующие классы или речевые образы фонем $X_r = \{x_{r,j}\}$, $r = \overline{1, R}$, вокруг некоторого центра — эталонной метки данного образа [7]. В информационной теории восприятия речи указанные эталоны определяются в строгом, теоретико-информационном смысле [8]. В соответствии с общесистемным принципом МИР речевая метка $x_r^* \subset X_r$ образует *информационный центр-эталон (ИЦ-эталон)* r -го речевого образа, если в пределах множества X_r она характеризуется минимальной суммой ИР по Кульбаку—Лейблеру относительно всех других его меток-реализаций $x_{r,j}$, $j = \overline{1, J_r}$, т. е. $x_r^* = x_{r,v}$:

$$J^{-1} \sum_{j=1}^J \rho_{r,v}(x_{r,j}) = \min_{i \leq J} J^{-1} \sum_{j=1}^J \rho_{r,i}(x_{r,j}). \quad (1)$$

Нетрудно увидеть, что именно в понятии ИЦ-эталона r -го множества X_r дается наиболее информативное описание свойств соответствующей фонемы. Одновременно становится очевидным и механизм постановки произношения по каждой отдельной фонеме языка: тестируемый сигнал x в процессе обучения конкретного диктора должен войти в состав множества допустимых реализаций X_r по критерию ограничения сверху

$$\rho_{r,v}(x) \leq \rho_r \quad (2)$$

на величину своего ИР относительно соответствующего ИЦ-эталона x_r^* из выражения (1). В зависимости от выбора в (2) значения порогового

уровня ρ_r меняется вероятность ошибки первого рода

$$P\{\rho_{r, v}(x) > \rho_r | x \in X_r\} = \alpha_r \quad (3)$$

и вместе с ней — уровень значимости $(1 - \alpha_r)$ алгоритма принятия решения. При учете особенностей нашего контингента учащихся уровень значимости на практике варьируется в весьма широких пределах: 0,7...0,99. Здесь надо учитывать только одно ограничение — на допустимую вероятность ошибки второго рода:

$$P\{\rho_{r, v}(y) < \rho_r | y \notin X_r\} \leq \beta_r, \quad r = \overline{1, R}, \quad (4)$$

где y — заведомо неприемлемый вариант тестируемого сигнала, принадлежащий альтернативному (не r -му) множеству реализаций $X_s, s \neq r$.

Предположим, что $\beta_r = \beta_0 = \text{const } \forall r = \overline{1, R}$. Тогда порог ρ_r в (4) будет зависеть от ИР $\rho_{r, v}(y)$ между тестируемым сигналом y и заданным ИЦ-эталоном (1) для r -й фонемы. Чем меньше значение $\rho_{r, v}(y)$, тем меньше порог ρ_r и тем более жесткими устанавливаются требования (2) к качеству тестового сигнала x . Нетрудно понять физический смысл сделанного вывода: требования к качеству тестового сигнала для разных фонем — разные. Сделанный вывод можно проиллюстрировать следующим образом. Отобразим отдельные реализации каждой фонемы языка в виде точек на плоскости (рис. 1). Тогда точка с минимальной суммой расстояний (1) выступает в роли своеобразного "центра массы" заданного множества (отмечена на рисунке жирной точкой). В этом состоит физический смысл ИЦ-эталона. Тогда множество рассматриваемых реализаций данной фонемы — это его допустимые вариации

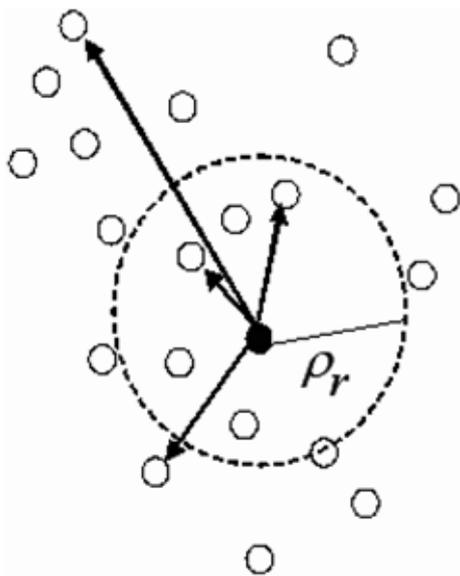


Рис. 1. Реализации фонем и информационный центр-эталон

с точки зрения своих акустических свойств и характеристик. А пороговый уровень ρ_r из выражения (2) — это по сути количественные требования к допустимому уровню упомянутых вариаций в теоретико-информационном смысле. Легко убедиться, что в данном механизме реализуется все тот же принцип МИР из работы [3]. Его адаптивный (по выборке тестового сигнала) вариант реализации рассматривается ниже.

Адаптивный алгоритм

В работе [9] показано, что для дискретных распределений вероятностей критерий МИР эквивалентен общесистемному критерию максимального правдоподобия (МП), т. е. является оптимальным в байесовском смысле. Для непрерывных сигналов свойство оптимальности критерия МИР сохраняется, как известно [10], в Гауссовом семействе распределений. Во многом именно поэтому в ИТВР для описания элементарных речевых единиц (ЭРЕ) типа отдельных фонем используется Гауссов (нормальный) закон с нулевым математическим ожиданием и автокорреляционной матрицей (АКМ) K_r размера $n \times n, n \geq 1$, где $r = \overline{1, R}$. В указанном случае решающая статистика МИР из выражения (1) раскрывается следующим образом [3, 4]:

$$\rho_{r, v}(x) = 0,5[\text{tr}(K \cdot K_v^{-1}) - \ln|K \cdot K_v^{-1}| - n].$$

Здесь K_x — выборочная АКМ тестируемого сигнала x ; $\text{tr}(\cdot), |\cdot|$ — соответственно след и определитель квадратной ($n \times n$)-матрицы. А при дополнительном актуальном [11] условии нормировки элементарных речевых сигналов по энтропии, когда определители всех АКМ равны некоторой константе, можно записать

$$\rho_{r, v}(x) = 0,5[\text{tr}(K \cdot K_v^{-1}) - n].$$

Или, следуя линейной авторегрессионной модели (АР-модели) сигналов типа фонем [8] и переходя в частотную область обработки сигналов [10], окончательно будем иметь следующее выражение для оптимальной решающей статистики МИР [3]:

$$\rho_{r, v}(x) = (F + 1)^{-1} \times \sum_{f=1}^F \frac{\left| 1 + \sum_{m=1}^p a_r(m) \exp(-j\pi mf/F) \right|^2}{\left| 1 + \sum_{m=1}^p a_x(m) \exp(-j\pi mf/F) \right|^2} - 1. \quad (5)$$

Здесь $a_x(m), a_r(m)$ — векторы АР-коэффициентов тестируемого сигнала x и r -го ИЦ-эталоны x_r^* , оба одного порядка $p > 1$. Выражение в числителе оп-

ределяет квадрат амплитудно-частотной характеристики r -го обеляющего фильтра, настроенного на r -ю фонему x_r^* , $r = \overline{1, R}$. Это стандартная формулировка метода обеляющего фильтра (МОФ) в частотной области ($f = 0, 1, \dots, F$ — дискретная частота). Преимуществом данной интерпретации критерия МИР является, прежде всего, возможность его эффективной реализации в адаптивном варианте на основе быстрых вычислительных процедур АР-анализа [12].

Анализ эффективности

Эффективность предложенного алгоритма обучения языку (1), (2), (5) может быть охарактеризована в общем случае по вероятностям его ошибок первого и второго рода. Ошибка первого рода — это когда тестируемый сигнал x выходит за рамки условия (2) при том, что представляет собой одну из немногих (в статистическом смысле на уровне значимости $1 - \alpha_r$) допустимых реализаций множества X_r . Ее вероятность (3) для Гауссова распределения речевого сигнала с нормировкой в частотной области (5) сводится, как известно [10], к выражению

$$\alpha_r = P\{\chi_M^2 > M(\rho_r + 1)\}, \quad (6)$$

где χ_M^2 — случайная величина, распределенная по закону χ_M^2 -Пирсона с M степенями свободы. Значение M ограничено сверху объемом выборки речевого сигнала T на интервале его стационарности, а также порядком p обеляющего фильтра: $M = T - p$. Чем выше порог ρ_r , тем выше становится уровень значимости принимаемого согласно правилу (2) решения и тем более мягкими, следовательно, становятся требования (2) в отношении качества обучения r -й фонеме. И, наоборот, более жестким требованиям к качеству речи отвечают низкий порог и пониженный уровень значимости $1 - \alpha_r$.

Нетрудно понять, что указанные требования могут сильно меняться от одной фонемы к другой. В самом деле, если учитывать ограничение (4) и ошибку второго рода с вероятностью [13]

$$\beta_r = 1 - P\{\chi_M^2 < M(\rho_r + 1)(\rho_{r, v}(y) + 1)^{-1}\}, \\ r = \overline{1, R}, \quad (7)$$

то из равенства $\beta_r = \beta_0 = \text{const } \forall r \leq R$ получаем соответствующее предельное (ограниченное сверху) значение порогового уровня ρ_r . Чем больше степень информационных различий между тестируемым сигналом y и фонемой x_r^* , тем выше это

значение и тем ниже требования к качеству звучания r -й фонемы. Иными словами, при определении требований к степени близости тестируемого сигнала к заданному набору эталонов фонем $\{x_r^*\}$ необходимо учитывать их взаимную различимость в теоретико-информационном смысле. Поэтому особые требования при обучении следует предъявлять к тем фонемам, которые имеют близкие аналоги по своему звучанию среди других фонем звукового строя национального языка. Именно в этом выводе состоит главный результат проведенного исследования.

В общем случае можно утверждать, что эффективность МОФ в задачах обучения языку зависит, прежде всего, от матрицы ($R \times R$) значений ВИР $\rho_r(x_k^*) \triangleq \rho_{r, v}(x_k^*)$, $\forall k, r \leq R$, между сигналами ЭРЕ в пределах обучающего словаря $\{x_r^*\}$. Чем больше ИР между сигналами, тем меньше вероятность их перепутывания при восприятии речи. Пары фонем (x_i^*, x_j^*) , для которых величина $\rho_j(x_i^*)$ минимальна среди всех других пар, будем называть в дальнейшем проблемными парами — по признаку максимальной вероятности их перепутывания. Именно на таких парах фонем должно быть сосредоточено, в первую очередь, наше внимание в задаче обучения речи. Сделанные выводы были в дальнейшем подтверждены и обобщены по результатам экспериментальных исследований. При этом использовался лабораторный образец информационной системы для фонетического анализа слитной речи [14, 15].

