

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

9(157)
2009

ТЕОРЕТИЧЕСКИЙ И ПРИКЛАДНОЙ НАУЧНО-ТЕХНИЧЕСКИЙ ЖУРНАЛ

Издается с ноября 1995 г.

УЧРЕДИТЕЛЬ
Издательство "Новые технологии"

СОДЕРЖАНИЕ

БЕЗОПАСНОСТЬ ИНФОРМАЦИИ

Васенин В. А. Критическая энергетическая инфраструктура: кибертеррористическая угроза 2

ОПТИМИЗАЦИЯ И ПРИНЯТИЕ РЕШЕНИЙ

Ратманова И. Д., Коровкин С. Д., Железняк Н. В. Информационная модель топливно-энергетического комплекса как основа анализа энергетической безопасности региона 9
Девятов Д. Х., Файнштейн С. И., Белявский А. Б., Торчинский В. Е. Эвристическая оптимизационная модель для задач с ограничениями, возникающих при оперативном планировании листопрокатного производства 16

МОДЕЛИРОВАНИЕ

Кухаренко Б. Г. Анализ независимых компонент записей колебаний в технологии спектрального анализа на основе быстрого преобразования Фурье 20
Нечаев В. В. Концептуальное модельное представление задачи как системы 26

ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ

Брындин Е. Г. Теоретические аспекты технологии непрерывной обработки на виртуальной памяти 33

ИНФОРМАЦИОННО-ИЗМЕРИТЕЛЬНЫЕ СИСТЕМЫ

Гребнов С. В. Разработка и реализация двухуровневого метода голосового управления на основе скрытых марковских моделей 40
Мальхина Г. Ф., Меркушева А. В. Восстановление формы сигналов по измерительной информации об их смеси. Концепции, модификации алгоритмов и приложения 46
Асадов Х. Г., Чобанзаде И. Г. Информационный метод радиометрической коррекции и калибровки бортовых спектрометров в ходе полета 54

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В СОЦИАЛЬНЫХ И ЭКОЛОГИЧЕСКИХ СИСТЕМАХ

Бождай А. С. Комплексная инфраструктура территории: методы и модели информационного мониторинга 57
Филиппова А. С., Брусиловский Д. П. Прогнозирование в ситуации непостоянного спроса для оптимального управления складскими запасами производственно-сборочных компаний 63

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ОБРАЗОВАНИИ

Штрик А. А. Индикаторы информационной грамотности 67
Пономарев Д. А. Двухэтапный алгоритм обучения речи слабослышащих на основе информационной теории восприятия речи 73
Contents 78
Приложение. Норенков И. П. Содержание и информационная поддержка подготовки специалистов в области информатики и вычислительной техники.

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 620.9:004.056

В. А. Васенин, д-р физ.-мат. наук, проф.,
Институт проблем
информационной безопасности
Московский государственный
университет имени М. В. Ломоносова,
e-mail: vasenin@msu.ru

Критическая энергетическая инфраструктура: кибертеррористическая угроза

Настоящая публикация, первая из двух, посвященных вопросам, связанным с защитой объектов критической энергетической инфраструктуры от деструктивных компьютерных воздействий с кибертеррористическими целями. В ней основное внимание уделено идентификации этой инфраструктуры и составляющих ее объектов, как критически важных для государства, систематизации угроз и способов их реализации. Данную предметную область изучают с использованием методов информатики, подходов к обеспечению информационной безопасности больших систем, и содержание публикации отражает результаты этих исследований. Для того чтобы сделать изложение понятным для более широкого круга заинтересованных лиц, многие специальные технические термины опущены или заменены на более широко употребляемые.*

Ключевые слова: энергетическая инфраструктура, топливно-энергетический комплекс, гидроэнергетика, атомная энергетика, технологический процесс, автоматизированная система управления, критически важный объект, деструктивное воздействие, информационная безопасность, классификация угроз.

Что такое энергетическая инфраструктура?

Прежде чем переходить к выяснению, систематизации и анализу угроз энергетической инфраструктуре на разных уровнях ее организации, необходимо определиться с тем, что будем иметь в виду, употребляя этот термин. Под энергетической инфраструктурой будем понимать совокупность составляющих ее элементов — объектов, включая здания и сооружения, технические средства и технологии; персонал, обслуживающий их, в целях получения энергоресурсов и их предоставления пользователям. *Энергоресурсами* будем име-

* Следующая статья, продолжающая эту тему, планируется к публикации в следующем номере.

новать продукты первичной переработки сырья, полученного в результате эксплуатации нефтяных и газовых месторождений, а также генерации электрической и тепловой энергии на тепловых, атомных станциях и гидроэлектростанциях. К числу задач, определенных назначением отдельных секторов энергетической инфраструктуры национального хозяйственного комплекса любой страны, относится организация производственно-хозяйственной и финансово-экономической деятельности, направленных на добычу (производство) и переработку, хранение и транспортировку:

- нефтегазовых (топливно-энергетических) ресурсов;
- ресурсов электроэнергетики;
- ресурсов атомной энергетики.

В некоторых странах, которые именуются добывающими, может решаться весь комплекс перечисленных задач, включая добычу ресурсов и их первичную переработку. В других странах решается только часть таких задач, например, переработка, хранение и транспортировка. В разных странах эти задачи близки по целям, ради достижения которых энергетические инфраструктуры создаются. Они похожи по способам их решения и, как следствие, по проблемам, которые при этом возникают. Однако в каждой из стран, имеющей такие инфраструктуры, целевые установки на организацию их деятельности в той или иной форме диктует, а условия и способы их достижения, как правило, контролирует, государство в лице уполномоченных на эту деятельность структур и отдельных лиц. Причина подобного внимания к энергетической инфраструктуре в том, что в отличие от целого ряда других инфраструктур национального хозяйственного комплекса, эта инфраструктура является критически важной для государства, а основные (ключевые) системообразующие ее объекты относятся к категории критически важных.

К числу критически важных, как и в работе [1], будем относить следующие инфраструктуры и их элементы, которые в той или иной мере упоминаются в данной статье.

Критически важная инфраструктура — набор взаимодействующих сегментов и составляющих их (входящих в их состав) объектов национального хозяйственного комплекса, поддерживающих сферы жизнедеятельности, частичная деградация и полная потеря функциональности которых способна прямо и в течение относительно короткого

интервала времени влиять на состояние тех или иных составляющих национальной безопасности, приводить к чрезвычайным ситуациям определенного уровня и масштаба.

Критически важная информационная инфраструктура — совокупность инфраструктурных элементов, включая информационно-вычислительные и коммуникационные ресурсы, обеспечивающие управление объектами критически важных инфраструктур.

Критически важные составляющие национальной информационно-телекоммуникационной инфраструктуры — такие элементы (сегменты, объекты) национальной информационно-телекоммуникационной инфраструктуры, которые обеспечивают управление отдельными частями (компонентами) критически важных для государства инфраструктур.

Критически важный объект — объект критически важной инфраструктуры, который в случае частичной деградации или полной потери функциональности способен прямо и в течение относительно короткого интервала времени влиять на состояние национальной безопасности, тех или иных ее составляющих, приводить к чрезвычайным ситуациям определенного масштаба и уровня.

Критически важный сегмент — совокупность критически важных объектов, объединенных по одному или нескольким квалификационным признакам таким, например, как использование в одном технологическом процессе, ведомственная принадлежность, требования к безопасности поддерживаемой ими информации и другие.

Экономическая эффективность и, как следствие, безопасность страны на международном уровне определяются эффективностью работы отдельных инфраструктур, поддерживающих те или иные секторы национального хозяйственного комплекса. К их числу можно отнести, например, различные отрасли промышленности, транспортную сферу, науку и медицину. Эффективность в значительной степени зависит от обеспеченности этих инфраструктур энергоресурсами. В этом смысле наличие необходимых для их функционирования энергоресурсов определяет одну из важнейших составляющих национальной безопасности — безопасность энергетическую.

В мире годами сложились взаимные обязательства стран в области приобретения и поставки тех или иных видов энергоресурсов. Объемы и сроки таких поставок рассчитаны на запланированные в этих странах объемы их потребления различными отраслями национального хозяйственного комплекса. Любое нарушение этих обязательств, вызванное теми или иными деструктивными факторами (обстоятельствами), способно изменить сложившееся положение. В таком контексте (смыс-

ле, понимании) энергетическую безопасность можно и нужно рассматривать как категорию на мировом, транснациональном уровне. Сложившуюся совокупность взаимодействующих на условиях упомянутых поставок национальных энергетических инфраструктур естественно интерпретировать и именовать **мировой (транснациональной) энергетической инфраструктурой**.

Основные принципы, архитектурные и технологические особенности организации деловых процессов

Исследования, направленные на поиск уязвимостей и средств защиты подконтрольного объекта от деструктивных информационных воздействий, как правило, должны начинаться с анализа среды его окружения [2—7]. В целом, для критически важных объектов это сложное, многопараметрическое понятие, требующее строгого системного анализа. К числу основных сущностей такой среды относятся:

- отдельные составляющие элементы объекта защиты (информационные активы, вычислительные или коммуникационные ресурсы) их архитектурно-технологические особенности;
- регламенты (электронные) их сопровождения и реализующий эти регламенты персонал;
- субъекты потенциальных деструктивных воздействий и возможные способы их реализации.

Несмотря на определенные различия, основные принципы организации добычи (получения, производства) и первичной переработки, хранения и транспортировки энергоресурсов в нефтегазовом, электроэнергетическом, атомно-энергетическом секторах одинаковы. Этот факт объясняется тем, что добыча энергоресурсов осуществляется в отдельных локально ограниченных местах, распределенных по территории России. Эти места определяются месторождениями нефтегазовых ресурсов, нахождением тепловых электростанций и гидротехнических сооружений, расположением атомных станций.

Транспортировка и хранение энергоресурсов в России связаны с необходимостью их эффективной доставки до потенциального потребителя, в том числе — зарубежного. Этим обстоятельством обусловлена необходимость создания большой, высокоразвитой, разветвленной сети транспортировки энергоресурсов. В узлах сети для перекачки нефти и газа строятся мощные, поддерживающие надлежащие уровни давления компрессорные станции, станции перекачки и другие технологические объекты, сопровождающие процессы транспортировки и хранения ресурсов [8—10]. Отличие электроэнергетического и атомного комплексов от нефтегазового связано с отсутствием

необходимости в аккумулировании и хранении больших объемов ресурсов. Однако организация их транспортировки и оперативной раздачи потребителям также требует построения и обслуживания большой, разветвленной транспортной сети. Для этого создаются высоковольтные сети передачи электроэнергии.

В современных условиях должный уровень организации всех взаимосвязанных друг с другом процессов добычи и первичной переработки сырья, хранения, транспортировки и доставки энергоресурсов потребителям, которые образуют непрерывный технологический цикл, возможен при условии их эффективного административно-хозяйственного и финансово-экономического сопровождения. Такое сопровождение, в свою очередь, возможно только на основе четкой регламентации (в том числе с использованием электронных регламентов) всех процессов непрерывного технологического цикла и создания компьютерных систем, обеспечивающих автоматизированную поддержку этих регламентов.

С позиций административно-хозяйственных и финансово-экономических перечисленный комплекс мер в России обеспечивается крупными хозяйствующими субъектами негосударственной формы собственности, однако со значительной долей участия в них государства. К их числу относятся ОАО "ГАЗПРОМ", ОАО "Лукойл", ОАО "Роснефть", ОАО РАО "ЕЭС России", а также государственное предприятие Концерн "Росэнергоатом". Активное участие государства в этих структурах объясняется, во-первых, необходимостью оперативного контроля за процессами, происходящими в этих критически важных составляющих национального хозяйственного комплекса, а во-вторых, возможностью таким образом влиять на них, используя ресурсы, которыми обладает государство.

Системы управления единым технологическим циклом в различных секторах энергетического комплекса страны на сегодня находятся в стадии активного формирования. В структурах (корпорациях), работающих в нефтегазовой отрасли [8–10], поэтапно создаются новые, пополняются и совершенствуются уже существующие регламенты поддержки единого технологического цикла, реализующие их бизнес-процессы, инструментальные средства и системы. Созданы и функционируют иерархически организованные информационные корпоративные системы управления технологическими процессами и объектами. Такие системы на нижележащих уровнях архитектурной иерархии включают автоматизированные системы управления предприятиями, а те, в свою очередь, системы управления линейными производственными участками и конечным оборудованием.

К технологическим процессам, подлежащим исполнению в цикле добычи, переработки, хранения и транспортировки нефтяных и газовых ресурсов, с полным правом относится и контроль за состоянием среды, окружающей объекты отрасли. Завершается первый этап работ по созданию автоматизированных систем корпоративного управления финансово-экономической и административно-хозяйственной деятельностью.

Аналогичные системы управления технологическими процессами генерации, транспортировки и сбыта энергоресурсов тепловых и гидроэлектростанций, работающих в составе единой энергетической системы страны, построены и в течение ряда лет (до 01.07.2008 года) эксплуатировались ОАО РАО "ЕЭС России" [11]. Так же, как и в нефтегазовой отрасли, здесь созданы и эксплуатируются: система управления процессами в рамках единого технологического цикла получения электроэнергии с использованием тепловых и гидроэлектростанций, ее транспортировки и сбыта; системы корпоративного управления административно-хозяйственной и финансово-экономической деятельностью. В течение последних семи лет последовательно реализовывался план реформирования ОАО РАО "ЕЭС России", основной этап которого завершился 1 июля 2008 года. Не останавливаясь на нормативно-правовой базе реформирования ОАО РАО "ЕЭС России" [11, 12], монополюльно представлявшего сектор электроэнергетики в энергетической инфраструктуре страны, отметим основные идеи этого процесса.

В целях повышения эффективности работы предприятий электроэнергетики, обеспечения их надежности и бесперебойного электро- и теплоснабжения потребителей, для совершенствования системы корпоративного управления осуществляется разделение функций генерирования (производства) энергоресурсов, оперативно-диспетчерское управление их распределением и сбытом. Вместо прежних вертикально интегрированных компаний в составе ОАО РАО "ЕЭС России" создаются структуры (компании, холдинги), специализирующиеся на отдельных видах деятельности. Менее конкурентоспособные (привлекательные для инвестиций) на настоящее время генерирующие, сбытовые и ремонтные компании становятся преимущественно частными и начинают конкурировать друг с другом. В естественно-монопольных видах деятельности, таких как оперативно-диспетчерское управление распределением ресурсов, происходит усиление государственного контроля. Таким образом, создаются условия для создания и развития конкурентоспособного рынка электроэнергии. Однако и это имеет определяющее значение в контексте целей настоящей публикации, после реформы в техническом отно-

шении информационно-телекоммуникационная инфраструктура автоматизации единого технологического цикла генерации, транспортировки и сбыта электроэнергии в стране в целом не изменилась. Изменяются условия и способы ее эксплуатации, которые в меньшей степени влияют на вопросы защиты объектов электроэнергетики от деструктивных информационных воздействий.

В меньших масштабах Концерном "Росэнергоатом" [13] ведется реформирование и поэтапная модернизация в сфере атомной энергетики. Разрабатываются регламенты (в том числе, электронные), реализующие их бизнес-процессы в рамках единого технологического цикла получения энергоресурсов с использованием атомных станций, их транспортировки до потребителя и сбыта. Идет поиск адекватных средств информатизации этих процессов, разработка эффективных структур автоматизированного управления технологическими процессами, административно-хозяйственной и финансово-экономической деятельностью. На этом направлении используется положительный опыт, полученный в смежных секторах энергетики — нефтегазовом и электроэнергетическом.

С учетом изложенных выше соображений можно прийти к выводу, что условия эксплуатации объектов всех трех рассматриваемых секторов энергетического комплекса страны и способы управления ими близки. С этих позиций их можно отнести к объектам одного класса.

Используя корпоративные и региональные сети передачи данных, в каждом из секторов национальной энергетической инфраструктуры осуществляется централизованное управление со стороны каждой из корпораций технологическим циклом, который находится в ее ведении, путем координации деятельности отдельных составляющих ее предприятий. Основную нагрузку управления производственными процессами при этом несут на себе отдельные предприятия, обеспечивающие технологический цикл добычи (получения) сырья, его первичной переработки, хранения и транспортировки энергоресурсов. Программно-технические средства, автоматизирующие производственные процессы, формируются при этом на основе открытых стандартов. Такой подход позволяет обеспечить возможность подключения новых объектов и модернизацию уже эксплуатирующихся без существенной модификации базового аппаратно-программного обеспечения.

Подходы к выявлению угроз объектам критически важных инфраструктур, к определению способов их реализации и к разработке методов противодействия в значительной степени определяются их архитектурно-технологическими особенностями. По этой причине особую значимость приобре-

тают вопросы классификации (кластеризации) множества таких объектов. Анализ архитектурно-технологических особенностей объектов нефтегазового, электроэнергетического комплексов и атомной энергетики показывает, что их можно отнести к объектам одного класса. В качестве систем для оперативного контроля за состоянием основных технологических процессов в автоматизированных системах всех секторов энергетического комплекса используются SCADA-системы [14, 15]. Для управления механизмами на нижнем уровне реализации технологических процессов, как правило, используются те или иные типы контроллеров, которые представлены на мировом IT-рынке и хорошо зарекомендовали себя на практике. Спектр информационных систем, используемых для управления административно-хозяйственной и финансово-экономической деятельностью, значительно шире. На этом направлении используются как информационные системы средней производительности, в том числе отечественные, так и сложно организованные, ресурсоемкие высокопроизводительные системы уровня R-3 SAP.

Объектами деструктивных информационных воздействий на критическую энергетическую инфраструктуру являются:

- автоматизированные системы управления технологическими процессами на нижнем уровне реализации и их компоненты (серверы, в первую очередь, серверы SCADA, автоматизированные рабочие места, микропроцессорные контроллеры, средства телемеханики);
- информационно-телекоммуникационные сети, поддерживающие автоматизированные системы управления технологическими процессами;
- объекты информатизации процессов добычи (получения), переработки (преобразования) и транспортировки энергоресурсов (объекты, поддерживающие компрессорные системы, перекачку топлива, электроснабжение и подобные им).

Угрозы и уязвимости. Терроризм и кибертерроризм

Угрозы деструктивных информационных воздействий на объекты критической энергетической инфраструктуры могут исходить:

- от отдельных преступников или криминальных групп, преследующих цели, направленные против интересов отдельных предприятий в рамках корпорации или против корпорации в целом;
- от террористических групп, преследующих цели дестабилизации социально-политического или экономического положения в стране, соз-

дания чрезвычайной ситуации национального уровня и масштаба.

Под **терроризмом**, следуя [1, 16, 17], будем понимать проявление крайнего экстремизма в действиях, основанных на разногласиях (национальных, транснациональных) отдельных групп лиц с государственными интересами и институтами (в политике, в социальной сфере, на религиозной и криминальной почве) и направленных на создание в обществе атмосферы страха и напряженности, на формирование факторов, прямо или косвенно дестабилизирующих состояние национальной безопасности, в целях выдвижения к властным структурам требований, которые не могут быть удовлетворены в рамках существующего правового поля.

В принятом контексте **кибертерроризм** — одно из направлений терроризма, которое в качестве объектов деструктивного информационного воздействия для достижения своих целей использует информационно-вычислительные комплексы и сетевые сегменты, поддерживающие системы, критически важные с точки зрения национальной безопасности, и в качестве предмета воздействия использует средства вычислительной техники и их программное обеспечение.

Кибертеррористический акт — террористический акт, реализуемый с помощью компьютерных средств, применение которых непосредственно повлекло или потенциально могло повлечь за собой угрозу жизни или здоровью людей, значительный по масштабам и уровню ущерб материальным объектам, а также другие последствия, влияющие на национальную безопасность.

Под **деструктивным информационным воздействием** на автоматизированную систему управления технологическим процессом (АСУ ТП) понимается несанкционированное информационное воздействие на отдельные ее информационные активы, вычислительные и телекоммуникационные ресурсы. Такие воздействия приводят к нарушению штатных (допустимых регламентом) режимов функционирования системы в результате нарушения или полного разрушения поддерживающей ее информационно-технологической инфраструктуры. Деструктивные воздействия при этом могут быть направлены на отдельные компоненты АСУ ТП, такие как сервер SCADA, архивный сервер, АРМ диспетчера, АРМ специалиста, локальная система автоматического управления, контроллер, устройство связи с объектом и другие, включая информационные активы, которые такие компоненты поддерживают.

Деструктивные информационные воздействия могут быть направлены на достижение следующих целей:

- нарушение конфиденциальности информации АСУ ТП;

- нарушение целостности информации АСУ ТП (несанкционированная модификация информации);
- нарушение работоспособности (доступности) информационных ресурсов АСУ ТП.

Нарушение конфиденциальности информации, которая хранится либо обрабатывается АСУ ТП, подразумевает раскрытие информации лицами, не имеющими прав доступа к ней. Нарушением целостности информации является несанкционированная модификация информации. Примером деструктивного воздействия, целью которого является нарушение конфиденциальности и целостности информации, может служить типовая удаленная атака, в ходе которой создается "ложный" объект АСУ ТП.

Принципиально другой целью деструктивного информационного воздействия является нарушение работоспособности (доступности) ресурсов АСУ ТП. В этом случае не предполагается получение атакующим несанкционированного доступа к информации. Его основная цель — добиться, чтобы ресурс на объекте, который подвергается атаке, стал недоступен для всех остальных элементов АСУ ТП и, как следствие, доступ к информационным активам этого объекта, а также к управляемым им технологическим устройствам был бы невозможен. Следует заметить, что нарушение работоспособности системы может быть также достигнуто методами физического или иного воздействия на аппаратуру и программное обеспечение АСУ ТП.

При реализации угроз нарушения конфиденциальности данных, как правило, используются следующие типы уязвимостей АСУ ТП:

- ошибки в реализации механизмов операционных систем, обеспечивающих разграничение доступа к данным;
- отсутствие необходимой физической защиты каналов передачи данных;
- уязвимости сред передачи данных, позволяющие осуществить несанкционированное подключение к каналу передачи данных;
- уязвимости управляющих сетевых протоколов, позволяющие осуществить перенаправление пакетов на другой хост локальной вычислительной сети;
- отсутствие стойких методов шифрования передаваемых данных.

В случае реализации угроз нарушения целостности в дополнение к перечисленным выше необходимо наличие уязвимости соответствующих сетевых протоколов. Такие уязвимости позволят злоумышленнику изменять данные по своему усмотрению.

К числу уязвимостей, которые, как правило, используются при реализации атак на отказ в обслуживании, относятся:

- ошибки в реализации механизмов операционных систем, обеспечивающих разграничение доступа к данным;
- отсутствие необходимой физической защиты каналов передачи данных;
- уязвимости сред передачи данных к возможным помехам.

Реализация атаки возможна в случае отсутствия оперативных способов противодействия возможным распределенным DDoS-атакам (*Distributed Denial of Service*). Такие способы, как правило, предоставляются системами мониторинга состояния инфокоммуникационной среды на предмет обнаружения деструктивных воздействий и оперативного реагирования на них.

Важнейшим элементом, обеспечивающим информационное взаимодействие в рамках АСУ ТП, является информационно-телекоммуникационная (сетевая) инфраструктура. Сетевую инфраструктуру АСУ ТП удобно рассматривать на трех уровнях: коммуникационном, уровне приложений, уровне устройств сопряжения с технологическими объектами.

Коммуникационный уровень включает в себя каналы передачи данных, набор необходимого коммуникационного оборудования, систему управления и мониторинга состояния коммуникационного оборудования (включая программное обеспечение, поддерживающее работу управляющих сетевых протоколов и сервисов автоматизированной системы) на предмет работоспособности и/или наличия деструктивных воздействий.

Уровень приложений (сетевых) включает программное обеспечение, функционирующее в АСУ ТП поверх управляющих протоколов (на более высоких уровнях модели OSI) коммуникационного уровня (например, Web-серверы, серверы электронной почты, СУБД или иные хранилища данных, системы электронного документооборота).

Уровень устройств сопряжения с технологическими объектами включает датчики, контроллеры и иное управляющее оборудование, устанавливаемое непосредственно на технологическом объекте и являющееся неотъемлемой частью этого объекта.

Рассмотрим основные классы уязвимостей элементов сетевой инфраструктуры АСУ ТП и связанные с ними классы деструктивных информационных воздействий на эти элементы. Уязвимостью сетевой инфраструктуры АСУ ТП называется свойство ее элементов, которое, в некотором окружении, позволит злоумышленнику нарушить регламентированные политикой безопасности режимы функционирования данной системы. В первую очередь, рассматриваются уязвимости, способные привести к нарушению конфиденциальности или целостности обрабатываемой информации, к потере доступности сервисов, функ-

ционирующих в АСУ ТП, либо к физическому повреждению элементов АСУ ТП. Необходимо отметить, что значительную часть уязвимостей в сетевой инфраструктуре АСУ ТП составляют ошибки реализации программного обеспечения, отвечающего за функционирование ядра операционной системы или пользовательских приложений, которые работают на отдельных хостах автоматизированной системы. Как правило, такие ошибки не зависят от назначения самой программы. В случае критических объектов, к которым относятся системы управления технологическими процессами, производственно-хозяйственной и финансово-экономической деятельностью в сфере национального энергетического комплекса, к числу перечисленных выше добавляются еще две угрозы — потери доверия к такой системе, а также зависимости от поставщика средств вычислительной техники.

Угроза потери доверия к системе обусловлена потенциальной возможностью ее компрометации перед пользователями в связи с поступающей к ним (объективной или намеренно искаженной) информацией о частых отказах, о недостатках качества реализуемых функций и других подобных фактах. Повышению уровня доверия к системе, как правило, способствует выполнение ее разработчиками, создателями и теми, кто ее эксплуатирует, положений всех нормативно-регулирующих документов [2—7] в области обеспечения безопасности ее информационных технологий, а также нормативное информирование пользователей о мерах, которые предусмотрены и предпринимаются для поддержания ее работоспособности и качества функционирования [18].

Угроза зависимости от поставщика средств вычислительной техники обусловлена возможностью отказа, вызванными теми или иными причинами, от поставки таких средств, которые заказчик не может получить по другим каналам. Устранение этой угрозы заключается в правильно выстраиваемой политике взаимодействия заказчиков с поставщиками средств вычислительной техники, импортозамещения тех средств вычислительной техники, которые являются ключевыми для критически важных информационно-телекоммуникационных систем.

Необходимость консолидации действий

Анализ угроз деструктивных информационных воздействий с террористическими целями на производственные объекты энергетической инфраструктуры позволяет сделать следующие выводы.

1. К числу вероятных сценариев кибертеррористической атаки на объекты энергетической инфраструктуры следует отнести те, которые обеспечивают не только временную или полную поте-

рю их функциональности, но и, как следствие (вторичный эффект), создают широкомасштабную чрезвычайную ситуацию с высоким уровнем потерь (материальных, человеческих и других) и/или угрозу национальной безопасности.

2. Реализация таких сценариев с высокой вероятностью будет осуществляться группой координирующих свои действия агентов из разных точек сетевой среды, которые обеспечивают доступ к ресурсам объекта атаки, в том числе, из точек, расположенных вне страны, объекты которой подвергаются атаке.

3. Наиболее значимыми с позиции величины потенциального ущерба (социального, материального, политического) являются:

- сложно предотвращаемые в оперативном порядке распределенные атаки на отказ в обслуживании (DDoS);
- комплексные атаки, результатом которых является получение контроля над управляемым производственным объектом или важными технологическими процессами, обеспечивающими его функционирование по назначению.

Организация эффективного противодействия перечисленным атакам требует детального исследования условий, которые оказывают влияние на их подготовку и проведение. Необходим учет всех аспектов киберпротивоборства — от побудительных мотивов субъектов атаки до программно-технических возможностей средств противодействия. В этой связи представляется, что в настоящее время необходим переход от преимущественно вербального описания кибертеррористического противоборства на политическом уровне к его формально строгому описанию и изучению на уровне научно-техническом и технологическом. Отличительной особенностью каждой из представленных выше атак является тот факт, что сценарии их реализации предполагают многоэтапную подготовку и скоординированный характер действий группы агентов, распределенных по сетевой среде. Это обстоятельство указывает на необходимость детального изучения подобных сценариев, требует разработки средств противодействия им на всех уровнях обеспечения информационной безопасности критически важных объектов. Эффективное использование таких знаний и средств противодействия возможно только на основе скоординированных действий всех организаций, задействованных на обслуживании информационно-вычислительных систем и средств телекоммуникаций, которые могут быть использованы для проведения кибертеррористических атак. Такая консолидация необходима на всех этапах — от анализа возможности организации тех или иных атак, сценариев их реализации, мониторинга состояния инфокоммуникационной

инфраструктуры на предмет наличия деструктивных воздействий до совместных действий на этапе выработки мер оперативного реагирования и средств противодействия.

Принимая во внимание транснациональный характер сетевой среды и энергетической инфраструктуры, который эта среда поддерживает, большую роль в создании эффективной системы противодействия кибертеррористической угрозе призвана сыграть консолидация сил на международном уровне.

Список литературы

1. **Критически** важные объекты и кибертерроризм. Часть 1. Системный подход к организации противодействия // О. О. Андреев и др. Под ред. В. А. Васенина. М.: МЦНМО, 2008. 398 с.
2. **Средства** вычислительной техники. Защита от несанкционированного доступа к информации. Показатели защищенности от несанкционированного доступа к информации. Руководящий документ ФСТЭК от 30 марта 1992 года / ФСТЭК. 1992.
3. **Автоматизированные** системы. Защита от несанкционированного доступа к информации. Классификация автоматизированных систем и требования по защите информации. Руководящий документ ФСТЭК от 30 марта 1992 года / ФСТЭК. 1992.
4. **ГОСТ Р ИСО/МЭК 17799—2005**. Информационная технология. Практические правила управления информационной безопасностью. М.: ИПК Издательство стандартов, 2005.
5. **ГОСТ Р ИСО/МЭК 15408-1—2002**. Информационная технология. Методы и средства обеспечения безопасности. Критерии оценки безопасности информационных технологий. Часть 1. Введение и общая модель. М.: ИПК Издательство стандартов, 2002.
6. **ГОСТ Р ИСО/МЭК 15408-2—2002**. Информационная технология. Методы и средства обеспечения безопасности. Критерии оценки безопасности информационных технологий. Часть 2. Функциональные требования безопасности. М.: ИПК Издательство стандартов, 2002.
7. **ГОСТ Р ИСО/МЭК 15408-3—2002**. Информационная технология. Методы и средства обеспечения безопасности. Критерии оценки безопасности информационных технологий. Часть 3. Требования доверия к безопасности. М.: ИПК Издательство стандартов, 2002.
8. **Никоненко И. С., Шуфчук Ю. Б., Майоров М. М.** Пути прогресса. История развития автоматизации газовой промышленности / Под. ред. М. М. Золочевского. М.: ООО "Недра-Бизнесцентр", 2000. 245 с.
9. **Новиков Е.** Принципы организации сетевой инфраструктуры ООО "Лукойл — Информ". URL: <http://www.connect.ru/article.asp?id=8579>
10. URL: <http://www.newsdesk.pcmag.ru/node/6163>.
11. **Реформирование:** основные итоги на 1 июня 2008 года ОАО РАО "ЕЭС России" // Информационный бюллетень. С. 4—14.
12. **IT** — будущее РАО ЕЭС. www.cnews.ru
13. URL: <http://www.rosenergoatom.ru>.
14. **ClearScada** — инструмент объектного проектирования систем управления технологическими процессами. ПЛК Системс. URL: <http://www.asutp.ru/?p=600635>
15. **Кутевич Н. А.** От SCADA-систем к SCADA-продуктам и MES-компонентам // Мир компьютерной автоматизации. 2003. № 4.
16. **Васенин В. А.** Информационная безопасность и компьютерный терроризм // Сб. "Научные и методические проблемы информационной безопасности" / Под. ред. В. П. Шерстюка. М.: МЦНМО, 2002. С. 67—83.
17. **Васенин В. А.** Научные проблемы противодействия кибертерроризму // Материалы международной научной конференции "Проблемы безопасности и противодействия терроризму". Интеллектуальный центр МГУ, 2—3 ноября 2005 г. М.: МЦНМО, 2006. С. 49—63.
18. **Липаев В. В.** Программная инженерия. Методологические основы: учебник. М.: ТЕИС, 2006. 608 с.

ОПТИМИЗАЦИЯ И ПРИНЯТИЕ РЕШЕНИЙ

УДК 004.6:620.9

И. Д. Ратманова, д-р техн. наук,
С. Д. Коровкин, канд. техн. наук,
Н. В. Железняк, аспирант,
e-mail: natasha@osi.ispu.ru,
Ивановский государственный
энергетический университет

Информационная модель топливно-энергетического комплекса как основа анализа энергетической безопасности региона

Излагается подход к построению информационной модели топливно-энергетического комплекса в целях организации информационной поддержки региональной энергетической политики. В основу модели положена метауровневая архитектура хранилища данных, разработанная с учетом особенностей исследуемой предметной области.

Ключевые слова: энергетическая безопасность, топливно-энергетический комплекс (ТЭК), системы поддержки принятия решений (СППР), хранилище данных, метаданные, концептуальная модель, многомерная аналитическая модель, топливно-энергетический баланс (ТЭБ).

Актуальность информационной поддержки принятия решений в энергетической сфере

Происходящие в России экономические реформы и реструктуризация энергетики обострили проблему энергетической безопасности, являющейся одним из ведущих факторов национальной энергетической политики и играющей важную роль при принятии решений по функционированию и развитию топливно-энергетического комплекса (ТЭК) страны и ее регионов. При этом под термином "энергетическая безопасность" (ЭБ) понимается состояние защищенности страны, ее граждан, общества, государства, экономики от угроз надежному топливно- и энергообеспечению [1].

Следует отметить растущую роль регионов в формировании собственной энергетической политики, которая призвана обеспечить необходимые и достаточные условия для социально-экономического развития территории за счет перехода к

энергоэффективным технологиям повышения надежности энергоснабжения потребителей.

Во избежание ошибок в принятии концептуальных решений при формировании оптимальной региональной энергетической политики необходимо осуществлять основанный на количественных оценках многовариантный анализ возможных путей развития электро- и теплоснабжающих систем региона. Этот процесс затрудняется многоуровневой иерархической структурой систем энергетики, их территориальной распределенностью, большой размерностью, а также многоаспектностью решаемых задач, многосвязанностью исследуемых объектов, нечеткостью и неоднозначностью критериев, неполнотой исходной информации. Такое положение наряду с ограниченностью формализованных подходов к управлению обуславливает необходимость и значимость организации информационной поддержки принятия решений на основе мониторинга и оценки состояния в рамках систем поддержки принятия решений (СППР).

Возрастает роль организации мониторинга ЭБ. Состав и структура объектов мониторинга, а также конкретных индикаторов ЭБ должны давать возможность гибкого и взаимосогласованного анализа и оценки ситуации на различных уровнях отраслевой (по видам деятельности ТЭК), территориальной (страна — регион — территория — объект), временной и управленческой иерархии. Принципиально важными задачами при этом являются определение и нормирование пороговых значений индикаторов ЭБ, что позволяет внести количественную определенность в критерии оценок и обоснования решений.

Кроме того, для эффективной реализации комплекса работ по энергосбережению требуется значительное повышение уровня информационного обеспечения всех планируемых мероприятий. В этом случае целесообразно накопление информации по потреблению топливно-энергетических ресурсов (ТЭР), исследование типологии энерго- и ресурсопотребления для определенных групп потребителей, выделение дефектных зон, лимитирование ТЭР в целях обоснованного планирования энергоаудита и энергосберегающих мероприятий, экономии материальных средств.

Вместе с тем, в целях выявления дефектных зон, определения наиболее вероятных причин аварийных ситуаций и их предотвращения необходимо иметь достоверную информацию по показателям надежности электрооборудования. Для

модернизации оборудования и прогнозирования его остаточного ресурса, а также для организации целевого реновационного инвестирования необходим мониторинг и оценка технического состояния и уровня эксплуатации оборудования — технический контроллинг [2].

Следует отметить, что с точки зрения мониторинга показателей ЭБ, проведения эффективной политики энергосбережения, а также тарифной и инвестиционной политики представляет интерес организация ведения топливно-энергетических балансов (ТЭБ) территориальных образований. ТЭБ — это главный инструмент, определяющий порядок функционирования энергетического сектора государства, в частности, структуру, объем добычи ресурсов, их изменение, структуру поставок; он является основой для перспективного прогнозирования спроса и предложения энергоресурсов с учетом оценки тенденций развития ТЭК.