Пример. Для иллюстрации выводов проведенного исследования был поставлен следующий эксперимент. Сначала были выбраны двадцать наиболее распространенных фонем русского языка: "а", "о", "у", "э", "ш", "щ", "р", "с", "в", "з", "ж", "и", "л", "л'", "ф", "х", "ч", "е", "ы", "м". Каждая из них была многократно проговорена с записью на диктофон группой из трех дикторов (мужчин среднего возраста). Полученные сигналы с помощью стандартной аппаратуры и встроенного АЦП (частота дискретизации 8 кГц) были записаны в память ПК в виде соответствующих звуковых файлов. При этом объем выборки $T = 120$ по каждой ЭРЕ был установлен исходя из существующих представлений в отношении интервала стационарности речевого сигнала (10–15) мс. В результате по каждому диктору был составлен экспериментальный (рабочий) словарь (ЭС) из множества одноименных ЭРЕ. Далее по каждому сигналу ЭРЕ из ЭС был рассчитан вектор АР-коэффициентов фиксированного порядка $p = 12$. Для

	а	ш	щ	в	з	л'	ф	х	ч
щ	95,72	4,281	0	22,73	6,161	18,90	9,157	146,3	0,126
р	2,908	3,787	28,88	0,839	0,812	5,672	0,527	3,121	32,57
с	21,09	0,448	2,815	30,58	5,916	38,35	6,888	14,28	3,002
в	1,485	4,832	27,23	0	0,745	3,342	0,308	4,029	33,86
з	8,092	1,816	8,948	1,198	0	2,539	0,265	5,852	9,549
ж	65,76	0,511	2,737	11,81	1,778	8,652	4,788	45,05	3,556
и	49,00	6,165	16,49	6,537	4,039	6,970	4,258	123,3	26,24
л	1,312	10,85	40,62	0,977	2,598	15,03	1,716	8,584	46,59
л'	26,60	3,664	15,65	2,978	1,451	0	1,719	58,45	21,64
ф	2,610	2,685	11,95	0,378	0,139	2,661	0	5,790	13,88
х	2,177	6,586	31,44	1,017	0,861	5,284	0,597	0	31,13
ч	106,7	3,491	0,207	52,21	12,48	53,12	16,16	124,5	0

этого применялась высоко скоростная рекуррентная процедура Берга—Левинсона [12]

$$a_m(i) = a_{m-1}(i) + c_m a_{m-1}(m-i), i = \overline{1, m};$$

$$c_m = S_{m-1}^{-2} \sum_{n=m}^{N-1} \eta_{m-1}(n) v_{m-1}(n-1);$$

$$S_{m-1}^2 = 0,5 \sum_{n=m}^{N-1} [\eta_{m-1}^2(n) + v_{m-1}^2(n-1)];$$

$$\eta_m = \eta_{m-1}(n) - c_m v_{m-1}(n-1);$$

$v_m(n) = v_{m-1}(n-1) - c_m \eta_{m-1}(n), n = 0, 1, \dots, N-1, m = \overline{1, p}$, при ее инициализации классифицированной последовательностью отсчетов речевого сигнала $\forall r = \overline{1, R}: v_0(n) = \eta_0(n-1) = x_r(n), n = 0, 1, \dots, T$. Финальные значения рекурсий (8) при $m = p$ и определяли, в конечном счете, базу априорных данных в задаче обучения речи по критерию МИР (1), (2). По ней в дальнейшем были настроены на параллельную обработку входного сигнала 20 линейных (трансверсальных) обеляющих фильтров [13] — все одинакового порядка p . С их помощью при равенстве $x = x_r, v \leq R$ согласно выражению (5) была рассчитана матрица ИР $\|\rho_{v,r}\|$ размера 20×20 . Ее фрагмент представлен в виде табл. 1.

В ней каждый столбец соответствует обеляющему фильтру, настроенному на сигнал X_r , а каждая строка — входному сигналу X_r . Здесь нулевые значения ИР — это признак эквивалентности сигналов в теоретико-информационном смысле. Проблемные пары ЭРЕ выделены в таблице полужирным шрифтом. В частности, это две фрикативные фонемы "щ"- "ч". Видно, что даже в худшем случае значение ИР между фонемами находится в пределах 0,13...0,20. Причем, было показано, что полученный результат практически не зависит от диктора. А при увеличении порядка p до 20 минимальное значение ИР возрастает до 0,20...0,30.

Основные результаты

Описанный эксперимент был проведен 1000 раз по отношению к $J_r = 1000$ независимым реализациям каждой фонемы из указанной проблемной пары. Сначала из множества реализаций фонемы "ч" согласно правилу (1) была отобрана ее эталонная реализация. Затем по ней методом Берга—Левинсона были построены АР-модель порядка $p = 12$ и соответствующий обеляющий фильтр. Выборочная дисперсия его отклика с точностью до константы и множителя 0,5 и определяет значение решающей статистики по МОФ (5). Поэтому, подавая на вход такого фильтра разные реализации фонемы "ч", мы получали последовательность реализаций минимальной решающей статистики, или статистики МИР. Ее график для 100 значений показан на рис. 2, кривая 1. Здесь по оси абсцисс отложены номера реализаций данной фонемы. А для сравнения на том же рисунке представлена кривая 2 — это последовательность из ста разных реализаций альтернативного варианта решающей статистики (5) при подаче на вход обеляющего фильтра множества реализаций фонемы "щ".

Видно, что между двумя кривыми в среднем существует значительный "зазор", что говорит о достаточно надежном, даже в проблемном случае, распознавании по МОФ двух рассматриваемых фонем. При этом полученные результаты хорошо сочетаются (даже визуально) с моделью стационарного временного ряда. Воспользуемся данным фактом для экспериментального оценивания по выборкам сигналов (рис. 2) вероятностей ошибок первого и второго рода (3) и (4) соответственно. Задаваясь приемлемым пороговым значением для вероятности ошибки второго рода, например $\beta_r = 0,3$, и следуя классической формуле Лапласа, для вероятности ошибки первого рода получаем $\alpha_r = 0,27$. Это говорит о довольно жестких требованиях к качеству произношения фонемы "ч". Аналогичный результат: $\beta_r = 0,3, \alpha_r = 0,21$ получим и для фонемы "щ". Но это фонемы из проблемной пары.

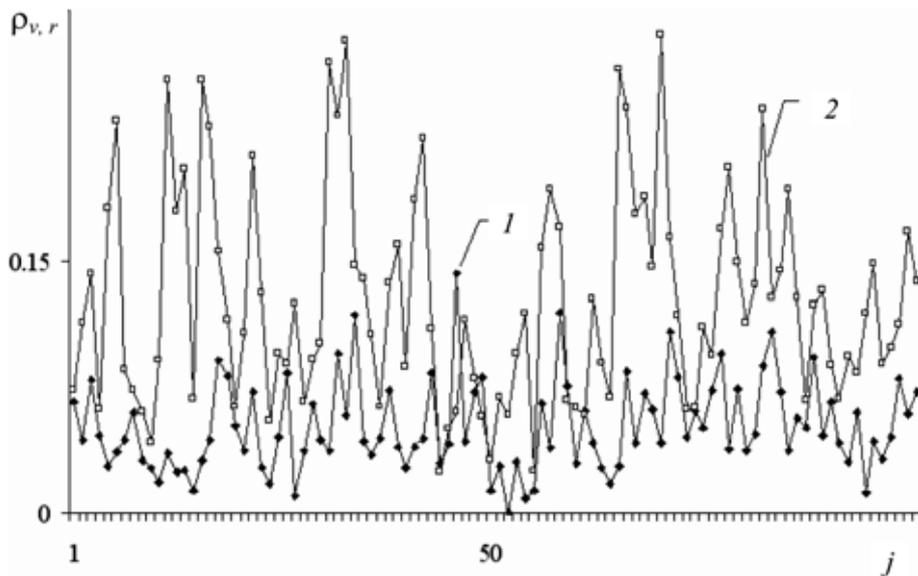


Рис. 2. Решающая статистика МИР

Аналогичный эксперимент, но проведенный в отношении всех остальных фонем из нашего списка, привел к существенно более мягким результатам. Это видно из представленной ниже табл. 2, в которой перечислены 20 рассматриваемых фонем и 20 соответствующих вероятностей α_r при заданной вероятности $\beta_r = 0,3 = \text{const} \forall r = \overline{1, 20}$. Например, для фонемы "а" будем иметь $\alpha_r = 0,001$, что соответствует уровню значимости принимаемого согласно (2) решения самого высокого порядка 0,999.

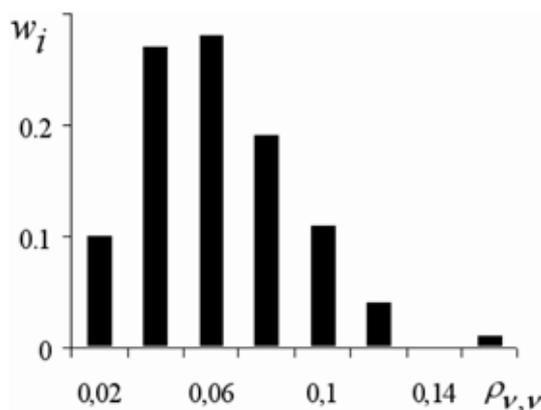


Рис. 3. Гистограмма распределения решающей статистики МИР

Таблица 2

Фонемы	"а"	"о"	"у"	"э"	"ш"	"щ"	"р"
Вероятности α_r	0,001	0,007	0,003	0,004	0,083	0,21	0,076
Фонемы	"с"	"в"	"з"	"ж"	"и"	"л"	"л'"
Вероятности	0,057	0,094	0,25	0,091	0,004	0,059	0,047
Фонемы	"ф"	"х"	"ч"	"е"	"ы"	"м"	
Вероятности α_r	0,12	0,034	0,27	0,004	0,009	0,087	

Отметим, что полученные результаты хорошо согласуются с теоретическими зависимостями (6) и (7). Например, для той же фонемы "а" при равенстве $M = T - p = 120 - 12 = 108$ и по данным табл. 1 — $\min_k \rho_r(x_k^*) = 1,54$ — с помощью таблиц χ_M^2 -распределения [16] получим примерно ту же вероятность ошибки первого рода $\alpha_r = 0,001$, что и в результате эксперимента.