Метауровневая архитектура СППР

В настоящий момент на рынке платформ для построения СППР имеется несколько десятков продуктов, причем преобладают в основном разработки зарубежных фирм. Однако существует ряд проблем, ограничивающих их применение в условиях нашей страны. К ним можно отнести следующие: преимущественная ориентация на задачи управления торгово-промышленными корпорациями; высокая стоимость; преимущественное отсутствие русификации; сложности адаптации к конкретному применению; вопросы национальной безопасности. В связи с этим актуальна разработка информационно-аналитических систем (ИАС), предназначенных для поддержки принятия решений и удовлетворяющих требованиям задач управления и проектирования в крупных отраслевых и региональных системах, и в частности в энергетической сфере.

Современный подход к автоматизации информационной поддержки принятия решений основан на использовании концепции хранилищ данных (ХД) корпоративных ИАС. Целесообразно также применение техники многомерного моделирования агрегированных данных в сочетании с гибкой навигацией по детализированным регистрационным сведениям. Поскольку консолидация больших объемов данных приводит к сложности их обработки, возникает необходимость поместить данные в определенный контекст. Подобная проблема управления данными решается при помощи метаданных.

Представляемая базовая структура метаданных хранилища включает системные метаданные, поддерживающие структуру данных хранилища;



Рис. 1. Метауровневая архитектура хранилища данных

метаданные источников информации, связанные с преобразованием и загрузкой данных; аналитические и навигационные метаданные, поддерживающие доступ и правила представления информации; метаданные управления, поддерживающие функционирование хранилища. При этом в терминах системных метаданных описывается концептуальная модель "сущность — связь". Она, в свою очередь, является основой описания информационно-аналитической модели СППР (аналитических и навигационных метаданных). Информационно-аналитическая модель — основа описания слоя операционных метаданных, поддерживающих обработку информации (рис. 1).

К настоящему времени разработаны стандартные нотации форматов обмена метаданными хранилищ, модели репозитория в рамках соответствующих платформ СППР крупных фирм-разработчиков (базирующихся, как правило, на конкретном сервере баз данных). Вместе с тем, стандартные модели метаданных сильно обобщены, что обедняет семантику модельного представления хранилища. Отработка единой промышленной технологии проектирования, создания и сопровождения метаданных практически невозможна. Вопросы, связанные с организацией метаданных, следует рассматривать индивидуально в рамках реализуемой технологической платформы с учетом особенностей исследуемой сферы применения.

Настоящая статья посвящена описанию информационно-аналитической модели ТЭК региона, которая положена в основу организации метаданных корпоративных ИАС в сфере энергетики.

Концептуальное моделирование предметной области

Как уже отмечалось, системы энергетики — это системы многоотраслевые, территориально

распределенные, включающие в свой состав объекты различных видов деятельности, форм собственности и технологической функциональности. И как любой структурно сложной системе, топливно-энергетическому комплексу присуще свойство иерархичности. Это свойство находит свое отражение и в подсистемах, которые можно выделить в составе энергетического комплекса, — структура отраслей иерархична; территориальные образования различных уровней находятся в отношении подчинения; энергетические предприятия (энергоснабжающие организации, электросетевые компании) состоят из совокупности энергетических объектов (котельных, подстанций), которые, в свою очередь, имеют в своем составе энергетические установки (котлы, трансформаторы, линии электропередач); потребители ТЭР завязаны с определенными энергетическими объектами, т. е. в определенном смысле "подчиняются" им. Следовательно, в структуре ТЭК можно выделить отраслевую, территориальную и управленческую иерархии. Если взглянуть на строение энергетической системы с более обобщенных позиций и свести воедино все перечисленные виды подчиненности, то структуру комплекса можно представить в виде иерархии: отрасль/территория — энергетическое предприятие — энергетический объект — энергетическая установка — потребитель (рис. 2). Эти уровни не изолированы, они тесно взаимодействуют, согласуются и дополняют друг друга. Так, с одной стороны, показатели мониторинга ЭБ, полученные, например, на уровне энергетических установок, могут быть агрегированы до уровня энергетических объектов, предприятий и выше — до уровня территорий и отраслей. С другой стороны, в ходе решения задач ЭБ и выявления дефектных зон требуется спус-

таться вниз по иерархии объектов ТЭК для выявления причин, вызвавших отклонение индикаторов ЭБ от допустимых значений. Важно также учитывать, что для проведения грамотной региональной энергетической политики необязательно выстраивать всю пятиуровневую систему объектов энергетики — зачастую бывает достаточно ограничиться двумя—четырьмя уровнями исходя из исследуемых проблем с учетом специфики ТЭК региона.

Каждый уровень иерархии специфичен с точки зрения постановки задач, подходов к их решению и набора исходных данных. Так, к уровню отраслей и территорий относятся такие задачи, как индикативный анализ и диагностика ЭБ, идентификация угроз, прогнозирование энергопотребления. Для решения этих задач требуется, во-первых, информация о производстве, покупке, структуре поступления и расходования, распределении и остатках различных видов ТЭР в территориальном и отраслевом разрезах; во-вторых, необходимы сведения об определенных социальных и экономических аспектах деятельности этих отраслей и территорий. Первая, техническая, группа показателей в полной мере покрывается сведениями ТЭБ, а вторая, экономическая, статистикой социально-экономического положения в отраслевом и территориальном разрезах.

В свою очередь, на уровне энергетических предприятий ставятся и решаются задачи энергосбережения, прогнозирования потребления ТЭР, анализа балансовой структуры и оптимизации баланса. Решение этих задач основывается преимущественно на сведениях технического характера, которые, как и в предыдущем случае, в достаточной мере покрываются показателями ТЭБ уровня предприятия. Помимо этого используется информация о производственных показателях деятельности, характеризующих объем, сроки и себестоимость выпуска продукции.

Уровню энергетических объектов и установок присущи задачи технического контроллинга, технико-экономического обоснования тарифов и инвестиций. Здесь необходим большой объем технической информации, включая паспортные, нормативные, фактические данные и экспертные оценки, причем набор показателей различен для определенных классов объектов и установок.

На последнем уровне иерархии объектов ТЭК — уровне конечных потребителей — основной задачей является энергоаудит, суть которого заключается в анализе результатов

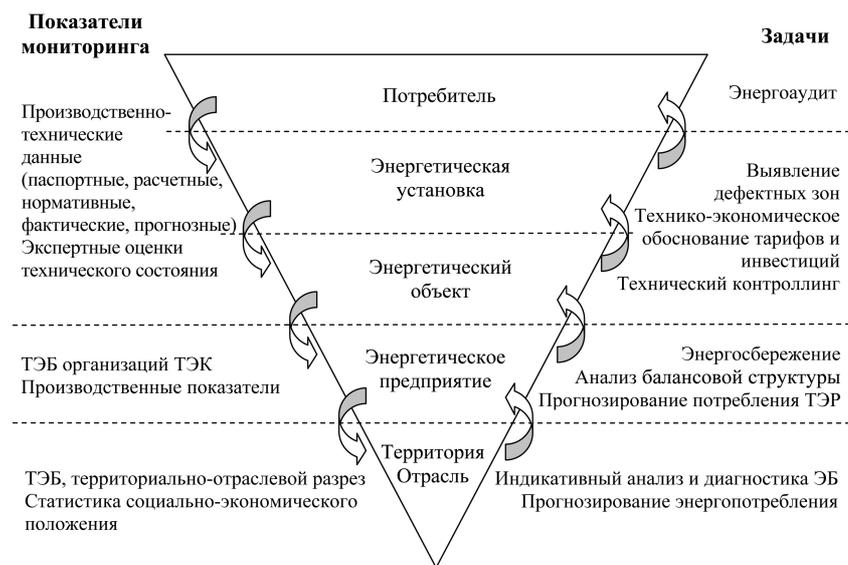


Рис. 2. Иерархия объектов топливно-энергетического комплекса

мониторинга деятельности потребителей в целях определения энергосберегающих мероприятий, выявления типологии потребления ТЭР и установления лимитов. Как и в предыдущем случае, состав показателей мониторинга включает нормативные, паспортные, фактические и оценочные производственно-технические данные.

Таким образом, с позиций организации данных в ориентации на их анализ, множество данных на каждом уровне иерархии объектов ТЭК можно свести к трем типам:

- паспортные (регистрационные) данные объектов ТЭК;
- перечень показателей мониторинга, необходимых для решения поставленных задач; показатели могут образовывать иерархии, группироваться по смысловой нагрузке и характеру решаемой задачи;
- ретроспектива значений показателей мониторинга (фактов) с достаточным уровнем временной детализации (мгновенные и совокупные мгновенные снимки).

Учитывая сказанное, концептуальная модель ТЭК представляет собой множество пересекающихся объектных областей, согласованных посредством общесистемных справочников (рис. 3). Каждая область является совокупностью сущностей хранилища, покрывающих определенный аспект исследуемой системы и используемых при

описании аналитических или навигационных метаданных хранилища, которые составляют основу информационно-аналитической модели СППР.

На основе исследования уровней иерархии ТЭК определены структуры объектных областей, которые ориентированы на решение задач диагностики ЭБ, прогнозирования энергопотребления, анализа балансовой структуры, энергосбережения, технического контроллинга.

Для двух верхних уровней иерархии объектов ТЭК — уровня территорий и отраслей и уровня энергетических предприятий — единым информационным базисом для решения задач ЭБ являются топливно-энергетические балансы соответствующих уровней. Однородный состав показателей мониторинга, используемых для решения задач ЭБ, а также возможность агрегации данных балансов энергетических предприятий до уровня отраслей и территорий позволяют объединить эти уровни иерархии ТЭК в одну объектную область. Рассматриваемая область включает несколько таблиц фактов, каждая из которых соответствует уровню объектов ТЭК и требуемой степени временной детализации данных. В таблицах фактов накапливаются сведения о производстве, покупке, распределении и потреблении первичных и вторичных ТЭР. Следует отметить, что именно область ТЭБ является ключевой с точки зрения решения большинства актуальных задач ЭБ, включая задачу ана-

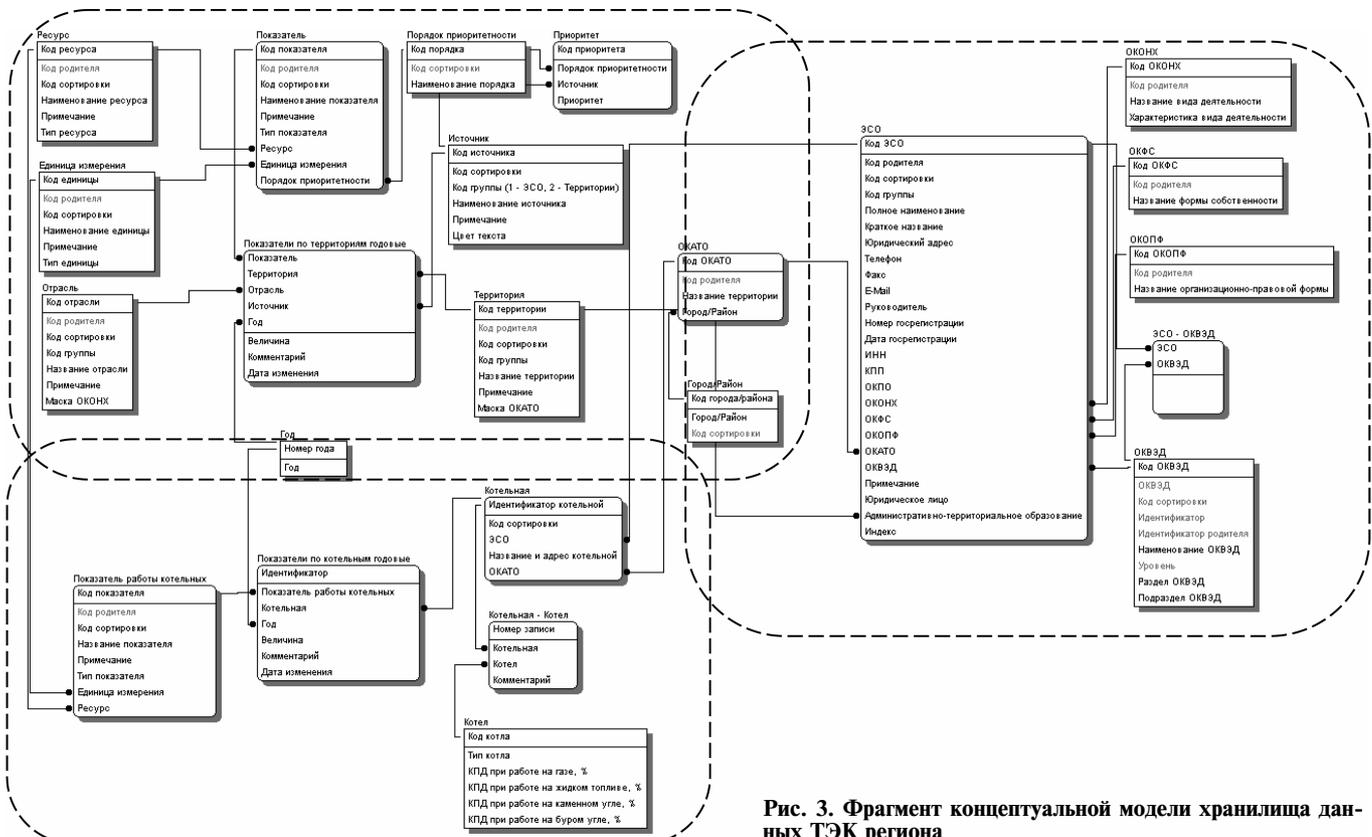


Рис. 3. Фрагмент концептуальной модели хранилища данных ТЭК региона

лиза социально-экономического положения территориальных образований в целях оценки уровня ЭБ и разработки стратегии развития ТЭК.

Различные виды энергетических объектов и установок составляют автономные объектные области вследствие принципиальных различий в наборе паспортных и мониторинговых показателей. На базе этих областей строят аналитические модели для решения задач технического контроллинга энергетических объектов. Уровни энергетических объектов и установок можно развивать "в ширину", добавляя в предлагаемую концептуальную модель аналогичные по структуре "звезды" по электрическим станциям, трансформаторам, линиям электропередач и т. п.

Для решения задач аудита потребления энергоресурсов в составе концептуальной модели ТЭК выделяется объектная область уровня конечных потребителей ТЭР.

Консолидация информации и взаимная согласованность объектных областей реализуется посредством ведения реестров энергетических предприятий, объектов, установок и потребителей с поддержанием истории регистрационных сведений. Объектные области, построенные на этих реестрах, ложатся на основу навигационных метаданных хранилища, которые служат для обеспечения нерегламентированного поиска детализированных данных, описывающих различные регистрационные сведения.

Информационно-аналитические модели, построенные на базе выделенных объектных областей, расширяют семантику модельного представления ХД корпоративной ИАС в целях организации эффективной информационной поддержки принятия решений ряда важнейших задач в сфере энергетики, включая мониторинг и оценку состояния исследуемых систем с точки зрения энергетической безопасности, проведения эффективной политики энергосбережения, тарифной и инвестиционной политики, а также повышения надежности функционирования оборудования. Многомерные аналитические модели являются основой эффективной интеграции данных и обеспечивают унифицированный доступ средств операционного окружения (средств интеграции данных и комплексной аналитической обработки).

Многомерная аналитическая модель ТЭК

Прикладные аналитические модели в исследуемых предметных приложениях — преимущественно гиперкубические, построены на основе совокупности таблиц фактов, объединяют показатели разных видов, с разными правилами получения значений при изменении уровней обобщения [3]. Структура многомерных аналитических моделей

должна учитывать следующие особенности, характерные для приложений в сфере энергетики:

1) представление каждого измерения (атрибута) многомерной информационной модели одним из возможных уровней обобщения данных с учетом многоаспектности решаемых задач и уровней иерархии управления в исследуемых системах;

2) объединение в одной многомерной модели нескольких таблиц фактов с учетом сложности исследуемых систем, содержащих множество объектов управления;

3) разделение множества экземпляров каждого справочника уровня обобщения данных по каждому атрибуту многомерной информационной модели на непересекающиеся подмножества для задания отдельных правил получения значений в ячейках гиперкуба с учетом территориальной распределенности исследуемых систем и многоаспектности их параметров и характеристик;

4) возможность выбора в разных запросах к многомерной информационной модели разных подмножеств существующих измерений с учетом многоаспектности решаемых задач;

5) совмещение в рамках одной информационной модели значений показателей ячеек различной природы, включая значения полей численного типа, а также различные функции агрегирования;

6) возможность организации нерегламентированного поиска в сложно структурированных справочниках измерений — ведомственных реестрах.

Формализация аналитической модели предполагает описание каркаса модели в виде совокупности атрибутов, уровней обобщения, групп однородных значений с соответствующими правилами извлечения значений показателей в ячейках гиперкуба, а также областей получения значений показателей. При этом каждая область определяется совокупностью правил извлечения значений показателей по группам однородных значений, соответствующих определенным уровням обобщения данных. Каждому уровню обобщения соответствуют непересекающиеся подмножества, составляющие множество групп однородных значений уровня обобщения. Предлагается подход к унификации видов групп однородных значений уровня обобщения с учетом специфики правил извлечения значений показателей гиперкуба, а также правил получения значений из соответствующих ячеек. Помимо этого, предлагаемая формализация позволяет обосновать целостность модели с точки зрения ее последующего использования в целях мониторинга и оценки ЭБ.

Суть работы аналитической модели заключается в трансляции запросов эксперта-аналитика в серию SQL-запросов к реляционной базе данных хранилища. Запрос пользователя на поиск информации включает: подмножество атрибутов

модели, в пространстве которых должны находиться ячейки запрашиваемых показателей; подмножество уровней обобщения, задающее для каждого атрибута уровень обобщения извлекаемых по нему значений; подмножества экземпляров для каждого выбранного в запросе уровня обобщения. Доступ к модели осуществляется посредством функционального ядра оперативной аналитической обработки информации. Полученное на выходе отношение имеет вид таблицы фактов структуры "звезда". В определенном формате (как API-интерфейс) она предоставляется для использования методами операционного окружения СППР.

Рассмотрим более детально состав, структуру и основные приемы работы с аналитической моделью топливно-энергетического баланса, являющейся информационной основой для полноценной многоаспектной оценки состояния, анализа и прогноза развития не только ТЭК, но и экономики региона в целом с позиций обеспечения его ЭБ.

Гиперкуб, лежащий в основе аналитической модели ТЭБ (рис. 4), строит на основе нескольких таблиц фактов, при этом оси гиперкуба определяют систему координат многомерного пространства показателей баланса. В ячейках куба данных в роли количественных мер выступают технические, производственные и экономические показатели ТЭБ; один и тот же показатель, в зависимости от способа получения его значения, может быть фактическим, расчетным, прогнозным, паспортным, нормативным или являться результатом экспертной оценки.

Структура аналитической модели ТЭБ включает шесть измерений.

1. Измерение "Показатель" позволяет консолидировать большие объемы разноплановой информации, характеризующей процессы в энергетической сфере, совмещая в рамках одной модели значения показателей различной природы. Вследствие этого на измерении вводятся группы однородных значений, которые позволяют задавать различные правила агрегации для суммируемых, усредняемых и других показателей; например, показатель топливно-энергетического баланса "Производство тепловой энергии" суммируется в процессе движения по уровням обобщения, а показатель "Удельный расход топлива на производство тепловой энергии" — усредняется.

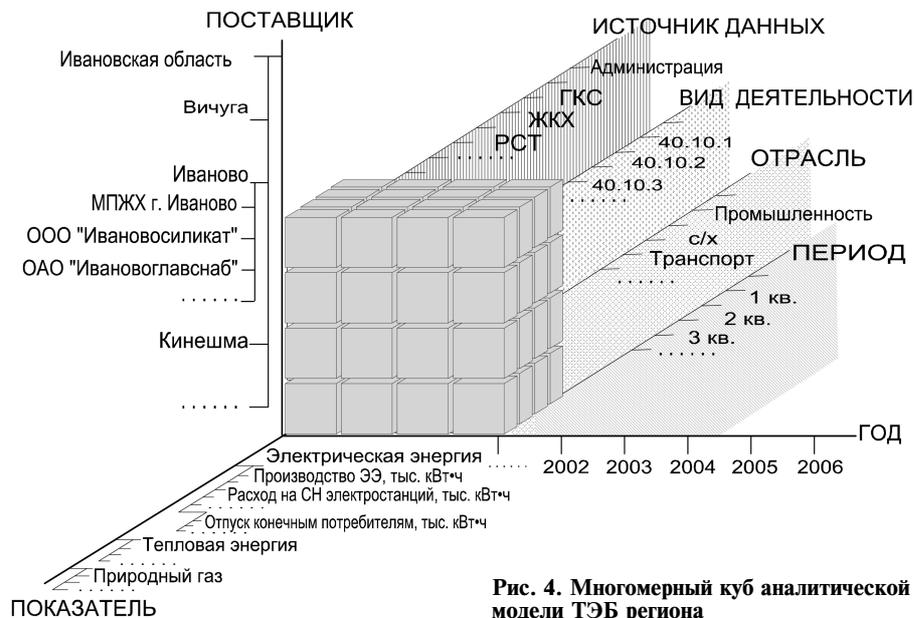


Рис. 4. Многомерный куб аналитической модели ТЭБ региона

2. Измерение "Поставщик ТЭР" определяет уровни объектов ТЭК, для которых в ХД поддерживается ретроспектива технико-экономических показателей. В роли поставщиков ТЭР выступают энергетические предприятия, территории и отрасли, образующие уровни обобщения по данному измерению, причем содержимое двух последних уровней обобщения разбивается на группы однородных значений, что связано со специфической иерархической структурой классификаторов территориальных образований и отраслей народного хозяйства.

3. Измерение "Источник" позволяет сравнивать значения показателей мониторинга, полученные из различных источников информации; каждый источник имеет приоритет, который характеризует степень доверия к нему; при наличии одновременно нескольких значений одного показателя предпочтение отдается значению, полученному из источника с наивысшим приоритетом.

4. Измерение "ОКВЭД" позволяет представить анализируемую информацию в соответствии с классификацией видов экономической деятельности энергетических предприятий.

5. Измерения "Год" и "Период" позволяют рассматривать значения показателей мониторинга в разрезах годовой и квартальной периодичности.

Каркас многомерной аналитической модели ТЭБ построен на основе четырех таблиц фактов, состоит из 144 областей получения значений и содержит 45 различных правил получения значений показателей: "пустое" правило (присутствует в тех областях, где сочетание групп однородных значений является бессмысленным с точки зрения логики предметной области), 10 правил на простое извлечение значений показателей из таблиц фак-

тов, 28 функций агрегирования и шесть правил-формул. Последние используются в случаях, когда по составу вектора групп однородных значений не удастся однозначно определиться с таблицей фактов, к которой должен адресоваться запрос. Логика работы формулы заключается в просмотре нескольких таблиц фактов и выборе в них значения показателя, поступившего из источника с наивысшим приоритетом. Например, требуется получить сведения о поступлении каменного угля в область в некотором году. Такая информация может в явном виде содержаться в таблицах фактов уровня территорий и отраслей или же может быть получена агрегированием соответствующих данных энергетических предприятий. В этом случае формула строит запрос на выборку значения показателя "Поступление каменного угля, т у.т." за некоторый год из всех таблиц фактов модели и возвращает в качестве своего результата единственное значение этого показателя, поступившее из источника с большим приоритетом.

Подготовка информации для принятия решений осуществляется посредством организации многомерного аналитического интерфейса, функционирующего на основе информационно-аналитической модели СППР. При этом реализуется представление агрегированной численной информации в виде кросстабличных отчетов, диаграмм деловой графики, геоинформационных карт. Поддерживается открытый интерфейс подключения модулей интеллектуального анализа данных, а также средств формирования шаблонов регламентированных отчетов со сценариями их заполнения (рис. 5, см. четвертую сторону обложки). Эти средства положены в основу реализации многоаспектного мониторинга и оценки состояния ЭБ в целях идентификации и анализа угроз, создания на уровне типовых решений ИАС технического контроллинга, ведения ТЭБ региона, регионального центра энергосбережения [4, 5, 6, 7].

Заключение

Таким образом, в отличие от существующих подходов и средств организации информационной поддержки принятия решений в рамках пространственных, преимущественно зарубежных, платформ СППР, предложенная модель обеспе-

чивает комплексное решение задач информационной поддержки процессов стратегического управления и проектирования применительно к крупным системам энергетики. В основу положена метауровневая архитектура хранилищ данных, разработанная с учетом особенностей исследуемых систем и базирующаяся на взаимодействии согласованных слоев метаданных, определяющих информационно-аналитическую модель СППР и операционное окружение, построенное на ее основе. Формируемые в процессе комплексной аналитической обработки накопленной информации корпоративные справочники, методики экспертной оценки состояния энергетических объектов и установок, аналитические доклады, системы отчетов, планы организационно-технических мероприятий и т. д. составляют важный информационный ресурс отраслевого масштаба, поддерживающий принятие решений при проведении технической, бюджетной, инвестиционной и тарифной политики.

Список литературы

1. **Энергетическая** безопасность. Термины и определения / Под. ред. Н. И. Воробай. М.: ИАЦ Энергия, 2005. 60 с.
2. **Паул В. К.** Состояние, проблемы и задачи энергокомпаний в области обеспечения надежности и безопасности в условиях реформирования электроэнергетики // Электронный журнал "Новое в Российской электроэнергетике". — 2006. — URL: http://www.rao-ees.ru/ru/news/news/magazin/show.cgi?index_06.htm.
3. **Решение** проблемы комплексного оперативного анализа информации хранилищ данных / С. Д. Коровкин, И. А. Левенец, И. Д. Ратманова, В. А. Старых, Л. В. Щавелев // СУБД. 1997. № 5—6. С. 47—51.
4. **Ратманова И. Д., Савельев В. А., Чичинский М. И.** Корпоративная информационная система ОАО "ФСК ЕЭС" для анализа технического состояния и уровня эксплуатации электросетевого оборудования // Методические вопросы исследования надежности больших систем энергетики. Методические и практические задачи надежности систем энергетики. Иркутск: ИСЭМ СО РАН, 2005. Вып. 55. С. 190—199.
5. **Коровкин С. Д., Ратманова И. Д., Щавелев Л. В., Левенец И. А.** Система ведения топливно-энергетического баланса как среда для поддержки принятия решений по управлению топливно-энергетическим комплексом региона // Вестник ИГЭУ. 2005. Вып. 4. С. 60—63.
6. **Ратманова И. Д., Железняк Н. В.** Подход к организации информационной поддержки государственной политики в сфере управления топливно-энергетическим комплексом региона // Вестник ИГЭУ. 2006. Вып. 4. С. 57—61.
7. **Коровкин С. Д., Ратманова И. Д., Щавелев Л. В., Левенец И. А., Козырев М. А.** Система мониторинга и анализа социально-экономического положения региона // Информационные технологии. 2004. № 3. С. 53—59.

Д. Х. Девятов, д-р техн. наук, проф., зав. каф.,
 С. И. Файнштейн, ст. преп.,
 А. Б. Белявский, доц.,
 В. Е. Торчинский, доц.,
 Магнитогорский государственный
 технический университет им. Г. И. Носова,
 e-mail: swetlana@mgn.ru

Эвристическая оптимизационная модель для задач с ограничениями, возникающих при оперативном планировании листопрокатного производства

Предложенная модель позволяет преобразовать некоторый класс NP-полных задач с ограничениями, возникающих, в частности, при оперативном планировании листопрокатного производства металлургического предприятия, к задаче безусловной оптимизации. Это достигается путем замены ограничений на штрафную функцию, а также начислению премий за следование внешним критериям оптимизации. Размеры штрафов и премий, а также ряд дополнительных параметров являются настраиваемыми константами, позволяющими настраивать эвристический пороговый SF-алгоритм под реальные данные и оперативно менять критерии оптимизации.

Ключевые слова: методы оптимизации, задачи с ограничениями, эвристические алгоритмы, оперативное планирование, листопрокатное производство.

Введение

При решении NP-трудных задач часто возникают проблемы, которые делают невозможным применение точных алгоритмов. Первая из них общеизвестна и связана с возникновением экспоненциального взрыва при большой размерности исходных данных. Вторая проблема характеризует отличие модельных NP-трудных задач от их производственных аналогов. Рассмотрим задачу формирования производственного графика, удовлетворяющего многочисленным ограничениям (техническим инструкциям). Допустимого решения, как правило, не существует; приемлемое решение всегда содержит некоторое число нарушений.

Нам представляется, что наилучший подход к решению подобных задач заключается в отказе от каких-либо ограничений вообще и превращения их в стоимости. Функция, вычисляющая стоимость размещения элемента в текущем решении, должна иметь смысл эвристической оценочной функции, начисляющей штрафы и премии по

всем параметрам операции. Наши исследования в течение последних трех лет [1–3] показали, что предложенная модель позволяет решать широкий класс задач оперативного планирования, возникающих при производстве, складировании и отгрузке листопрокатной продукции.

1. Общая задача упорядоченного разбиения с ограничениями

Сформулируем задачу упорядоченного разбиения с ограничениями (I).

Условие. Дано конечное множество A . Каждый элемент $a \in A$ имеет целую положительную длину $l(a)$ и k целых положительных характеристик $\langle g_1(a), g_2(a), \dots, g_k(a) \rangle$, $k \geq 2$. Также задано ограничение на суммарную длину $L_0 \in Z^+$, где Z^+ — положительное число, и два вектора целых положительных ограничений $\langle D_1^1, \dots, D_k^1 \rangle$, $\langle D_1^2, \dots, D_k^2 \rangle$. На множестве A задано несимметричное бинарное отношение β , задающее совместимость двух элементов, $\beta(a_i, a_j) = 0$, если a_i и a_j совместимы, $\beta(a_i, a_j) = 1$ — в противном случае.

Вопрос. Существует ли разбиение множества A на m попарно непересекающихся упорядоченных множеств A_1, A_2, \dots, A_m , $m \geq 1$ со следующими ограничениями:

- ограничением на суммарную длину элементов множеств A_j , $j = 1 \dots m$,

$$\sum_{i=1}^{\text{count}(j)} l(a_i^j) \leq L_0, \quad (1)$$

где $\text{count}(j)$ — число элементов множества A_j ;

- ограничениями на порядок следования элементов внутри множеств A_j :

$$0 \leq g_r(a_{i+1}^j) - g_r(a_i^j) \leq D_r^1; \quad (2)$$

$$0 \leq g_r(a_i^j) - g_r(a_{i+1}^j) \leq D_r^2; \quad (3)$$

$$\beta(a_i^j, a_{i+1}^j) = 0, \text{ где } i = 1 \dots \text{count}(j) - 1, \\ j = 1 \dots m, r = 1 \dots k. \quad (4)$$

Эта задача является NP-полной задачей распознавания. Покажем, что все ее частные подзадачи также являются NP-полными.

Отбросим все ограничения, кроме ограничения (4) на совместимость элементов, и положим $m = 1$. Соединим совместимые элементы ориентированным ребром. К этой задаче полиномиально сводится задача о существовании ориентированного гамильтонова пути [4].

Отбросим все ограничения, кроме ограничения на порядок следования элементов (2) и (3). Положим $m = 1$, $k = 2$ и $D_r^1 = D_r^2$ для всех $r = 1 \dots k$. К этой задаче полиномиально сводится задача о пути геометрического коммивояжера с

ограничением на длину ребра в смысле метрики $L_\infty: d((x_1, y_1), (x_2, y_2)) = \max(|x_1 - x_2|, |y_1 - y_2|)$ [4]. Этот частный случай является для нас очень важным, поэтому рассмотрим его отдельно.

Геометрический коммивояжер с ограничением на длину ребра

Условие. Заданы множество $P \subseteq Z \times Z$ точек на плоскости и положительное целое число B .

Вопрос. Существует ли путь, проходящий через все точки и не содержащий ребер длины больше B , где $d((x_1, y_1), (x_2, y_2)) = \max(|x_1 - x_2|, |y_1 - y_2|)$?

Рассмотрим содержательную интерпретацию этой задачи. Пусть каждой точке с координатами (x, y) соответствует элемент с параметрами (x, y) . Перед обработкой требуется упорядочить элементы таким образом, чтобы разность соответствующих параметров любой пары соседних элементов составляла не более B .

Теперь сформулируем задачу иначе. Пусть нам разрешено пополнять множество P произвольным образом, в количестве не более чем K точек. Может показаться, что от возможности пополнения множества произвольными элементами задача становится проще. Покажем, что это не так. Рассмотрим задачу, где точки множества P заданы с повторениями, т. е. для каждой точки указана ее кратность, необходимо пройти через точку число раз, равное ее кратности, и разрешено пополнять множество P произвольными точками так, чтобы сумма кратностей добавленных точек не превосходила K .

Геометрический коммивояжер с повторениями (ГКП)

Условие. Заданы множество $P^* \subseteq Z \times Z$ точек на плоскости, целое положительное число экземпляров (кратность) $c(p)$ каждой точки p и положительные целые числа K и B^* .

Вопрос. Существует ли на плоскости множество $Q \subseteq Z \times Z$ такое, что сумма кратностей всех точек множества Q не превосходит K и существует путь, который проходит $c^*(p)$ раз через каждую точку p множества $P^* \cup Q$ и не содержит ребер длины, большей B^* в смысле метрики L_∞ , где $c^*(p)$ — кратность точки p с учетом добавленных точек?

Теорема. Задача ГКП NP -полная.

Доказательство. В силу целочисленности выбранной метрики задача принадлежит классу NP . Покажем, что ПС полиномиально сводится к ГКП.

Выберем в качестве известной NP -полной задачи путь геометрического коммивояжера [4]. (Заданы множество $P \subseteq Z \times Z$ точек на плоскости и положительное целое число B , существует ли путь коммивояжера стоимости $\leq B$?). Рассмотрим произвольную индивидуальную задачу I задачи ГК.

Пусть $|P| = n$. Положим $P^* = P$, кратности всех точек $c(p) = 1$, $K = B - n + 1$, $B^* = 1$. Докажем, что полиномиальная сводимость сохраняет ответы "да" и "нет".

1. Пусть ГК имеет ответ "да", т. е. в модельной задаче существует Гамильтонов путь длины $\leq B$, проходящий через все точки P . Этот путь содержит

$n - 1$ ребро с длинами d_1, \dots, d_{n-1} , $\sum_{i=1}^{n-1} d_i \leq B$.

С помощью $d_i - 1$ целочисленной точки заменим каждое i -е ребро ломаной, состоящей из $d_i - 1$ единичных звеньев. При совпадении вершины ломаной с уже существующей точкой будем увеличивать кратность этой точки. Таких точек по-

требуется ровно $\sum_{i=1}^{n-1} (d_i - 1) \leq B - n + 1$.

2. Пусть ГКП имеет ответ "да", т. е. после добавления K точек существует путь из единичных ребер, проходящий через все точки $P \cup Q$ с учетом кратностей, и $K \leq B - n + 1$. Будем проводить редукцию этого пути, пока не получим Гамильтонов путь, проходящий только через множество P . Рассмотрим возможные варианты вхождения в путь добавленных точек y_1, \dots, y_k . Если существует начальный и/или конечный участок пути, состоящий только из добавленных точек, удалим эти точки из пути. Если существуют внутренние участки, состоящие из ребер нулевой длины y_i, y_j, \dots, y_i , удалим нулевые ребра и уменьшим кратность точки y_i . В результате редукции путь начинается и заканчивается точками из множества P , вхождения добавленных точек представляют собой последовательности типа $p_i, y_{j+1}, \dots, y_{j+s}, p_{i+1}$, где любая пара соседних вершин различна. Удалим из пути последовательность добавленных точек y_j, \dots, y_{j+s} , заменив ее ребром (p_i, p_{i+1}) . Так как удалена ломаная из $s + 1$ единичного звена, то длина ребра (p_i, p_{i+1}) не превосходит $s + 1$. Пусть таких участков будет r , в каждом i -м участке находится s_i

точек, $\sum_{i=1}^r s_i \leq K \leq B - n + 1$. Таким образом, сум-

марная длина неединичных ребер не превосходит

$\sum_{i=1}^r (s_i + 1) \leq B - n + 1 + r$. Единичных ребер в пути

остается $n - r - 1$, и суммарная длина пути не превосходит $B - n + 1 + r + n - r - 1 \leq B$.

2. Эвристическая модель пространства состояний

Рассмотрим задачу упорядоченного разбиения с ограничениями (I). Чтобы избавиться от огра-

ничений (1)—(4), опишем задачу в терминах модели "пространство состояний" [5]. Для этого нужно указать начальное состояние задачи, ее терминальные состояния, а также набор допустимых операций, переводящих задачу из одного состояния в другое.