Для полноты картины на рис. 3 представлена гистограмма распределения решающей статистики МИР (5) из рассматриваемого примера, рассчитанная по экспериментальным данным, фрагмент которых отображен кривой 1 на рис. 2. Нетрудно заметить, что качественно, т. е. с точностью до множителя в интервале значений 100—110, она имеет вполне регулярный вид, напоминающий χ_M^2 -распределение с числом степеней свободы $M = 108$. Важно отметить, что законы распределения статистики МИР для всех других фонем русского языка мало отличаются от рассмотренного в примере.

Заключение

Таким образом, благодаря предлагаемому реторико-информационному подходу удается впервые охарактеризовать качество произношения строгой, информационной мерой и, тем самым, впервые преодолеть ряд принципиальных противоречий в самой постановке задачи обучения речи глухонемых и слабослышащих. При этом достижимость эталонного произношения обеспечивается использованием не одного, а целой группы "эталонов", включающей в себя и лучшие образцы произнесений от группы слабослышащих, успешно прошедших ранее обучение.

Полученные результаты могут найти практическое применение не только в процессе обучения, но и в исследовательских целях, например, при сравнительном анализе качества произношения разных дикторов, анализе зависимости их произношений от внешних факторов, в частности, эмоционального состояния и т. п. Круг потенциальных пользователей разработанной информационной системы охватывает как армию частных (физических) лиц, испытывающих острые потребности в постановке или улучшении своего произношения (по существующим оценкам это

почти 10 % населения страны), так и многочисленные коллективные центры психолого-социального сопровождения слабослышащих детей и подростков, такие как Общество ГЛАГОЛ, г. Нижний Новгород.

Список литературы

1. **Комаров К. В.** Методика обучения русскому языку в школе для слабослышащих детей. М.: Оникс 21 век. 2005. 223 с.
2. **Рау Ф. Ф., Слезина Н. Ф.** Методика обучения произношению в школе глухих. М.: Просвещение. 1981. 191 с.
3. **Савченко В. В.** Информационная теория восприятия речи // Изв. вузов России. Радиоэлектроника. 2007. Вып. 6. С. 10—14.
4. **Савченко В. В.** Анализ качества речи на основе информационной теории восприятия // Изв. вузов России. Радиоэлектроника. 2008. Вып. 5. С. 22—32.
5. **Кукушкина О. И., Королевская Т. К., Зеленская Ю. Б.** Информационные технологии в обучении произношению. М.: Полиграф-Сервис. 2004. 158 с.
6. **Кульбак С.** Теория информации и статистика. М.: Наука, 1967. 408 с.
7. **Савченко В. В., Акатьев Д. Ю.** Автотестирование качества произношения речи по принципу минимального информационного рассогласования // Сб. науч. тр. "Современные тенденции компьютеризации процесса изучения иностранных

языков". Луганск: Восточно-украинский национальный университет. 2005. Вып. 3. С. 205—206.

8. **Савченко В. В.** Фонема как элемент информационной теории восприятия речи // Изв. вузов России. Радиоэлектроника. 2008. Вып. 4. С. 3—11.

9. **Савченко В. В., Савченко А. В.** Принцип минимального информационного рассогласования в задаче распознавания дискретных объектов // Изв. вузов России. Радиоэлектроника. 2005. Вып. 3. С. 10—18.

10. **Савченко В. В.** Различение случайных сигналов в частотной области // Радиотехника и электроника. 1997. Т. 42. № 4. С. 426—431.

11. **Савченко В. В., Акатьев Д. Ю., Карпов Н. В.** Автоматическое распознавание элементарных речевых единиц методом обесцараженного фильтра // Изв. вузов России. Радиоэлектроника. 2007. Вып. 4. С. 35—42.

12. **Марпл С. Л.-мл.** Цифровой спектральный анализ и его приложения: Пер. с англ. М.: Мир, 1990. 584 с.

13. **Савченко В. В.** Обнаружение и прогнозирование разладки случайного процесса на основе спектрального оценивания // Автометрия. 1996. № 2. С. 77—84.

14. **Савченко В. В. и др.** Программа для ЭВМ № 2008615442 (Роспатент). Информационная система фонетического анализа слитной речи / Заявка № 2008614233 от 15.09.2008. Зарегистрирована в государственном реестре программ для ЭВМ.

15. **Савченко В. В., Акатьев Д. Ю.** Патент № 80000 (Роспатент). Устройство для фонетического анализа речи / Заявка № 2008136513 от 10.09.2008 г. Опубл. 20.01.2009. Бюл. № 2.

16. **Мюллер П., Нойман П., Шторм Р.** Таблицы по математической статистике: Пер. с нем. М.: Финансы и статистика, 1982. 278 с.

КОНТРОЛЬ И ДИАГНОСТИКА СИСТЕМ

УДК 004.891:004.031.43

Е. Т. Дюндиков, канд. техн. наук, ст. науч. сотр.,
А. А. Качкин, канд. техн. наук, ст. науч. сотр.,
ФГНИИЦ РЭБ ОЭСЗ Минобороны России,
г. Воронеж, e-mail: centtreb@mail.ru

Технология динамической интеграции и представления разнородных данных для анализа и оценки состояния многопараметрических объектов

Рассматриваются последовательность выполнения и содержание унифицированных процедур, обеспечивающих формализованное интегральное представление результатов оценки соответствия заданным требованиям состояния многопараметрических объектов по разнородной измерительной информации, полученной при проведении различных видов контроля.

Ключевые слова: процесс динамической интеграции и представления разнородных данных, цветотрафический образ состояния многопараметрического объекта контроля

Введение

При принятии решений в различных сферах деятельности и условиях (прежде всего в чрезвычайных) для оперативного формирования перечня, содержания управляющих воздействий необходимо иметь сведения об изменениях обстановки и состояния контролируемых объектов (процессов). В целях повышения оперативности информационного обеспечения процесса анализа разнородных данных и принятия решений осуществляется интенсивное внедрение новейших способов, средств контроля, диагностики и информационного обмена. Одной из основных проблем при функционировании или сопряжении информационно-управляющих и диагностических систем является обеспечение динамической интеграции всех имеющихся разнородных данных в единую и наглядную форму, отражающую происходящие изменения.

Очевидно, что существенная часть содержания интегрированной формы должна формироваться на основе сведений о состоянии, тенденциях изменений характеристик объекта контроля (ОК) и его элементов. Несмотря на значительные успехи, достигнутые в теории и практике распознавания

образов, формализации и представления результатов обработки больших массивов информации [1, 2], окончательно решить задачу компактного и наглядного отображения результатов различных видов контроля для оперативного анализа соответствия или причин несоответствия заданным требованиям состояния многопараметрических объектов пока не удастся именно вследствие недостаточной эффективности используемого методического обеспечения интеграции разнородных данных [3, 4]. Представленная в статье технология, обеспечивающая формализованное, интегральное представление результатов оценки соответствия заданным требованиям состояния многопараметрических объектов по разнородной измерительной информации, может найти применение в различных технических и социальных областях деятельности, в которых требуется оперативная проверка соответствия установленным нормам. Основу технологии составляет приводимая ниже методика.

Методика формирования и интегрального представления результатов оценки соответствия состояния контролируемых объектов заданным требованиям

Методика может быть использована в составе специального программного обеспечения аппаратно-программных комплексов (АПК) центров обработки и управления (ЦОУ), средств различных видов контроля (СК), информационно-управляющих и диагностических систем в виде унифицированного модуля формирования и отображения результатов оценки соответствия состояния объектов заданным требованиям.

При использовании методики обеспечивается:

- оценка соответствия установленным нормам фактических значений характеристик контролируемых объектов в масштабе времени, близком к реальному;
- унифицированное формирование и отображение результатов оценки соответствия состояния объектов установленным органами управления требованиям независимо от типа контролируемых объектов, числа и физической сущности их параметров;
- комплексирование (интеграция) результатов оценки состояния объектов, полученных от экспертов и средств,

осуществляющих различные виды контроля в различные моменты времени;

- определение тенденций и динамичности изменения значений параметров и состояния ОК.

В методике в качестве исходных данных используются экспертные оценки и результаты измерений значений контролируемых параметров ОК, зафиксированные средствами различных видов контроля.

Привлекаемые для контроля средства могут осуществлять несколько видов контроля, используя в процессе функционирования дистанционные и/или контактные методы измерения значений параметров ОК. Однако в общем случае для реализации предлагаемой технологии привлекаемые для проведения контроля средства должны обеспечивать:

- получение и преобразование командной и измерительной информации о значениях параметров ОК, по которым оценивается соответствие его состояния установленным требованиям;
- передачу в ЦОУ информации о состоянии ОК для ее использования при формировании решений;
- своевременность оповещения ЦОУ о наличии признаков несоответствия;
- оперативное формирование и доведение до ЦОУ предложений (в виде рекомендаций) по профилактике и устранению нарушений установленных норм;
- интеграцию и документирование результатов проведенных видов контроля.

Структурная схема аппаратно-программного комплекса средств контроля представлена на рис. 1. В структуре аппаратно-программного ком-

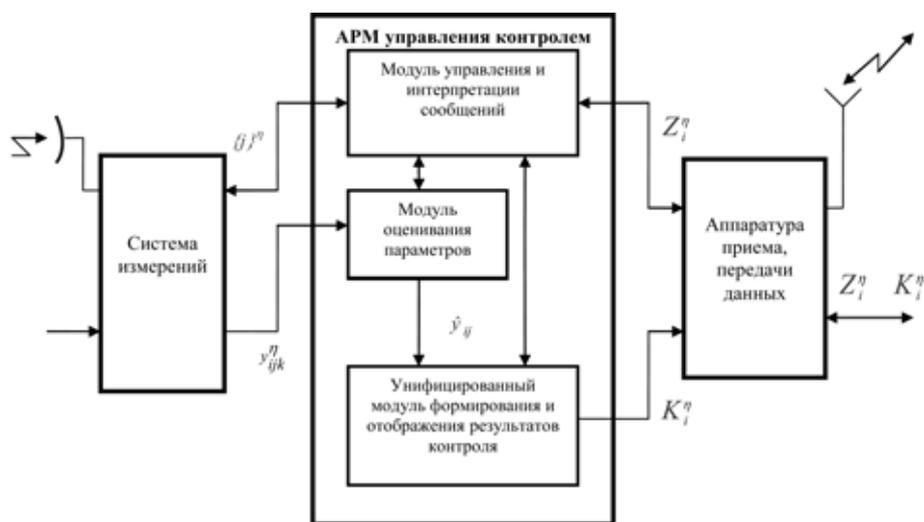


Рис. 1. Структурная схема аппаратно-программного комплекса средств контроля: \hat{y}_{ijk}^{η} — измеренные значения контролируемых параметров; $\{j\}^{\eta}$ — номера контролируемых η -м средством параметров; \hat{y}_{ij} — оцененные значения параметров; Z_i^{η} — задание η -му средству на проведение контроля; K_i^{η} — матрица оценок контролируемых η -м средством параметров

плекса ЦОУ может отсутствовать система измерений. В случаях ведения контроля в каком-либо регионе и оснащенности СК аппаратурой приема-передачи данных с требуемыми характеристиками для обеспечения устойчивого информационного обмена функции сбора данных и их обработки интеграции результатов и управления могут возлагаться на одно из средств контроля.

Наличие в методике унифицированных процедур формализации, интерпретации заданий средствам контроля, результатов контроля и логического вывода рекомендаций на основе анализа значений признаков несоответствия фактических (оцененных) и допустимых значений контролируемых параметров обеспечивает сокращение массивов передаваемых данных при информационном обмене между средствами контроля и управления, что существенно повышает оперативность, а следовательно, и эффективность процесса принятия решений и управления.

В методике используются унифицированные приемы обработки данных на средствах контроля и в ЦОУ для обеспечения в заданные моменты времени визуально ориентированного, интегрально-дифференциального представления результатов оценивания соответствия фактических и допустимых значений параметров и их изменений (т. е. представление результатов оценивания состояния ОК лицу, принимающему решение, проводится одновременно по каждому и по всей совокупности контролируемых параметров).

Методика не требует специальных процедур адаптации для отображения результатов отдельных видов контроля состояния ОК.

Структура методики представлена на рис. 2. Методика состоит из трех частей. В первой части выполняются импортирующие и вычислительные процедуры, позволяющие:

- осуществить сбор необходимой информации в соответствии с заданием и дополнить содержание базы данных D_i об ОК данными по интервалам допустимых значений контролируемых параметров, по которым оценивается соответствие состояния ОК установленным нормам;
- сформировать по измеренным в процессе контроля значениям параметров, контролируемых средствами контроля, матрицу $\hat{Y}_i = \{\hat{y}_{ij}\}$ (среднее значение параметра, дисперсия, абсолютная скорость изменения и т. п.) значений их оценок.

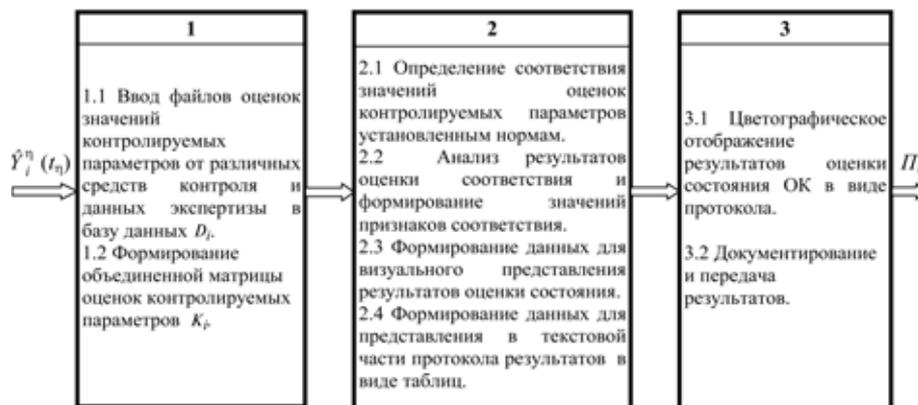


Рис. 2. Структура методики и последовательность выполнения процедур

Во второй, расчетной части методики проводятся сравнивающие и вычислительные операции, а также операции логического выбора, позволяющие:

- сравнить значения $\{\hat{y}_{ij}\} | \hat{y}_{ij} = f(y_{ijk})$ оценок контролируемых параметров (и/или интегральных показателей) ОК, полученных на средствах контроля, с допустимыми значениями;
- зафиксировать факты соответствия и несоответствия значений оценок контролируемых параметров ОК их допустимым значениям;
- получить значение признаков $\delta_i = \{\delta_{ij}^0, \delta_{ij}^1\}$ несоответствия и соответствия значений оценок контролируемых параметров установленным нормам (допускам);
- сформировать цветографическую экранную форму "Протокола результатов оценки состояния ОК" для визуального интегрального представления результатов сравнения и фактов наличия признаков несоответствия;
- сформировать текстовую часть "Протокола результатов оценки состояния ОК" в виде таблицы, содержащей идентификаторы параметров ОК, которые имеют недопустимые фактические значения, и их отклонения от допустимых;
- сформировать по совокупности зафиксированных фактов отклонений от допустимых значений рекомендации (предложения) по устранению нарушений.

В третьей части методики выполняются процедуры, обеспечивающие отображение на мониторе АРМ средств контроля и/или ЦОУ:

- цветографического образа состояния объекта;
- таблицы, содержащей номера, идентификаторы параметров, величины (относительные и/или абсолютные) их несоответствия допускам, а также рекомендации.

Далее осуществляется документирование экранной формы "Протокола результатов оценки состояния ОК" в виде жестких копий.

Исходные данные и требования к системе интеграции и представления результатов

Задан момент времени $t_k \in t$ представления результатов комплексного контроля пространственно распределенного (или сосредоточенного) объекта O_i , i — номер объекта.

В течение интервала времени $\tau_\eta = [t_0, t_\eta]$ проведен контроль объекта, где $\eta = 1, \dots, \eta^*$ — номера средств, привлекаемых для контроля. Для O_i априорно известен состав параметров $\{y_{ij}\}$, где $j = 1, \dots, j^*$ — номера параметров ОК, доступных для измерения средствами контроля. При этом также известны интервалы допустимых значений каждого параметра (и/или интегрального показателя) $[\hat{y}_{ij}^H, \hat{y}_{ij}^B] \in D_j$. Нижние и верхние значения $\hat{y}_{ij}^H, \hat{y}_{ij}^B$, как правило, задаются в виде требований в руководящих документах, конструкторской или эксплуатационной документации.

С помощью средств различных видов контроля по измеренным значениям (и/или экспертами) в соответствии с правилами $\Psi = \{\Psi_j\}$ определяются оценки $\hat{Y}_i^n = \{\hat{y}_{ij}^n\}$ значений контролируемых параметров.

По результатам контроля требуется сформировать и передать в ЦОУ необходимые для анализа и принятия решений сведения:

- о времени t_η окончания и длительности τ_η проведения контроля объекта каждым средством;
- о виде состояния ОК и местоположении ОК и СК;
- о параметрах ОК, фактические значения которых не соответствуют допустимым;
- о величинах несоответствия δ_{ij}^0 фактических (оцененных) и допустимых значений контролируемых параметров;
- о тенденциях и динамичности изменения характеристик и состояния ОК.

Совокупность перечисленных сведений должна быть включена как в содержание формируемых средствами протоколов Π_i^n результатов отдельных видов контроля, так и в содержание протокола Π_i результатов комплексного контроля состояния объекта, который формируется в центре обработки и управления путем интеграции разнородных данных о результатах, полученных при проведении различных видов контроля. При этом содержание протоколов всех видов должно представляться не более чем одной экранной формой, обеспечивающей оперативное восприятие и однозначную интерпретацию результатов оценки соответствия характеристик контролируемого объекта заданным требованиям.

Методические основы интеграции и представления разнородных данных

При сопряжении информационно-управляющих систем с системами контроля и диагностиро-

вания для обеспечения единой интерпретации семантического содержания сообщений при информационном обмене одним из основных требований к подсистеме представления информации является обеспечение на всех уровнях иерархии сопрягаемых систем универсальности, оперативности и компактности отображения результатов контроля при инвариантности к физической сущности и числу параметров, по которым оценивается состояние ОК. Для выполнения этих требований и обеспечения информационного сопряжения должна быть директивно осуществлена адресация средств контроля и управления, единая нумерация характеристик (параметров), по которым проводится оценка соответствия ОК установленным нормам и требованиям. При этом должны быть установлены единые правила формирования и использования:

- словаря терминов $T = \{T_1, T_2\}$ для формирования текстовой части протоколов $T = \{T_1, T_2\}$, где $(T_1) \vdash (T_{1i}) \in T_1$ — текстовая часть, единая форма которой при контроле всех типов ОК устанавливается для отображения даты и времени контроля, идентификаторов СК, координат местоположения ОК и СК $X_i = (\Lambda_i, \Phi_i)$, $X_\eta = (\Lambda_\eta, \Phi_\eta)$; $(T_2) \vdash (T_{2i}) \in T_2$ — текстовая часть, содержащая идентификаторы параметров ОК, значения которых не соответствуют допустимым, величины несоответствия δ_{ij}^0 фактических и допустимых значений контролируемых параметров, а также рекомендации. Содержание T_2 формируется в зависимости от результатов анализа выявленной совокупности признаков δ_{ij}^0 .

Схема организации процесса интеграции и представления разнородных данных при проведении комплексного контроля (мониторинга) многопараметрического объекта представлена на рис. 3.

Процесс интеграции и унифицированного представления разнородных данных при формировании результатов комплексного контроля (мониторинга) может быть представлен в виде последовательности выполнения во времени логических и вычислительных процедур $\Phi = \{\Phi_p\}$, $p = 1, \dots, 6$, по совокупности правил, составляющих основу продукционной базы знаний.