В начальном состоянии все элементы находятся во множестве A , и множества A_1, A_2, \dots, A_m — пустые; в терминальном состоянии множество A — пусто, его элементы размещены в упорядоченных множествах A_1, A_2, \dots, A_m . Имеется следующий набор допустимых операций: размещение элемента a в одно из множеств A_j ; удаление элемента a из множества A_j . Через них определяется операция перекаладывания элемента из одного множества в другое. Текущее состояние задачи u полностью определяется координатами уже размещенных элементов. Под координатами элемента будем понимать пару (j, p) , где j — номер множества, p — позиция элемента внутри множества.

Припишем каждой допустимой операции вещественную стоимость. Стоимость будет иметь вид аддитивной функции от положительных штрафов и отрицательных премий, начисленных по всем параметрам операции. Штрафы есть стоимостная форма ограничений (1)—(4), градация штрафов будет соответствовать степени тяжести нарушений. Премии начисляются исходя из какого-то внешнего критерия оптимизации [2], [3].

Пусть ограничениям (1)—(4) соответствует система целых положительных штрафов $\{\text{VeryBig}F, F_r^i (i = 1, 2; r = 1 \dots k), F_0\}$, где $\text{VeryBig}F$ — штраф за превышение допустимой суммарной длины последовательности; F_r^i — штрафы за превышение допустимой разности характеристик соседних элементов; F_0 — штраф за несовместимость соседних элементов. Тогда стоимость операции размещения элемента a в j -ю последовательность можно определить следующим образом:

$\text{Cost}(a, j, 1) = 0$ при размещении a в пустую последовательность;

$\text{Cost}(a, j, p)$ при размещении a в p -ю позицию последовательности длины $p - 1$:

$$\begin{aligned} & \sum_{i=1}^k (F_r^1 \cdot \text{sg}(g(a_p^j) - g(a_{p-1}^j)) - D_r^1) + \\ & + F_r^2 \cdot \text{sg}(g(a_{p-1}^j) - g(a_p^j) - D_r^2)) + \\ & + \text{VeryBig}F \cdot \text{sg}\left(\sum_{i=1}^p l(a_i^j) - L_0\right) + F_0 \cdot \beta(a_{p-1}^j, a_p^j), \end{aligned}$$

где sg — функция знака, $\text{sg}(x) = 1$ при $x > 0$ и $\text{sg}(x) = 0$ при $x \leq 0$. Аналогично можно вычислить стоимость всех остальных операций. Определим стоимость последовательности A_j : $\text{Cost}(A_j) =$

$$\begin{aligned} & = \sum_{r=1}^k \sum_{i=1}^{\text{count}(j)-1} (F_r^1 \cdot \text{sg}(g(a_{i+1}^j) - g(a_i^j) - D_r^1) + \\ & + F_r^2 \cdot \text{sg}(g(a_i^j) - g(a_{i+1}^j) - D_r^2)) + \text{VeryBig}F \times \\ & \times \text{sg}\left(\sum_{i=1}^{\text{count}(j)} l(a_i^j) - L_0\right) + F_0 \cdot \sum_{i=1}^{\text{count}(j)-1} \beta(a_i^j, a_{i+1}^j), \end{aligned} \quad (1)$$

тогда стоимость полного разбиения равна $\sum_{j=1}^m \text{Cost}(A_j)$. Таким образом, на всем пространстве состояний U задачи (I) мы определили эвристическую оценочную функцию $f(u) = \sum_{j=1}^m \text{Cost}(A_j)$,

равную нулю, если ни одно из ограничений (1)—(4) не было нарушено. Так как премия приписывается отрицательная стоимость, то задачу распознавания (I) можно свести к задаче оптимизации (II): $f(u) \rightarrow \min$, где $u \in U$.

Сделаем важное замечание о размерах штрафов и премий. Числовые значения штрафов и премий не являются наперед заданными величинами, а играют роль настроечных констант, подбираемых во время прогона программы на реальных данных. Меняя значения настроечных констант, мы изменяем систему приоритетов и генерируем решение согласно выбранным критериям оптимизации.

3. Задача о перестройках прокатного стана

Перестройки прокатного стана аналогичны переналадкам оборудования. Из набора заказов на текущие сутки (около 250 заказов) необходимо сформировать не более семи упорядоченных монтажных партий (монтажей) — суточный график работы прокатного стана. Заказы имеют следующие характеристики: длина, ширина, толщина, марка стали, ГОСТ, группа выкатываемости и др. В зависимости от характеристик двух соседних заказов перестройка стана проводится или не проводится. Кроме того, имеется обширный список перестроек, недопустимых согласно требованиям технологических инструкций. Формированием монтажей занимается высококвалифицированный диспетчер, различающий нарушения по степени тяжести и избегающий наиболее тяжких нарушений. В силу большого количества технологических инструкций целевая функция f^* имеет довольно сложный вид [1], поэтому рассмотрим только несколько основных ограничений.

1. Длина: нельзя превышать предельное значение суммарной длины монтажа LMP .

2. Ширина: перестройка от узкого к широкому с шагом свыше 250 мм считается сверхгрубым на-

рушением, от узкого к широкому с шагом ≤ 250 мм — грубым нарушением, перестройка от широкого к узкому с шагом свыше 250 мм — легким нарушением, от широкого к узкому с шагом ≤ 250 мм допустима.

3. Толщина разбита на диапазоны, внутри каждого диапазона установлен допустимый шаг. Превышение шага внутри диапазона считается легким нарушением, скачки из диапазона в диапазон — нарушением средней степени тяжести, скачки через диапазон — грубым нарушением.

4. Группа выкатываемости: для каждой пары заказов известна их совместимость/несовместимость, т. е. задана функция β (ограничение (4) задачи (I)).

Определим стоимость перестройки от заказа a_i к a_{i+1} внутри j -го монтажа:

$$\text{Cost}(a_i, a_{i+1}, j) = \text{VeryBig}F \cdot \text{sg}(\text{MLength}[j] + L[a_{i+1}] - \text{LMP}) + F_0 \cdot \beta(a_i, a_{i+1}) + Cd(a_i, a_{i+1}) + Cw(a_i, a_{i+1}),$$

где $\text{MLength}[j]$ — суммарная длина j -го монтажа, $L[a_{i+1}]$ — длина добавляемого заказа, LMP — предельное значение длины монтажа, $\text{VeryBig}F$ — очень большой штраф, Cd и Cw — стоимости перестроек по толщине и ширине, определяемые следующим образом:

$$Cd(a_i, a_{i+1}) = \begin{cases} \text{Const1} + \text{Const2}(\text{abs}(dt) - \text{Step})/\text{Step} \\ 0, \text{ если } Cd(a_i, a_{i+1}) < \text{Const1} \end{cases}$$

где dt — разность толщин a_i и a_{i+1} , Step — шаг диапазона толщины заказа a_{i+1} ,

$$Cw(a_i, a_{i+1}) = \begin{cases} F_3, \text{ от узкого к широкому с шагом } > 250 \text{ мм} \\ F_2, \text{ от узкого к широкому с шагом } \leq 250 \text{ мм} \\ F_1, \text{ от широкого к узкому с шагом } > 250 \text{ мм} \\ 0, \text{ от широкого к узкому с шагом } \leq 250 \text{ мм}, \end{cases} \quad (2)$$

Const1 , Const2 , F_0 , F_1 , F_2 , F_3 — настроечные константы.

Определив стоимость перестройки от заказа a_i к заказу a_{i+1} , можно определить стоимость j -го монтажа (1) и целевую функцию f^* .

4. Эвристический пороговый SF-алгоритм (Sorted Fines-algorithm)

Задача о перестройках прокатного стана имеет очень большой размер пространства поиска ($250! \approx 10^{453}$). Для ее решения целесообразно воспользоваться методом разбиения задачи на подзадачи. В качестве алгоритма минимизации целевой функции используется эвристический пороговый SF-алгоритм с выделенной настроечной константой MinCut , имеющей смысл стоимостного предела размещения в текущем решении нового элемента или последовательности элементов. Значение MinCut на каждой итерации алгоритма посте-

пенно увеличивается от минимального штрафа до некоторого стоимостного предела. Отсортированные в порядке возрастания значения штрафа и премии играют роль естественных энергетических уровней, позволяющих алгоритму выходить из "ловушек" локальных оптимумов.

SF-алгоритм

Шаг 1. Сортируем штрафы в порядке возрастания F_1, \dots, F_s . Полагаем $j = 1$. Устанавливаем значение минимального разреза MinCut (настроечная константа), $\text{MinCut} < F_1$ и значения стоимостного предела Limit (настроечная константа), $F_1 \leq \text{Limit} \leq F_s$.

Шаг 2. Объединяем все неразмещенные элементы в одну последовательность, сортируем сортировкой Sort (тип сортировки также является настроечной константой). Если стоимость перестройки между элементами $F_0 \cdot \beta(a_i, a_{i+1}) + Cd(a_i, a_{i+1}) + Cw(a_i, a_{i+1}) > \text{MinCut}$, то "разрезаем" между элементами a_i и a_{i+1} . Затем полученные подпоследовательности (столбцы) еще раз сортируем "стоимостным пузырьком" (перестановка двух соседних элементов проводится только в случае уменьшения суммарной стоимости столбца).

Шаг 3. Размещаем по монтажах целые столбцы. Столбец присоединяется к последнему элементу текущего монтажа, если разность между новой и старой стоимостью монтажа $\leq \text{MinCut}$.

Шаг 4. Столбцы, не размещенные на шаге 3, делятся на отдельные заказы и размещаются поэлементно. Для каждого элемента вычисляется стоимость и позиция оптимального размещения внутри каждого монтажа, затем выбирается монтаж, который соответствует размещению минимальной стоимости. Если стоимость $\leq \text{MinCut}$, то заказ размещается.

Шаг 5. $\text{MinCut} = F_j$, $j = j + 1$. Если $\text{MinCut} \leq \text{Limit}$, то переход на шаг 2. Иначе переход на шаг 6.

Шаг 6. Все элементы, не размещенные на предыдущих итерациях, размещаются согласно шагу 5 со стоимостным пределом, равным $+\infty$.

Вычислительная сложность алгоритма равна $O(sn^2)$, где n — число размещаемых заказов; s — число жестких ограничений.

5. Результаты вычислительного эксперимента

Для идентификации алгоритма был проведен сравнительный анализ планов работы стана 2000 горячей прокатки в феврале 2006 г., составленных диспетчером и программой. Планы диспетчера и программы сравнивались по числу нарушений каждого типа, а также по максимальной длине монтажа (при коротком монтаже перевалка проводится раньше, чем валки исчерпают свой ресурс). Ре-

Таблица 1. Число нарушений монтажа

Нарушения	Диспетчер	Программа
По ширине	3	0
По толщине	0	2
По группе выкатываемости	9	1
Итого	12	3

Таблица 2. Длина монтажа, км

№ монтажа	1	2	3	4	5	6	7
Диспетчер	46	96	95	96	101	96	105
Программа	168	150	112	70	125	5	6

зультаты сравнения для монтажей, прокатанных 14.02.2006 г., приведены в табл. 1.

Распределение длин монтажа по планам, составленным диспетчером и программой, представлено в табл. 2.

Анализ показал, что применение программы позволяет увеличить длину монтажа без увеличения общего числа нарушений. Таким образом, возникает задача дополнительного выбора со склада K заказов, которые могут быть прокатаны за те же сутки, что и основной набор заказов. Как было показано в теореме, эта задача также является NP -полной.

Выводы

1. Всей эвристической информации из предметной области задачи придан единообразный числовой вид. Превращение жестких ограничений в целевую штрафную функцию стоимости

позволяет генерировать решения с некоторым числом нарушений в случае, если допустимого решения не существует.

2. Гибкая система настроечных штрафов и премий позволяет адаптировать алгоритм к реальным данным и оперативно менять критерии оптимизации.

3. Построение решения происходит не только из отдельных элементов, но и из отсортированных последовательностей, что улучшает качество решения и уменьшает время его генерации.

4. Изменение текущего стоимостного предела размещения позволяет алгоритму выходить из ловушек локальных минимумов целевой функции. Отсортированные в порядке возрастания штрафы и премии играют роль естественных энергетических уровней.

5. Генерация решения происходит за время $O(sn^2)$, что позволяет осуществлять оперативное управление производством.

Список литературы

- Каплан Д. С., Девятков Д. Х., Белявский А. Б., Файнштейн С. И., Торчинский В. Е. Алгоритм оперативного планирования погрузки металла в печи листопрокатного стана // *Сталь*. 2007. № 2. С. 130–133.
- Каплан Д. С., Девятков Д. Х., Файнштейн С. И., Тутарова В. Д., Калинтаев А. Н. Эвристический полиномиальный алгоритм оперативного планирования размещения готовой продукции на складах металлургических предприятий // *Автоматизация и современные технологии*. 2009. № 6. С. 35–39.
- Девятков Д. Х., Файнштейн С. И., Тутарова В. Д., Калинтаев А. Н. Оперативное планирование отгрузки готовой продукции со складов металлургических предприятий // *Мехатроника, автоматизация, управление*. 2008. № 4. С. 36–40.
- Гэри М., Джонсон Д. Вычислительные машины и труднорешаемые задачи. Пер. с англ. М.: Мир, 1982.
- Нильсон Н. Принципы искусственного интеллекта. М.: Радио и связь, 1985.

МОДЕЛИРОВАНИЕ

УДК 541.182

Б. Г. Кухаренко, канд. физ.-мат. наук, вед. науч. сотр.,
Институт машиноведения РАН, г. Москва, e-mail: kukharenko@imash.ru

Анализ независимых компонент записей колебаний в технологии спектрального анализа на основе быстрого преобразования Фурье

Описана совместная предварительная обработка составляющих многомерных временных рядов посредством анализа независимых компонент. Она повышает точность и увеличивает частотное разрешение спектрального метода Фурье. В качестве примера рассматриваются записи коллективных колебаний лопаток ротора при флаттере в турбокомпрессоре. Предварительный анализ независимых компонент записей обеспечивает исследование по методу Фурье процесса синхронизации неустойчивых колебаний лопаток.

Ключевые слова: многомерные временные ряды, анализ независимых компонент, спектральный анализ, метод Фурье, флаттер лопаток компрессоров, синхронизация колебаний.

Совместная предварительная обработка составляющих многомерных временных рядов

Статья посвящена описанию средств совместной предварительной обработки составляющих многомерных временных рядов в технологии спектрального анализа на основе метода Прони (Prony) [1]. Главной предпосылкой использования метода Прони для спектрального анализа нестационарных временных рядов является кусочно-линейная модель регистрируемого колебания. Этот метод использует разложение сегментов временных рядов, в результате которого для суммы временных зависимостей вида $Ae^{-\delta t} \cos(2\pi f t + \varphi)$ определяются все четыре независимых параметра этих временных зависимостей: A , φ , δ , f . В результате для каждого последовательного сегмента временного ряда по методу Прони определяется его собственный спектр частот и демпфирование на этих собственных частотах.

В работе [2] метод Прони используется для спектрального анализа неустойчивых коллективных колебаний (флаттера) лопаток ротора турбокомпрессора. Эти колебания описываются линейной моделью и характеризуются относительно стабильным спектром собственных частот. Зависящее от времени демпфирование на собственных частотах позволяет характеризовать эволюцию коллективных колебаний (флаттера) лопаток компрессора. Для этого используется сегментирование записей нестационарных колебаний (составляющих многомерных нестационарных временных рядов).

Пусть $x[\overline{1, N}]$, $N \ll N_0$, представляет собой некоторый сегмент полного временного ряда (записи колебания) $x[\overline{1, N_0}]$ (индекс временного ряда опущен, поскольку метод Прони одномерный, и каждая составляющая многомерного временного ряда обрабатывается по отдельности). Для краткости выбран шаг дискретизации времени $\Delta t = 1$. Спектральная декомпозиция Прони сегмента $x[\overline{1, N}]$ имеет вид

$$x[k] = \sum_{l=1}^p r[l]z[l]^{k-1} + n[k], \quad k = \overline{1, N}. \quad (1)$$

Здесь p — число полюсов сегмента; $z[l] = \exp(\delta[l] + j2\pi f[l])$, $l = \overline{1, p}$, — полюса сегмента, где $\delta[l]$ и $f[l]$ — соответственно фактор демпфирования и частота; $r[l] = A[l] \exp(j\varphi[l])$, $l = \overline{1, p}$, — вычеты в полюсах, $A[l]$ и $\varphi[l]$ — соответственно амплитуда и фаза; $n[k]$ — аддитивный шум.

Когда определены главные полюса $z[l]$, $l = \overline{1, p}$, вычеты $r[l]$, $l = \overline{1, p}$, в полюсах определяются из формулы (1) по методу наименьших квадратов. Оценка зависящих от времени спектров факторов демпфирования и частот и соответ-

ствующих спектров амплитуд и фаз для полного временного ряда (записи колебания) $x[\overline{1, N_0}]$ получается посредством последовательных сдвигов временного окна фиксированной длины $N\Delta t$.

При исследовании динамических систем на основе нестационарных временных рядов по методу Прони учитывается возможная (потенциальная) корреляция отдельных составляющих многомерного временного ряда, но только на этапе оценки результатов спектрального анализа [2]. Например, для многомерного временного ряда, состоящего из одновременных записей колебаний лопаток ротора, отдельные составляющие многомерного временного ряда — это различные проекции одного и того же коллективного колебательного процесса. Поэтому они представляют собой различные смеси мод лопатки ротора. То есть для полных спектров частот отдельных составляющих многомерного временного ряда их общей частью (с принятой при анализе точностью или толерантностью) могут быть спектры собственных частот мод лопатки ротора. Априорное знание, что составляющие многомерного временного ряда представляют собой различные смеси мод исследуемой распределенной системы (лопаток ротора), может быть использовано при предварительной совместной обработке этих составляющих посредством анализа независимых компонент (который и учитывает факт многомерности временных рядов). В результате могут быть выделены независимые компоненты, представляющие моды распределенной системы (если они возбуждены таким образом, что являются статистически независимыми). Для спектрального анализа таких независимых компонент достаточно одночастотной аппроксимации, т. е. $p = 2$ в формуле декомпозиции Прони (1). Это существенно повышает точность определения спектральных параметров собственных мод колебаний распределенной системы (лопаток ротора).

Анализ независимых компонент

Пусть $y = (y_1, \dots, y_m)^T$ — наблюдаемый m -мерный случайный вектор с нулевым средним. Наблюдаемые значения y соответствуют m -мерным записям колебаний (наблюдаемым m -мерным временным рядам)

$$y[i] = y(i\Delta t), \quad i = \overline{1, N_0},$$

т. е. временной зависимости $y = y(t)$ в дискретном времени $t = i\Delta t$, $i = \overline{1, N_0}$ (ниже для краткости выбран шаг дискретизации времени $\Delta t = 1$). Задача анализа независимых компонент состоит в оценке линейной модели данных:

$$y = A \cdot s, \quad (2)$$

где $\mathbf{s} = (s_1, \dots, s_n)^T$ — n -мерный (скрытый) случайный вектор, компоненты которого предполагаются взаимно независимыми, и \mathbf{A} — постоянная матрица размерности $m \times n$. Независимые компоненты \mathbf{s} должны быть не Гауссовыми случайными переменными (анализ независимых компонент — это не Гауссова версия факторного анализа) [3]. Численные эксперименты показывают, что в случае независимых компонент с простыми и различными спектрами частот (например, моды распределенных систем) следует использовать алгоритм SOBI (Second-Order Blind Identification) [3, 4]. Алгоритм SOBI основан на диагонализации ковариационных матриц с запаздыванием, что является альтернативой не-Гауссовости независимых компонент. Алгоритм SOBI состоит из двух шагов.

Первый шаг — это линейное преобразование \mathbf{W} (отбеливание) сигнальной части $\mathbf{y}(t)$ наблюдаемых данных. Отбеленные данные не коррелированы и имеют единичную ковариационную матрицу

$$E[\mathbf{W}\mathbf{y}(t)\mathbf{y}^+(t)\mathbf{W}^+] = \mathbf{I}, \quad (3)$$

где $+$ обозначает Эрмитово сопряжение (транспонирование и комплексное сопряжение).

С учетом (2) можно записать

$$E[\mathbf{W}\mathbf{y}(t)\mathbf{y}^+(t)\mathbf{W}^+] = \mathbf{W}\mathbf{A}\mathbf{A}^+\mathbf{W}^+, \quad (4)$$

т. е. $\mathbf{W}\mathbf{A}\mathbf{A}^+\mathbf{W}^+ = \mathbf{I}$. Таким образом, из формул (3) и (4) следует, что каждой матрице \mathbf{W} соответствует унитарная матрица $\mathbf{U} = \mathbf{W} \cdot \mathbf{A}$, $\mathbf{U} \cdot \mathbf{U}^+ = \mathbf{I}$. В присутствии шума для наблюдаемых данных $\mathbf{x}(t)$ в дискретном времени $t = i\Delta t$, $i = \overline{1, N_0}$, $\mathbf{x}(t) \neq \mathbf{y}(t)$, и для отбеленных данных $\mathbf{z}(t) = \mathbf{W} \cdot \mathbf{x}(t)$

$$\begin{aligned} E[\mathbf{z}(t)\mathbf{z}^+(t)] &= E[\mathbf{W}\mathbf{x}(t)\mathbf{x}^+(t)\mathbf{W}^+] = \mathbf{W}\mathbf{A}\mathbf{A}^+\mathbf{W}^+ + \\ &+ \mathbf{W}\sigma^2\mathbf{W}^+ = \mathbf{W}(\mathbf{R}_x(0) - \sigma^2\mathbf{I})\mathbf{W}^+ + \\ &+ \mathbf{W}\sigma^2\mathbf{W}^+ = \mathbf{W}\mathbf{R}_x(0)\mathbf{W}^+, \end{aligned} \quad (5)$$

где σ^2 — вариация x .

Таким образом, матрица отбеливания \mathbf{W} может быть определена на основе ковариационной матрицы $\mathbf{R}_x(0) = E[\mathbf{x}(t) \cdot \mathbf{x}^+(t)]$.

Второй шаг состоит в определении унитарной матрицы \mathbf{U} . Для отбеленных данных $\mathbf{z}(t) = \mathbf{W} \cdot \mathbf{x}(t)$ с заранее выбранными значениями запаздывания τ ковариационные матрицы

$$\begin{aligned} \mathbf{R}^w(\tau) &= E[\mathbf{z}(t + \tau)\mathbf{z}^+(t)] = \mathbf{W}E[\mathbf{x}(t + \tau)\mathbf{x}^+(t)]\mathbf{W}^+ = \\ &= \mathbf{W}\mathbf{A}E[\mathbf{s}(t + \tau)\mathbf{s}^+(t)]\mathbf{A}^+\mathbf{W}^+, \quad \forall \tau \neq 0. \end{aligned} \quad (6)$$

одновременно приблизительно диагонализуются с помощью унитарной матрицы \mathbf{U} :

$$\mathbf{R}^w(\tau) = \mathbf{U}\mathbf{R}_s(\tau)\mathbf{U}^+, \quad \forall \tau \neq 0, \quad (7)$$

где

$$\mathbf{R}_s(\tau) = E[\mathbf{s}(t + \tau)\mathbf{s}^+(t)] \quad (8)$$

— это ковариационная матрица для независимых компонент с временным запаздыванием τ , т. е. матрица $\mathbf{R}_s(\tau)$ — диагональная. Таким образом, унитарная матрица \mathbf{U} определяется из условия, что с ее помощью несколько ковариационных матриц отбеленных данных $\mathbf{R}^w(\tau)$ с заранее выбранными запаздываниями τ одновременно приводятся к диагональному виду. Например, это могут быть четыре ковариационные матрицы (8) с запаздываниями $\tau = 1, 2, 3, 4$.

Кроме алгоритма JAD (Joint Approximation Diagonalization), который основан на вращении Гивенса [5], алгоритм SOBI реализуется с использованием других приближительных диагонализаций, таких как диагонализация FAJD (Fast Approximate Joint Diagonalization) [6] и диагонализация SAD (Sequential Approximate Diagonalization) [7].

После определения матрицы \mathbf{U} матрица смешивания \mathbf{A} определяется из формулы $\mathbf{U} = \mathbf{W} \cdot \mathbf{A}$, а независимые компоненты $\mathbf{s} = (s_1, \dots, s_n)^T$ определяются из (2). Алгоритм SOBI хорошо работает в случае независимых компонент с простыми различными спектрами частот (или различными автокорреляционными функциями). Для того чтобы выделить определенное число независимых компонент с простыми и различными спектрами частот, необходим многомерный временной ряд наблюдаемых данных с таким же числом составляющих [8, 9]. При сокращении числа составляющих многомерного временного ряда наблюдаемых данных его спектральные компоненты остаются произвольным образом смешанными в определяемых независимых компонентах. Исключение составляют многомерные записи колебаний составных распределенных систем. При сокращении размерности наблюдаемого многомерного временного ряда независимые компоненты аккумулируют парциальные моды колебаний отдельных составляющих распределенной системы [10].

Синхронизация колебаний лопаток ротора при флаттере в турбокомпрессоре

На рис. 1 представлены одновременные записи колебаний лопаток ротора турбокомпрессора с ростом числа оборотов ротора.

На рис. 2 представлены четыре независимых компоненты записей колебаний лопаток ротора турбокомпрессора (см. рис. 1), полученные с помощью алгоритма SOBI.

На рис. 3 представлены зависимости от времени частот главных спектральных составляющих независимых компонент s_1, s_2, s_3 и s_4 (см. рис. 2), полученные на основе декомпозиции Прони (1)

для $p = 2$. Этот прецизионный спектральный анализ независимых компонент позволяет различать близкие частоты 1 и 3. Их интерпретация приводится ниже.

При флаттере формируются круговые моды, в которых синхронные колебания соседних лопаток в одной собственной моде лопатки (например, крутильной) имеют постоянный сдвиг фазы между колебаниями соседних лопаток, обеспечивающий периодичность на колесе ротора

$$\sigma_m = \frac{2\pi m}{N_b}, m = 0, \pm 1, \pm 2, \dots, \pm(N_b - 1), \quad (9)$$

где N_b — число лопаток в роторе турбокомпрессора [11]. Порядок m определяет число узлов круговой моды синхронных колебаний лопаток ротора и называется также числом ее узловых диаметров. Пусть $\mathbf{u}(t) = [\mathbf{u}_1(t), \dots, \mathbf{u}_{N_b}(t)]$ — вектор, представляющий степени свободы колеблющихся лопаток ротора. При флаттере периодическое колебание на крутильной частоте f_T отдельной лопатки представляется декомпозицией на круговые моды синхронных колебаний лопаток:

$$\mathbf{u}_i(t) = \sum_{m=-N_b+1}^{N_b-1} A_m \exp(j(2\pi f_T t + (i-1)\sigma_m)). \quad (10)$$

Как видно на рис. 2, амплитуда независимых компонент меняется при $t \approx 46$ с (момент наступления флаттера), т. е. две независимые компоненты s_1 и s_3 представляют неустойчивую моду лопатки (крутильную), связанную с явлением флаттера на рис. 1. Среди частот на рис. 3 наиболее низкие частоты соответствуют независимым компонентам s_1 и s_3 . Согласно (9) и (10), а при флаттере неустойчивые колебания лопаток должны синхронизоваться на частоте f_T крутильной моды. Но колебания лопаток ротора (см. рис. 1) представлены двумя (статистически) независимыми компонентами s_1 и s_3 , и частота крутильной моды f_T оказывается расщепленной (рис. 3). Более детально это видно на рис. 4 (размер N последовательных сегментов независимых компонент соответствует временному интервалу 1,5 с).

Возникает вопрос о причине появления двух независимых компонент и расщеплении частоты крутильной моды лопатки, неустойчивость которой связана с появлением флаттера. Оказывается,

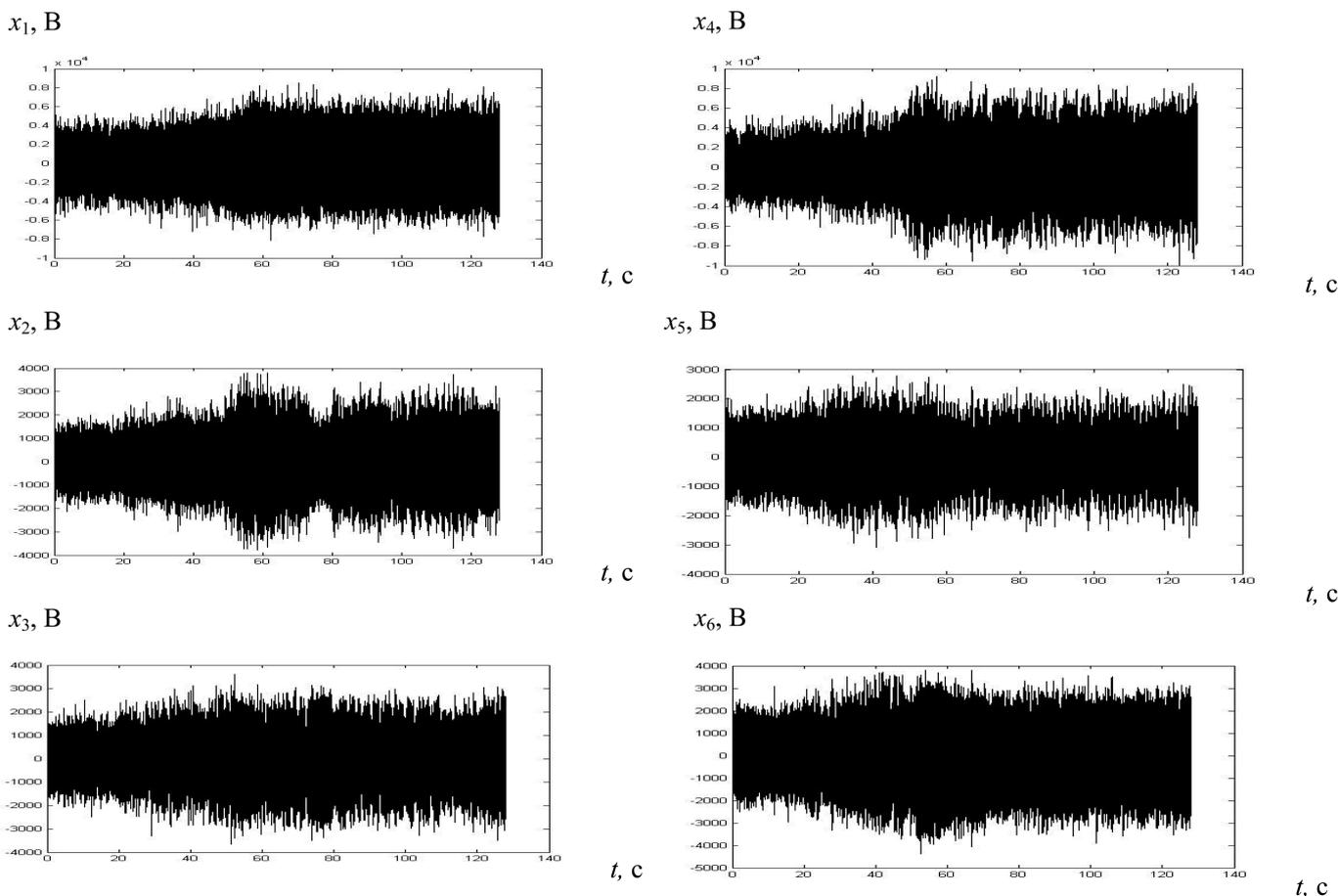


Рис. 1. Одновременные записи колебаний лопаток ротора:
 x_1 — 1-й лопатки, x_2 — 6-й лопатки, x_3 — 11-й лопатки, x_4 — 12-й лопатки, x_5 — 13-й лопатки, x_6 — 17-й лопатки

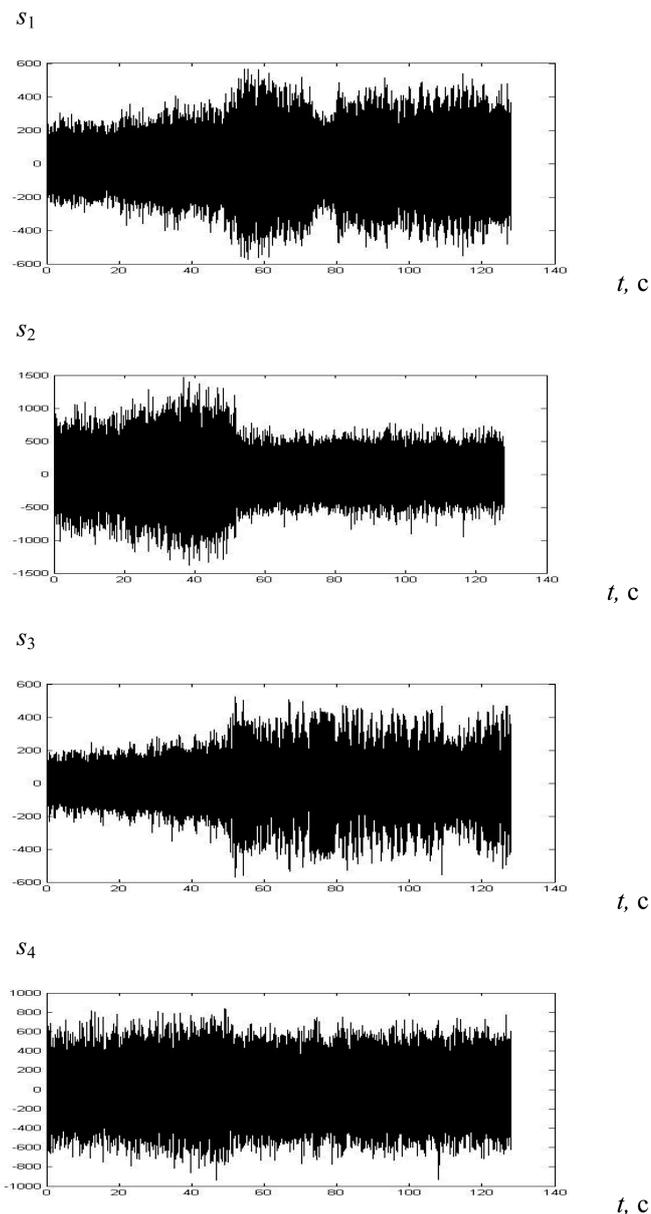


Рис. 2. Независимые компоненты записей колебаний лопаток, приведенных на рис. 1

что предварительный анализ независимых компонент одновременных записей неустойчивых колебаний лопаток при флаттере позволяет оценить полноту синхронизации этих неустойчивых колебаний. На рис. 5 представлена зависимость от времени крутильной частоты, определенная непосредственно по трем записям колебаний лопаток, представленным на рис. 1 (размер N последовательных сегментов временных рядов соответствует временному интервалу 0,5 с).