База знаний, включающая базы данных, правил [5] и формируемая на средствах контроля и в ЦОУ, обеспечивает:

- формирование заданий Z_i^n на проведение контроля средствами контроля;
- интерпретацию заданий средствам контроля, формирование команд управления системами измерения и приема-передачи данных;
- логическое управление процессом преобразования измеренных и оценивания значений контролируемых параметров;

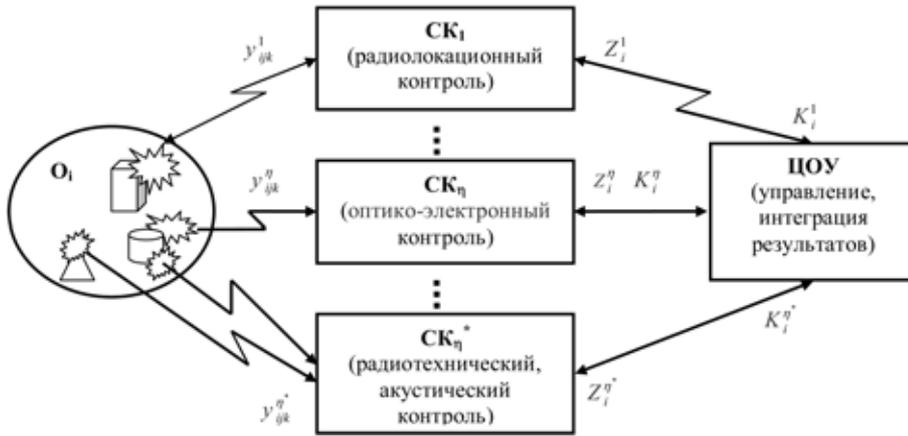


Рис. 3. Схема организации комплексного контроля многопараметрического объекта

- формирование и интеграцию результатов оценки состояния ОК;
- формирование по правилам U рекомендаций по устранению выявленных нарушений установленных норм при наличии фактов несоответствия фактических и допустимых значений параметров.

В соответствии с первой процедурой

$$\Phi_1: \forall j, \eta \exists \{((k_{ij}^z = 1), (\hat{y}_{ij}^n), (\hat{y}_{ij}^b), T, ID, \Psi, \Omega, U, W) \in D_i, i = \text{const} \mid t < t_0; \\ \forall j, \eta \exists (y_{ijk}^n) \in D_i \mid t = t_{in}^*$$

(где k_{ij}^z — эталонные значения признаков соответствия параметров, t_{in} — момент времени окончания измерений) в базы данных и правил АРМ средств контроля и ЦОУ до момента t_0 — начала измерений — должны быть введены исходные данные:

- совокупность идентификаторов $ID = \{IK, IO, IP\}$, где $IK = \{IK_k\}$ — идентификаторы СК, $IO = \{IO_i\}$ — идентификаторы ОК и $IP = \{IP_j\}$ — идентификаторы контролируемых параметров;
- матрицы значений верхних и нижних границ $(\hat{y}_{ij}^n)_{1 \times j^*}, (\hat{y}_{ij}^b)_{1 \times j^*}$ интервалов допустимых изменений характеристик (параметров) объектов, подлежащих контролю;
- правила $\Psi = \{\Psi_j\}$ оценивания значений контролируемых параметров и/или вычисления значений интегральных показателей;
- правила Ω формирования значений признаков $\delta_i = \{\delta_{ij}^0, \delta_{ij}^1\}$;
- правила W цветографического отображения состояния ОК в полярной системе координат $P(\rho_{ij}, \theta_{ij})$, где ρ_{ij}, θ_{ij} — полярные радиусы и углы их поворота относительно полярной оси;
- словарь терминов $T = \{T_1, T_2\}$ и правила $U = \{U_1, U_2\}$ для формирования текстовой части протоколов.

Кроме этого должна быть известна последовательность моментов времени $\{t_n\}, n = 1, \dots, n^*$ представления докладов средствами контроля в ЦОУ. При этом $t_n = t_0 + n\Delta_n$, где $\Delta_n = \text{var}$ — заданный для n -го средства интервал времени проведения контроля, значение которого может варьироваться в зависимости от динамики изменения состояния объектов контроля и оперативной обстановки.

В соответствии с процедурой

$$\Phi_2: \{((y_{ijk}^n) \xrightarrow{\Psi} (\hat{y}_{ij}^n) \xrightarrow{\Omega} \delta_i^n), \\ (T \xrightarrow{U} T_i^n) \mid t_{in} < t < t_{in}^*$$

после момента t_{in} (окончания измерений) на средствах каждого вида контроля по совокупности (y_{ijk}^n) измеренных значений рассчитываются $\hat{Y}_{ij}^n = (\hat{y}_{ij}^n)$ оценки значений параметров (и/или значения интегральных показателей), по результатам сравнения которых с допусками вычисляются значения признаков соответствия δ_i^n . По совокупности значений признаков осуществляется выбор необходимых терминов для формирования рекомендаций.

В соответствии с процедурой

$$\Phi_3: \{((\theta_{ij}, \delta_i^n) \xrightarrow{W} M^n \rightarrow (\text{Fr } M^n) \subset P(\rho_{ij}, \theta_{ij}), \\ (K_i^n = (\hat{y}_{ij}^n)_{1 \times j^*}) \mid t = t_{in}^*$$

отображаются в полярной системе координат значения признаков соответствия в виде множества меток $M^n = (M_j^n) \mid M_j^n = f(\delta_{ij}^n, \theta_{ij})$, соединенных линией $\text{Fr } M^n$, являющейся границей фигуры [6], которую интерпретируют как цветографический образ (ЦГО) состояния объекта контроля.

Оценки значений параметров формируются в момент t_{in} в виде файла K_i^n данных и затем передаются в ЦОУ, где после преобразования в аппаратуре приема-передачи данных (АППД) вводятся в АРМ. Порядок расположения данных в файлах передаваемых при информационном обмене между средствами контроля и управления представлен табл. 1, 2.

В АРМ ЦОУ в соответствии с процедурой $\Phi_4: (K_i = \sum_{\eta} K_i^n) \mid t = t_n$ значения оценок K_i^n , полученные от СК, осуществляющих различные виды контроля, объединяются в матрицу $(K_i)_{1 \times j^*}$ оценок контролируемых параметров.

Порядок распоряжения данных в файлах заданий, передаваемых средству контроля

1	2	3	4	5	6	7	8	...	$j-1$	j	$j+1$	$j+2$	$j+3$...	j^*-1	j^*
0	0	0	0	0	0	0	0	...	1	1	1	1	1	...	0	0

Порядок распоряжения данных (оценок) в файлах, передаваемых от средства контроля

1	2	3	4	5	6	7	8	...	$j-1$	j	$j+1$	$j+2$	$j+3$...	j^*-1	j^*
0	0	0	0	0	0	0	0	...	\hat{y}_{ij-1}^n	\hat{y}_{ij}^n	\hat{y}_{ij+1}^n	\hat{y}_{ij+2}^n	\hat{y}_{ij+3}^n	...	0	0

В соответствии с процедурами

$$\Phi_5: (K_i \xrightarrow{\Psi, \Omega} \delta_i) | t_n < t < t_k;$$

$$\Phi_6: \{((\theta_{ij}, \delta_i) \xrightarrow{W} M \rightarrow (\text{Fr}M) \in G_i),$$

$$((\theta_{ij}, k_{ij}^z) \xrightarrow{W} M^z \rightarrow (\text{Fr}M^z) \in G_i) \subset P(\rho_{ij}, \theta_{ij}),$$

$$(T \xrightarrow{U} T_i) | t = t_k,$$

где M^z — множество меток эталонных значений признаков соответствия, на основе правил Ψ , Ω , U , W определяются значения элементов матрицы $(\delta_i)_{1 \times j^*}$, отображаются в полярной системе координат значения признаков соответствия в виде множества меток $M = (M_j)$, соединенных линией $\text{Fr}M$, которая является границей фигуры, разделенной по числу видов контроля, осуществляемых всеми средствами контроля на цветные поля. Полученную фигуру интерпретируют как интегрированный цветографический образ состояния объекта в момент t_k окончания комплексного контроля.

Для формирования фигуры выполняются следующие операции:

- определяются данные в виде упорядоченных пар значений радиусов и углов поворота радиусов (ρ_{ij}, θ_{ij}) , необходимые для формирования и отображения в полярной системе координат множества M .

Значения углов θ_{ij} вычисляются по формуле

$$\theta_{ij} = \frac{\Theta}{\omega} j,$$

где $\Theta = \frac{360^\circ}{j^*}$; ω — градусная мера радиана;

- затем соотносятся значения элементов с одинаковым номером из матриц $(\hat{y}_{ij}^H)_{1 \times j^*}$, $(\hat{y}_{ij}^B)_{1 \times j^*}$, $(k_{ij})_{1 \times j^*}$ вычисляются величины δ_{ij}^0 , являющиеся признаком несоответствия уста-

новленным нормам, и δ_{ij}^1 , являющиеся признаком соответствия установленным нормам. При этом предварительно проверяется выполнение ряда условий. В случае если оцененное значение параметра (или рассчитанное значение интегрального показателя) находится в поле допуска, т. е. $\hat{y}_{ij}^H \leq k_{ij} \leq \hat{y}_{ij}^B$, в полярной системе координат $\mathfrak{R}(\rho_{ij}, \theta_{ij})$ фиксируются метки M_j с координатами (ρ_{ij}, θ_{ij}) на окружности, где $\rho_{ij} = \delta_{ij}^1 = 1$ — единичный радиус и θ_{ij} — угол поворота радиуса относительно полярной оси. Число меток при отображении ЦГО равно числу j^* контролируемых параметров.

В случае если оценка значения параметра находится вне поля допуска, фиксируются метки с координатами $(\delta_{ij}^0, \theta_{ij})$, не находящиеся на окружности, так как $\rho_{ij} = \delta_{ij}^0$, а $\delta_{ij}^0 \neq 1$. При такой ситуации на мониторе АРМ наблюдается "выпадение" меток из упорядоченного расположения на окружности, соединяющей все метки граничной линии, т. е. происходит образование излома. Направление и величина излома формируются на основе анализа ситуации, при которой реализовалось нарушение $k_{ij} < y_{ij}^H$ нижней или верхней $k_{ij} > y_{ij}^B$ границы допуска. Вполне очевидно, что направление излома определяет вид действий, необходимых для обеспечения выполнения установленных норм. Если $\delta_{ij}^0 < 1$, то требуется выполнить организационно-технические мероприятия, обеспечивающие увеличение значения контролируемого параметра с номером j . Если $\delta_{ij}^0 > 1$, то требуется осуществить действия по уменьшению значения контролируемого параметра. При этом термины, соответствующие рекомендуемому виду действий, определяются с помощью операций логического выбора из словаря терминов и включаются в содержание рекомендаций.