На рис. 5 показан процесс синхронизации колебаний лопаток при крутильном флаттере. Как и на рис. 4, крутильные частоты лопаток уменьшаются по мере развития флаттера. Собственные

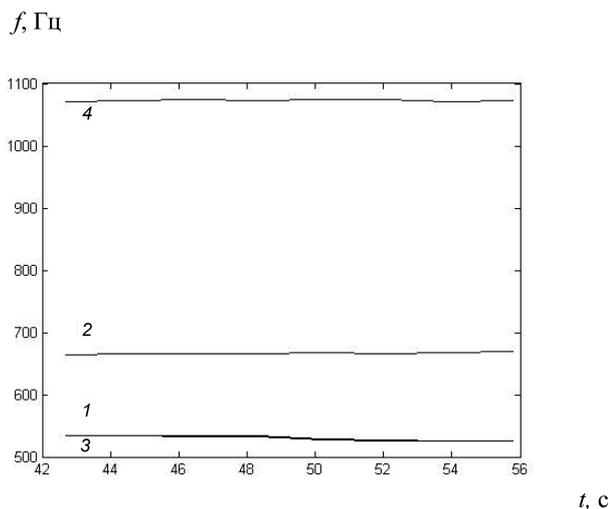


Рис. 3. Зависимость от времени частот главных спектральных составляющих независимых компонент s_1, s_2, s_3 и s_4

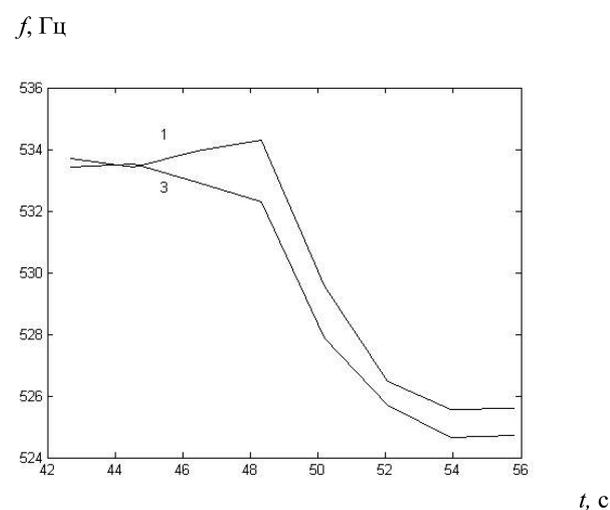


Рис. 4. Зависимость от времени частот главных спектральных составляющих независимых компонент s_1 и s_3

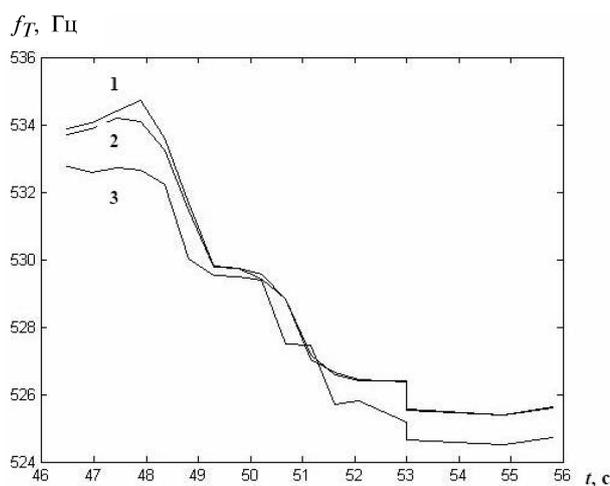


Рис. 5. Зависимости от времени крутильных частот f_T лопаток ротора:
1 — лопатки 1, 2 — лопатки 17 и 3 — лопатки 12

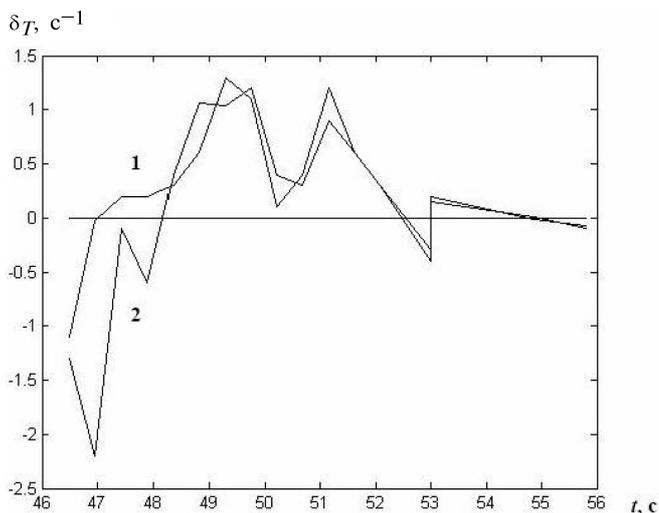


Рис. 6. Зависимости от времени демпфирования δ_T на крутильной частоте лопаток ротора:
1 — лопатки 1 и 2 — лопатки 17

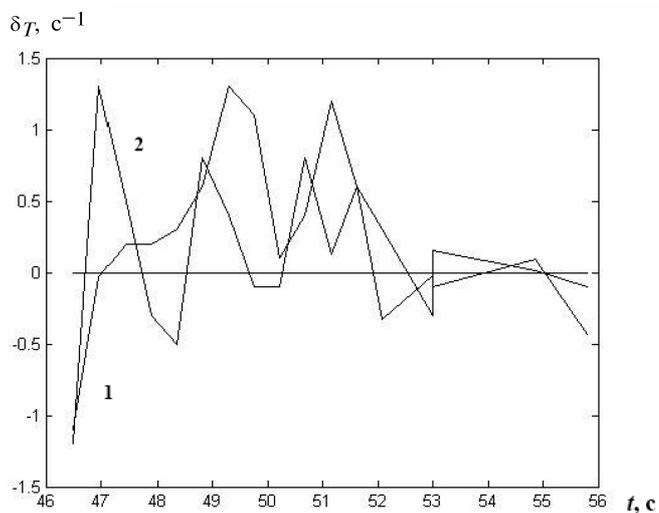


Рис. 7. Зависимости от времени демпфирования δ_T на крутильной частоте лопаток ротора:
1 — лопатки 1 и 2 — лопатки 12

частоты f_T крутильной моды 1-й и 17-й лопаток более близкие, поэтому процесс синхронизации их колебаний при развитии крутильного флаттера более ярко выражен. На рис. 6 представлены зависимости демпфирования δ_T от времени на крутильных частотах лопаток 1 и 17 (двух лопаток с более ярко выраженным процессом синхронизации колебаний). На рис. 5 и 6 видно, что приблизительное равенство крутильных частот f_T и близость демпфирования δ_T на крутильных частотах лопаток наступает в одном и том же временном интервале.

На рис. 7 для сравнения показаны зависимости от времени демпфирования δ_T на крутильной час-

тоте лопаток ротора: 1 — лопатки 1 и 2 — лопатки 12 при развитии крутильного флаттера. Собственная крутильная частота лопатки 12 заметно меньше собственных крутильных частот лопаток 1 и 17 (близких к номинальной) (см. рис. 5), и в процессе развития неустойчивых колебаний лопаток (флаттера) полная синхронизация колебаний лопатки 12 с колебаниями остальных лопаток, представленными на рис. 1, не наступает — ее частота крутильных колебаний остается примерно на 0,9 Гц меньше частот крутильных колебаний всех остальных лопаток ротора. Понятие постоянного сдвига фаз (9) неприменимо к лопатке 12. За один период колебаний этого расстроенного каскада лопаток колебание лопатки 12 проходит все сдвиги фаз с колебаниями соседних лопаток от 0 до 360° [12]. В результате ее фактор демпфирования оказывается усредненным по всем сдвигам фаз (9). Колебания лопатки 12 не синхронизируются с колебаниями остальных лопаток ротора, и она единственная не демонстрирует тенденцию к сближению временной зависимости фактора демпфирования на крутильной частоте по мере развития неустойчивости (флаттера) (рис. 7).

Список литературы

1. Weiss L., McDonogh R. N. Prony's method, Z-transform, and Padé approximation // SIAM Review. 1963. V. 9. № 2. P. 145—149.
2. Кухаренко Б. Г. Технология спектрального анализа на основе быстрого преобразования Прони // Информационные технологии. 2008. № 4. С. 38—42.
3. Cichocki A., Amari S. Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications. John Wiley & Sons. 2002.
4. Belouchrani A., Abed-Meraim K., Cardoso J. F., Molines E. A blind source separation technique using second-order statistics // IEEE Transactions on Signal Processing. 1997. V. 45. P. 434—444.
5. Cardoso J.-F., Souloumiac A. Jacobi angles for simultaneous diagonalization // SIAM Journal of Matrix Analysis and Applications. 1996. V. 17. N 1. P. 161—164.
6. Ziehe A., Laskov P., Nolte G., Müller K.-R. A fast algorithm for joint diagonalization with non-orthogonal transformations and its application to blind signal separation // Journal of the Machine Learning Research. 2004. V. 5. P. 777—800.
7. Li X.-L., Zhang X.-D. Sequential blind extraction adopting second-order statistics // IEEE Signal Processing Letters. 2007. V. 14. N 1. P. 58—61.
8. Kerschen G., Poncelet F., Golinval J.-C. Physical interpretation of independent component analysis in structural dynamics // Mechanical Systems and Signal Processing. 2007. V. 21. P. 1561—1575.
9. Poncelet F., Kerschen G., Golinval J.-C., Verhelst D. Output-only modal analysis using blind source separation techniques // Mechanical Systems and Signal Processing. 2007. V. 21. P. 2335—2358.
10. Фролов К. В., Кухаренко Б. Г., Минаев А. Я. Идентификация мод колебаний с использованием анализа независимых компонентов // Проблемы машиностроения и надежности машин. 2008. № 5. С. 13—18.
11. Crawley E. F. Aeroelastic formulation for tuned and mistuned rotors // AGARD Manual on Aeroelasticity in Axial-Flow Turbomachines. AGARD-AG-298. 1988. V. 2. Structural Dynamics and Aeroelasticity, Chapter 19.
12. Sadeghi M., Liu F. Computation of mistuning effects on cascade flutter // AIAA Journal. 2001. V. 39. N 1. P. 22—28.

В. В. Нечаев, д-р физ.-мат. наук (ВМАКК),
зав. каф. МИРЭА,
e-mail: nechaev@mirea.ru

Концептуальное модельное представление задачи как системы

Рассматривается проблема создания и формализованного представления концептуальной модели задачи (КМЗ) как целостной системы. На основе результатов системно-комплексного анализа предлагается версия КМЗ, ориентированная на системные приложения. Необходимость такой модели обусловлена настоящей потребностью автоматизации процессов формирования, описания, постановки и решения системных задач в различных предметных областях деятельности.

Ключевые слова: задача, метазадача, модель, метамодель, система, данные, потребность, решатель задач, метод, адекватность, релевантность, интеллект, алгоритм, информация, технология, системная задача, концептуальная модель, информационная технология, интеллектуальный решатель задач.

Введение

Деятельность человека имеет две основные формы: умственную — "деятельность мозга" и физическую — "деятельность рук". Очевидно, что и умственная, и физическая деятельность сопряжены с постоянной необходимостью решения разнообразных задач [1]. Спектр таких задач определяется многочисленными факторами как существенными, так и малозначимыми. К существенным факторам прежде всего следует отнести причины возникновения задачи. В качестве первопричины, порождающей задачу, выступают *потребности человека*. Удовлетворение потребностей осуществляется в процессе взаимодействия человека не только с внешней, но и с собственной внутренней средой, т. е. с природой, с другими людьми или самим собой. Реализация потребностей всегда связана с решением соответствующих этим потребностям задач. Практически весь спектр потребностей человека определяется тремя сферами его деятельности: материальной, социальной и духовной [2]. Следовательно, можно констатировать, что возникновение задач обусловлено материальными, социальными и духовными факторами. К существенным факторам, определяющим содержание и свойства задач, следует также отнести: предметную область существования задачи, ее размерность и уровень сложности,

степень определенности, возможности формализации, полноту исходных данных, требования к конечному результату, а также ряд других свойств, характеристик, параметров и особенностей.

За длительный период эволюционного развития человечество накопило богатый опыт решения задач различных классов и типов, который обобщен, формализован и представляется в форме методов, способов, приемов, алгоритмов и т. п., иными словами, в виде теоретического и эмпирического знания. Тем не менее, практическая деятельность человека постоянно порождает новые, все более сложные в качественном и количественном аспектах задачи, решение которых зачастую определяет прогресс общества. Сложность и масштабы такого рода задач, как правило, превосходят физиологические возможности восприятия и обработки необходимых объемов информации не только индивидуального — "единичного головного мозга" человека — *первый информационный барьер* (по В. М. Глушкову), но и социального — "коллективного мозга" человека, определяющего *второй информационный барьер*. Для преодоления информационных барьеров, отражающих ограниченные возможности живого (витального) мозга к восприятию, переработке и хранению информации, человек создал универсальную электронную вычислительную машину — ЭВМ. Основанные на модели фон Неймана (1945 г.) ЭВМ с последовательным принципом обработки информации, как показал в своих исследованиях академик Э. В. Евреинов [3], имеют принципиально неустранимые ограничения — физические барьеры — на скорость и объемы обрабатываемой информации (*третий информационный барьер*). Только модель коллектива вычислителей, предложенная Э. В. Евреиновым в 1954 г. и реализующая принцип параллельной обработки информации, дала возможность создавать вычислительные средства, полностью снимающие принципиальные физические ограничения на объемы и скорости обработки информации.

В настоящее время технический уровень развития программно-аппаратных средств электронной вычислительной техники (ЭВТ), компьютерных и информационных технологий дают возможность успешно разрешать проблемы накопления, обработки, передачи, запоминания, хранения и представления информации (данных и знаний), формирования в компьютерных средах значительных информационных ресурсов в целях многократного их использования для решения различных, в том числе очень больших и сложных, задач. Современные подходы к решению таких задач реализуются в рамках технологий баз и хранилищ данных (знаний), информационных и автоматизированных, а также экспертных и интеллекту-

альных систем. Инструментальные и информационные средства поддержки решений сложных задач развиваются по пути их универсализации для различных приложений. Иными словами, эволюция идет в направлении индустриализации компьютерных и информационных технологий решения сложных системных задач. Можно констатировать, что в триаде "Задача — Данные — Решатель" компоненты "Решатель" и "Данные" на сегодняшний день уже имеют явно выраженную универсальную форму реализации, т. е. ориентированы на многократные, не обязательно однородные, приложения. Однако ведущая компонента этой триады — "Задача", ради которой в конечном счете и существуют две другие — "Решатель" и "Данные", до настоящего времени остается "индивидуально представленной", "кустарно реализуемой" в форме конкретного алгоритма и машинной программы. Такое положение объясняется, по меньшей мере, следующими причинами:

- чрезвычайным разнообразием задач, высокой степенью проблемной ориентации каждой из них, а также масштабами (размерностью) и сложностью;
- сложностью создания универсальной формальной модели задачи и ее реализации, ориентированной на информационные и компьютерные технологии;
- узостью спектра профессиональных знаний конкретных специалистов, для которых решаются на ЭВМ задачи, а как следствие — априорным отсутствием у таких специалистов представлений о возможности индустриализации и автоматизации технологий и средств решения сложных задач.

Для разрешения существующей проблемы и создания *индустриальной технологии решения* задач (ИТР-задач) необходимо в триаде "Задача—Данные—Решатель" устранить до сих пор существующее "узкое место" — противоречие между универсальными инструментальными средствами, реализующими компоненты "Решатель" и "Данные", и индивидуально-конкретным "кустарным" представлением компонента "Задача". Разрешение указанного противоречия возможно посредством создания такой модели задачи, которая легко бы адаптировалась к конкретным прикладным задачам (КПЗ) пользователей независимо от их содержания и сложности. По существу, речь идет о создании концептуальной модели задачи (КМЗ), выступающей в качестве метазадачи или метаметазадачи по отношению к конкретной пользовательской задаче [4, 5]. В прикладном аспекте такая концептуальная модель задачи должна выступать в качестве своеобразного каркаса интеллектуального задачного интерфейса между конкретной поль-

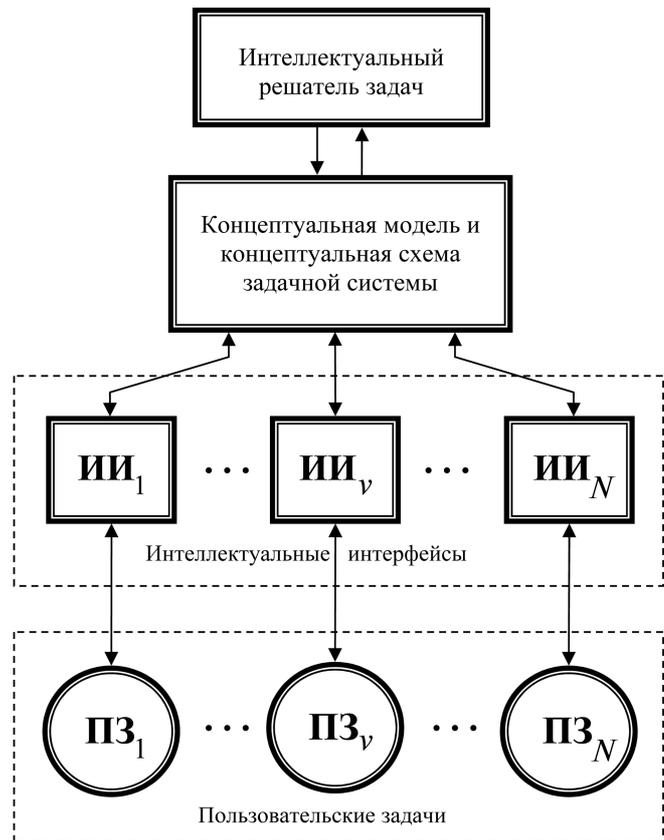


Рис. 1. Архитектура задачного решающего комплекса

зовательской задачей и соответствующими инструментальными средствами ее решения (рис. 1).

Рассматриваемый подход к индустриализации технологии решения задач дает возможность формировать в машинной среде на основе КМЗ машинную форму ее представления, независимую от конкретной прикладной задачи. КМЗ и ее концептуальная схема, представленные в форме интеллектуальной программно-информационной оболочки, дают возможность разделить и сделать независимыми друг от друга конкретные пользовательские задачи и инструментальные средства, осуществляющие решение таких задач (рис. 1).

Более того, возникает возможность формирования инвариантных к КПЗ *однородных задачных структур* (ОЗС), погруженных в среду задачной оболочки и ориентированных на реализацию в *однородных вычислительных средах* (ОВС). Предлагаемый подход дает возможность ввести новый вид внутреннего параллелизма и представления задач в форме однородных заданных структур, органически сопрягаемых с ОВС-решателем.

Таким образом, концептуальная модель задачи, представленная в решающей вычислительной среде в виде специфического инструментального средства, осуществляет функции *универсального задачного интерфейса* (УЗИ), согласующего представление КПЗ с решающей инструментальной

средой (решателем). При таком подходе возникает возможность представления КМЗ в совокупности с УЗИ в качестве стандартного программно-информационного обеспечения, образующего вместе с ЭВМ (или решающей средой) *задачный решающий комплекс* (ЗРК). Адаптация ЗРК к конкретным прикладным задачам на основе УЗИ требует заполнения соответствующих баз данных и знаний, а также создания механизмов (алгоритмов) их использования на основе предметно-ориентированных интерпретаторов. Принимая во внимание изложенное, можно сделать заключение о том, что речь идет о создании принципиально нового интеллектуального компьютерного средства — *интеллектуального решателя задач* (ИРЗ).

В настоящей статье рассматривается авторская версия концептуальной модели задачи. КМЗ разрабатывалась на основе совокупности методов системно-комплексного анализа (СК-анализа) "*внутреннего устройства задачи*" как целостной системы и ее внешнего окружения — среды существования задачи. Для успешного разрешения проблемы создания КМЗ использовались:

- системно-комплексный подход и СК-анализ;
- основы теории концептуального моделирования;
- методологические аспекты метамоделирования;
- технология концептуального метамоделирования (КММ-технология), а также метод раскрытия неопределенности системной задачи на уровне ее концептуального представления, основанный на метазадачной технологии и рекурсивных механизмах, приводящий к рекурсивным и взаимно рекурсивным структурам системной КМЗ.

Изложенный подход к созданию и использованию КМЗ по своей сути может быть определен как концептуальная метамодельная технология автоматизированного описания, представления, постановки и решения системных задач. Перейдем к рассмотрению компонент и этапов КММ-задачной технологии.

Концептуальные модели описания и постановки системной задачи

Возникновение задачи. Задача — продукт интеллектуальной деятельности мозга человека. Следовательно, задача — это, прежде всего, информационный объект. Первопричиной, порождающей задачу, выступает *неудовлетворенная потребность*. Весь обширный спектр потребностей человека в общем случае может быть представлен тремя взаимосвязанными и взаимодействующими сферами деятельности (рис. 2): материальной (М), социальной (С) и духовной (Д) [2], в соответствии с которыми и формируются задачи.

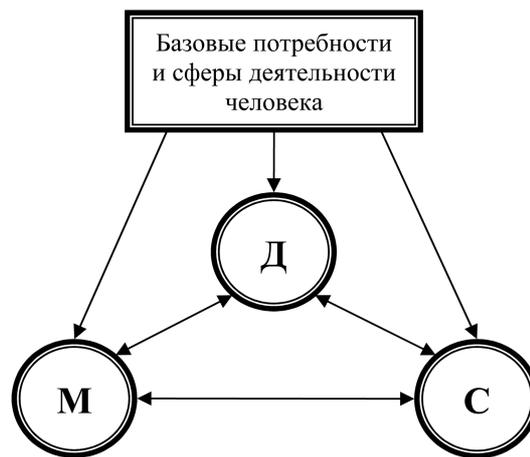


Рис. 2. Граф-модель базовых потребностей человека в контексте сфер его деятельности

Неудовлетворенная потребность, в свою очередь, приводит к противоречию. Противоречие в данном случае определяется как различие (несовпадение) действительного и желаемого состояний [6]. При этом степень неудовлетворенности потребности характеризуется величиной различия действительного и желаемого состояний, т. е. степенью "невязки" этих состояний, которая и определяет *глубину противоречия*. С точки зрения *проблемологии* — науки о задачах — несовпадение состояний, приводящих к противоречию, определяется как *проблемная ситуация*. Содержательное формулирование проблемной ситуации с указанием всех известных на данном этапе факторов дает возможность трансформировать проблемную ситуацию в проблему. Отметим, что глубина противоречия характеризует степень сложности проблемы. В тех случаях, когда в рамках проблемы оказывается возможной ее конкретизация и, как минимум, определяется среда существования проблемы, а также цель, на достижение которой направлено разрешение проблемы, последняя трансформируется в задачу.

Понятие задачи. В русском языке, по-видимому, трудно найти более обширное понятие, чем то, которое определяется термином *задача*. Существует широкий спектр определений понятия "*задача*". Приведем две формулировки, в определенном смысле удовлетворяющих контексту настоящей работы. Известный математик Д. Пойа [7] определяет задачу как "*необходимость сознательного поиска соответствующего средства для достижения ясно видимой, но непосредственно недоступной цели*". Второе определение несет в себе некоторый психологический оттенок. *Задача* — это заданная в определенных условиях цель деятельности, которая должна быть достигнута преобразованием этих условий согласно определенной процедуре. Приведенные формулировки понятия за-

дачи включают конструктивные элементы и могут быть использованы как рабочие.

Задача как информационный объект.

Исходя из представления, что задача — это информационный объект, рассмотрим ее определение на основе информационного подхода. С информационной точки зрения под задачей будем понимать осознанную в соответствии с целью деятельности необходимость определения непосредственных связей (отношений) между двумя или более информационными совокупностями (множествами), которые на начальном этапе рассмотрения (в исходном состоянии) в явном виде между собой не связаны, однако декларативно известно, что при определенных условиях такие связи — отношения существуют и могут быть определены. В развитие приведенного определения отметим, что одна из информационных совокупностей идентифицирует область определения задачи — ее исходные данные, вторая — определяет область значений — результаты решений задачи, а третья отражает условия, при которых устанавливаются связи, т. е. конкретизируются отношения между исходными данными и целью задачи — результатом ее решения (рис. 3). Штриховыми линиями на схеме отражены опосредованные (не прямые) связи (влияния).

Замечание 1. Задача как продукт интеллектуальной деятельности мозга всегда является информационным объектом. Однако в зависимости от сферы деятельности, в которой реализуется эта задача, результат ее решения может быть как идеальным, так и материальным. Для сложных задач материальный и идеальный результаты ее решения могут существовать одновременно, дополняя друг друга в различных соотношениях.

Концептуальная модель описания системной задачи. Создание концептуальной модели задачи основано на ее представлении как системы. Такое представление базируется на методологии теории систем, системно-комплексном (СК) подходе и, в частности, СК-анализе [8]. Исходя из общих представлений о системах, введем понятие системной задачи.

Определение. Системная задача — это внутризадачная целенаправленно организованная совокупность взаимосвязанных компонентов, функционирование которых направлено на достижение конечного результата — цели задачи.

Представление задачи совокупностью взаимосвязанных и взаимодействующих (функционирующих) компонентов возможно осуществить в результате ее анализа на основе методов стратификации и декомпозиции. По существу речь идет

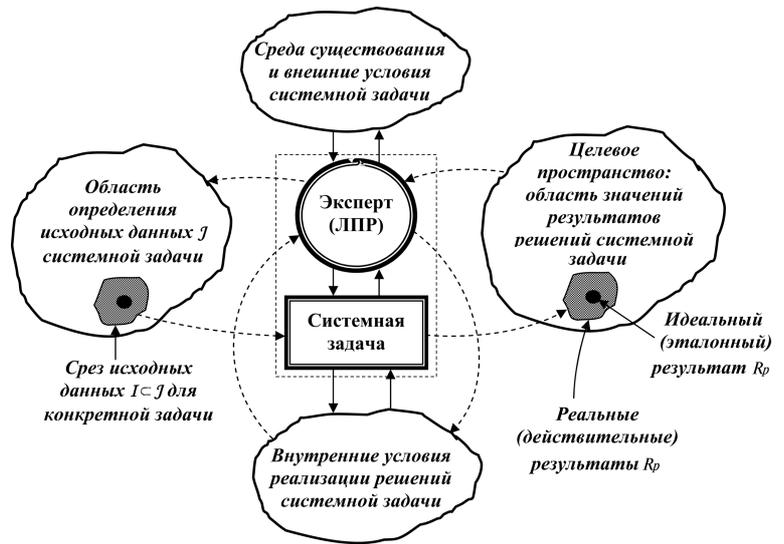


Рис. 3. Структурная модель информационного представления системной задачи

о СК-анализе задачи на концептуальном уровне ее представления.

СК-анализ предполагает введение в рассмотрение двух взаимосвязанных объектов: системной задачи и окружающей эту задачу внешней среды — среды существования задачи (рис. 3). Взаимосвязанное комплексное использование методов стратификации и декомпозиции дает возможность в рамках СК-анализа выявить наиболее существенные аспекты и компоненты как внешней среды существования задачи, так и внутреннего "устройства" самой задачи.

Концептуальная модель внешнего описания системной задачи

Первым шагом на пути к концептуальному описанию системной задачи P_{Σ} , а как следствие — к формированию КМЗ, должно быть признание факта ее существования, т. е. восприятие и осознание наличия проблемной ситуации и проблемы. Задача P_{Σ} как система существует в окружающей ее среде Σ_p . Если свойства и характеристики среды Σ_p , в которой порождается и существует задача P_{Σ} , известны, то они задаются при внешнем описании задачи. Будем исходить из представлений, что задача P_{Σ} имеет место в тех случаях, когда (рис. 3):

- определена среда Σ_p существования задачи P_{Σ} ;
- известны или определены: потребность, противоречие, проблемная ситуация, проблема, а также сфера, предметная область и объект деятельности, в рамках которых рассматривается задача P_{Σ} ;
- определена и описана цель Z_p , под которой будем понимать модель желаемого идеализиро-

ванного (эталонного) результата R_p^z решения задачи P_Σ ;

- в рамках области определения (рис. 3) выявлена, целесообразным образом сгруппирована и в удобной форме представлена необходимая для решения задачи P_Σ информация I_p , которая включает исходные данные D_p и исходные знания K_p , используемые для получения результата R_p , определяемого целью Z_p ;
- известны, определены или сформулированы условия C_p , конкретизирующие отношения между исходной информацией $I_p = \langle D_p, K_p \rangle$ и целью Z_p — идеализированным (эталонным) результатом R_p^z решения рассматриваемой задачи.

С учетом перечисленных аспектов, определяющих внешние условия существования задачи P_Σ , ее символическое (семиотическое) представление в форме концептуальной модели определится кортежем (четверкой) вида

$$P_\Sigma = \langle \Sigma_p, Z_p, C_p, I_p \rangle. \quad (1)$$

Концептуальная модель внутреннего описания системной задачи

Концептуальная модель, представленная записью (1), отражает внешний "взгляд" на задачу как целостную систему. Входящая в эту модель компонента C_p определяет условия задачи и отражает (с внешней точки зрения) факт необходимости существования связей или отношений между целью Z_p и информацией I_p — исходными данными задачи P_Σ . Однако при внешнем анализе содержание этих условий раскрыть не представляется возможным. Определение структуры и содержания условий C_p — проблема внутреннего (интроспективного) анализа задачи P_Σ .

Условия C_p являются одним из определяющих компонент системной задачи P_Σ . Как уже отмечалось, они конкретизируют связи (отношения) между требуемым результатом R_p , характеризующимся целью Z_p задачи P_Σ и исходной информацией $I_p = D_p \cup K_p$. Четко сформулированные условия C_p декларируют существование связей (отношений) между известным — исходной информацией I_p и неизвестным — результатом R_p ; констатируют, что необходимо для получения желаемого результата R_p решения задачи P_Σ , а также определяют требования к результату R_p^z , т. е. каким условиям, задаваемым показателем адекватности T_p , этот результат должен удовлетворять.

С учетом перечисленных особенностей рассмотрим условия C_p системной задачи P_Σ на основе анализа ее внутреннего "устройства", т. е. таких внутренних компонент, которые на концептуальном уровне представления инвариантны конкретным задачам $\{P_v: v = 1, \dots, N_p\}$.

СК-анализ условий C_p решения системной задачи P_Σ дает возможность в качестве базовых инвариантных внутренних компонент, определяющих такие решения, выделить: метод — M_p , алгоритм — A_p , программу — P_p и адекватность — T_p . Рассмотрим каждую из этих компонент.

Метод M_p решения системной задачи P_Σ определяет, что надо делать для получения результата R_p . Однако метод M_p не отражает возможности и средства реализации требуемого решения для получения результата R_p системной задачи P_Σ . В содержательном аспекте метод M_p решения задачи P_Σ представляет совокупность принципов, способов и приемов, направленных на достижение теоретических или практических результатов решения системной задачи P_Σ .

Алгоритм A_p предназначен для реализации метода M_p решения задачи P_Σ . По определению, алгоритм представляется как некоторое предписание — логически заверченный процесс, ведущий от варьируемых начальных данных (исходной информации I_p) к искомому результату R_p . Алгоритм A_p определяет механизмы получения требуемого результата R_p решения задачи P_Σ на основе метода M_p . Формы представления алгоритма могут быть различными. Они определяются типом задачи, ее содержанием, методом и средствами практической реализации решения.

Программа P_p реализации алгоритма A_p , как правило, ориентирована на компьютерные технологии. Программа представляется последовательностью предложений определенного языка программирования, описывающих алгоритм A_p решения задачи P_Σ . По отношению к ЭВМ программа определяет упорядоченную последовательность инструкций (или команд), подлежащих выполнению.

Адекватность T_p — еще одна компонента, входящая в состав условий C_p , отражающая качественные и количественные требования к результату R_p решения задачи P_Σ . Адекватность T_p — комплексный показатель, определяющий максимально допустимое несоответствие между идеальным (эталонным) R_p^z и действительным (реальным) R_p^d результатами решения системной задачи P_Σ .

С учетом выделенных и описанных в рамках условий C_p компонент, отражающих внутреннее содержание задачи P_Σ , определим концептуальную модель внутреннего описания системной задачи кортежем (четверкой) вида

$$C_p = \langle M_p, A_p, P_p, T_p \rangle. \quad (2)$$

Концептуальная модель полного описания системной задачи

Проведенный СК-анализ, результаты которого представлены концептуальными моделями внеш-

него (1) и внутреннего (2) описаний системной задачи P_{Σ} , дают возможность сформировать системно-комплексную концептуальную модель полного описания задачи P_{Σ} . Такая модель представляется семиотической системой, состоящей из кортежей (1) и (2), взаимосвязь между которыми реализуется через компоненту C_P — условия задачи P_{Σ} :

$$\left. \begin{aligned} P_{\Sigma} &= \langle \Sigma_P, Z_P, C_P, I_P \rangle; \\ C_P &= \langle M_P, A_P, P_P, T_P \rangle. \end{aligned} \right\} \quad (3)$$

Концептуальную модель, представленную системой (3), возможно и целесообразно доопределить посредством конкретизации элементов кортежей. Как уже отмечалось, информация I_P включает данные D_P и знания K_P . Следовательно, она может быть задана двойкой:

$$I_P = \langle D_P, K_P \rangle. \quad (4)$$

Адекватность T_P определяется через два комплексных показателя: качества — Q_{WP} и эффективности — E_{fP} . Показатель качества Q_{WP} характеризует степень близости реального результата R_P^d решения задачи P_{Σ} к теоретически обоснованному эталонному (идеальному) R_P^z . Показатель эффективности E_{fP} определяет комплексные затраты — "цену", необходимую для достижения заданного качества Q_{WP} . Следовательно, адекватность T_P может быть представлена кортежем, включающим две компоненты:

$$T_P = \langle Q_{WP}, E_{fP} \rangle. \quad (5)$$

Таким образом, концептуальная модель полного описания задачи P_{Σ} , полученная на основе СК-анализа, представляется системой кортежей (3), (4) и (5), взаимосвязанных друг с другом через соответствующие компоненты:

$$\left. \begin{aligned} P_{\Sigma} &= \langle \Sigma_P, Z_P, C_P, I_P \rangle; \\ C_P &= \langle M_P, A_P, P_P, T_P \rangle; \\ I_P &= \langle D_P, K_P \rangle; \\ T_P &= \langle Q_{WP}, E_{fP} \rangle. \end{aligned} \right\} \quad (6)$$

Структурная схема концептуальной модели системной задачи P_{Σ} , определяемая системой кортежей (6) и сформированная на основе стратифицированного подхода, представлена на рис. 4.

Замечание 3. Система кортежей (6) как по своему составу, так и по содержательному описанию на концептуальном уровне, удовлетворяет условиям полноты и целостности, а также необходимости и достаточности концептуального модельного представления системной задачи P_{Σ} .

Концептуальная модель постановки задачи

Постановка задачи — вторая фаза ее развития на пути к достижению цели. Важность этой фазы

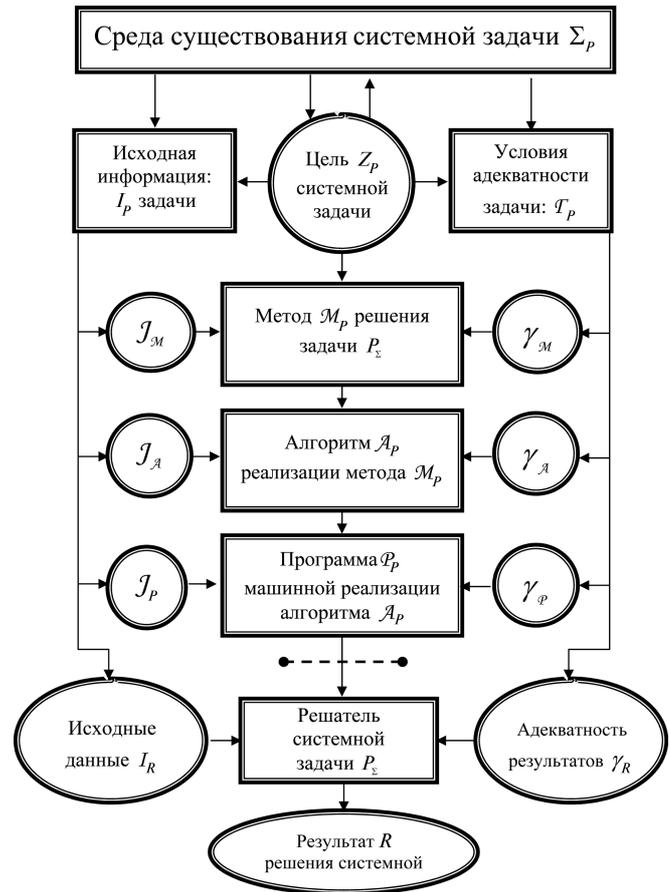


Рис. 4. Структурная схема КМ системной задачи

подчеркивали акад. В. М. Глушков, чл.-корр. РАН В. Г. Хорошевский и многие другие исследователи. Джон Дьюи в своих работах отмечал, что "хорошо поставленная задача уже наполовину решена". В традиционных постановках задач формулируется требование к достижению цели, т. е. определению конечного результата. Однако невозможно определить результат решения задачи, если о нем ничего не известно. Следовательно, в постановке задачи необходимо также указать и что известно (или что задано) для достижения цели. Таким образом, концептуальная модель постановки задачи должна включать два компонента:

- что известно (или что задано);
- что требуется определить.

С учетом приведенных условий определим концептуальную модель постановки системной задачи на основе концептуальной модели ее описания, представленного в форме (6), для наиболее простого — рутинного случая, т. е. когда задача P_{Σ} полностью определена.

Замечание 4. Системные задачи P_{Σ} , для которых полностью определены условия C_P и задана исходная информация I_P , а требуется только выполнение вычислительных процедур и операций, реализация ко-

торых однозначно приводит к конечному результату R_P^d , будем определять как рутинные.