Формирование значений признаков $\delta_i = (\delta_{ij}^0, \delta_{ij}^1)$ как на средствах контроля, так и в ЦОУ осуществляется следующим образом:

$$\delta_{ij}^0 = \begin{cases} \left(\frac{k_{ij}}{y_{ij}^B}\right), k_{ij} > y_{ij}^B; \\ \left(\frac{k_{ij}}{y_{ij}^H}\right), k_{ij} < y_{ij}^H; \end{cases}$$

$$\delta_{ij}^1 = \begin{cases} \delta_{ij}^1 = 1, k_{ij} \in [y_{ij}^H, y_{ij}^B]; \\ \delta_{ij}^0 \neq 1, k_{ij} \notin [y_{ij}^H, y_{ij}^B]. \end{cases}$$

То есть если оценки параметров совпадают с эталонными значениями или находятся в границах допуска контролируемой характеристики, то признаку соответствия δ_{ij} присваивается 1. В противном случае при формировании данных для отображения ЦГО используется вычисленное значение $\delta_{ij} = \delta_{ij}^0$.

В соответствии с процедурой Φ_6 в полярных координатах $P(\delta_{ij} = \rho_{ij}, \theta_{ij})$ строится линия $k_i^3 = f(\theta_{ij})$, ограничивающая фигуру, которая интерпретируется как эталонный образ состояния ОК и совпадает с фигурой, представленной в задании на комплексный контроль средствам контроля (рис. 4, см. четвертую сторону обложки). На рис. 5 (см. четвертую сторону обложки) представлен ЦГО состояния объекта контроля в моменты времени t_1, t_2, t_3 .

В результате отображения образов и их сопоставления в одной системе координат на мониторе АРМ фиксируются:

- факты наличия несовпадений (признаков несоответствия);
- номера параметров, имеющих недопустимые значения;
- направления изменений значений параметров и относительная величина их несоответствия допускам.

При наличии хотя бы одного факта $\delta_{ij} \neq 1$ совместно с ЦГО на экранной форме в текстовой части T_2 "Протокола..." отображаются оцененные значения k_{ij} тех параметров, значения которых не соответствуют допустимым, а также значения отклонений от границ допуска в абсолютных или относительных величинах. Цветовые поля в ЦГО формируют в виде частей фигуры, ограниченных полярными радиусами, номера которых соответствуют максимальным и минимальным номерам параметров, контролируемых при осуществлении конкретного вида контроля, и граничной линией, соединяющей метки значений признаков соответствия, вычисляемые при ведении этого вида контроля. При этом используют цветокодую матрицу, номера строк которой соответствуют номерам видов контроля, а номера столбцов — но-

мерам контролируемых параметров. Элементами этой матрицы являются коды цветов, которыми выделяют поля при отображении.

Для определения характера изменений во времени состояния ОК выполняются операции:

- объединения табличных и совмещения графических форм унифицированных протоколов комплексного контроля, полученных центром обработки и управления в моменты $t_k, k = 1, 2, \dots, k^*$;
- определения по графическим формам предыдущих и последнего из сформированных протоколов комплексного контроля фактов наличия изменения значений признаков несоответствия для каждого параметра;
- определения по значениям признаков, полученных в различные моменты времени, величин и тенденций изменения контролируемых параметров;
- фиксирования по объединенной табличной форме оцененных значений параметров и моментов t_n времени окончания измерений на средствах контроля;
- формирования из зафиксированных значений временных рядов, которые далее используются как исходные данные при необходимости определения корреляционных свойств, а также при прогнозировании изменений значений параметров и состояния многопараметрического объекта во времени.

В целях подтверждения возможности достижения ожидаемых технических эффектов при практическом внедрении предлагаемой технологии динамической интеграции разнородных данных проведено макетирование и моделирование функционирования фрагмента системы мониторинга пространственно распределенного ОК. Макет фрагмента строился на основе комплексирования средств радиотехнического и оптико-электронного контроля. В процессе функционирования макета информационный обмен между его элементами осуществлялся по каналам радиосвязи. Установлено, что при передаче результатов оптико-электронного контроля различных типов ОК объем данных, передаваемых от средств контроля, составлял до 100 Кбайт, что соответствует уменьшению (по сравнению с существующими способами сбора данных для их анализа и оценки состояния ОК) объема получаемой в ЦОУ измерительной информации на $10^6 - 10^7$ байт. При этом время получения "Протокола результатов комплексного контроля состояния объекта", содержащего текстовую часть и интегрированный образ состояния ОК, составило 3—5 с. Результаты анализа функционирования макета системы комплексного контроля подтвердили эффективность разработанных методических приемов, возмож-

ность и экономическую целесообразность практической реализации на существующей технической базе предлагаемой технологии для обеспечения информационного сопряжения ведомственных и государственной систем комплексного мониторинга.

Таким образом, использование унифицированных методических приемов интеграции, визуально ориентированной формализации разнородных данных о состоянии объектов обеспечивает:

- интегрально-дифференциальное представление результатов оценки соответствия заданным требованиям состояния ОК в масштабе времени, близком к реальному;
- сокращение времени анализа представленных данных о состоянии ОК и принятия решений. Основные технические эффекты, ожидаемые в случае практической реализации технологии динамической интеграции разнородных данных при контроле состояния многопараметрических объектов:
- обеспечивается унификация и компактность представления результатов оценки состояния объектов по всем видам контроля независимо от числа и типов объектов, числа и физической сущности их характеристик, единиц измерения и ведомственной принадлежности средств контроля и управления;
- снижается уровень требований по пропускной способности к разрабатываемым средствам информационного обмена и уменьшается нагрузка используемых (арендуемых) линий связи для передачи данных о состоянии контролируемых объектов и команд управления;
- обеспечивается комплексная оценка соответствия установленным нормам состояния объектов по различным видам контроля в масштабе времени, близком к реальному;
- обеспечивается повышение точности распознавания вида состояния и определения причин аномального функционирования элементов объектов контроля;
- обеспечивается сопряжение существующих и перспективных информационно-управляющих ведомственных систем, получающих данные от подчиненных систем различных видов контроля (мониторинга);
- минимизируется влияние человеческого фактора на достоверность представляемой инфор-

мации в процессе формирования результатов контроля.

Выводы

Реализация предложенной технологии динамической интеграции разнородных данных на основе унификации порядка и содержания информационного обмена между органами управления и средствами контроля обеспечит:

- использование в полном объеме сложившегося порядка проведения комплексного мониторинга, состава методического обеспечения и комплекта средств контроля, имеющихся в различных ведомствах, без их существенной модернизации;
- создание единой информационно-методической основы систем комплексного мониторинга в различных ведомствах;
- семантическое сопряжение ведомственных (корпоративных) систем при существующем уровне технического и программного обеспечения их функционирования;
- уточнение концепции и плана модернизации системы комплексного мониторинга государства, а также содержания и формы информационного обмена в глобальной системе;
- существенное сокращение временных затрат, оптимизацию финансовых ресурсов на модернизацию, развертывание ведомственных и государственной систем комплексного мониторинга.

Список литературы

1. **Везенов В. И., Новиков Ю. А.** Организация вычислительного процесса в системах сбора измерительной информации // Информационные технологии. 2007, № 1. С. 69—73.
2. **Чернышов А. В.** Система ввода, обработки и документирования измерительной информации рабочего места контроля бортовой телеметрической аппаратуры // Информационные технологии. 2007, № 2. С. 58—60.
3. **Патент** на изобретение, Россия. № 2296421, МПК Н 04 В 7/185, 2007. Система автоматизированного контроля состояния потенциально опасных объектов Российской Федерации в интересах обеспечения защиты от техногенных, природных и террористических угроз.
4. **Патент** на изобретение, Россия. № 2138849, МПК G06F19/00, 1999. Способ анализа состояния многопараметрического объекта или процесса.
5. **Алиев Р. А.** и др. Производственные системы с искусственным интеллектом. М.: Радио и связь, 1990.
6. **Каазик Ю. Я.** Математический словарь. М.: ФИЗМАТЛИТ, 2007.

УДК 620.193.2:197.3

Е. С. Углова, науч. сотр.,
Л. С. Моисеева, д-р техн. наук,
зав. лаб. новых материалов,
ООО "Объединенный центр исследований
и разработок", г. Москва,
e-mail: uglovaes@rambler.ru

Математическая модель и информационная система поддержки принятия решений в задачах коррозионной защиты

Рассматривается разработанная математическая модель расчета скорости коррозии стали в водной среде для информационной системы поддержки принятия решений в задачах противокоррозионной защиты для конкретного типа коррозии — углекислотной.

Ключевые слова: система, коррозия, защита, среда

Математическое моделирование и связанный с ним компьютерный эксперимент позволяют значительно упрощать, убыстрять и унифицировать решения многих проблем, возникающих на производственных объектах, например, таких как предупреждение аварий трубопроводов на нефтепромысле, возникающих по причине коррозии металла. Еще в 70—80-х годах прошлого века за рубежом были начаты работы по созданию компьютерных систем обработки и хранения информации в области коррозии и защиты (CORROS, EXPERT, ESCORT, BEASY-CP, PRIME, CORREAU, ACHILLES, MENTOR) [1]. Отечественный опыт в плане создания таких автоматических систем невелик, и первые описания или упоминания о разработке компьютерных программ в области коррозии и противокоррозионной защиты, в частности, касающихся ингибиторной защиты от коррозии, относятся к 80—90-м годам прошлого века (КОРСАПР, РЕАГЕНТ, ЭПИС, Экстра 6.0) [2, 3].

В разрабатываемой авторами информационной системе поддержки принятия решений в задачах противокоррозионной защиты проводятся оценка агрессивности среды (пластовой воды) и необходимости ее ингибирования, анализ типа коррозии и расчет прогнозируемой скорости коррозии, а также выбор ингибитора по десяти при-

знакам. На данном этапе разработки программы рассматривается углекислотная коррозия, так как ее вклад в коррозионные разрушения трубопроводов месторождений Западно-Сибирского мегабассейна, являющегося одним из крупнейших в мире, велик (скорость достигает 8 мм/год), и проблема борьбы с ней остается крайне актуальной [4—6].

Основой системы является расчет прогнозируемой скорости коррозии, поскольку все дальнейшие экспертные оценки агрессивности среды и необходимости ее ингибирования зависят от полученного значения. Авторами разработана математическая модель расчета скорости коррозии углеродистой стали. Для этого на основе результатов проведенных модельных экспериментов процесса коррозии и сопоставления их с результатами анализа сред и осадков, выделены десять значимых физико-химических факторов среды (рН, общая минерализация, содержание CO_2 и др.), влияющих на скорость и интенсивность коррозионных разрушений нефтегазопромысловых трубопроводов, контактирующих с агрессивной средой. Для каждого фактора определены значения веса фактора g и веса значения фактора w . К оценке скорости коррозии применена линейная факторная модель, оценочная функция которой имеет вид

$$F_i = \sum_j w_j^i g_j^i, \quad (1)$$

где i — среда; j — фактор; g — вес фактора; w — вес значения фактора.