Постановка рутинной задачи. Рассмотрим вариант постановки задачи P_Σ , когда для получения конечного результата R_P ее решения необходимо и достаточно выполнить над исходной информацией I_P совокупность процедур и операций (команд), определяемых программой P_P . Иными словами, рассматривается полностью определенная задача P_Σ , т. е. такая, для которой полностью определены условия C_P и задана исходная информация I_P . С учетом приведенных требований концептуальная модель постановки рутинной задачи P_Σ будет иметь следующий вид:

Известно:

$$\left. \begin{aligned} P_\Sigma &= \langle \Sigma_P, Z_P, C_P, I_P \rangle; \\ C_P &= \langle M_P, A_P, P_P, T_P \rangle. \end{aligned} \right\} \quad (7)$$

Требуется определить $Z_P = ?$

Замечание 5. В тех случаях, когда сформулированные в "Замечании 3" условия не выполняются, т. е. помимо искомого конечного результата R_P^d имеется одно или более других неизвестных, системную задачу P_Σ будем относить к не полностью определенным или недоопределенным. Для сведения недоопределенных задач P_Σ к полностью определенным, т. е. рутинным, в данной работе использован метод рекурсивного оформления КМЗ для каждого неизвестного исходной задачи P_Σ [5].

Заключение

Подводя итог изложенному выше, необходимо отметить: в соответствии с заявленной в статье целью настоящей работы, а именно — включение в архитектуру решателя системных задач концептуального задачного уровня, разделяющего внешнее (пользовательское) и внутреннее (компьютерное) представление решаемой задачи, автором предложена и разработана версия концептуальной модели задачи, которая выступает в качестве ядра и схемы концептуального уровня решателя системных задач.

Предлагаемая КМЗ релевантна достаточно широкому спектру приложений. Она, подобно концептуальным моделям данных и знаний, дает возможность продуктивно использовать компьютерные и информационные технологии на всех фазах жизненного цикла системной задачи от возникновения потребности в ее решении до получения конечного результата. Использование на протяжении ряда лет рассмотренной КМЗ в теоретических и прикладных исследованиях при выполнении НИР, курсовых и дипломных проектов, а также диссертационных работ [9—12] дает возможность сделать заключение о ее практической полезности.

Список литературы

1. Спиридонов В. Ф. Психология мышления: Решение задач и проблем: Учеб. пос. М.: Генезис, 2006. 319 с.
2. Нечаев В. В., Дарьин А. В. Информация и личность: персональные информационные ресурсы // Матер. Международного форума информатизации МФИ-97. М.: Междунар. изд-во "Информациология", 1997. С. 150—157.
3. Евреинев Э. В. Однородные вычислительные системы, структуры и среды. М.: Радио и связь, 1981. 208 с.
4. Нечаев В. В. Творческая задача и интеллектуальный решатель: концептуальный подход // Сб. науч. тр. "Модели и системы представления знаний." М.: МИРЭА, 1990. С. 53—68.
5. Нечаев В. В. Концептуальное метамодельное представление творческих задач // Сб. науч. тр. "Методы анализа и оптимизации сложных систем". М.: ГосИФТП, 1993. С. 118—124.
6. Горский Ю. М. Системно-информационный анализ процессов управления. Новосибирск: Наука, 1988. 327 с.
7. Пойа Д. Математическое открытие. Решение задач: основные понятия, изучение и преподавание. М.: Наука, 1976. 448 с.
8. Нечаев В. В. Системно-комплексный анализ индустриальной технологии моделирования // Тез. докл. XLIII Всесоюзной научной сессии, посвященной дню радио. Ч. I. М.: Радио и связь, 1988. С. 18—19.
9. Грознов К. Ю. Психологический интеллектуальный интерфейс адаптивного общения пользователя с вычислительной системой // Дисс. на соискание ученой степени канд. техн. наук. М.: МИРЭА (ТУ), 1995. 173 с.
10. Ильин С. Н. Инструментальная среда поддержки решения системных задач // Дисс. на соискание ученой степени канд. техн. наук. М.: МИРЭА (ТУ), 1997. 134 с.
11. Мартынова Т. Л. Система поддержки процесса формирования интеллектуальных умений в области прикладной информатики // Дисс. на соискание ученой степени канд. техн. наук. М.: ИНЭУМ, 1999. 150 с.
12. Дарьин А. В. Концептуальное метамодельное представление персональных информационных ресурсов // Дисс. на соискание ученой степени канд. техн. наук. М.: МИРЭА (ТУ), 2000. 167 с.

УДК 529.58

Е. Г. Брындин,

Исследовательский центр

"ЕСТЕСТВОИНФОРМАТИКА", Новосибирск,

e-mail: bryndin@ngs.ru

Теоретические аспекты технологии непрерывной обработки на виртуальной памяти

Рассматриваются теоретические аспекты, решающие проблему непрерывной автоматизированной обработки больших объемов информации на виртуальной памяти без задержки получения результатов из-за обменов.

Ключевые слова: программа с детерминированно-связанными модулями, непрерывная обработка на виртуальной памяти, упреждающее перемещение общих данных, упреждающий анализ связей модулей, упреждающее замещение модулей программы на оперативной памяти.

1. Непрерывная обработка на виртуальной памяти программ с детерминированно-связанными модулями

Вычислительные процессы осуществляют обработку, поиск, перемещение, замещение, запись информации, а также управление ресурсами. Вычислительные процессы используют ресурсы памяти для хранения, операционные устройства центрального процессора для обработки, устройства коммутаций для коммутации, устройства обмена для перемещения и замещения, устройства поиска, устройства записи, устройства чтения и другие.

Пространственно-временные акты вычислительного процесса использования ресурсов ЭВМ определяются принципами их организации, методом обработки и методом управления памятью.

Будем рассматривать вычислительные процессы, которые используют упреждающий метод управления виртуальной памятью.

Пусть $P = \{P_v, \dots, P_n\}$ — совокупность модулей программы P с детерминированно-связанными модулями [1–3], и пусть в оперативной памяти для размещения модулей программы имеется m мест, причем $1 < m < n$.

В упреждающем методе управления памятью используются:

- детерминированный, динамический метод распределения ресурсов оперативной памяти;

- упреждающий запрос и параллельный метод замещения модулей программы P . Пусть t_3 — момент начала замещения. Обработанный модуль PI_0 замещается на PI_t , где PI_t — модуль следующий по использованию. Упреждающее обращение к виртуальной памяти для обмена модулей программы осуществляется после того, как появляется обработанный модуль в оперативной памяти;
- упреждающие акты анализа связности модулей программы;
- потоковое упреждающее перемещение значений общих данных между модулями на оперативных сегментах памяти.

Программы с детерминированно-связанными модулями допускают упреждающие анализ цепочек связей, поиск модулей, перемещение модулей и значений общих данных глубины m по текущим данным, где m — число сегментов оперативной памяти.

Упреждающий метод управления позволяет обрабатывать непрерывно программу с детерминированно-связанными модулями на виртуальной памяти. Модули программы обрабатываются последовательно, обработка каждого модуля параллельная.

Вычислительные процессы с упреждающим методом управления памятью непрерывно обрабатывают программу с детерминированно-связанными модулями, если время обработки каждой последовательности из m связанных модулей больше времени анализа связей, перемещения общих данных и параллельного замещения m модулей на $2m$ сегментах оперативной памяти.

Метод замещения модулей программ с детерминированно-связанными модулями на оперативной памяти, упреждающий их обработку, позволяет обеспечить непрерывность обработки на виртуальной памяти по критерию непрерывности, в отличие от всех используемых методов замещения модулей программ со случайными обращениями к модулям.

Вычислительные процессы с упреждающим методом управления памятью — это новое направление в информатике автоматизированной обработки больших программ с детерминированно-связанными модулями.

2. ЭВМ с упреждающим управлением памятью

ЭВМ с упреждающим управлением памятью реализуют параллельно-асинхронное взаимодей-

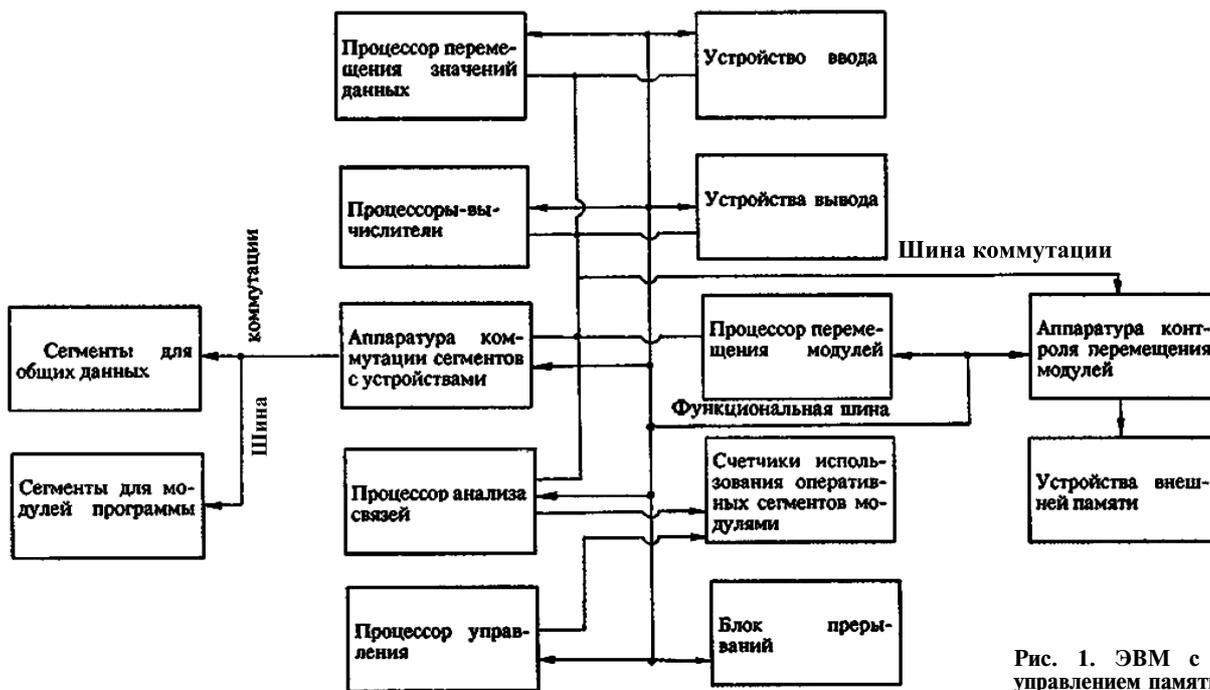


Рис. 1. ЭВМ с упреждающим управлением памятью

ствии актов использования ресурсов ЭВМ под управлением программ с детерминированно-связанными модулями.

ЭВМ содержит: процессор управления, процессоры обработки, оперативные сегменты для общих данных и модулей программы, аппаратуру коммутации оперативных сегментов с устройствами, блок прерываний, устройства внешней памяти, ввода и вывода, процессор анализа связей между модулями, процессор перемещения модулей программы на неоднородной памяти, счетчик использования сегментов оперативной памяти модулями, процессор перемещения значений общих данных по модулям, находящихся на сегментах оперативной памяти, аппаратуру контроля перемещения модулей, шины коммутации, функциональную шину (рис. 1).

ЭВМ содержит новые устройства: процессор анализа связей между модулями программы, счетчики использования сегментов оперативной памяти модулями, процессор перемещения модулей по неоднородной памяти, процессор перемещения общих данных модулей.

Процессор анализа проводит упреждающий анализ связей модулей программ с детерминированно-связанными модулями. Процессор анализа реализует процесс вычисления номера внешнего модуля перемещения на оперативную память по программе связи SPP_i модуля PI_i , процесс корректировки значения счетчика использования сегментов оперативной памяти модулями программы, процесс запуска процессора перемещения модулей программы с внешней памяти на сегменты оперативной памяти.

Процессор перемещения модулей реализует процессы перемещения модулей между устройствами внешней и оперативной памяти, обеспечивая наличие необходимых модулей на оперативной памяти по заявке процессора анализа связей. Процессор перемещения модулей реализует: процесс коммутации сегментов оперативной памяти с сегментами накопителя, процессы передачи и контроля информации модулей программы, процессы прерываний по передаче информации.

Процессор перемещения общих данных реализует перемещение общих данных между модулями. Общие переменные имеют последовательности адресов перемещения из текущих значений. По последовательности адресов перемещения организуются потоки значений общих данных с доставкой их на место использования в модулях на оперативных сегментах. Обращение к модулям происходит по их номерам. Для модулей внешней памяти значения общих переменных переносятся в резидент общих данных при замещении модуля, содержащего общие данные.

Активизация устройств ЭВМ осуществляется процессором управления по пользовательской программе с детерминированно-связанными модулями. Непрерывный вычислительный процесс начинается с загрузки начальных модулей программы с внешней памяти на сегменты оперативной памяти с перемещением общих данных и анализом связей между модулями. При загрузке модулей устанавливаются значения на счетчиках использования оперативных сегментов модулями. Для модулей данных, ввода, вывода значение

счетчика равно единице. При загрузке модуля на сегмент оперативной памяти в управляющую информацию предшествующего по поиску модуля заносится номер оперативного сегмента. Затем процессор управления запускает процессоры обработки на обработку первого модуля. После обработки каждого операционного модуля из значения соответствующего счетчика использования памяти вычитается единица и проверяется его значение на ноль. Для модулей данных значение счетчика обнуляется по команде "обнулить счетчик" из операционного модуля.

Когда значение счетчика становится равным нулю и определен указатель на внешний модуль, тогда иницируется процессор перемещения модулей для замещения использованного модуля на оперативной памяти внешним модулем по указателю. После замещения указатель на внешний модуль обнуляется. Если в модуль на оперативной памяти не было записи, то он не переносится на внешнюю память. Процессор перемещения общих данных по адресным последовательностям переносит накопившиеся значения общих данных с резидента общих данных в замесивший модуль. Процессор анализа корректирует счетчик использования оперативной памяти замесившим модулем и определяет по программе связи замесившего модуля указатель на внешний модуль.

Процессор управления организует обработку, перемещение общих данных, анализ связей и замещение модулей. Он совмещает работу устройств над одним модулем за разные циклы обработок к оперативному сегменту.

Число оперативных сегментов для непрерывной обработки программы с детерминированными модулями определяется в процессе ее трансляции или компиляции.

При технической реализации непрерывное присутствие текущих модулей в оперативной памяти достигается путем баланса скорости упреждающего обработку многоканального параллельного перемещения модулей с внешней памяти в оперативную со скоростью обработки информации из оперативной памяти.

ЭВМ с упреждающим управлением памятью обеспечивает непрерывный процесс обработки программы с детерминированно-связанными модулями упреждающим параллельным перемещением общих данных модулей и упреждающим параллельным замещением модулей на оперативной памяти через многоканальный обмен.

Время получения результата на целевой ЭВМ с упреждающим замещением сокращается на $T = tK$, где t — время одного обмена; K — число замещений модулей на оперативной памяти.

Поток значений общих данных. Готовые к последующей обработке значения общих данных перемещаются по модулям программы, находящимся в оперативной памяти. Для каждого значения общего данного d определяются последовательность использующих его модулей, места использования их в этих модулях и относительные моменты использования значений d в модулях. По множеству модулей использования d составляется дополнительное множество модулей, через которые перемещаются значения данного d .

Значения общих данных перемещаются по модулям, находящимся на сегментах оперативной памяти, динамически, образуя поток данных.

Общие данные модулей, находящихся не на оперативной памяти, перемещаются в резидентные модули ROD. В резидентном модуле общих данных значения хранятся вместе с указателями перемещения. Значения, перемещаемые в один модуль, располагаются подряд. В начале последовательности указывается их количество. После записи новых значений в модуль общих данных перемещается его указатель свободного места (записи), если счетчик модуля общих данных не превышает допустимое число значений.

Значения снабжаются признаками перевычисления. Если признак принимает состояние неизменяемости, то значение перемещается во все используемые модули.

Значения помещаются в модуль общих данных в порядке их перемещения в модули, поступающие с внешней памяти на оперативную. В модуле общих данных значения могут снабжаться несколькими указателями. После перемещения всех значений в модуль программы в нем устанавливается признак "перемещено", который указывает, что модуль готов к обработке.

Пусть имеется k модулей последовательности исполнения и n сегментов оперативной памяти. Пусть первый модуль имеет переменные. Для каждой переменной выпишем номера последующих модулей, в которых она используется. Для второго модуля выпишем все переменные, которых нет в первом модуле. Для каждой переменной выпишем номера последующих модулей, в которых она используется. Для последующих модулей аналогично выпишем последовательности использования переменных, которые не указаны в предыдущих модулях.

Для каждой переменной определим внешние модули. Выпишем последовательности номеров внешних использующих модулей. Переменные будут храниться в резидентном модуле общих данных согласно последовательной нумерации внешних модулей, использующих переменные.

Замещение модулей программы на оперативной памяти. Для хранения модулей программ исполь-

зуется внешняя память. Пусть каждый модуль размещается на отдельном сегменте. И пусть размер сегмента и модуля определяется максимальной порцией обмена информации между устройствами внешней и оперативной памяти.

Сегмент оперативной памяти является локальным адресным пространством для модуля. Областью действия команд модуля, кроме команд процессора перемещения данных, является адресное пространство оперативного сегмента, на котором находится соответствующий модуль. Соответствие между сегментами оперативной памяти и модулями, находящимися на оперативных сегментах, устанавливается через ассоциативные регистры аппаратуры коммутации.

Если используются два сегмента оперативной памяти, то на одном сегменте хранится обрабатываемый модуль, на другом сегменте модуль готовится к обработке. Если используются три оперативных сегмента, то на одном сегменте происходит замещение обработанного модуля на модуль по внешней ссылке, на другом сегменте хранится обрабатываемый модуль, на третьем — модуль готовится к обработке.

Замещение модулей на сегментах оперативной памяти определяется состоянием счетчиков использования сегментов модулями программы P и указателем внешнего модуля аппаратуры контроля перемещения модулей.

Разделение оперативной памяти на независимые сегменты делается для упреждающей подкачки модулей с внешней памяти на оперативную. Перемещаться между устройствами внешней и оперативной памяти одновременно могут несколько модулей.

Аппаратная реализация замещения модулей на оперативной памяти сводит время управления этим процессом практически к нулю, в сравнении с временем их замещения и обработки.

3. Примеры программ с детерминированно-связанными модулями

Рассмотрим несколько программ с детерминированно-связанными модулями.

Сортировка данных. Пусть имеются данные a_1, \dots, a_m . Эти данные нужно пересортировать в неубывающей последовательности. Разместим последовательно исходные данные в k модулях. Программа сортировки перебором сравнений исходных данных следующая:

$i := 0; j := 0; t_1 := 1; t_2 := 1;$
цикл пока $t_1 = k$ **цикл пока** $j \neq c$ **цикл пока** $t_2 \neq k$
цикл пока $i \neq c$ **установить** $\leq (a_{t_1, i}, a_{t_2, i+1});$
 $i := i + 1$ **конец**
сохранить $(t_2); t_2 := t_2 + 1; i := 0$ **конец** $j := j + 1;$
 $t_2 := t_1;$

Структура связи модуля PI_{k+1} с остальными модулями

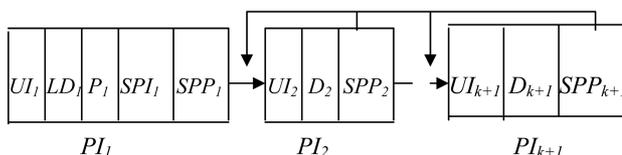


Рис. 2. Структура программы сортировки

$i := j$ **конец сохранить** $(t_1); t_1 := t_1 + 1; t_2 := t_1;$
 $i := 0; j := 0;$
 $i := j$ **конец сохранить** $(t_1); t := t + 1; t := t; i := 0;$
 $j := j + 0$ **конец**

t_1, t_2 — счетчики модулей.

Оператор **установить** $\leq (a_{t_1, i}, a_{t_2, i+1})$ размещает значения указанной пары согласно отношению \leq . По адресу $a_{t_1, i}$ размещается меньшее значение, а по адресу $a_{t_2, i+1}$ большее значение.

Оператор **сохранить** (t_1) отказывается от модуля с номером t_1 с сохранением его.

Структура программы с детерминированно-связанными модулями, определяющая использование операционного модуля и модулей данных, изображена на рис. 2.

Вывод модулей данных назначается в операционном модуле. Программа сортировки размещается в первом модуле, в остальных модулях размещаются данные. Модули модульной канонически связанной программы связаны по номерам, кроме модуля с номером $k + 1$. Структура связи модуля $k + 1$ с остальными модулями с возвратами. Она определяется программой SPP_{k+1} : **если** $t_1 = 1$ **то** PI_2 **если** $t_1 = 2$ **то** PI_3 ... **если** $t_1 = (k - 2)$ **то** PI_{k-1} .

Номера связи модуля и адрес программы связи модуля $k + 1$ хранятся в управляющей информации.

Работа программы сортировки со случайными обращениями к модулям на современной ЭВМ занимает 4 ч, а замещение модулей на оперативной памяти в среднем — 20 ч. Результаты пользователь получает через 24 ч. При непрерывной обработке программ с детерминированно-связанными модулями на ЭВМ с упреждающим управлением памяти пользователь получит результат через 4 ч.

Обработка банков данных. Пусть в банке данных имеются анкеты пяти миллиардов людей, в которых сообщены следующие данные: вес, рост, возраст, национальность, пол, социальное положение, специальность, должность, адрес и т. д. Нужно из пяти миллиардов людей выбрать всех, кто отвечает по своим анкетным данным эталону. Анкетные данные последовательно размещены в k модулях. В каждом модуле размещаются q анкет. Программа выбора анкет по эталону следующая:

$i := 0; t_1 := PODj := 0; t_2 := 1; n := 0; t_3 := k + 1;$
цикл пока $t \neq k$ М: цикл пока $i \neq q$ сравнить (эталон, анкета (t_2, i)) если да то
 записать (t_1, j, t_2, i); $j := j + 1$
 если $j > q$, то сохранить (t_1, t_3); $t_3 := t_3 + 1; n := n + 1$
 если $n = 2$ то $t_1 := t_1 - 2$ иначе $t_1 := t_1 + 1; j := 0$
 на M
 иначе $i := i + 1$ конец освободить (t_2); $t_2 := t_2 + 1$
конец все
 t_1, t_2, t_3 — счетчики модулей.

Подпрограмма **сравнить (эталон, анкета (t_2, i))** сравнивает анкету i из модуля t_2 с эталоном.

Подпрограмма **записать (t_1, j, t_2, i)** записывает анкету i из модуля t_1 в модуль t_2 по месту j .

Структура программы с детерминированно-связанными модулями, определяющая использование первого операционного модуля, k модулей анкет и с $k + 2$ по $k + p$ модулей эталонов, изображена на рис. 3.

Программа обработки хранится в первом модуле. Далее идет последовательность модулей с анкетами. В модулях PI_2, \dots, PI_{k+1} хранятся исходные анкеты. В модулях $PI_{k+2}, \dots, PI_{k+p}$ хранятся эталонные анкеты. Модули связаны по номерам. Номера модулей хранятся в управляющей информации. В резидентном модуле общих данных накапливаются текущие исходные анкеты, совпадающие с эталоном.

Программа синтаксического разбора выражений. Нужно провести синтаксический контроль правильности записи выражения, находящегося в k модулях вводных данных, если понятие выражения определяется следующими синтаксическими правилами:

ВЫРАЖЕНИЕ ::= ТЕРМ I ТЕРМ + ВЫРАЖЕНИЕ
 ТЕРМ ::= МНОЖИТЕЛЬ I МНОЖИТЕЛЬ * ТЕРМ
 МНОЖИТЕЛЬ ::= ИМЯ I ЧИСЛО I ВСКОБКАХ
 ИМЯ БУКВА I < ИМЯ > < БУКВА > I < ИМЯ >
 < ЦИФРА >
 ЧИСЛО ::= ЦИФРА I < ЦИФРА > ЧИСЛО
 ВСКОБКАХ ::= (ВЫРАЖЕНИЕ)

Программа контроля синтаксической правильности выражения составляется из подпрограмм: ВСКОБКАХ, ЧИСЛО, ИМЯ, МНОЖИТЕЛЬ, ТЕРМ, ВЫРАЖЕНИЕ.

Эти подпрограммы определяют правильность выражения соответствующей его части. Синтаксически правильные понятия выделяются из текста последовательно. Подпрограммы дают результат "да", если текст состоит из синтаксически правильных понятий. Анализ ведется до тех пор, когда счетчик модулей будет удовлетворять условию $t = k + 1$.

При анализе используется модуль общих данных.

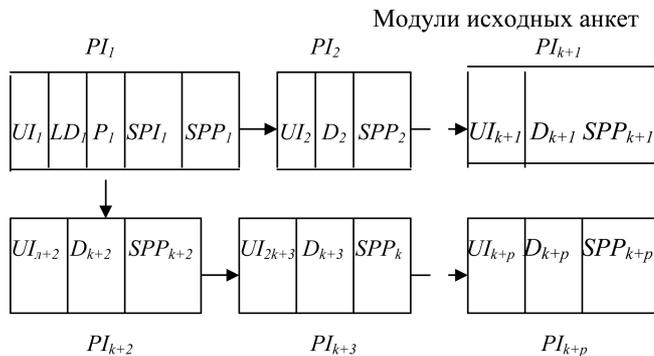


Рис. 3. Структура программы обработки банков данных

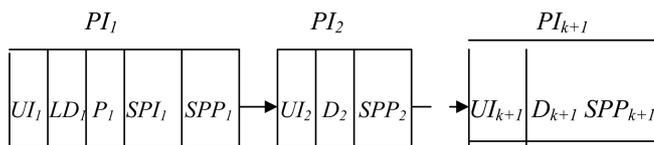


Рис. 4. Структура программы синтаксического разбора выражений

Синтаксическому анализу легко поддаются тексты за один проход. Структура программы с детерминированно-связанными модулями, определяющая использование первого операционного модуля и остальных текстовых модулей, изображена на рис. 4.

Программа анализа располагается в первом модуле. Далее располагаются модули с входным текстом. Модули в последовательности связаны линейно по номерам. Номера модулей хранятся в управляющей информации.

4. Практические аспекты

Рассмотренные в статье теоретические аспекты технологии непрерывной обработки на виртуальной памяти позволяют разрабатывать модели вычислительных процессов, технические архитектуры и структуры ЭВМ на основе современных электронных нанотехнологий, методы решения больших задач, системное и математическое обеспечение ЭВМ, приступить к разработке проектной документации ЭВМ нового поколения для ее дальнейшей реализации, начать разрабатывать языки детерминированного программирования высокого уровня, новые системы автоматизации для обработки больших объемов информации, развивать новое направление вычислительной техники, программирования и алгоритмов обработки сверхбольших и долговременно исполняющихся инженерно-научных задач в различных предметных областях знаний.

Речь идет прежде всего о задачах, предполагающих интенсивную обработку транзакций для сверхбольших баз данных, приложений ядерной

физики, например, при обработке результатов экспериментов на ускорителях элементарных частиц, физики плазмы и статистической механики, физики конденсированных сред, молекулярной и атомной физики, газовой динамики, теории турбулентности, астрофизики, квантовой химии, химической кинетики, химии твердого тела, задача долгосрочного прогноза погоды, задача автоматизации игры в шахматы и многие другие.

Технология непрерывной обработки на виртуальной памяти является интеллектуальным продуктом. Превращение интеллектуального продукта в рыночный является предметом инновационной экономики. В условиях, когда в России инновационная экономика начинает только складываться, большой интеллектуальный проект можно реализовать через государственную корпорацию, в которой есть научно-исследовательское подразделение, опытно-конструкторское, производственное и маркетинговое. После превращения интеллектуального продукта в рыночный государственная корпорация может стать акционерным обществом с сохранением инфраструктуры для развития полученного инновационного рыночного продукта и реализации спроса на него.

Предлагаемый подход к выводу интеллектуального продукта на рынок, начиная от фундаментальных исследований, научно-исследовательских работ, опытно-конструкторских разработок и кончая серийным или массовым производством, позволяет приступить к реализации новой технологии непрерывной обработки на виртуальной памяти и для развития самой инновационной экономики в России.

Научно-исследовательские работы по созданию целевой суперЭВМ. СуперЭВМ с упреждающим управлением памятью можно создать путем программно-аппаратной модернизации современной суперЭВМ, например СКИФ, с использованием механизмов виртуализации и ресурсных возможностей системы Symmetrix V-Max. При скорости последовательно-параллельной обработки информации на оперативной памяти менее или равной 192 Гбайт/с обработка целевой суперЭВМ на виртуальной памяти будет непрерывной. При скорости последовательно-параллельной обработки информации на оперативной памяти более 192 Гбайт/с можно обрабатывать в режиме упреждающего замещения несколько программ. Обработка информации состоит из нескольких последовательно-параллельных действий: считывания команды и данных из оперативной памяти, исполнения команды и записи результата в оперативную память. Она меньше, чем пиковая производительность суперЭВМ.

Механизм виртуализации Symmetrix V-Max компании EMC позволяет объединить до восьми модулей V-Max Engine. Максимальная комплектация имеет объем общей памяти 1024 Гбайт. Виртуальная скорость обмена достигает 192 Гбайт/с. СуперЭВМ с упреждающим управлением памятью, пиковая производительность которых находится в пределах 192 Гбайт/с, при наличии виртуальной памяти Symmetrix V-Max, обеспечат непрерывную обработку на виртуальной памяти.

Общее число портов для подключения дисков составляет 128. Емкость дисков может варьироваться от 146 Гбайт до 1 Тбайт, а флэш-накопителей — от 200 до 400 Гбайт. В максимальной конфигурации система будет состоять из одной управляющей стойки и восьми стоек с накопителями. Если говорить об объемах хранения, то в них может быть установлено от 96 до 2400 дисков. В зависимости от типов дисков и избранного типа RAID полезный объем хранения колеблется от 6 до 2026 Тбайт. Существенно то, что можно сочетать диски разных типов, а значит, оптимизировать состав в зависимости от решаемой задачи, подбирая соответственно флэш-диски — быстрые, но малые по объему, и большие, но более медленные.

Память Symmetrix V-Max может расти в процессе своего жизненного цикла, по мере появления более производительных V-Max Engine. Таким образом, создается эволюционирующая среда, которая позволяет плавно развивать виртуальную память, не вызывая перерывов в обслуживании.

С технической точки зрения виртуализация V-Max Engine стала возможной во многом благодаря тому, что в качестве соединения между модулями и для прямого доступа в память использована коммутационная инфраструктура. Помимо скорости, коммутационная инфраструктура имеет прямой доступ к памяти, без которого невозможно создать систему, работающую в режиме реального времени. Число V-Max Engine может быть существенно больше восьми, до 256, что позволит создавать большую виртуальную память.

Виртуальная память (*virtual memory*) организуется аппаратной частью операционной системы за счет использования дополнительно к оперативной памяти некоторого объема памяти на внешнем запоминающем устройстве. Такая организация памяти позволяет данные, не используемые в настоящий момент времени, спрятать на внешнем носителе, освободив таким образом участок в быстрой оперативной памяти. Упреждающе необходимые страницы из внешней памяти будут перенесены в освободившуюся оперативную память. Данные между внутренней и внешней памятью пересылаются блоками определенного размера — *страницами*. Быстрота виртуальной памяти

во многом определяется скоростью обмена с внешним устройством и скоростью самого устройства. Для хранения временно неиспользуемых данных на диске резервируется место, обычно называемое *файл подкачки, пэйдж-файл (pagefile)* или *своп-файл (swapfile)*.

Пока в мире не существует универсальных вычислительных систем с непрерывной обработкой программ и данных на виртуальной памяти супер-ЭВМ. Конкуренты предлагаемого проекта пока не наметились и не предвидятся. Инновационный проект реализуется на новых теоретических основах. Без этих знаний не удастся вступить в конкурентную борьбу на рынке вычислительных услуг многосуточной обработки информации, особенно в режиме реального времени. Выигрыш во времени ожидания результата многосуточной работы программы составляет до 12 часов в сутки в сравнении с существующими вычислительными системами на современных суперЭВМ, которые постоянно прерывают вычисления для подкачки недостающей информации в оперативную память с внешней памяти. Команду из целеустремленных и умных людей, которая сможет инновационный проект развить в успешный международный бизнес, предстоит формировать путем кооперации нескольких успешных компаний в соответствующих направлениях вычислительной техники. Такие компании, которые в рамках международной ассоциации могут реализовать инновационный проект в успешный бизнес, уже есть.

Интеграция усилий компаний НКО Исследовательский центр "ЕСТЕСТВОИНФОРМАТИКА", Т-Платформы, ИПС РАН, ЕСМ, ALT Linux по реализации инновационного проекта супер-ЭВМ с упреждающим управлением памятью для непрерывной обработки программ с детерминированно-связанными модулями на виртуальной памяти позволит создать серийный образец целевой суперЭВМ из узлов и компонентов Intel, IBM, ЕСМ. Протоколы Fibre Channel, FICON, iSCSI обеспечат работу целевой суперЭВМ в сетевой конфигурации.

Целевые компьютеры можно создавать по технологии Unified Computing System компании Cisco. Тогда потребуются интеграция НКО Исследовательский центр "ЕСТЕСТВОИНФОРМАТИКА", компаний Cisco, ЕСМ, Red Hat.

Система Unified Computing System представляет собой новую модель компьютерных технологий, использующую интегрированное управление, и сочетает единую структуру со стандартными отраслевыми вычислительными платформами для оптимизации виртуализации, улучшает ИТ-производительность, увеличивает расширяемость без дополнительной сложности.

Red Hat и Cisco предложили ОС Enterprise Linux и виртуализацию для новой унифицированной среды вычисления. Унифицированная среда вычисления Cisco, операционная система Red Hat Enterprise Linux и система виртуализации Red Hat Enterprise Virtualization дополнены совместными услугами и технической поддержкой. Унифицированная среда вычислений Cisco объединяет средства вычислений, сетевые механизмы и средства виртуализации в единую систему с простой настройкой. Сочетание унифицированной среды вычислений Cisco и программных решений Red Hat предоставляет заказчикам платформу для строительства вычислительных инфраструктур, использующих преимущества обеих компаний, включая уникальные средства Cisco для интегрированного управления устройствами, унифицированную матрицу коммутации и технологию расширения памяти, а также унифицированную среду вычислений Cisco и проверенное на практике программное обеспечение Red Hat. Операционная система Red Hat Enterprise Linux и технология корпоративной виртуализации Red Hat Enterprise Virtualization поддерживают множество высокопроизводительных вычислительных систем по всему миру. Стратегия компании Cisco для среды унифицированных вычислений предоставляет заказчикам платформу для внедрения и долгосрочного использования мощных вычислительных структур.

Список литературы

1. **Брындин Е. Г.** Квантодетерминированная информационная технология. Новосибирск: Наука. Сибирское отделение, 1992. 150 с.
2. **Брындин Е. Г.** Квантодетерминированная информационная технология // Известия Томского политехнического университета. 2003. № 3. С. 28—32.
3. **Брындин Е. Г.** Теоретические основы имитации мышления и непрерывной обработки на виртуальной памяти. Новосибирск: ИЦЕ, 2009. 257 с.

УДК 004.522

С. В. Гребнов, аспирант,
Ивановский государственный
энергетический университет,
e-mail: sergei.grebnov@gmail.com

Разработка и реализация двухуровневого метода голосового управления на основе скрытых марковских моделей

Предложена система голосового управления на основе скрытых марковских моделей, использующая двухуровневый метод распознавания команды. Она включает в себя собственные разработки автора, а также современные методы и идеи из области речевых технологий. Предложенная система имеет достаточную скорость работы и хорошее качество, что позволяет использовать ее на практике в режиме реального времени.

Ключевые слова: скрытые марковские модели, распознавание речи, голосовое управление, метод скользящего окна.

Введение

Построение системы голосового управления является в настоящее время актуальной задачей. Такие системы способны существенно облегчить взаимодействие пользователя с компьютерной системой. Особенно эта идея развита в концепции так называемых "умных домов". Более того, иногда голосовой интерфейс является необходимым компонентом, например, когда речь идет о людях с нарушениями опорно-двигательного аппарата.

Задачей таких систем является выделение и распознавание из потока звукового сигнала (как речевого, так и нет) заранее определенного набора речевых команд. Примером такой команды может служить фраза "Включить свет". При этом система не должна реагировать на другие участки речевого сигнала, включая и те, которые содержат отдельные слова предопределенных команд.

При создании подобной системы разработчик сталкивается с определенными проблемами:

1) отсутствие математической модели семантики речевого сигнала, в связи с чем для определения семантики речевого сигнала могут применяться только вероятностные и эвристические методы, не дающие точного результата, их точ-

ность обратно пропорциональна числу смысловых единиц, на которые они рассчитаны;

2) индивидуальные характеристики говорящего — специфика произношения, акценты, ударения, хезитации¹;

3) работа со спонтанной речью и необходимость определения присутствия ключевого слова;

4) различия в акустической обстановке, шумы.

В работе рассматривается разработанная автором система голосового управления на основе скрытых марковских моделей (СММ), использующая двухуровневый метод распознавания команды, новый алгоритм распознавания ключевого слова и модификацию одного из существующих методов шумоочистки.

В разделе 3 приведены результаты тестирования системы, а также сравнение показателей разработанных модификаций и методов с существующими аналогами.

1. Типичная архитектура системы

Большинство современных систем автоматизированного распознавания имеют модульную архитектуру [12] с использованием блока шумоочистки (*speech enhancement*), детектора голоса (*VAD*), преобразователя сигнала в векторы особенностей² (*front end*) и главного модуля (*search engine*), включающего алгоритм распознавания ключевого слова. Цифровой сигнал сначала поступает в модуль шумоочистки; здесь повышается качество сигнала вследствие удаления шумов и внесенного каналом искажения. Затем детектор голоса выделяет участки сигнала, содержащие речь. Эти участки с помощью модуля преобразования сигнала в векторы особенностей превращаются в наборы коэффициентов, которые поступают в главный модуль, в котором происходит непосредственное определение наличия и распознавания команды. Таким образом, на выходе главного модуля мы получаем информацию о команде или ее отсутствии.

Существующие методы. В качестве метода распознавания в большинстве современных систем используется метод скрытых марковских моделей

¹ Хезитация — речевое колебание, связанное со спонтанностью речи: речевое сбой, заминка в речи, колебание в выборе слова или конструкции. Чаще всего хезитация выражается в виде паузы, нелексических вставных звуках, словах-паразитах.

² Под вектором особенностей будем понимать фиксированный набор коэффициентов, характеризующий участок звукового сигнала.

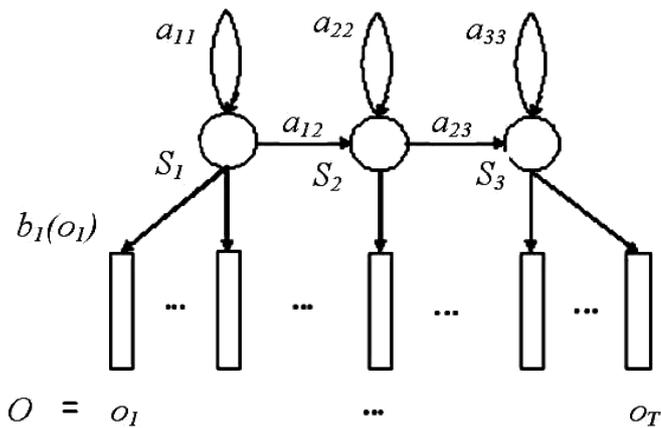


Рис. 1. СММ с тремя состояниями

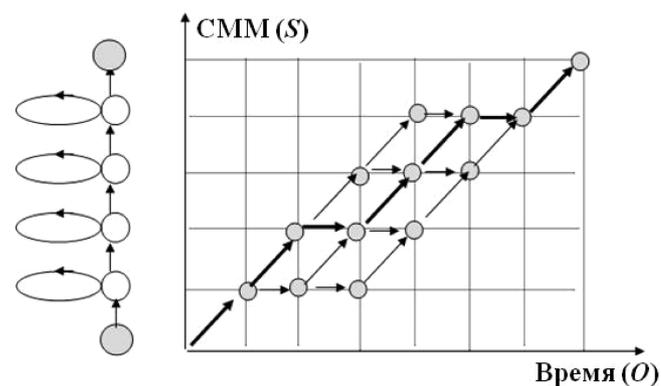


Рис. 2. Пример работы алгоритма Витерби. Утолщенная линия соответствует наиболее вероятной последовательности СММ

[3, 4, 5]. Анализ применимости СММ для распознавания речи приводится в работах [6, 7]. Использование СММ для распознавания речи базируется на следующих предположениях: речь может быть разбита на сегменты (состояния), внутри которых речевой сигнал можно рассматривать как стационарный, переход между этими состояниями осуществляется мгновенно; вероятность символа наблюдения, порождаемого моделью, зависит только от текущего состояния модели и не зависит от предыдущих. Чаще всего используются СММ с тремя состояниями (рис. 1).

СММ представляет собой конечный автомат, изменяющий свое состояние в каждый дискретный момент времени t . Переход из состояния s_i в s_j осуществляется случайным образом с вероятностью a_{ij} . В каждый дискретный момент времени модель порождает вектор наблюдений o_t (который в конкретной задаче является вектором особенностей, полученном в преобразователе сигнала) с вероятностью $b_j(o_t)$. Распределение плотности вероятности наблюдений моделируется конечной гауссовской смесью с четырьмя компонентами. Каждая такая модель обозначает один из звуков русского языка или отсутствие звука (одна из моделей).

В алгоритмах распознавания ключевого слова [8, 9, 10] используются эти модели для определения команд в потоке речи. Наиболее часто эта задача решается с помощью метода скользящего окна (*sliding window*) [11] и метода моделей-заполнителей (*filler models*) [12].

В данной работе в качестве метода для сравнения выбран метод скользящего окна, так как он не требует подробного дополнительного моделирования посторонней речи, что позволяет проводить сравнение алгоритмов (данного и разработанного) на одних и тех же моделях.

Метод скользящего окна. Суть метода скользящего окна заключается в определении вхождения ключевого слова с помощью алгоритма Витерби (*Viterbi*) [5], который широко применяется для распознавания слитной речи (CSR). Этот алгоритм решает следующую задачу: дан вектор наблюдений o , требуется определить наиболее подходящую последовательность СММ (s) и переходов между их состояниями для этого вектора наблюдений (рис. 2). Далее будем называть такую последовательность путем³. Так, на рис. 2 изображены все возможные пути для данного участка сигнала и определенной последовательности СММ; утолщенной линией обозначен наиболее вероятный путь. Так как ключевое слово может начинаться и заканчиваться в любом месте сигнала, то этот метод перебирает все возможные пары начала и конца вхождения ключевого слова и находит самый вероятный путь для ключевого слова и этого отрезка, как если бы ключевое слово присутствовало в нем. Для каждого найденного вероятного пути ключевого слова применяется функция правдоподобия, основанная на срабатывании, если значение пути, рассчитанное в соответствии с применяемым методом оценки пути, больше предопределенного значения. Часто для оценки пути используется значение вероятности, полученное с помощью алгоритма Витерби.

Главным недостатком такого подхода является необходимость перебора всех возможных вариантов вхождения ключевого слова, что создает большую вычислительную сложность. В разделе 2 будет рассмотрен подход, который позволяет решить эту проблему.

Типичный метод распознавания команды. Типичный метод распознавания голосовой команды заключается в применении алгоритма распознавания ключевого слова (*Keyword spotting*) ко всему речевому участку для каждой возможной команды из словаря команд. Такой подход имеет два существенных недостатка:

- большая вычислительная сложность;

³ Под путем будем понимать возможную последовательность СММ и их состояний для определенного участка сигнала.

- низкое качество распознавания некоторых команд.

Первая проблема возникает ввиду необходимости применения алгоритма распознавания ключевого слова для каждой возможной команды из словаря; вторая проблема — по следующим двум причинам:

- составные части команды содержат сложные для распознавания фонемы языка;
- существуют дефекты в некоторых моделях фонем, полученные в силу несбалансированности речевой базы данных, на которой проводилось обучение, или же вследствие неправильного процесса обучения.

Ниже будет рассмотрен подход, который позволяет решить эти проблемы.

2. Предложенная система

Разработанная система базируется на типичной архитектуре, описанной выше, со следующими изменениями:

1. Разработан и реализован двухуровневый метод распознавания голосовой команды, основанный на разделении команды на две части: ключевого слова (с помощью него система понимает, что обращаются к ней) и собственно команды (действия, которое хочет выполнить пользователь). Метод позволяет решить проблемы, описанные в разделе "Типичный метод распознавания команды".

2. Разработан и реализован новый однопроходный алгоритм распознавания ключевого слова.

3. Разработаны и реализованы модификации одного из существующих методов шумоочистки. Далее рассмотрим каждую модификацию более подробно.

Двухуровневый метод распознавания команды. Для решения проблем, описанных в разделе 1, был разработан следующий подход: в команду добавляется заранее предопределенное ключевое слово, чтобы система могла понять, что обращаются к ней, и применяется разработанный двухуровневый метод распознавания, который использует алгоритм распознавания ключевого слова только для этого вводного (ключевого) слова, а для оставшейся части команды стандартный алгоритм распознавания слитной речи (CSR).

Таким образом использование предложенной системы заключается в произнесении команды, которая состоит из двух частей: ключевого слова (с помощью него система понимает, что обращаются к ней) и собственно команды (действия, которое хочет выполнить пользователь). Например: "Агент, включить свет". В данном случае "Агент" — ключевое слово, а "включить свет" — непосредственно команда.

Выделение двух частей выполнено для увеличения скорости работы, а также выигрыша в качестве распознавания.

Увеличение скорости достигается за счет сокращения числа распознаваемых ключевых слов до одного, заранее предопределенного, и использования для распознавания непосредственно команды алгоритма распознавания слитной речи (CSR), который на порядок быстрее алгоритма распознавания ключевого слова и временем работы которого в данном случае можно пренебречь.

Улучшение же качества распознавания достигается за счет введения заранее предопределенного, единственного ключевого слова, что дает возможность в качестве такого слова выбрать легко-распознаваемое слово или провести дополнительные меры для улучшения его распознавания: подготовку дополнительного речевого материала, калибровку параметров алгоритма распознавания ключевого слова.

Кроме этого, использование названия системы или же имени как ключевого слова имеет следующие преимущества:

- является для человека более интуитивно привычным и удобным способом произношения команд;
- добавляет одинаковую для всех пользователей командную интонацию, что уменьшает число интонационных вариаций произношения команды и тем самым упрощает процесс распознавания;
- уменьшается число ложных срабатываний за счет сокращения числа ключевых слов до одного.

Недостатком предложенного подхода распознавания команды являются некорректные срабатывания системы, если после ключевого слова следуют команды или фразы не из словаря команд. Например, если пользователь скомандует "Агент, включить телевизор" и такой команды нет в словаре команд, то алгоритм распознает эту команду как команду из словаря, наиболее созвучную с произнесенной, например, "Агент, включить свет". Это происходит вследствие применения алгоритма распознавания слитной речи для распознавания непосредственно команды. Этот алгоритм лишь распознает голосовую фразу в соответствии с заранее определенным словарем и не может определить, является ли данная фраза не из словаря (OOV, Out Of Vocabulary); такую задачу решает алгоритм распознавания ключевых слов. Таким образом, предлагаемый подход вносит следующее ограничение на использование системы: пользователь должен знать набор поддерживаемых системой команд и запрещать угадывание команд.

Архитектурной отличительной особенностью подхода является объединение двух различных ал-

```

для каждого участка входного сигнала  $o_t \in \mathcal{O}$ 
создать новый путь для состояния  $s_1$  ключевого слова
для всех текущих путей  $p \in \mathcal{P}$ 
  для всех возможных переходов  $s_{new}$  из текущего состояния пути  $s_{current}$ 
  если (Фонема[ $s_{new}$ ] == Фонема[ $s_{current}$ ]) // переход внутри одной фонемы
    осуществить переход  $p$  в  $s_{new}$ 
    увеличить длину пути  $p$ 
    пересчитать правдоподобие  $p$ 
  иначе
    Score = правдоподобие( $p$ )
    если (Score > порог [Фонема[ $s_{current}$ ]])
      если (Фонема[ $s_{current}$ ] == последняя фонема ключевого слова)
        вернуть true
      создать новый путь  $p_{new}$  для состояния  $s_{new}$  ключевого слова
      добавить  $p_{new}$  в  $\mathcal{P}$ 

сокращение путей
вернуть false

```

Рис. 3. Псевдокод алгоритма распознавания ключевого слова

горитмов распознавания речи в единый блок (*search engine*). В этом блоке сначала делается вывод о наличии вводного слова, и затем, если это слово присутствует, происходит распознавание самой команды. Таким образом на выходе главного модуля мы получаем информацию о команде или ее отсутствии.

Алгоритм распознавания ключевого слова. Данная система использует свой собственный метод определения ключевого слова, включающий новый алгоритм распространения и новую функцию правдоподобия. Псевдокод алгоритма изображен на рис. 3. В разделе 3 приведены результаты сравнения данного метода с методом скользящего окна.

В используемом алгоритме определения ключевого слова применяются те же модели, что и в алгоритме распознавания команды, но в нем не учитываются вероятности перехода. В каждый дискретный момент t этого алгоритма:

а) стартует новый путь из текущей позиции сигнала o_t в начальное состояние ключевого слова (в начальное состояние СММ первой фонемы ключевого слова s_1);

б) каждый существующий путь дублируется: один остается в текущем состоянии, второй переходит в следующее состояние ключевого слова;

в) происходит сокращение путей на основе функции правдоподобия.

Как показатель соответствия в разработанном методе используются не только сами вероятности наблюдения $b_j(o_t)$, но и отношение вероятности наиболее подходящего состояния из всего набора моделей и текущего состояния ключевого слова $b_j(o_t)/b_{best}(o_t)$. Функция же правдоподобия основана на показателях худшей модели в ключевом слове, а не

на общих показателях всего слова. Экспериментальные данные, приведенные далее, показали, что такой подход уменьшает ошибку, связанную с неправильным срабатыванием на слове, близком по звучанию. Кроме этого экспериментальные данные показывают, что некоторые фонемы распознаются лучше, чем другие. Поэтому для каждой фонемы ключевого слова экспериментально был определен свой собственный порог срабатывания. Использование более высокого порога для хорошо распознающихся фонем и более низкого для плохо распознающихся позволило существенно улучшить качество распознавания. Кроме того, для сокращения заведомо ложных путей в функции правдоподобия используется специальный общий порог на каждое состояние пути. Путь удаляется из списка текущих путей, если не выполняется хотя бы одно из следующих условий:

- показатель соответствия всех состояний пути больше общего порога;
- среднее значение показателя соответствия текущей фонемы больше порога для этой фонемы.

Шумоочистка. Шумоочистка решает проблему различия в акустических условиях и наличия шумов. Об эффективных методах борьбы с различными шумами и их влиянии на системы распознавания речи можно узнать в работах [13—16]. Наиболее существенные результаты шумоочистки показывают алгоритмы, работающие в спектрально-временной области. В основе многих таких алгоритмов используется двухшаговый подход: оценка шумовой составляющей (*noise psd estimation*) и переоценка речевой составляющей на основе оцененного шума и статистической модели речи (*gain function*).

В данной системе для оценки уровня шума используется улучшенный алгоритм минимума рекурсивного контролируемого среднего [17] (*Improved Minimum Controlled Recursive Averaged*), в качестве алгоритма для функции усиления выбран алгоритм, основанный на средней квадратичной ошибке [18, 19] (OM-LSA). Блок-схема алгоритма шумоочистки показана на рис. 4.

Такой подход дает отличные результаты в борьбе даже с сильно нестационарными шумами, при слабых речевых компонентах и низком отношении сигнала к шуму.

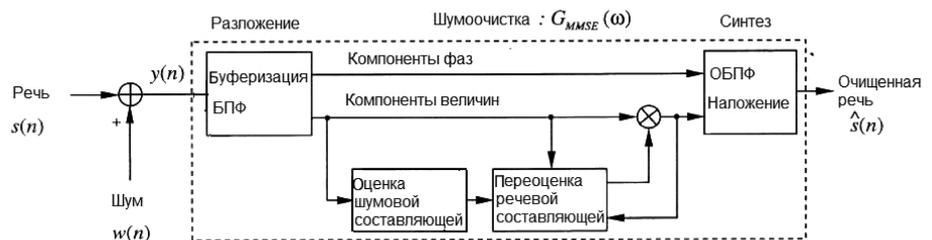


Рис. 4. Блок-схема алгоритма шумоочистки

3. Экспериментальные данные

Так как в качестве характеристик работоспособности алгоритмов шумоочистки используются, как правило, меры качества звучания, а не повышения распознаваемости в системах распознавания речи, то в ходе исследований возможности адаптации описанного подхода к решению поставленной задачи был разработан дополнительный блок постобработки передаточной функции частотного спектра, включающий функцию масштабирования и пороговую функцию.

Функция масштабирования. Введем следующие обозначения:

$SNR_{prior}(f)$ — априорное значение отношения сигнал/шум для спектральной полосы f , рассчитанное на шаге оценки шумовой составляющей;

$G(f)$ — коэффициент усиления (уменьшения, если $G(f) < 1$) амплитуды сигнала для спектральной полосы f , рассчитанный на шаге оценки речевой составляющей.

Тогда разработанная функция масштабирования примет вид:

$$G_m(f) = \begin{cases} kG(f), & SNR_{prior}(f) > \delta, \\ G(f), & SNR_{prior}(f) \leq \delta, \end{cases}$$

где k — коэффициент усиления ($k = 1,33$, получено экспериментально); δ — пороговое значение отношения сигнал/шум для усиления сигнала при масштабировании ($\delta = 0,7$, получено экспериментально).

Функция масштабирования выделяет и усиливает спектральные участки сигнала с высоким соотношением сигнал/шум, тем самым предохраняя алгоритм от удаления важных для распознавания компонентов речи вместе с присутствующим в этой полосе незначительным шумом.

Пороговая функция. Пороговая функция (G_f) служит для предотвращения нежелательного полного подавления слабых речевых компонентов в зашумленных речевых участках, что также приводит к сохранению важных для распознавания компонентов сигнала:

$$G_f(f) = \begin{cases} G(f), & G(f) > G_{floor}, \\ G_{floor}, & G(f) \leq G_{floor}, \end{cases}$$

где параметр G_{floor} — минимальное (пороговое) значение передаточной функции ($G_{floor} = 0,35$, получено экспериментально).

Таким образом обе разработанные эвристические функции позволяют за счет сохранения некоторых слышимых искажений сигнала и шума добиться сохранения важных для распознавания компонентов речи. В разделе 3 будут приведены результаты испытания данного метода и разработанных эвристик.

Речевая база данных. Численные эксперименты выполнялись на речевом корпусе sSpeech-Corp, разработанном в НИИ СпецЛаб и ООО "Спецлаборатория" [25]. Эта речевая база данных (РБД) содержит 10-часовой речевой материал, записанный 40 людьми возрастом от 18 до 50 лет и транскрибированный вручную. Для транскрибирования использовался алфавит Russian SAMPA (*speech assessment methods phonetic alphabet*) и разделение гласных фонем на ударные (*stressed*) и безударный (*unstressed*), всего 50 фонем.

Для проведения тестирования вся РБД была разбита на два блока: один блок содержал 90 % речевого материала и использовался для обучения (применялся алгоритм Баума—Уелша (Baum—Welch) [5]), другой блок — 10 % речевого материала и использовался для тестирования. Блоки содержали различных дикторов.

Конфигурация системы. Для тестирования применялась следующая конфигурация системы.

Детектор голоса состоит из двух частей, которые в сумме дают представление о наличии речевой составляющей в сигнале. Первая часть основана на изменении энергии сигнала, вторая часть — на периодичности, которая рассчитывается с помощью алгоритма, основанного на методе наименьших квадратов [20].

Для преобразования сигнала [23] в векторы особенностей за основу взят алгоритм, предложенный Европейским институтом стандартов телекоммуникации (ETSI) [21], блок-схема которого изображена на рис. 5.

Для сбора особенностей в алгоритме используются свойства человеческого восприятия. Кроме того, согласно работе [22] в него были внесены следующие модификации, направленные на улучшение качества работы: лифтинг (*liftering*); вычет среднего кепстрального значения (*cepstral mean subtraction, CMS*), нормализация энергии (*energy normalization*). На выходе данного блока для каждого участка сигнала в 25 мс формируется вектор из 39 параметров. Первые 13 из них — 12

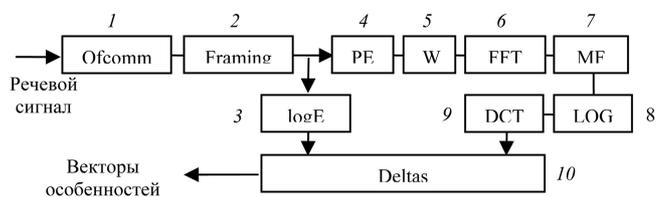


Рис. 5. Блок-схема алгоритма преобразования сигнала: 1 — компенсация смещения; 2 — разбиение на блоки; 3 — вычисление энергии; 4 — предварительная коррекция; 5 — применение оконной функции; 6 — разложение на частотные составляющие; 7 — применение фильтров; 8 — нелинейное преобразование; 9 — косинус-преобразование; 10 — вычисление 1-й и 2-й производной по времени

Таблица 1

Результаты работы алгоритма распознавания слитной речи для разных вариантов модуля шумоочистки

Вариант Модуля шумоочистки	Правильность распознавания, %	
	Уровень шума 25 дБ	Уровень шума 5 дБ
Шумоочистка отсутствует	99	25
OM-LSA	96	94
OM-LSA + разработанные эвристики	99,5	99

Таблица 2

Сравнение показателей разработанного алгоритма распознавания ключевого слова с методом скользящего окна

Показатель	Метод скользящего окна	Разработанный алгоритм
Время работы алгоритма (полученное для входного участка продолжительностью 7,4 с)	20 с	0,5 с
Качество распознавания, %	86,1/5	98/0,01

кепстральных коэффициентов (12 *mel-frequency cepstral coefficient*) и логарифм энергии ($\log E$), а остальные — производные (1-го и 2-го порядка) этих коэффициентов (они показывают динамику изменения).

Результаты тестирования. Тестирование проводилось для распознавания слитной речи для словаря из 20 слов. В качестве материала для тестирования работы модуля шумоочистки использовался шум из специального набора шумов NOISEX92 [24]. На первом этапе было проведено сравнение работы алгоритма распознавания слитной речи для трех различных конфигураций модуля шумоочистки:

- без шумоочистки;
- с использованием алгоритма шумоочистки OM-LSA;
- с использованием алгоритма шумоочистки OM-LSA и с применением разработанных эвристик.

В табл. 1 приведены результаты экспериментов. Видно, что лучшие показатели распознавания показала система, использующая алгоритм OM-LSA с применением разработанных эвристик.

Второй этап тестирования заключался в экспериментальном сравнении разработанного алгоритма распознавания ключевого слова с методом скользящего окна. В качестве показателей для сравнения были выбраны скорость работы и качество распознавания. При этом качество распознавания измерялось двумя величинами:

- правильностью срабатывания и распознавания голосовой команды;
- неправильным срабатыванием на участке сигнала, не содержащем голосовой команды.

Результаты экспериментов (табл. 2) показывают превосходство разработанного метода определения ключевого слова, как в плане скорости, так и в плане качества распознавания.

Заключение

В данной статье рассмотрена система голосового управления, использующая новый алгоритм распознавания ключевого слова и построенная на самых современных методах распознавания речи и шумоочистки. Отличительной чертой данной системы является использование двухуровневого алгоритма распознавания команды: определение наличия ключевого слова, непосредственное распознавание команды. Экспериментальные данные подтверждают эффективность такого подхода: срабатывание на ключевом слове — 98 %, не на ключевом слове — 0,01 %, распознавание команды — 99,5 %. Система также имеет высокую скорость работы, дикторнезависимость, хорошо работает в условиях шума.

Было проведено экспериментальное сравнение разработанного метода распознавания ключевого слова с методом скользящего окна [22]. Результаты показывают превосходство разработанного метода, как в плане скорости, так и качества.

В дальнейшем предполагается использовать в качестве моделей гибриды СММ/нейронная сеть, разделять речевой сигнал на потоки, каждый из которых соответствует одному источнику, усовершенствовать алгоритмы в плане качества, скорости и борьбы с речеподобным шумом.

Список литературы

1. **Rose R.** Robust speech recognition techniques applied to a speech in noise task // Proc. of the European Conference on Speech Communication and Technology, Aalborg, Denmark, 3–7 Sept. 2001.
2. **Ahadi S.** An Efficient front-end for automatic speech recognition // IEEE Trans. on Speech and Audio Processing, 2003.
3. **Demuynek K.** Extracting, modeling and combining information in speech recognition // PhD thesis, ESAT, 2001.
4. **Rosti I.** Linear gaussian models for speech recognition // PhD thesis, University of Cambridge, 2004.
5. **Couvreur Chr.** Hidden Markov Models and Their Mixtures. DEA Thesis, Department of Mathematics, Catholic University of Louvain, 1996.
6. **Rabiner R. L.** A tutorial on Hidden Markov Models and selected applications in speech recognition. Proceedings of the IEEE, 1989.
7. **Morgan N.** Neural Network for Statistical Recognition of Continuous Speech // Proc. of the IEEE, 1995.
8. **Xhenyu X.** Comparison and combination of confidence measures in IWR // ISCSLP, 2002.
9. **Hazen T.** Recognition confidence scoring and its use in speech understanding systems // Computer Speech and Language, 2002.

10. **Mengusoglu E.** Use of acoustic prior information for confidence measure in ASR // Proc. of the European Conference on Speech Communication Technology. 2005.

11. **Bridle J.** An efficient elastic template method for detecting given words in running speech // British Acoustical Society Meeting, Apr. 1973.

12. **Higgins A.** Keyword recognition using template concatenation. Acoustics, Speech, and Signal Processing // IEEE International Conference on ICASSP, 1985.

13. **Martin R.** Statistical methods for the enhancement of noisy speech // International Workshop on Acoustic Echo and Noise Control, 2003.

14. **Rangachari S.** Noise estimation algorithms for highly non-stationary environments // Theses in Speech Processing Lab at UT-Dallas, 2004.

15. **Acero A.** Acoustical and environmental robustness in automatic speech recognition // Ph. D. Thesis, Carnegie Mellon University, 1990.

16. **Stern R.** New directions in robust speech recognition // Proc. of the International Conference on Spoken Language Processing, 2006.

17. **Cohen I.** Noise spectrum estimation in adverse environments: Improved MCR // Acoustics, Speech, and Signal Processing. Proc. IEEE International Conference on ICASSP, 2003.

18. **Ephraim Y.** Speech enhancement using MMSE LSA estimator // Proc. of the IEEE, 1985.

19. **Cohen I.** On speech enhancement under signal presence uncertainty // Proc. of the 26th IEEE International Conference on Acoustics, Speech, and Signal Processing, 2001.

20. **Tucker R.** Voice activity detection using a periodicity measure // Proc. of the IEEE. 1992. Vol. 139.

21. **European Telecommunications Standards Institute.** ES 201 108 Distributed Speech Recognition Encoding // Proc. of the ETSI, 2003.

22. **Parihar N.** Performance analysis of advances front ends on the Aurora LV evaluation: M. S. Dissertation, Mississippi State University. 2003.

23. **Кос А.** Acoustic feature analysis for robust speech recognition. M. S. Thesis, Bilkent University, 2002.

24. **NOISEX92** noise database. URL: http://spib.rice.edu/spib/select_noise.html.

25. **ООО "Спецлаборатория"** URL: <http://www.goal.ru>.

УДК 681.51;621.391;519.21

Г. Ф. Малыхина, д-р техн. наук, проф.,
А. В. Меркушева, канд. техн. наук,
инженер-исследователь,
Санкт-Петербургский политехнический
университет,
e-mail: g_f_malychina@mail.ru

Восстановление формы сигналов по измерительной информации об их смеси. Концепции, модификации алгоритмов и приложения

Представлены методы восстановления вида сигналов по измерительной информации об их смеси. Алгоритмы и нейросетевые структуры, реализующие эту задачу, основаны на различных концепциях, которые отражают особенности формирования смеси сигналов: наличие статистических характеристик шума при регистрации сигналов в процессе их многоканального смешивания и возможность снижения уровня шума в восстановленных сигналах.

Ключевые слова: смеси сигналов, восстановление формы, концепция независимых компонент, нейронные сети, адаптивная фильтрация, структуры преобразования, разделение сигналов, приложения.

Введение

Задачи восстановления формы сигналов (ВФС), регистрируемых средствами нескольких каналов в виде смеси с различными соотношениями, имеет

значительное число приложений, поскольку чаще всего первичный информационно значимый вектор — сигнал (т. е. группа сигналов по нескольким каналам) — не является доступным в используемой системе измерений. В частности, в системах беспроводной связи методы уплотнения каналов с временным или частотным разделением используют наиболее простой вид ВФС, когда компоненты смеси занимают не накладывающиеся интервалы в частотной или временной области. При перекрытии спектров компонент неполное ВФС возможно только при частичном их наложении в области время-частотных преобразований. Среди потенциальных приложений ВФС — обработка многокомпонентных речевых сигналов, геофизических данных, анализ многоканальных биомедицинских сигналов (ЭЭГ, ЭМГ, ЭКГ) и анализ сигналов элементов антенны при приеме множественных целей.

Возможные формы постановки задачи ВФС имеют ряд вариантов:

- выделение группы сигналов-компонент, определяющих максимальную долю дисперсии регистрируемых смесей сигналов источника (СИ);
- ВФС по критерию независимости ее компонент;
- ВФС без информации о виде и числе компонент и пропорциях их в смесях.

Общая постановка задачи ВФС предполагает, что наблюдается m -компонентный сигнал $\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_m(t)]^T$, который является выходом нелинейной динамической системы, имеющей n входов и m выходов. Целью является получение обратной системы (в виде нейронной сети или другой адаптивной структуры), оценивающей

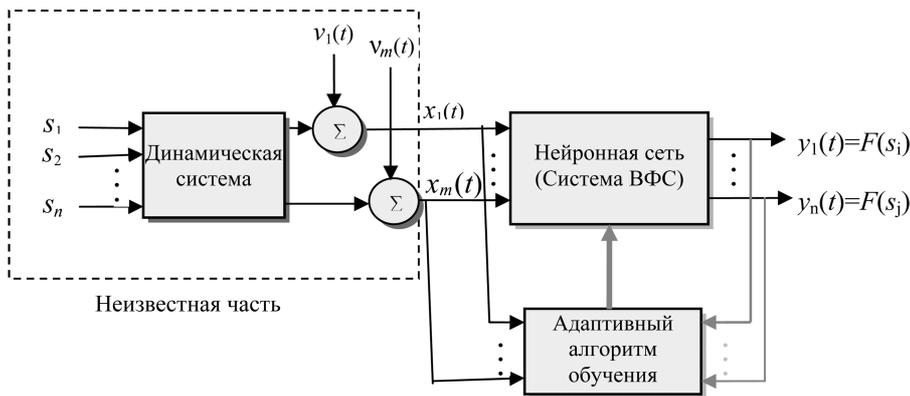


Рис. 1. Схема обработки сигналов для восстановления их формы по данным регистрации смеси на выходе линейной динамической системы

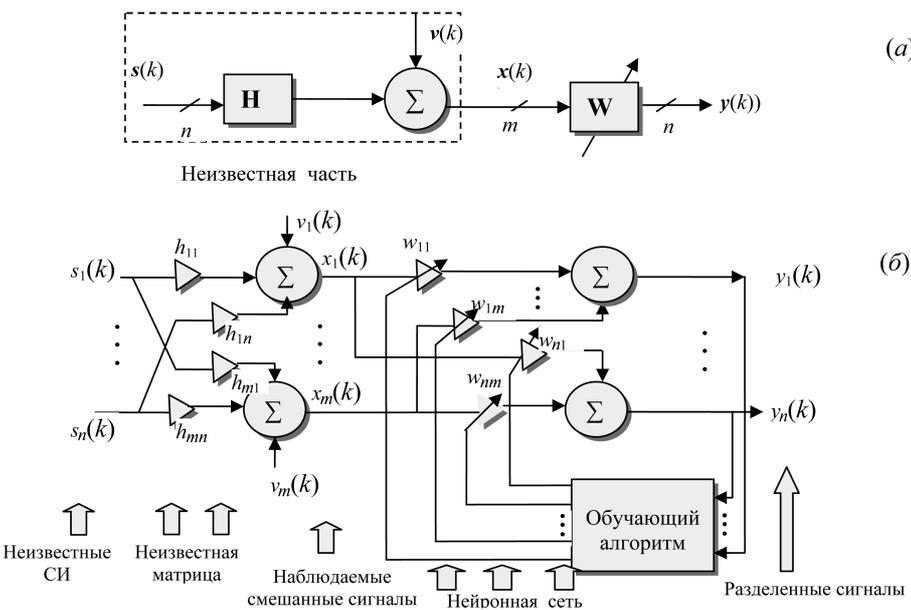


Рис. 2. Векторно-матричная структура (а) и логика (б) системы разделения сигналов (ВФС)

первичный вектор-сигнал $\mathbf{s}(t) = [s_1(t), s_2(t), \dots, s_n(t)]^T$, поступающий на сенсоры. Оценка $\mathbf{s}(t)$ осуществляется на основе n -мерного вектора-сигнала $\mathbf{y}(t) = [y_1(t), y_2(t), \dots, y_n(t)]^T$ на выходе обратной системы (рис. 1). Адаптивная обратная система способна отслеживать нестационарные изменения $\mathbf{s}(t)$ и поступающего на ее вход сигнала $\mathbf{x}(t)$. Если обратную систему получить трудно или число регистрируемых смесей меньше, чем число первичных СИ, то вместо прямой оценки СИ сначала идентифицируют систему, смешивающую сигналы.

Элементы общей концепции разделения сигналов смеси

Решение задачи ВФС базируется на методе анализа главных компонент [1–3] и на более новых подходах к разделению сигналов, таких как последовательное извлечение СИ из смеси на вы-

ходе системы с преобразованием типа свертки [3–6]. В зависимости от особенностей приложения задачи ВФС процессы смешивания и фильтрации неизвестных СИ могут соответствовать различным математическим или физическим моделям. В наиболее простом случае m смешанных сигналов $x_i(k)$ ($i = 1, \dots, m$) являются линейной комбинацией n ($n \leq m$) неизвестных, статистически взаимно независимых СИ $s_i(k)$ ($i = 1, 2, \dots, n$), которые имеют нулевые средние и "искажены" присутствием сопутствующего шума. Этот случай описывается соотношением (1) или его матричной формой (2):

$$x_i(k) = \sum_{j=1}^n h_{ij}s_j(k) + v_i(k) \quad (1)$$

($i = 1, 2, \dots, m$);

$$\mathbf{x}(k) = \mathbf{H} \cdot \mathbf{s}(k) + \mathbf{v}(k), \quad (2)$$

где $\mathbf{x}(k) = [x_1(k), x_2(k), \dots, x_m(k)]^T$ – вектор сигналов сенсоров; $\mathbf{s}(k) = [s_1(k), s_2(k), \dots, s_n(k)]^T$ – вектор СИ; $\mathbf{v}(k) = [v_1(k), v_2(k), \dots, v_m(k)]^T$ – вектор аддитивного шума; \mathbf{H} – неизвестная смешивающая матрица размерности $m \times n$ и полного ранга (т. е. ранг матрицы \mathbf{H} равен n)⁽¹⁾.

Предполагается, что число СИ n неизвестно, доступен только вектор-сигнал сенсоров $\mathbf{x}(k)$. Требуется использовать нейронную сеть (прямого распространения или рекуррентную) и адаптивный алгоритм обучения, позволяющий оценить СИ, идентифицировать смешивающую матрицу \mathbf{H} или разделяющую матрицу \mathbf{W} (рис. 2). При этом достигается адаптация значений получаемых матриц \mathbf{H} и \mathbf{W} к изменениям СИ.

Задача ВФС может трактоваться как анализ независимых компонент. При этом для разделения вектора-сигнала $\mathbf{x}(k) = [x_1(k), x_2(k), \dots, x_m(k)]^T$ и восстановления формы первичных сигналов (СИ) требуется найти линейное преобразование в

⁽¹⁾ При таком значении ранга матрица \mathbf{H} имеет псевдообратную матрицу \mathbf{H}^+ [7], что в принципе позволяет восстановить СИ. Однако в практических приложениях матрица \mathbf{H} , смешивающая первичные сигналы, не бывает известна.

виде ("разделяющей") матрицы \mathbf{W} размерности $n \times m$, полного ранга и такой, чтобы выходной вектор-сигнал $\mathbf{y}(k) = [y_1(k), y_2(k), \dots, y_n(k)]^T$, определенный соотношением $\mathbf{y} = \mathbf{W} \cdot \mathbf{x}$, содержал компоненты, которые максимально статистически независимы. Степень независимости компонент проверяется с помощью информационно-теоретического критерия Кульбака — Лейблера [8, 9]. При таком подходе к задаче ВФС требуется адаптация весов w_{ij} — компонент матрицы \mathbf{W} линейной системы $\mathbf{y} = \mathbf{W} \cdot \mathbf{x}$, которая реализуется в виде однослойной нейронной сети. Эта сеть комбинирует наблюдаемые и измеряемые $x_i(k)$ и дает оценку $\hat{s}_j(k)$ СИ:

$$\hat{s}_j(k) = y_j(k) = \sum_{i=1}^m w_{ji} x_i(k), \quad (j = 1, 2, \dots, n). \quad (3)$$

Оптимальные веса соответствуют статистической независимости выходных сигналов $y_j(k)$ (рис. 2, б).