Далее скорость коррозии рассматривается как один из видов выхода продукта комплекса химических и электрохимических реакций, протекающих при коррозионном процессе. При этом учитывается, что значения всех факторов изменяются в определенных границах. Отсюда получаем следующую математическую модель расчета скорости коррозии:

$$V_{\text{кор}} = \frac{F - F_{\text{min}}}{F_{\text{max}} - F_{\text{min}}} V_{\text{max}}, \quad (2)$$

где $V_{\text{кор}}$ — скорость коррозии по результатам факторного анализа для среды, основу которой составляет пластовая вода, мм/год; V_{max} — максимальная скорость коррозии, измеренная на объекте (реальная), мм/год; F — оценочная функция, складывающаяся из суммы произведений факторических (определенных для анализируемой среды)

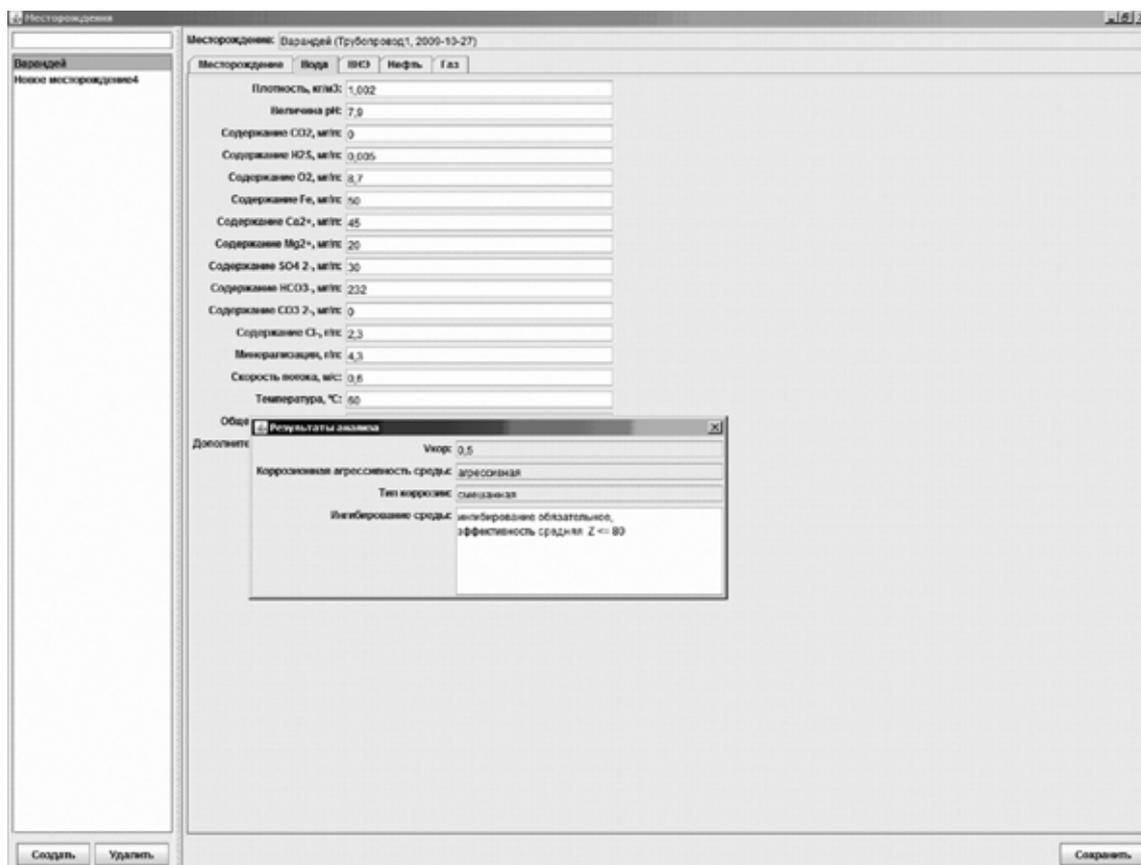


Рис. 1. Окно программы с заполненными значениями для среды "Вода" одного из северных месторождений РФ

веса фактора и веса значения фактора; F_{\min} и F_{\max} — соответственно минимальная и максимальная оценочные функции, складывающиеся из суммы произведений соответственно минимальных и максимальных весов фактора и веса значения фактора (1).

Предложенная математическая модель легла в основу разработки авторами информационной системы, условно названной "Барьер-СО₂", представляющей собой клиент-серверное многопользовательское приложение. Клиентская часть системы "Барьер-СО₂" предназначена для работы под управлением ОС Windows 2000, XP, Vista, а также ОС семейства Unix. Серверная часть системы

"Барьер-СО₂" предназначена для работы на серверах под управлением ОС Windows 2000, XP, 2003 или ОС семейства Unix. Окно программы с заполненными значениями для среды "Вода" одного из северных месторождений РФ представлено на рис. 1.

Для проверки адекватности полученной математической модели реальности были проведены модельные эксперименты по определению значений равномерной скорости коррозии низкоуглеродистой стали в реальной среде — пластовой воде из единой системы трубопроводов, транспортирующих минерализованную воду для заводнения нефтяных пластов в целях поддержания

Физико-химические показатели, расчетные и реальные значения скоростей коррозии участков системы ППД

Участок системы	Фактор								$V_{\text{кор}}$, мм/год		
	pH	M	C_{Cl^-}	$C_{\text{H}_2\text{S}}$	C_{O_2}	$C_{\text{HCO}_3^-}$	t , °C	$\vartheta_{\text{потока}}$	$V_{\text{кор.расчетная}}$	$V_{\text{кор.реальная}}$	$V_{\text{кор.мах реальная}}$
B1.1	7,9	4,2	2,3	0,005	8,6	256	25	1	0,9	0,8	3,2
B01	6,4	23,8	13,9	1,5	4,2	610	25	1	0,4	0,3	0,6
B03	7,9	4,3	2,3	0,005	8,7	232	25	1	0,5	0,5	1,8

Примечание: pH — водородный показатель среды; M — общая минерализация; C — концентрация элемента (хлоридов, сероводорода, растворенного кислорода, гидрокарбонатов); t — температура; $\vartheta_{\text{потока}}$ — скорость движения потока

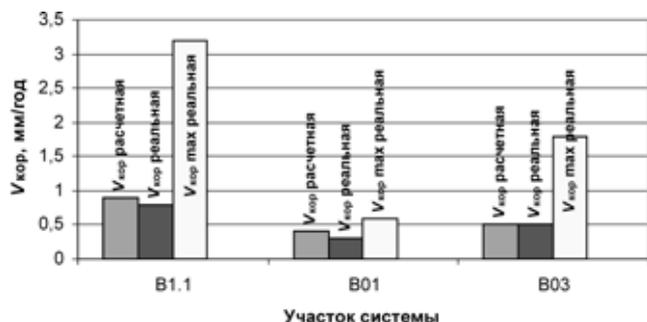


Рис. 2. Расчетные и реальные значения скоростей коррозии на участках системы ППД

пластового давления (система ППД) одного из месторождений Западной Сибири. Рассмотрены три участка системы, условно обозначенные B1.1, B01, B03, различающиеся физико-химическими параметрами среды. Получены расчетные значения скорости коррозии по результатам факторного анализа. При этом известны максимальные реальные значения скорости коррозии на исследуемых участках. Результаты приведены в таблице и на рис. 2.

Разработанная математическая модель расчета скорости коррозии стали и создаваемая на ее основе информационная система поддержки принятия решений в задачах коррозионной защиты позволяют своевременно получать важные сведе-

ния о коррозионной ситуации на конкретном участке трубопроводной системы нефтепромыслового объекта. Для их использования необходимо знать только определенный набор физико-химических показателей, характеризующих протекающую в объекте среду, что позволит предупредить отказы, нередко связанные со значительными выбросами углеводородного сырья, наносящими огромный экономический и экологический ущерб.

Список литературы

1. Тимонин В. А. Применение экспертных систем в противокоррозионной технике (обзор) // Защита металлов. 1992. Т. 28. № 1. С. 3—30.
2. Моисеева Л. С., Задко И. И., Чалых А. Е. Как выбрать реагент? // Нефть России. 2002. № 2. С. 54—57.
3. Моисеева Л. С., Загайтов А. И., Чалых А. Е. Экспертно-поисковая система "Реагенты для процессов добычи и переработки нефти и газа" // Коррозия: материалы, защита. 2004. № 10. С. 9—23.
4. Моисеева Л. С., Кузнецов Ю. И. Углекислотная коррозия нефтегазопромыслового оборудования и вопросы ее ингибирования // Защита металлов. 1996. Т. 32. № 6. С. 561—568.
5. Моисеева Л. С., Кузнецов Ю. И. Влияние ингибиторов коррозии на пассивацию анодно-поляризованной низкоуглеродистой стали в карбонат-бикарбонатных буферных средах // Журнал прикладной химии. 1998. Т. 71. Вып. 6. С. 950—955.
6. Моисеева Л. С., Куксина О. Д. Прогнозирование коррозионной агрессивности сред нефтяных и газовых скважин, содержащих CO_2 // Химическое и нефтегазовое машиностроение. 2000. №5. С. 42—45.



ПОЗДРАВЛЯЕМ ЮБИЛАРА!

Заведующему кафедрой нейрокомпьютеров Московского физико-технического института (ТУ), руководителю Российской секции IEEE CIS, члену редколлегии журнала "Информационные технологии", лауреату премии Правительства, известному специалисту в области современных информационных технологий, доктору технических наук, профессору

Александр Иванович Галушкин исполнилось 70 лет

А. И. Галушкин является одним из пионеров нейросетевых технологий, управления и обработки информации, основателем одной из ведущих отечественных школ в области нейроинформатики. Под руководством А. И. Галушкина в СССР активно продвигалась и реализовывалась идеология унифицированных систем передачи данных и унифицированных локальных сетей персональных ЭВМ, а также сформировалось новое направление отечественной вычислительной техники — транспьютерные системы. Он возглавил ра-

боты по теории нейронных сетей, нейроматематике, проектированию нейрочипов, созданию и применению нейрокомпьютеров, а также работы по созданию первых образцов отечественного суперкомпьютера.