Модификации метода анализа независимых компонент

Применение метода анализа независимых компонент (АНК) для получения оценок сигналов, регистрируемых в виде их смеси, имеет некоторые особенности, которые связаны со спецификой приложений.

- При АНК зашумленного (m -мерного) сигнала $\mathbf{x}(k)$ определяется разделяющая матрица \mathbf{W} размерности $n \times m$ ($n < m$), которая по соотношению АНК $\mathbf{y} = \mathbf{W} \cdot \mathbf{x}$ дает вектор-сигнал $\mathbf{y}(k) = [y_1(k), y_2(k), \dots, y_n(k)]^T$, содержащий оценку компонент СИ $s(k)$, причем эти компоненты (насколько возможно) статистически независимы.
- Для зашумленного вектора-сигнала $\mathbf{x}(k)$, определяемого выражением $\mathbf{x}(k) = \mathbf{H}\mathbf{s}(k) + \mathbf{v}(k)$, метод АНК реализуется путем оценивания как смешивающей матрицы \mathbf{H} , так и независимых компонент $\mathbf{s}(k) = [s_1(k), s_2(k), \dots, s_n(k)]^T$.
- АНК может формулироваться как задача оценки всех СИ, их числа и идентификации смешивающей матрицы $\hat{\mathbf{H}}$ (или псевдообратной разделяющей матрицы $\mathbf{W} = \hat{\mathbf{H}}^+$). При этом единственным допущением является статистическая независимость первичных СИ и линейная независимость столбцов матрицы \mathbf{H} . (Последнее допущение равноценно тому, что ранг \mathbf{H} равен n .) В контексте метода АНК модель смешивания СИ, объединяющая набор из N векторов-сигналов $\mathbf{X} = [\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(N)]^T \in R^{m \times N}$ и соответствующий набор N сигналов

источника $\mathbf{S} = [s(1), s(2), \dots, s(N)]^T \in R^n \times N$, может быть представлена в виде

$$\mathbf{X} = \mathbf{H} \cdot \mathbf{S}. \quad (4)$$

- В реальных приложениях данные на сенсорах представляют сложные нелинейные структуры из компонент СИ, поэтому применение АНК ко всей совокупности данных может привести к неудовлетворительным результатам. В этом случае используется предварительная обработка данных с группировкой их в кластеры (в подгруппы, однородные по особенностям данных). После этого АНК применяется последовательно (индивидуально) к каждому кластеру.

В отдельных приложениях, связанных с задачами ВФС, бывает необходимо оценить только одну или несколько желаемых компонент (вектора СИ), которые обладают особыми статистическими признаками или свойствами, и не рассматривать остальные компоненты и шумы. В этом случае используется метод (близкий к модификации АНК), условно называемый извлечением сигналов с восстановлением их вида (ИСВВ). Для его реализации необходимо идентифицировать соответствующие векторы \mathbf{h}_j матрицы смешивания $\hat{\mathbf{H}}$ и \mathbf{w}_j , являющиеся строками разделяющей матрицы $\mathbf{W} = \hat{\mathbf{H}}^+$.

Дополнительные замечания по концепции разделения сигналов смеси

Пока нет вполне сформировавшейся методологии в области прикладных задач, связанных с ВФС или ИСВВ, особенно при большой размерности сигналов и нелинейном типе смешивания. В частности, встречается подмена понятий ВФС, АНК и ИСВВ, хотя они относятся к одной и той же (или очень близкой) проблематике современных алгоритмов обработки сложных и нелинейных комбинаций многомерных сигналов при статистической независимости их компонент. В реальных задачах использование АНК и ВФС направлено на несколько различающиеся цели.

- Целью ВФС служит получение оценки СИ, даже если они не полностью взаимно статистически независимы, в то время как цель АНК состоит в определении такого преобразования, при котором выходные сигналы были (максимально) статистически независимыми.
- Методы АНК, как правило, основаны на использовании статистик высокого порядка (кумулянт), а в базовом методе ВФС используются только статистики второго порядка (СВП). Метод на основе СВП предполагает, что СИ имеют некоторую временную структуру, в то время как использование статистик высокого

порядка предполагает полную статистическую независимость СИ. Другое различие состоит в том, что метод, основанный на статистиках высокого порядка, не может быть применен к Гауссовым сигналам (т. е. при нормальном распределении компонент СИ), а метод, основанный на СВП, не имеет такого ограничения. Таким образом, методы ВФС (в описанной выше их интерпретации) не замещают АНК и наоборот, так как каждый из методов основан на различных предположениях и часто на различных целевых установках.

- Наиболее часто в качестве критериев используется мера независимости сигналов (компонент СИ), не-Гауссовость, или разреженное распределение главных компонент. Когда компоненты СИ статистически независимы и не имеют структуры, для решения задачи разделения СИ применяются статистики высокого порядка. В этом случае допустимо не более одной компоненты и с Гауссовым распределением.
- Если СИ имеют временную структуру (т. е. каждая компонента СИ имеет временные корреляции), то это меньшее ограничение, чем статистическая независимость, и СВП достаточны для оценки смешивающей матрицы и СИ [6, 10–11]. Однако методы СВП не позволяют разделять СИ с одинаковым (идентичным) спектром мощности или независимые и одинаково распределенные СИ.

Использование метода АНК для зашумленных данных

В присутствии шума оценки разделяющей матрицы \mathbf{W} и матрицы смешивания $\hat{\mathbf{H}}$ достигаются более сложными процедурами, чем в случае отсутствия шума или при его несущественной значимости. Методическая схема, пригодная для ВФС, по зашумленным данным ориентируется на алгоритмы, робастные к шуму, хотя иногда оканчивается возможным интерпретировать шум как неизвестный СИ.

В отдельных прикладных системах возможно измерение или моделирование шума среды, в которой анализируются информационные сигналы. В этом случае шум называют опорным, или референтным ($v_p(k)$). Шум v_p (за счет среды его рас-

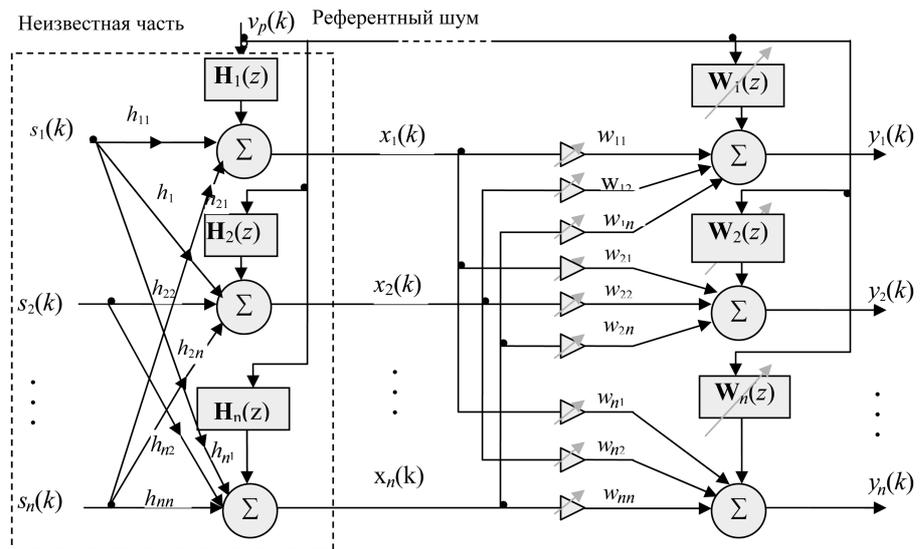


Рис. 3. Структура преобразований, обеспечивающая подавление шума и разделение сигналов при их линейном преобразовании в форме свертки

пространства) может влиять на каждый сенсор неизвестным образом: возможны эффекты задержки, реверберации, эха и нелинейных искажений. Можно считать, что перед тем как достигнуть сенсора, референтный шум обрабатывается неизвестной динамической системой. В простом случае система преобразования может описываться операцией свертки, где v_p обрабатывается некоторым фильтром (рис. 3). В этом случае одновременно происходит два процесса обучения:

- процедура несупервизорного обучения ("без учителя"), осуществляющая разделение сигналов;
- супервизорная ("с учителем") процедура, реализующая алгоритм снижения шума.

Этот подход использован Каруненом, Сичоки, Каспрзаком и Пезуненом (Karhunen, Sichoeki, Kasprzak, Pejunen) [12] и Сичоки, Каруненом, Каспрзаком и Вигарио (Vigario) [13] для снижения шума в предположении, что референтный шум доступен для оценки.

Для удаления шума (безотносительно к многоканальной задаче ВФС) разработан ряд методов [14–16], которые с некоторыми модификациями могут быть использованы для удаления шума при разделении СИ и ВФС. Однако в общем случае решение задачи удаления шума крайне трудно (а порой и невозможно), так как получается $(m + n)$ сигналов (n СИ и m сигналов шума).

Разделение сигналов при многоканальной смеси в форме свертки

Структурная схема процедур, связанных с восстановлением сигнала при его преобразовании в форме свертки, показана на рис. 4, а для однока-

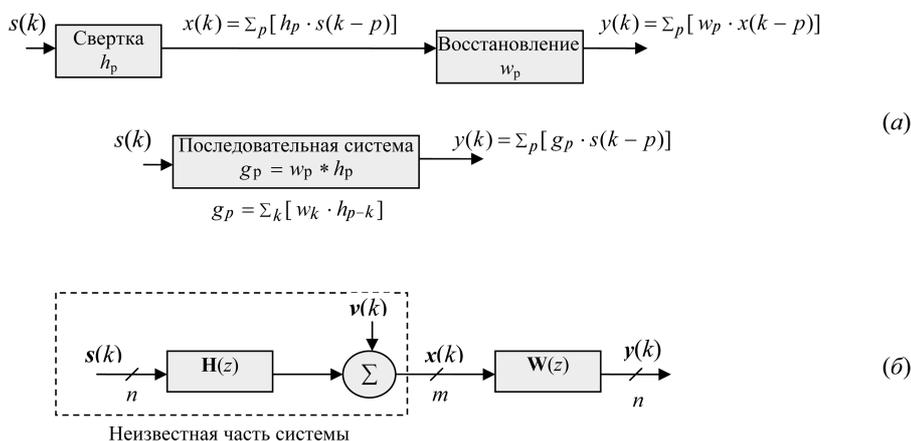


Рис. 4. Схема процедур, связанных с восстановлением сигнала при его преобразовании в виде свертки: одноканальный (а) и многоканальный (б) случаи

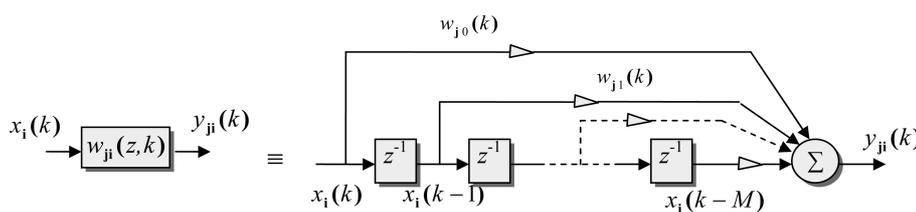


Рис. 5. Структура синаптических весов НС в составе адаптивной системы, показанной на рис. 2, б

нального случая (два эквивалентных вида схемы) и рис. 4, б — для многоканальной смеси компонент вектор-сигнала. В многоканальном случае измеряемый сигнал $\mathbf{x}(k) = [x_1(k), x_2(k), \dots, x_m(k)]^T$ образуется как смесь типа свертки из компонент $s(k) = [s_1(k), s_2(k), \dots, s_n(k)]^T$ ($n < m$). Это отражает устойчивая модель

$$\mathbf{x}(k) = \sum_{p=-\infty}^{\infty} \mathbf{H}_p s(k-p) = \mathbf{H}_p * s(k)$$

при $\sum_{p=-\infty}^{\infty} \|\mathbf{H}_p\| < \infty$, (5)

где $*$ — операция свертки; \mathbf{H}_p — $(m \times n)$ — матрица коэффициентов смешивания для СИ с временной задержкой p .

Выражение $\mathbf{H}(z) = \sum_{p=-\infty}^{\infty} \mathbf{H}_p \cdot z^{-p}$ (где z^{-1} — операция задержки на единицу времени) является Z -преобразованием для набора $\{\mathbf{H}_p\}$, т. е. матрицей передаточных функций системы. При этом соотношение (5) приобретает форму $\mathbf{x}(k) = [\mathbf{H}(z)] \cdot s(k)$.

Задача разделения сигналов многоканальной свертки состоит в определении (возможно, масштабированных и задержанных) СИ с использованием приблизительных знаний о распределении СИ их статистических свойствах. Предпола-

гается, что каждый СИ $s_i(k)$ является независимо и одинаково распределенной последовательностью (по k), которая статистически независима от всех других последовательностей — компонент СИ. Для восстановления СИ могут быть использованы модели нейронных сетей (НС), показанные на рис. 2, б и 3.

Задача разделения сигналов из смеси описывается многоканальным адаптивным фильтром с конечной импульсной характеристикой и дискретным временем (рис. 5).

Каждый вес НС (элемент матрицы \mathbf{W}) задается его Z -преобразованием:

$$W_{ji}(z, k) = \sum_{p=0}^M w_{jip}(k) \cdot z^{-p}.$$

Сигналы источника оцениваются с помощью модели НС, которая представляется соотношением

$$y_j(k) = \sum_{i=1}^{\infty} \sum_{p=-\infty}^{\infty} w_{jip} x_i(k-p) \quad (j = 1, 2, \dots, n)$$

или его матричной формой

$$\mathbf{y}(k) = \sum_{p=-\infty}^{\infty} \mathbf{W}_p(k) \mathbf{x}(k-p) = \mathbf{W}_p(k) * \mathbf{x}(k) = \mathbf{W}(z, k) \mathbf{x}(k),$$

где $\mathbf{y}(k) = [y_1(k), y_2(k), \dots, y_n(k)]^T$ — это n -мерный вектор на выходе системы; $\mathbf{W}(k) = \{\mathbf{W}_p(k), -\infty \leq p \leq \infty\}$ — последовательность $n \times m$ коэффициентов матриц; $\mathbf{W}(z, k)$ — Z -преобразование матричной передаточной функции в виде $\mathbf{W}(z, k) = \sum_{p=-\infty}^{\infty} \mathbf{W}_p(k) \cdot z^{-p}$.

Цель адаптивного разделения СИ из смеси в форме свертки достигается подстройкой матрицы $\mathbf{W}(z, k)$ таким образом, чтобы общая система (последовательность матриц смешивания $\mathbf{H}(z)$ и восстановления $\mathbf{W}(z, k)$) представляла СИ, преобразованные перестановкой компонент (\mathbf{P}) и масштабированием ($\mathbf{D}(z)$): $\mathbf{W}(z, k) \cdot \mathbf{H}(z) = \mathbf{P} \cdot \mathbf{D}(z)$ ⁽²⁾.

(2) \mathbf{P} — $(n \times n)$ — матрица перестановок; $\mathbf{D}(z)$ — диагональная $(n \times n)$ — матрица, (i, j) -й элемент которой равен $c_i z^{-\Delta_i}$; c_i — числовой множитель, а Δ_i определяет временную задержку компонент СИ.

Последовательное извлечение сигналов из смеси в задаче ВФС

При решении задачи разделения СИ и восстановления их формы в случае линейной смеси или смеси в виде свертки может быть использовано два подхода: в первом (описанном выше) разделяются одновременно все СИ, во втором — выделяются сигналы последовательно, один за другим. Второй подход целесообразен, когда в многоканальной системе используется много сенсоров (электродов, микрофонов или других преобразователей), а информационно значимым является только малое число сигналов, поступающих на сенсоры (рис. 6)⁽³⁾.

Сигналы могут извлекаться в определенном порядке в соответствии с их статистическими признаками, например, в порядке, определяемом величиной нормализованного эксцесса. Логика такой процедуры тесно связана с логикой восстановления СИ из смеси в форме многоканальной свертки [17] и вместе с тем может трактоваться как обобщение анализа главных компонент (АГК), где декоррелированные выходные сигналы извлекаются в порядке уменьшения значения их дисперсии. Последовательное извлечение СИ из смеси может выполняться с использованием последовательно связанных нейронных сетей аналогично тому, как это осуществляется при АГК.

Для таких систем целесообразна разработка алгоритмов обучения НС, устойчивых к небольшим уровням шума и позволяющих извлекать из смеси только небольшое число сигналов, которые представляют интерес и содержат полезную информацию.

Восстановление формы сигналов из смеси в форме многоканальной свертки: модель пространства состояния

В общем случае (при наличии СИ и шума) динамические системы смешивания сигналов и восстановления их формы могут описываться моде-

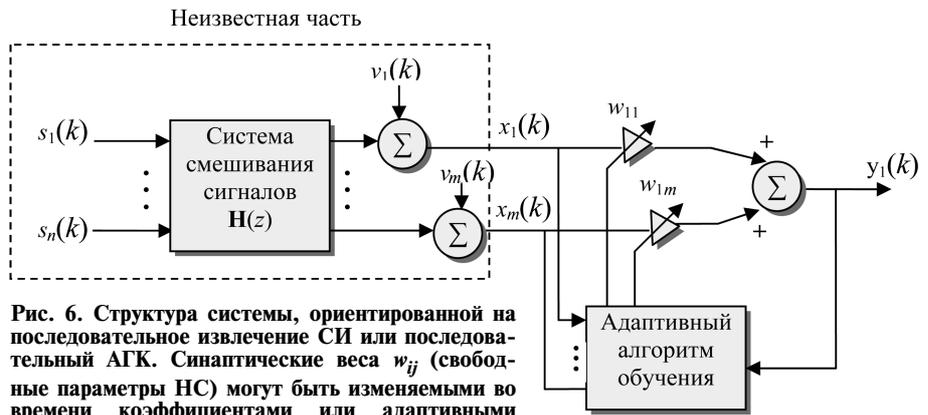


Рис. 6. Структура системы, ориентированной на последовательное извлечение СИ или последовательный АГК. Синаптические веса w_{ij} (свободные параметры НС) могут быть изменяемыми во времени коэффициентами или адаптивными фильтрами

люю пространства состояния (рис. 7), которая представлена соотношениями

$$\begin{aligned} \bar{\xi}(k+1) &= \bar{\mathbf{A}} \cdot \bar{\xi}(k) + \bar{\mathbf{B}} \cdot \mathbf{s}(k) + \bar{\mathbf{N}} \cdot \mathbf{v}_p(k); \\ \mathbf{x}(k) &= \bar{\mathbf{C}}(k) \cdot \bar{\xi}(k) + \bar{\mathbf{D}}(k) \cdot \mathbf{s}(k) + \mathbf{v}(k), \end{aligned} \quad (6)$$

где $\bar{\xi}$ — вектор состояния системы; $\mathbf{s}(k) (\in \mathbb{R}^n)$ — вектор неизвестных входных сигналов (предполагается, что компоненты его имеют нулевое среднее и имеют не-Гауссовы одинаковые и взаимно независимые распределения); $\mathbf{x}(k)$ — вектор сигналов сенсоров; $\mathbf{v}_p(k)$ — вектор шума; $(r \times r)$ -матрица $\bar{\mathbf{A}}$ — матрица состояния; $(r \times n)$ -матрица $\bar{\mathbf{B}}$ — входная матрица смешивания; $(m \times r)$ -матрица $\bar{\mathbf{C}}$ — выходная матрица смешивания; $(m \times n)$ -матрица $\bar{\mathbf{D}}$ — матрица смешивания вход-выход; $(r \times p)$ -матрица $\bar{\mathbf{N}}$ — матрица преобразования шума.

Передаточная функция $\mathbf{H}(z)$ является $(m \times n)$ -матрицей и имеет вид

$$\mathbf{H}(z) = \bar{\mathbf{C}} \cdot (z \cdot \mathbf{I} - \bar{\mathbf{A}})^{-1} \cdot \bar{\mathbf{B}} + \bar{\mathbf{D}},$$

где z^{-1} — символ операции задержки (т. е. $z^{-1} \cdot \mathbf{x}(k) = \mathbf{x}(k-1)$).

В аналогичной форме представляется модель системы, реализующей ВФС:

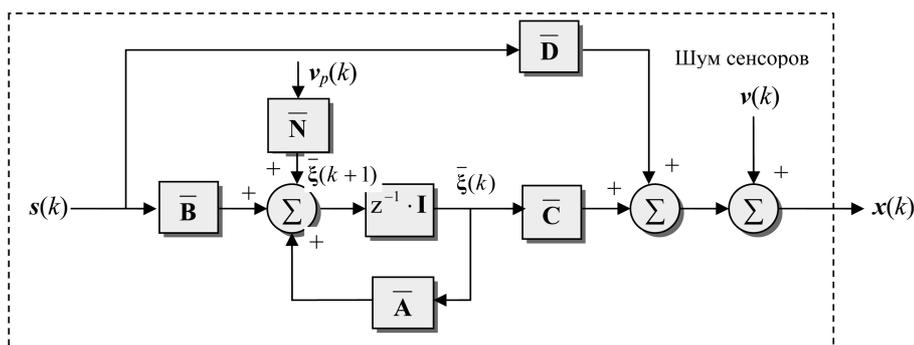
$$\begin{aligned} \xi(k+1) &= \mathbf{A} \cdot \xi(k) + \mathbf{B} \cdot \mathbf{x}(k) + \mathbf{L} \cdot \mathbf{y}_p(k); \\ \mathbf{y}(k) &= \mathbf{C}(k) \cdot \xi(k) + \mathbf{D}(k) \cdot \mathbf{x}(k). \end{aligned} \quad (7)$$

Соотношение (7) включает неизвестные матрицы пространства состояния: $(M \times M)$ -матрица \mathbf{A} ; $(M \times m)$ -матрица \mathbf{B} ; $(m \times M)$ -матрица \mathbf{C} ; $(m \times m)$ -матрица \mathbf{D} и $(M \times m)$ -матрица \mathbf{L} . При этом имеется в виду, что $M \geq r$, т. е. порядок системы ВФС должен быть не менее, чем порядок системы смешивания сигналов.

Концепция модели пространства состояния, отражающей общую форму смешивания сигналов источника и адаптивную процедуру их разделения при динамическом анализе независимых компонент, показана на рис. 7.

⁽³⁾ В системах регистрации электроэнцефалограммы и магнитоэнцефалограммы используются 64 (и более) сигналов сенсоров, но только небольшое их число представляет интерес, а остальные могут рассматриваться как шумы интерференции. Таким же образом бывает необходимо выделить речевые сигналы (РС) только нескольких человек из большого числа регистрируемых РС многих (и часто одновременно говорящих) участников конференции.

Динамический процесс смешивания



Динамический процесс смешивания

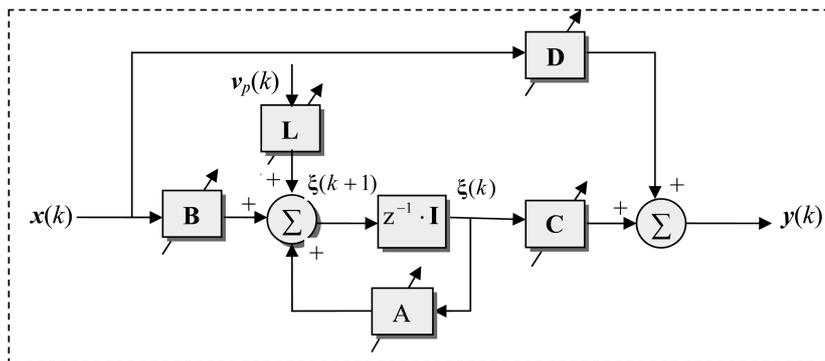


Рис. 7. Структура общей модели пространства состояния, включающей линейное смешивание (компонент вектора состояния) и адаптивную модель разделения для динамического анализа независимых компонент. Цель алгоритмов обучения состоит в оценке набора матриц $\{A, B, C, D, L\}$

Можно считать, что линейная модель пространства состояния является расширением модели одновременного разделения сигналов и ВФС. В частном случае, когда матрицы \bar{A} , \bar{B} , \bar{C} в модели смешивания и матрицы A , B , C в модели разделения сигналов являются нуль-матрицами, общая модель упрощается и отражает метод анализа независимых компонент. Элементы набора матриц $\Theta = [A, B, C, D, L]$ являются параметрами, которые определяются в процессе обучения на основе знания временной последовательности $x(k)$ и, возможно, некоторых априорных знаний о системе в целом.

Таким образом, динамическое ВФС реализует получение вида первоначальных СИ по регистрируемым значениям $x(k)$ без использования априорных знаний о СИ и о матрицах пространства состояния $[\bar{A}, \bar{B}, \bar{C}, \bar{D}]$. Передаточная функция, определяющая восстановление сигналов, имеет вид: $W(z) = C \cdot (z \cdot I - A)^{-1} \cdot B + D$. Обычным (и практически не ограничивающим) предположением является взаимная независимость СИ и равенство нулю их средних значений. В случае отсутствия шума СИ могут быть восстановлены по выходным сигналам $y(k) = [y_1(k), y_2(k), \dots, y_m(k)]^T$ в соответствии с соотношением

$$y(k) = [W(z) \cdot H(z)] \cdot s(k) = [D(z)] \cdot P \cdot s(k), \quad (8)$$

где P — обобщенная $(m \times n)$ -матрица перестановок (общего вида); $D(z) = \text{diag}\{D_{11}(z), D_{22}(z), \dots, D_{mm}(z)\}$ — диагональная матрица с передаточными функциями $D_{ii}(z)$, определяющими форму фильтров⁽⁴⁾.

В некоторых приложениях требуется, чтобы $D_{ii}(z) = \lambda_i z^{\tau_i}$, — τ_i , где λ_i — постоянный ненулевой коэффициент масштаба, а τ_i — временная задержка.

Область приложений обработки сигналов с разделением (квазиразделением) смеси

Задачи анализа независимых компонент (АНК), восстановление вида первичных сигналов источника (СИ) при линейной их смеси и смеси в форме многоканальной свертки потенциально могут иметь применение в следующих приложениях и областях исследований:

- обработка и анализ биомедицинских сигналов: электроэнцефалограмм (ЭЭГ), магнито-энцефалограмм (МЭГ), электрокардиограмм (ЭКГ);
- обработка геофизических данных;
- интеллектуальный анализ баз данных сложной структуры ("data mining");
- усиление речевых сигналов (ВФС при наличии шума);
- распознавание изображений и совершенствование системы беспроводной связи [10, 18—22].

В этих приложениях бывает доступным большее число сигналов сенсоров или данных, которые являются фильтрованной суперпозицией отдельных различных и независимых СИ, а цель со-

⁽⁴⁾ Матрица перестановок P состоит из n ненулевых элементов и имеет только по одному (ненулевому) элементу в каждом столбце. Размерность $(m \times m)$ -матрицы $D(z)$ предполагает, что число выходов системы равно числу сенсоров, хотя в практике бывает, что число СИ может быть меньше, чем число сенсоров (т. е. $m \geq n$). Используемая модель определяется двумя факторами: 1) число сенсоров может быть неизвестным и может изменяться со временем; 2) в практике иногда сигналы аддитивного шума удобно считать вспомогательными источниками и их целесообразно извлечь из смеси, сохраняя остальные информативные СИ. В идеальном случае отсутствия шума избыточные $(m - n)$ выходные сигналы y_i должны исчезать в ходе адаптивного процесса обучения и только n выходов будут соответствовать восстановленным СИ.

стоит в обработке наблюдений таким образом, чтобы выход соответствовал разделенным первичным СИ.

Разделение биомедицинских сигналов и ВФС осуществляется в следующих исследованиях:

- восстановление сигнала ЭКГ плода (удаление или фильтрация сигнала ЭКГ матери);
- отделение ЭКГ трансплантированного сердца от сигнала ЭКГ "старого" сердца [23];
- удаление артефактов и шумов из регистрируемых ЭЭГ и МЭГ. Усиление компонент ЭЭГ и МЭГ с низким уровнем.

Акустические приложения включают ситуации, в которых СИ — это сигналы от различных микрофонов в звуковом поле, производимом несколькими людьми (задача выделения индивидуального речевого сигнала), или СИ — это сигналы от нескольких акустических преобразователей в подводном звуковом поле от шума двигателей нескольких судов (задача гидроакустического контроля). Примеры радио- и беспроводной связи включают наблюдения, соответствующие элементам массива с выхода антенны (как отклик от нескольких передатчиков) и могут также включать эффект взаимной связи элементов. В радиолокационной технике встречается наложение сигналов за счет различных механизмов модуляции сигналов от целей, наблюдаемых несколькими приемниками с различной чувствительностью к компонентам поляризации.

Заключение

Рассмотрены методы, позволяющие получать форму отдельных сигналов по измерительной информации об их смеси. Основой алгоритмов и нейросетевых структур, реализующих восстановление формы сигналов (ВФС) — компонент смеси, служат концепции, связанные с особенностями формирования смеси сигналов, наличием статистических характеристик шума, сопровождающего регистрируемые сигналы на сенсорах, и с методологией снижения уровня шума в восстановленных сигналах.

Приведены элементы общей концепции смешивания первичных сигналов источника (СИ), их фильтрации и разделения при наличии шума или без него.

Даны структуры преобразований, обеспечивающих реализацию полного восстановления первичных сигналов источника (СИ) или последовательное выделение компонент вектора сигнала, который регистрируется сенсорами в форме смеси неизвестного вида. Дана трактовка задачи ВФС как задачи анализа независимых компонент и критерий Кульбака—Лейблера для оценки степени независимости компонент.

Проанализированы особенности задачи разделения СИ при многоканальной смеси в форме свертки и показана логика процедур для решения такой задачи.

Рассмотрена модификация задачи ВФС, решение которой основано на последовательном извлечении сигналов из смеси. Сигналы могут извлекаться в определенном порядке в соответствии с их статистическими признаками.

Дан метод решения задачи ВФС на основе модели пространства состояний. Модель отражает форму смешивания СИ и адаптивную процедуру их разделения при динамическом анализе независимых компонент.

Показана область приложений обработки сигналов с целью ВФС.

Идеи концепции и методы ВФС по регистрируемым данным смеси СИ на сенсорах ИИС разрабатывались Каруненом и Джоутсенсало (Karhunen, Joutsensalo) [3], Плумбле (Plumbley) [4], Белроучрани, Абед-Мераимом, Кардосо и Моулинесом (Berlouchrani, Abed-Meraim, Cardoso, Moulines) [6], Амари и Сичоки (Amari, Cichocki) [10], Сервьером и Баундосом (Serviere, Baudois) [20], Пхамом и Кардосо [24].

Список литературы

1. Айвазян С. А., Енюков И. С., Мешалкин Л. Д. Прикладная статистика. Исследование зависимостей. М.: Финансы и статистика, 1985. 488 с.
2. Comon P. Independent component analysis: A new conception? / Signal Processing. 1994. V. 36. P. 278—314.
3. Karhunen J., Joutsensalo J. Generalizations of principal component analysis, optimization problems and networks // Neural Networks. 1995. V. 8. P. 549—562.
4. Plumbley M. Conditions for nonnegative independent component analysis // IEEE Transactions on Neural Networks, 2003. V. 14. N 3. P. 534—543.
5. Jutten C., Herault J. Separation of sources: Adaptive algorithm based on neuron network architecture // Signal Processing. 1991. V. 24. N 1. P. 1—10.
6. Berlouchrani A., Abed-Meraim K., Cardoso J.-F., Moulines E. A source separation technique using second-order statistics // IEEE Transactions on Signal Processing. 1997. V. 45. N 2. P. 434—444.
7. Гантмахер Ф. П. Теория матриц. М.: Наука, 1988. 548 с.
8. Малыхина Г. Ф., Меркушева А. В. Адаптивный алгоритм на основе рекуррентной сети для задачи восстановления формы линейно смешанных сигналов // Нейрокомпьютеры. 2006. № 7. Часть 2. С. 3—13.
9. Кульбак С. Теория информации и статистика. М.: Наука, 1967. 408 с.
10. Amari S., Cichocki A. Adaptive blind signal processing — neural network approaches // Proceedings of IEEE. 1998. V. 86. P. 1186—1187.
11. Belouchrani A., Cichocki A. Robust whitening procedure in source separation context // Electronic Letters. 2000. N 36 (24). P. 2050—2053.
12. Karhunen J., Cichocki A., Kasprzak W., Pejunen P. On neural method for signal separations with noise suppression and redundancy reduction // International Journal of Neural Systems. 1997. V. 8. N 2 (April). P. 219—237.
13. Cichocki A., Karhunen J., Kasprzak W., Vigarrio R. Neural networks for signal separation with unknown number of sources // Neurocomputing. 1999. V. 24. N 1—3 (February). P. 55—93.

14. **Xie L.** Robust Kalman Filtering for uncertain Systems // System and Control Letters. 1994. V. 22. P. 123–133.

15. **Van Gerven S., Van Comprenolle D.** The use of de-correlation in scalar signal separation // International Conference on Acoustics, Speech and Signal Processing. 1994. V. III. P. 57–60.

16. **Малыхина Г. Ф., Меркушева А. В.** Сеть с симметричной функцией преобразования нейронов для подавления искажений и восстановления изображения // Научное приборостроение. 2008. Т. 18. № 2. С. 73–80.

17. **Shalvi O., Weinstein E.** New criteria for de-convolution of non-minimum phase systems (channels) // IEEE Transaction on Information Theory. 1990. V. 36. N 2. P. 312–321.

18. **Столингс В.** Беспроводные линии связи и сети. М.: Изд. дом "Вильямс", 2003. 637 с.

19. **Gardner W. A.** A new method of clinical identification // IEEE Transactions on Communications. 1991. V. 39. N 6. P. 813–817.

20. **Serviere C., Baudois D.** Source separation with noisy observations: a noise canceling application // Signal Processing, 1995. V. 42. N 1. P. 45–57.

21. **Shalvi O., Weinstein E.** New criteria for signal de-convolution of non-minimum phase systems (channels) // IEEE Transactions on Information Theory. 1993. V. 39. N 2. P. 504–519.

22. **Skarbek W., Cichocki A., Kasperzak W.** Principal subspace analysis for incomplete image data in one learning epoch // Neural Network World. 1996. V. 6. N 3. P. 375–382.

23. **Nanui A. K., Zaraoso V.** Forth order cumulant based source signal separation // Signal Processing Letters. 1996. V. 3. N 12. P. 312–314.

24. **Pham D. T., Cardoso J. F.** Separation of instantaneous mixtures of non stationary sources // IEEE Transaction on Signal Processing. 2001. V. 49. N 9. P. 1837–1848.

УДК 621.865

Х. Г. Асадов, д-р техн. наук,
нач. отдела атмосферных измерений,
И. Г. Чобанзаде, аспирант,
Азербайджанское национальное
аэрокосмическое агентство,
e-mail: asadzade@rambler.ru

Информационный метод радиометрической коррекции и калибровки бортовых спектрометров в ходе полета

Предложен информационный метод осуществления радиометрической коррекции бортовых спектрометров при полете. Показано, что используемые методы радиометрической коррекции могут нарушить достоверность гиперспектрометрических данных. Согласно предложенному методу в кадре съемки спектрометра следует выделить участок с максимальной информативностью исходя из некоторых заданных ограничений и далее для учета участков кадра, не включенных в данный контур, применить внутрисеточные и межсеточные корреляционные связи внутри кадра.

Ключевые слова: радиометрическая коррекция, калибровка, информация, спектрометр, атмосфера.

Общеизвестно, что в конечном счете эффективность использования бортовых спектрометрических систем определяется качеством предполетной и полетной коррекции и калибровки параметров функционирования. Например, как указано в [1], важной задачей обработки космических снимков с высоким динамическим диапазоном является осуществление предполетной и полетной радиометрической и геометрической калибровки и коррекции. Одним из путей радиометри-

ческой коррекции является использование масочных устройств с динамически управляемым коэффициентом пропускания для предотвращения насыщения фотоприемников при больших интенсивностях входного сигнала. В общем случае проблема приема и обработки зашумленных изображений прямо или косвенно связана как с радиометрической, так и с геометрической калибровкой бортовых спектрометрических систем, поскольку масочный вариант решения указанной проблемы в случае шумящего фотоприемника может быть заменен вариантом увеличения времени экспозиции за счет выбора определенной геометрической траектории полета. Например, как указывается в [2], в спектрометрах с множеством углов наблюдения, в частности в аппаратуре CHRIS/PROBA, для повышения отношения сигнал/шум спутник должен осуществлять медленное движение по вращательной траектории относительно объекта съемки. При этом дистанция, пройденная спутником, оказывается длиннее засканированной поверхности, хотя создается впечатление, что скорость движения спутника замедляется.