Талантливый ученый и организатор науки, высококвалифицированный специалист, чуткий, отзывчивый и обаятельный человек, Александр Иванович пользуется глубоким уважением учеников, коллег, товарищей.

*Сердечно поздравляем Александра Ивановича с 70-летием,
желаем ему крепкого здоровья и новых творческих успехов!*

*Редакционная коллегия и редакция
журнала "Информационные технологии"*



Поздравляем Юбилера!



Директору Института проблем проектирования в микроэлектронике (ИППМ) РАН, руководителю учебного центра МИЭТ при ИППМ РАН, члену Бюро Отделения нанотехнологий и информационных технологий РАН, лауреату Государственной премии РФ в области науки и техники, члену редколлегии журнала "Информационные технологии", академику РАН, доктору технических наук, профессору

Александр Леонидовичу

СТЕМПКОВСКОМУ

исполнилось **60 лет**

Александр Леонидович Стемпковский — видный ученый, крупный организатор, создатель одной из ведущих отечественных научных школ в области автоматизации проектирования микро- и нанoeлектронных устройств.

Академик А. Л. Стемпковский внес огромный вклад в развитие методов создания микроэлектронной аппаратуры и оптимального проектирования цифровых СБИС нанометрового диапазона, а также в разработку новых архитектур высокопроизводительных параллельных вычислительных систем для решения различных классов научно-технических задач. Он автор многочисленных трудов, изобретений, монографий и учебных пособий.

Александр Леонидович является заместителем председателя Экспертного совета ВАК Минобрнауки России, председателем докторского диссертационного совета при ИППМ РАН и членом докторского диссертационного совета при МИЭТ, членом редколлегий ряда научных журналов, инициатором проведения и председателем организационного комитета регулярных Всероссийских научно-технических конференций "Проблемы разработки перспективных микро- и нанoeлектронных систем". Представляет Россию в технических органах Международной федерации по обработке информации (IFIP-International Federation for Information Processing). Он — один из создателей и член правления крупнейшей европейской конференции по автоматизации проектирования и тестированию электронных схем и систем — DATE (Design, Automation and Test in Europe).

Талантливый ученый и организатор науки, высококвалифицированный специалист, чуткий отзывчивый и обаятельный человек, А. Л. Стемпковский пользуется глубоким уважением учеников, коллег, товарищей.

Редакционный совет, редакционная коллегия и коллектив редакции журнала сердечно поздравляют Александра Леонидовича Стемпковского с юбилеем и желают ему здоровья, благополучия и дальнейших творческих успехов на благо российской науки и техники.



CONTENTS

Balakirev N. E., Dobryshina E. V. *The Conceptual Model and Structure of the Text Information Processing System* 2

Problems of formalization and automatic text processing were analysed. The approach to natural language texts analysis which takes account of the most essential language particularities was defined. The conceptual model and structure of text processing system were offered.

Keywords: automatic text analysis, system of analysis, processing of natural language text, presentation of knowledge, formalization of natural language.

Golomazov D. D. *Term Extraction from a Collection of Documents Divided into Rubrics* 8

A new algorithm called Brainsterm is presented in this paper. It is intended to solve the term extraction task in the case when documents are divided into rubrics and terms consist of two words. The algorithm only uses statistical properties of words. To measure the efficiency of the algorithm within the vector model of document representation the documents have been mapped into points and then classification of the points has been performed. For the purpose of comparison the same procedure has been done using the popular algorithms in the field, TF-IDF and LSI. The Brainsterm algorithm has shown high classification accuracy (up to 85 %) exceeding the rates yielded by the other algorithms. Due to the results of the comparison the Brainsterm algorithm has proved to be of considerable practical importance.

Keywords: term extraction, automatic classification, clustering, vector space model, collections of documents, automatic processing, significance.

Shapkin P. A. *Models and Methods for the Development of Web-Applications Based on Domain Ontologies* 13

The article describes models and methods suitable for processing of information that is represented in the form of RDF graphs along with the domain ontology. Models are drawn up in a way similar with the processing of XML data and in compliance with the principles of the object-oriented programming. Architecture of the system based on the described approach is presented.

Keywords: ontologies, web programming, description logics.

Perevarukha A. Yu. *New Method for Computer Modeling of Chaotic Transient* 18

Author discusses issue of route to chaos and crises of chaotic attractors in models of biology systems. New method of population dynamics modeling, based on hybrid automaton technique was suggested. Elaborated model was investigated with application of computing environment, which allows studying of nonlinear phenomena in population dynamics. New dynamic system gives an opportunity for modeling of aperiodic fluctuation of population into a basin of transient chaos. Transfer to chaos occurs without period doubling cascade.

Keywords: simulation of population dynamic, nonlinear dynamic systems, chaotic transient.

Karpenko A. P., Seliverstov E. Yu. *Global Optimization by the Particle Swarm Optimization Method. Review* 25

This paper presents review of the particle swarm optimization method (PSO) for a global optimization problem and its certain serial and parallel modifications.

Keywords: global optimization, particle swarm optimization method, PSO method.

Limanova N. I., Mamzin E. A. <i>Discrete Mathematical Model of the Determined Cell Automatic Machine and its Practical Application</i>	34
---	----

The mathematical model of the determined hexagonal cell automatic machine and its program realization are suggested. The elaborated automatic machine peculiarity consists in step by step change of combined elements configurations. These elements are next to current cell. According to their conditions the central cell state is determined. The program is realized with the help of C language and OpenGL graphic library. Comparison with analogs shows more than twice increasing of calculations rate. This is achieved by using of parallel programming methods and step by step change of combined elements configuration.

Keywords: cell automatic machine, mathematical model, hexagonal cell, combined elements configuration, calculations rate.

Kovel I. V. <i>Probability Method of Construction of a Neighbourhood of a Statistical Global Optimum of NPC-Problems</i>	38
---	----

The probability method of deriving of the ranked suboptimal solutions of NPC-problems in a neighbourhood of a statistical global optimum is considered. The essence of a method consists in construction of solutions on the basis of use of generators of the pseudorandom numbers working on statistically justified geometrical function of probability of ranks of local solutions.

Keywords: NPC-problem, statistical global optimum, neighbourhood, rank, local optimum, problem of the direct-sales representative, optimization of cutting-packing.

Babak L. I., Dorofeev S. Yu., Peskov M. A., Cherkashin M. V., Sheyerman F. I., Baryshnikov A. S., Koshevoy S. E. <i>Intelligent Automated Design System for Microwave Devices INDESYS</i>	42
--	----

This paper describes INDESYS — intelligent design system for microwave devices synthesis. This system is intended for circuit generation to meet designer requirements with using precise element models. The system structure, synthesis techniques and algorithms are showed.

Keywords: microwaves, RF devices, synthesis, INDESYS, CAD, Visual calculations, genetic algorithms, decomposition, monolithic Microwave Integrated Circuits, UML, API, Locus-MMIC, Microwave Office, morphological tree.

Razumowsky A. I., Sizova L. N. <i>Design and Routing of Printed Circuit Boards Using the Program Complex "DESIGN — R"</i>	48
--	----

The article describes the possibilities of interactive software system "Graphics — TP", as an effective tool, designed for the design of electronic devices, communications equipment, the establishment of circuit documentation in accordance with ESKD, development of drawings and plotting of schemes of algorithms, tables, find and display the shortest path between objects.

Keywords: interactive software package, design, electronic devices, circuit documentation, drawings, principal electrical connections, wiring diagram, printed circuit board, automatic routing connections, the model scheme, the contact elements, list of items.

Nazarov L. E. <i>The Symbol-by-Symbol Decoding of Signals</i>	53
--	----

The description of symbol-by-symbol decoding algorithms has been presented in the article. The base of these developed algorithms is Fast Hadamard Transformation and its modification.

Keywords: signals, symbol-by-symbol decoding, Fast Hadamard Transformation.

Dvornikov S. V., Dvornikov S. S. *Detecting of Signals with Different Ranges of Amplitudes* 56

The detection method of signals with different ranges of amplitudes on the basis of poly-phase decision making algorithm is presented. One of the approaches on optimization of decision making procedure on the basis of maximum probability criterion is considered. Variant of automatic threshold selection without prior system training is proved. The results of a theoretical and practical experiment are shown.

Keywords: detecting of signals, poly-phase algorithm, different ranges of amplitudes, automatic threshold selection, without prior system training.

Savchenko V. V., Akatiev D. Yu. *Deaf-and-Dumb and Hearing Impaired Training of Language Audio System Based on Information Theory of Speech Perception* 60

It is put the task of hearing impaired training of spoken language at the basic phonetic level. Taking into account the distinctive features of this category of learners — their inability to evaluating their pronunciation quality themselves in the learning process — a new approach to the solution of the problem are suggested. This approach is based on the minimum information divergence criterion with visualization of it decision statistics.

Keywords: speech, speech training, information theory of speech perception, the information-theoretic approach, visualization of information error.

Dyundikov E. T., Kachkin A. A. *The Technology of Dynamic Integration and of Representation of Heterogenous Data for the Analysis and the Estimation of State of Multiparametric Objects* 66

The sequence of executing and essence of the unified procedures, ensuring formalised integral representation of results of meeting the given requirements of state of multiparametric objects with use of heterogenous measuring information, received when carrying out the various types of control is examined.

Keywords: the process of dynamic integration and of representation of heterogenous data; chromographical image of state of multiparametric objects of control.

Uglova E. S., Moiseeva L. S. *Mathematical Model and Information System for Support of Decision-Making in Corrosion Protection Problems* 74

The developed mathematical model for calculation of steel corrosion rate in water media for informational system of support and decision-making in problems of corrosion protection for specific type of corrosion (Carbon dioxide corrosion) is considered.

Keywords: system, corrosion, protection, media.

Адрес редакции:

107076, Москва, Стромьинский пер., 4

Телефон редакции журнала **(499) 269-5510**

E-mail: it@novtex.ru

Дизайнер *Т.Н. Погорелова*. Технический редактор *О. А. Ефремова*.

Корректор *Т. В. Зверева*

Сдано в набор 08.12.2009. Подписано в печать 25.01.2010. Формат 60×88 1/8. Бумага офсетная. Печать офсетная.

Усл. печ. л. 9,8. Уч.-изд. л. 11,38. Заказ 103. 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.

Отпечатано в ООО "Подольская Периодика"

142110, Московская обл., г. Подольск, ул. Кирова, 15