Таким образом, необходимость в радиометрической коррекции приводит к реализации сложных траекторных решений относительно сканируемого объекта. Как сообщается в работе [3], погрешности, влияющие на точность георектификации, могут быть разделены на три группы:

- погрешности статического наведения;
- погрешности динамического наведения;
- погрешности, связанные с топографией проекционной поверхности.

Для устранения этих погрешностей должны быть поэтапно приобретены следующие объемы информации в последовательном порядке:

1) данные для верного статического нацеливания внутренней геометрии (индивидуальной ориентации) камер;

2) вспомогательные данные для поддержания георектификации при динамических погрешностях нацеливания;

3) статистические параметры и базовые оценки для оценки качества георектификации.

Однако очевидно, что метод накопления (интегрирования) слабых сигналов также не может быть признан в качестве универсального решения проблемы применительно к динамическим космическим системам. Например, как указывается в работе [4], гиперспектральные системы осуществляют сканирование перпендикулярно к линии полета и вследствие нерегулярности топографии и/или траектории полета пиксели могут резко изменить свое место. Это обстоятельство указывает на особую важность точной динамической геометрической коррекции при работе гиперспектрометров. Для осуществления такой коррекции возможно использование бортовой навигационной информации. Например, в работе [5] была предложена серия уравнений для корректировки данных AVIRIS при движении на платформе ER-2. Однако при этом топографические эффекты (изменение высоты наземных объектов) не были учтены. В работе [6] была использована бортовая GPS — инерциальная навигационная система (ИНС) C-MIGITS-II — для получения координат x, y, z (от GPS) и трех осевых высотных данных (от ИНС) для полной модели фотометрической камеры. Эта модель использовала лучевое слежение для обнаружения местонахождения каждого пикселя на поверхности Земли и потенциально может быть применена к любым гиперспектральным данным при наличии соответствующей аппаратуры навигационной привязки.

Суммируя вышеизложенное, можно заключить, что стремление любой ценой извлечь информацию с зашумленных участков изображения зачастую оборачивается чрезмерным усложнением полетной корректировочной и калибровочной аппаратуры.

В такой ситуации общая проблема повышения информативности космических систем должна быть решена с позиций оптимизации исходной информативности бортовых спектрометров в смысле достижения максимальной информативности систем за счет отказа от высокоточной регистрации удаленных малоинформативных зашумленных участков кадра, полезность которых весьма сомнительна.

Таким образом, исходная проблема совместной радиометрической и геометрической калибровки в целях регистрации зашумленного сигнала трансформируется в задачу определения таких достаточных условий калибровки и коррекции, при которых извлекаемый кадр обладал бы максимальной информативностью.

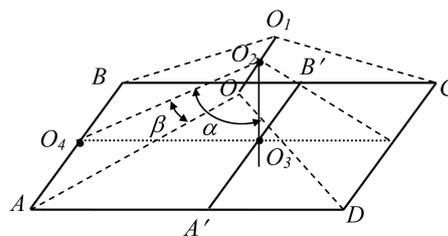


Рис. 1. Упрощенное схематическое представление кадра MISR. $A'B'$ — строка, формируемая при надирных измерениях. O_1O_2 — щель на входе спектрометра

Таким образом, предлагаемый подход к радиометрической коррекции при полете состоит в следующем:

- определение в кадровом формате контура потенциально наиболее информативного участка с учетом влияния атмосферы; определение малоинформативных зашумленных участков в кадре вне указанного контура;
- анализ целесообразности применения вышеописанных методов для коррекции и повышения информативности выделенных зашумленных участков;
- использование межстрочных и межэлементных связей при дешифровке малосканированных слабоинформативных участков.

Рассмотрим порядок применения указанного подхода к бортовому спектрометру MISR, размещенного на борту спутника TERRA [7].

Упрощенная схема формирования кадра в спектрометре MISR показана на рис. 1. Работа спектрометра MISR с многими углами обзора построена так, что в каждый фиксированный момент спектрометр с помощью девяти объективов осуществляет сканирование определенного участка $ABCD$ под девятью разными углами обзора, соответственно чему формируется кадр из девяти строк.

В соответствии с изложенным подходом к радиометрической коррекции и калибровке следует вычислить контуры кадра, реализация которого путем соответствующей конфигурации сканирования позволила бы достичь максимального количества информации в выделенном контуре кадра с учетом наложенных ограничений.

Обозначим промежуток O_2O_3 буквой H , а промежуток O_2O_4 как h . В этом случае имеем

$$h = \frac{H}{\sin(90^\circ - \alpha)}. \quad (1)$$

Вычислим длину строки $AB = l$. Имеем

$$l = 2h \sin \beta = \frac{2H \sin \beta}{\sin(90^\circ - \alpha)}. \quad (2)$$

В первом приближении, приняв все элементы кадра равноинформативными, информативность

одной строки длиной l вычислим следующим образом:

$$M = \frac{l}{\Delta l} \log_2[(S_0 - S'_h h)], \quad (3)$$

где S_0 — отношение сигнал/шум по направлению $O_4 O_2$ приема оптического сигнала; S'_h — первая производная S_0 по h .

Учитывая (1), (2), запишем:

$$M = \frac{2H \sin \beta}{\Delta l \sin(90^\circ - \alpha)} \log_2 \left[S_0 - S'_h \frac{H}{\sin(90^\circ - \alpha)} \right]. \quad (4)$$

Задачу вариационной оптимизации сформулируем следующим образом: следует вычислить оптимальную функцию $\beta = \beta(\alpha)$ при заданном ограничении

$$\int_0^{a_{\max}} \frac{2H \sin \beta(\alpha)}{\sin(90^\circ - \alpha)} d\alpha = C, \quad (5)$$

при котором нижеприведенный информационный функционал достигает максимального значения

$$M_{\text{и}} = \int_0^{a_{\max}} \frac{2H \sin \beta(\alpha)}{\Delta l \sin(90^\circ - \alpha)} \log_2 \left[S_0 - S'_h \frac{H}{\sin(90^\circ - \alpha)} \right] d\alpha + \lambda \int_0^{a_{\max}} \frac{2H \sin \beta(\alpha)}{\sin(90^\circ - \alpha)} d\alpha, \quad (6)$$

где λ — множитель Лагранжа.

Для вычисления оптимальной функции $\beta(\alpha)$ используем метод Эйлера, согласно которому должно быть

$$\frac{dM_{\text{и}}}{d[\beta(\alpha)]} = 0. \quad (7)$$

С учетом (7) из (6) получаем

$$\log_2 \left[S_0 - \frac{S'_h H}{\sin(90^\circ - \alpha)} \right] = -\lambda \Delta l. \quad (8)$$

Из (8) имеем

$$\frac{(S_0 - e^{-\lambda \Delta l})}{S'_h} = \frac{H}{\sin(90^\circ - \alpha)}. \quad (9)$$

С учетом (9) ограничительное условие (5) принимает вид

$$\frac{2(S_0 - e^{-\lambda \Delta l})}{S'_h} \int_0^{a_{\max}} \sin \beta(\alpha) d\alpha = \text{const}. \quad (10)$$

Из выражения (10) имеем

$$-\lambda = \frac{1}{\Delta l} \log_2 \left[S_0 - \frac{C S'_h}{2 \int_0^{a_{\max}} \sin \beta(\alpha) d\alpha} \right]. \quad (11)$$

Из формул (8) и (11) получаем

$$\frac{H}{\sin(90^\circ - \alpha)} = \frac{C}{2 \int_0^{a_{\max}} \sin \beta(\alpha) d\alpha}, \quad (12)$$

или

$$\int_0^{a_{\max}} \sin \beta(\alpha) d\alpha = \frac{C \sin(90^\circ - \alpha)}{H}. \quad (13)$$

Из полученного интегрального тригонометрического уравнения (13) ясно, что при увеличении α значение $\beta(\alpha)$ должно уменьшаться.

Таким образом, показано, что применение изложенного подхода к спектрометру MISR приводит к необходимости использования особой конфигурации мгновенного кадра съемки, где прямоугольный контур кадра должен приобрести овальный вид (рис. 2). При этом малоинформативные участки кадра (заштрихованные участки на рис. 2) могут быть учтены двумя способами:

- специальные мероприятия по уменьшению зашумленности (радиометрическая коррекция) осуществляются только применительно к этим участкам;
 - в указанных участках особые меры по радиометрической коррекции не предпринимаются. Недостающая информация для многокурсовой интерпретации изучаемых земных объектов извлекается на основе внутрисканной и межстрочной корреляции внутри кадра.
- Суммируя изложенное выше, в качестве заключения отметим следующее:
- оперативная радиометрическая и геометрическая коррекция и калибровка бортовых спектрометров в ходе полета неотделимы друг от друга для реальных земных объектов с учетом топографических, атмосферных и других эффектов;

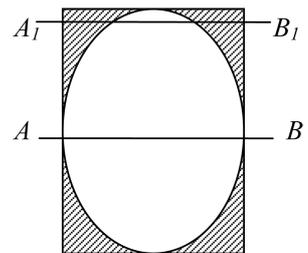


Рис. 2. Формируемый кадр спектрометра

- предлагаемый альтернативный подход к осуществлению необходимой коррекции в ходе полета заключается в выделении контура особо информативной части кадра, обеспечивающего максимальную информативность кадра при некоторых принятых ограничениях. Места исходного прямоугольного кадра, не включенные в контур максимально информативной части, могут быть подвергнуты дополнительной радиометрической коррекции или отброшены. Взамен этих участков используется информация, содержащаяся в межэлементных и межстрочных связях, имеющихся в кадре.

Список литературы

1. **Nayar Sh. K.** High Dynamic Range Imaging: Spatially Varying Pixel Exposures. http://www.dpreview.com/news/0009/382_nayar.pdf.

2. **Alonso L., Moreno J.** Quasi — automatic Geometric Correction and Related Geometric Issues in the Exploitation of Chris/Proba Data // Proceeding of the 2nd CHRIS/Proba Workshop, ESA / ESRIN, Frascati, Italy, 28—30 April (ESA SP-578, July 2004). ESA Publication Division. ESTEC. The Netherlands.

3. **Jovanovich V. M.** MISR Automatic Geometric Quality Assessment and In-Flight Geometric Calibration Updates // Pecora 15/Land Satellite Information IV/ISPRS Commissions I / FIEOS 2002 Conference Proceedings. 2002. <http://www.isprs.org/comrnission/proceedings02/paper/00060.pdf>.

4. **Aspinall R. J., Marcus W., Boardman J. W.** Considerations in collecting, processing and analyzing high spatial resolution hyperspectral data for environmental investigations // Journal of Geographical systems. 2002. N 4. P. 15—29.

5. **Clark R. N., Livo K. E., Kokalu R. F.** Geometric correction of AVIRIS imagery using on board navigation and engineering data // AVIRIS 1998 Proceedings, JPL, California. 1998. 9 p. http://makalu.jp./nasa.gov/docs/workshops/98_docs/9.pdf.

6. **Boardman J. W.** Precision geocoding of low altitude AVIRIS data: lessons learned in 1998 // AVIRIS 1999 Proceedings, JPL, Californiz. 1999. 6 p. http://makalu.jp./nasa.gov/docs/workshops/98_docs/7.pdf.

7. **Multi-angle Imaging SpectroRadiometer (MISR).** <http://www-misr.jpl.nasa.gov>.

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В СОЦИАЛЬНЫХ И ЭКОЛОГИЧЕСКИХ СИСТЕМАХ

УДК 004.652

А. С. Бождай, канд. техн. наук, доц.,

Пензенский государственный университет, e-mail: bozhday@yandex.ru

Комплексная инфраструктура территории: методы и модели информационного мониторинга

Обсуждается перспективный подход к информационному мониторингу социально-экономических систем. Вводится новое понятие комплексной инфраструктуры территории (КИТ). Для формализованного представления КИТ рассматривается метод на основе декомпозиции (для социальных инфраструктур низкой и средней сложности). При этом рассматривается возможность интеграции арсенала передовых информационных технологий (OLAP, ГИС, САПР, телекоммуникации) в рамках единой информационно-математической модели.

Ключевые слова: комплексная инфраструктура территории, информационный мониторинг, декомпозиция, социально-экономическая система.

Введение

Состояние и развитие современного общества определяется наличием ряда важных особенностей, которые обуславливают усиление взаимовлияния всех его составных частей на экономическом, политическом, научно-техническом и социокультурном уровнях. Совокупная среда жизнедеятельности человека представляет собой сейчас такую резонансную среду, где любая, незначительная в структурном плане, подсистема может своим функционированием повлечь глобальные системообразующие последствия.

Любой процесс, протекающий как в социуме, так и в природе, можно рассмотреть с двух основных позиций: вещественно-энергетической и информационной. В классической теории управления традиционно было принято выносить на первый план вещественную основу процессов. В связи с этим был детально проработан концептуальный и математический аппарат, подразумевающий приоритет жестких вещественно-энергетических воздействий на различные объекты со стороны внешней среды (регуляторов).

Ситуация начала меняться коренным образом на рубеже 80-х и 90-х годов XX века, когда нача-

лась всеохватывающая информатизация общества. Это явилось резким качественным изменением всего облика социума после длительного количественного наращивания энергетических, производственных и вычислительных мощностей. Суть нового облика среды жизнедеятельности человека определяется наступлением такой фазы развития, когда определяющую роль начали играть информационные ресурсы и взаимодействия. В итоге, имеются все основания для того, чтобы говорить об *обществе информационного типа*, представляющего собой открытую систему, функционирующую на основе преобразования и распределения информации в области "единого информационного поля" [1].

Другим важным фактором, определяющим характер жизнедеятельности человека, выступает территория. Многочисленные системы, образуемые вследствие этой жизнедеятельности, жестко привязаны к территориально-географическому аспекту. Отдельно взятый пространственный участок, даже небольшой протяженности, может одновременно объединять значительное количество антропогенных систем как взаимосвязанных, так и разрозненных. Таким образом, территориальный аспект позволяет лицам, принимающим решения (ЛПР), рассматривать в системном единстве всю совокупность естественно-природных, антропогенных и техногенных систем, объединяя их в единую целостность.

Все перечисленные свойства современной среды жизнедеятельности человека и связанные с ними проблемы управления актуализируют необходимость постановки задачи, связанной с разработкой общих принципов и моделей мониторинга этой среды с учетом факторов географического местоположения, информатизации и самоорганизационных проявлений.

Комплексная инфраструктура территории

При постановке задачи мониторинга той или иной социально-экономической системы (СЭС) в первом приближении создается иллюзия, что объектом исследования должна служить только лишь эта самая, отдельно взятая, СЭС. Большинство традиционных методов мониторинга базируются именно на такой позиции. Головные управляющие организации, занимаясь аналитикой работы своих подведомственных иерархий, зачастую прибегают к упрощенной схеме, заключающейся в периодическом сборе первичной информации (например, с помощью специально разработанных форм унифицированного образца), ее централизации, автоматической очистке и последующей обработке различными математико-статистическими методами. Органы управления обра-

зованием, здравоохранением и прочих СЭС проводят эти мероприятия обособленным друг от друга образом, не пытаясь осуществлять корреляцию форм первичной отчетности, не учитывая текущих состояний и тенденций смежных систем и т. п.

Даже оставив пока в стороне проблему достоверности первичной информации, предоставляемой учреждениями низших уровней иерархии, можно сделать заключение о том, что рассмотренный подход позволяет лишь *констатировать* общее текущее состояние СЭС, достигнутое за отчетный период. Важнейшие вопросы, связанные с нахождением факторов, которые способствовали переходу системы в это состояние (то есть *причин*), остаются за пределами видимости.

Очевидно, что любая СЭС регионального масштаба является открытой системой, входящей в состав общей инфраструктуры, более обширной как по территориальному, так и тематическому охвату. Административно-хозяйственные границы СЭС не могут препятствовать явным или косвенным воздействиям со стороны смежных подсистем инфраструктуры. Иными словами, инфраструктура жизнедеятельности человека в рамках выбранного территориального охвата является единым системным организмом и многие причинно-следственные закономерности следует искать именно в масштабах всей инфраструктуры.

Для терминологической идентификации такого системного единства предлагается ввести специальное понятие — комплексная инфраструктура территории (КИТ), определить которое можно следующим образом. *КИТ — совокупность антропогенных, техногенных и природно-географических систем, представляющих собой системную целостность в рамках выбранного пространственно-временного масштаба.*

Концепция неразрывной связи между природными и антропогенными системами неоднократно обсуждалась многими учеными в рамках различных научных дисциплин (география, геоэкология, физика, социология, информатика, нелинейная динамика, синергетика и др.). Особо стоит упомянуть работы Д. Л. Арманда, А. Д. Арманда, Ю. П. Михайлова, В. Б. Сочавы, В. Н. Солнцева и др. [3, 4]. Проведенный анализ всего терминологического многообразия данной концепции показал, что можно соотнести с КИТ два наиболее близких по сути и хорошо изученных понятия — *геосистема* и *природно-техническая система (ПТС)*.

Однако несмотря на видимое сходство, имеются серьезные причины различать геосистемы, ПТС и КИТ с точки зрения специфики и тематики решаемых задач. КИТ — это понятие, в котором основной акцент делается на информацион-

ной составляющей, описывающей социально-экономические инфраструктурные процессы для поддержки ЛПР и информационно-аналитических исследований. Комплексность территориальной инфраструктуры обеспечивается за счет пространственной привязки всех подсистем этой инфраструктуры к единой территории в единой системе координат. Границы КИТ определяются исходя из социально-экономического (инфраструктурного) аспекта. Сложность КИТ оценивается на основе информационной сложности соответствующей инфраструктуры.

В качестве основных элементов структуры КИТ выделим: участок территории; слой инфраструктуры; информационное пространство.

Под *участком территории* в структуре КИТ понимается участок земной поверхности, вмещающий всю материальную основу подсистем КИТ и ограниченный их пространственной протяженностью. Подсистемы КИТ и соответствующий участок территории имеют неразрывные связи различного характера и оказывают существенное воздействие друг на друга.

Классификация социально-экономических функций в привязке к территориальным, информационным и организационным аспектам приводит к понятию *слоя инфраструктуры*. Под слоем инфраструктуры будем понимать тематически обособленную сферу жизнедеятельности человека, неразрывно связанную с соответствующим участком территории, взаимодействующую с другими слоями КИТ и внешней средой, обладающую собственной материально-информационной структурой и органами управления, подлежащую исследованию в рамках мониторинга КИТ. Типичные слои инфраструктуры представлены на рис. 1.

Информационное пространство, в котором существует такая сложная система, приобретает особый статус и дает представление о внутренних состояниях КИТ, внешней среде и совокупности управляющих воздействий. Проектирование информационного пространства и механизмов его использования в рамках создания автоматизированных мониторинговых методов и систем приобретает первостепенную важность.

В зависимости от уровня, специфики и широты задачи мониторинга предлагаемая концепция определяет две точки зрения на построение формализованной модели КИТ — *модель на основе декомпозиции* и *модель на основе целостного анализа*.

Модель КИТ на основе декомпозиции

Данная точка зрения на технологию мониторинга близка к традиционным методам системной техники. Она подразумевает разбиение (декомпозицию) сложной комплексной системы на иерар-



Рис. 1. Обобщенная структура КИТ

хию более мелких подсистем нижнего уровня с последующим их локальным анализом. Далее на основе результатов локального анализа выстраивается общесистемная картина. Целостность процесса мониторинга обеспечивается за счет специальных, связанных в одну систему моделей, построенных на стыке целого ряда научных информационно-компьютерных технологий (теория графов, геоинформационные системы, многомерные хранилища данных — OLAP, методы интеллектуального анализа данных).

Преимуществом такого подхода является простота и наглядность моделей, хорошо отработанный и надежный математический аппарат, алгоритмическая простота, численная статистическая точность результатов мониторинга, наглядность мониторинговых отчетов.

Однако имеет место и ряд важных недостатков, ограничивающих применение этого метода в некоторых случаях. Прежде всего, мониторинг на основе декомпозиции ограничивается порогом структурной и поведенческой сложности КИТ. При анализе крупных, структурно сложных КИТ (например, в масштабе федерального округа) имеют место: резкое усложнение процесса декомпозиции; огромная размерность результатов локального анализа; повышенные требования к вычислительной и ресурсной емкости аппаратного обеспечения; необходимость учета склонности сложных систем к самоорганизации и хаотическим проявлениям.

Другой проблемой при использовании метода декомпозиции является возможная высокая степень информационной гетерогенности слоев КИТ (разные формы отчетности, форматы представления данных, программно-технические платформы и т. п.). Предлагаемая ниже модель, в принципе, позволяет решить эту проблему, но это может потребовать дополнительных временных, вычислительных и материальных расходов.

В рамках данного подхода КИТ рассматривается с точки зрения четырех различных аспектов,

охватывающих все стороны ее структурной и функциональной составляющих.

1. **Общесистемный аспект** — КИТ рассматривается с точки зрения гетерогенных информационных и организационных связей между ее подсистемами, а также с точки зрения их классификаций, возникающих с течением времени мониторинга.

2. **Логический аспект** — КИТ рассматривается с точки зрения иерархической декомпозиции общей структуры ее подсистем по критериям информационной совместимости.

3. **Пространственный аспект** — КИТ рассматривается с учетом вмещающей ее пространственно-географической среды.

4. **Физический аспект** — КИТ рассматривается с точки зрения присущих ее подсистемам информационных массивов, программно-аппаратных платформ, протоколов обмена, прав доступа к информации, телекоммуникационной среды.

В соответствии с перечисленными аспектами обобщенная информационно-математическая модель КИТ состоит из взаимосвязанных моделей четырех уровней: абстрактно-математического, информационно-логического, информационно-пространственного, физического.

Модель абстрактно-математического уровня. Главной особенностью модели абстрактно-математического уровня является представление КИТ в виде единой модели, в которой присутствуют все ее элементы и описываются некоторые возможные связи между ними. Модель представляется в виде гиперграфа, структура гиперребер которого является результатом классификации элементов КИТ в зависимости от конкретных задач мониторинга. На разных стадиях и в разных процедурах мониторинга классификация может существенно изменяться, вызывая динамическую реструктуризацию гиперребер. Таким образом, для модели абстрактно-математического уровня имеет место свойство информационной гетерогенности, т. е. единая модель в виде гиперграфа одновременно описывает все типы элементов, участвующих в процессе мониторинга, и предоставляет их динамические классификации в зависимости от конкретных задач мониторинга (рис. 2).

В теоретико-множественном представлении модель абстрактно-математического уровня (гиперграф AMG) состоит из двух множеств (V, U) и бинарного предиката P :

$$AMG = (V, U, P). \quad (1)$$

Множество V описывает структуру гиперграфа на уровне вершин:

$$V = \{v_i\}, i = 1, 2, \dots, N, \quad (2)$$

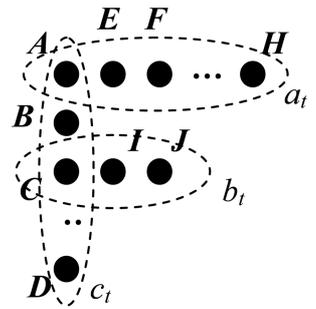


Рис. 2. Модель абстрактно-математического уровня с одним слоем гиперребер: $V = \{A, B, C, \dots, J, \dots\}$ — множество вершин гиперграфа (подсистем КИТ); $U = \{a_t(A, E, F, \dots, H), b_t(C, I, J), c_t(A, B, C, \dots, D), \dots\}$ — множество гиперребер (классов подсистем КИТ) в момент времени t

где N — общее число вершин, соответствующее числу элементов КИТ.

Множество U имеет переменную мощность и описывает многослойную структуру гиперграфа на уровне гиперребер:

$$U = \{u_j\}_f; j = 1, 2, \dots, K_t; f = 1, 2, \dots, F_t, \quad (3)$$

где K_t — число гиперребер в момент времени t , соответствующее текущим правилам взаимодействия между подсистемами КИТ; f — номер слоя гиперребер; F_t — число слоев.

Структура гиперребер однозначно отражает общую картину любых регламентированных взаимодействий между подсистемами КИТ или их различных классификаций. Множество слоев гиперребер позволяет классифицировать тип взаимодействий, представляя удобный механизм абстрагирования. Множество слоев гиперребер позволяет интерпретировать предлагаемую гиперграфовую модель в виде нескольких (F_t) гиперграфов с одинаковой структурой вершин и различной структурой гиперребер.

Порядок взаимодействий элементов может существенно меняться при прохождении различных стадий мониторинга, поэтому мощность множества U с течением времени мониторинга также может существенно меняться. Причем имеют место следующие граничные значения мощности M множества U и числа слоев f :

$$M_{\min}(U) = 1, M_{\max}(U) = N, f_{\min} = 1. \quad (4)$$

Бинарный предикат P определяет инцидентность вершин и гиперребер каждого слоя; P определен на множестве всех пар ($v \in V, u \in U$).

Модель информационно-логического уровня. Для разработки методов учета и анализа информационного обмена между подсистемами КИТ необходимо преобразовать полную гетерогенную модель в строго ранжированную, иерархическую структуру. Классы подсистем КИТ (образованные по типу их взаимоотношений — гиперребра), определенные на абстрактно-математическом уровне, логически объединяются в домены (контейнеры). Вся совокупность доменов объединяется в

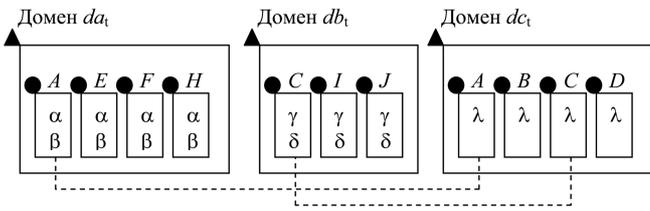


Рис. 3. Модель логического уровня и соответствующий многомерный куб: $a_t, b_t, c_t \in U$ — классы подсистем КИТ в момент t , логически представленные в виде доменов; $A, B, C, \dots, J \in V$ — подсистемы КИТ; $\alpha, \beta, \gamma, \delta, \lambda$ — массивы данных, предоставляемые подсистемами КИТ в ходе взаимодействия; ---- — междоменные связи через инцидентные подсистемы КИТ

единую информационную модель — многомерный куб и соответствует осям его измерений.

Теоретико-множественное представление модели логического уровня DM выглядит следующим образом:

$$DM = (D, Pd). \quad (5)$$

Здесь D — множество доменов, образованных исходя из текущей структуры гиперребер AMG [см. ф-лу (3)]:

$$D = \{d_i\}, \quad i = 1, 2, \dots, K_t; \quad (6)$$

Pd — бинарный предикат, определенный на множестве всех пар (d_i, d_j) , где $i \neq j, i = 1, 2, \dots, K_t; j = 1, 2, \dots, K_t$. Его истинность определяет наличие междоменного взаимодействия между доменами d_i и d_j .

На примере рис. 2 и 3 можно сформулировать правила преобразования структуры гиперребер графа AMG в доменную структуру DM :

$$a_t(va_t) \Rightarrow da_t(va_t(Q_a)) \quad (7)$$

где $a_t \in U_f$ — гиперребро слоя f ; $va_t \subset V$ — множество вершин, инцидентных a_t ; da_t — домен логического уровня, соответствующий a_t ; Q_a — множество данных, соответствующих va_t .

Как видно из рис. 3, домены da_t и dc_t инцидентны, благодаря общему элементу A (на гиперграфе рис. 1 эта вершина инцидентна гиперребрам a_t и c_t). Домены db_t и dc_t инцидентны благодаря общему элементу C .

Таким образом, на логическом уровне, с одной стороны, сохраняется единство математической модели (единый многомерный куб), а с другой стороны, снимается проблема гетерогенности за счет организации разнородных данных в разные

домены (т. е. по разным осям измерения куба) по критерию совместимости. Для анализа взаимодействий подсистем КИТ на логическом уровне предлагается использовать технологию OLAP (online analytical processing).

Модель информационно-пространственного уровня. Каждому домену логического уровня ставится в соответствие объект информационно-пространственного уровня, обладающий универсальной информационной составляющей — пространственной привязкой (например, реальные географические координаты), которая однозначно увязывает домен и все его содержимое (подсистему КИТ и присущую ему информацию) с другими смежными социально-экономическими системами и прочими слоями КИТ. Пространственная привязка доменов и вложенной в них информации осуществляется к цифровой картографической основе территории, которая вмещает КИТ. Для представления потоков данных снова предлагается использовать теорию графов (рис. 4).

В теоретико-множественном представлении модель КИТ информационно-пространственного уровня будет иметь вид:

$$ISG = (V', U'), \quad (8)$$

где $V' = \{v'_{i,x,y}(P_i)\}, i = 1, 2, \dots, N$ — множество вершин, каждой из которых соответствует пространственно-географическая привязка (x, y) и некоторый набор данных P_i ; N — число вершин;

$U' = \{u'_j(e)\}, j = 1, 2, \dots, M$ — множество ориентированных ребер, каждое из которых имеет вес e , соответствующий объему передаваемых данных

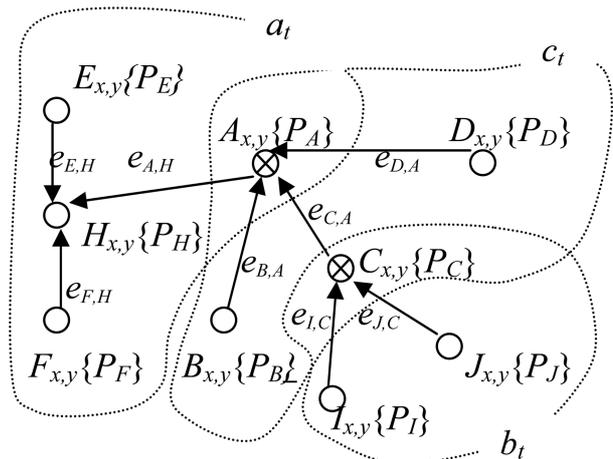


Рис. 4. Модель КИТ информационно-пространственного уровня: A, B, \dots, H — подсистемы КИТ; x, y — пространственная привязка элемента; $\{P_i\}$ — данные, которые описывают подсистему i ; $e_{i,j}$ — поток данных от элемента i к элементу j ; \otimes — инцидентные подсистемы

между инцидентными вершинами; M — число ребер (взаимосвязей между подсистемами КИТ).

Главной особенностью модели информационно-пространственного уровня является пространственная привязка статистики, обрабатываемой в ходе мониторинга. Это обеспечивает возможность обработки информации без отрыва от того пространственно-географического окружения, которому она соответствует. Поскольку пространственная привязка (в данном случае географические координаты) является неотъемлемым атрибутом каждой подсистемы КИТ, то таким образом мы получаем универсальную составляющую в разрозненном множестве данных, относительно которой любые процессы КИТ могут быть рассмотрены в системном единстве.

Кроме того, информационно-пространственный уровень обеспечивает решение задач формирования наглядной отчетной документации по результатам мониторинга в виде картографических атласов, схем, планов, тем самым существенно упрощая процессы административного принятия решений.

Модель физического уровня. Объекты информационно-пространственного уровня на физическом уровне позволяют рассматривать физические коммуникации между подсистемами КИТ. Например, они могут ассоциироваться с узлами телекоммуникационной сети, сохраняя свою доменную принадлежность логического уровня. Предложенные на логическом уровне инцидентные подсистемы устанавливают порядок междоменного шлюзования гетерогенных участков физической сети и политику информационной безопасности (рис. 5).

Также представляется эффективным использовать арсенал геоинформационных систем и наличие пространственной координации данных, определенной на информационно-пространственном уровне, для моделирования коммуникаций подсистем КИТ на физическом уровне.

В теоретико-множественном представлении модель физического уровня будет иметь вид:

$$PhG = (V'', U''), \quad (9)$$

где $V'' = \{v_i''(W_i)\}$, $i = 1, 2, \dots, N$ — множество вершин, каждая из которых моделирует программно-аппаратную платформу подсистемы КИТ; N — число вершин, W_i — совокупность данных, предоставляемая узлом v_i'' ;

$U'' = \{u_j''(h)\}$, $j = 1, 2, \dots, M$, — множество ориентированных ребер, каждое из которых имеет вес h , соответствующий характеристикам коммуникационной связи между инцидентными подсистемами КИТ, M — число ребер (телекоммуникационных взаимосвязей между элементами КИТ).

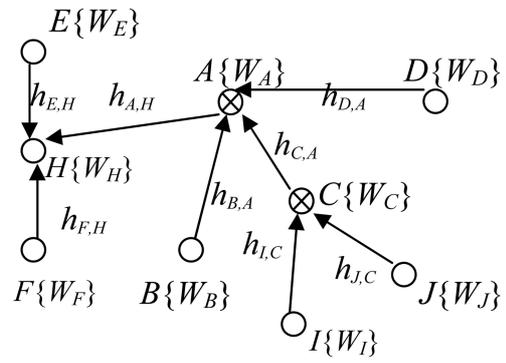


Рис. 5. Модель КИТ физического уровня: A, B, \dots, H — программно-аппаратные платформы элементов КИТ (узлы сети); $\{W_i\}$ — данные, которые хранятся на узле i ; $h_{i,j}$ — телекоммуникационная среда от узла i к узлу j ; \otimes — узел, выполняющий роль шлюза между двумя доменами сети

Таким образом, обобщенная математическая модель КИТ имеет вид:

$$CM = (AMG, DM, ISG, PhG), \quad (10)$$

где AMG — модель абстрактно-математического уровня (1); DM — модель логического уровня (5); ISG — модель информационно-пространственного уровня (8); PhG — модель физического уровня (9).

Предлагаемая модель на основе декомпозиции хорошо подойдет для систем наблюдения, оценивания и контроля за КИТ малой и средней сложности. При этом имеет место комплексная интеграция технологий САПР, ГИС, OLAP, интеллектуального анализа данных для создания универсальной системы мониторинга с централизованным автоматизированным управлением.

Существенные недостатки декомпозиционной модели проявляются при описании систем высокого уровня сложности (масштаба федерального округа и выше). В поведении таких систем начинают проявляться факторы самоорганизации, а информация хаотической природы может оказывать системообразующее воздействие. Такие сложные сферы мониторинга целесообразно рассматривать с позиции моделирования *на основе целостного анализа* [5] — как единое целое, и подвергать анализу колебания выборки ключевых показателей, совокупность трендов которых образует фазовый портрет КИТ и позволяет эффективно применять методы нелинейной динамики и синергетики.

Заключение

Очевидно, что ЛПР нуждаются в таком инструментарии, который, в масштабах определенного территориального охвата, позволил бы эффективно анализировать и наблюдать за управляемыми системами с учетом новых реалий информационного общества. Необходимо не просто

учитывать фактологию происходящего, но и иметь средства для реализации следующих управленческих задач.

1. Установка информационных связей, зависимостей и влияний между различными сегментами антропогенной, техногенной и природной сред.

2. Построение аналитических срезов и выборов для любого сочетания тематических слоев. Это будет способствовать решению экспертных проблем абстрагирования, генерализации, классификации, и тем самым позволит предельно минимизировать информационное пространство, в котором рассматривается управляемый процесс.

3. Нахождение скрытых тенденций, резонансов и взаимосвязей между различными процессами жизнедеятельности человека.

4. Построение отчетов, наглядных и оптимально насыщенных полезной информацией.

Рассмотренная концепция КИТ предоставляет новые расширенные возможности в решении перечисленных задач. Принципиальным моментом является возможность разработки универсальной модели КИТ, которая позволит увидеть целост-

ность и территориальную инвариантность сферы жизнедеятельности человека, выделить особенности взаимного влияния различных социально-экономических процессов друг на друга, выявить неявные закономерности, упорядочить в рамках единой модели гетерогенные массивы статистической информации.

Список литературы

1. Абдеев Р. Ф. Философия информационной цивилизации. М.: Владос, 1994.
2. Сочава В. Б. Введение в учение о геосистемах. Новосибирск: Наука, 1978.
3. Арманд А. Д. Самоорганизация и саморегулирование географических систем. М.: Наука, 1988.
4. Бершадский А. М., Бождай А. С. Информационный мониторинг глобальных сред жизнедеятельности человека // Приложение к журналу "Открытое образование": Материалы Международной конференции "Информационные технологии в науке, социологии, экономике и бизнесе. IT + SE'2004", Украина, Крым, Ялта — Гурзуф, 2004 г. — С. 59—61.
5. Бершадский А. М., Бождай А. С. Методы и модели информационного мониторинга социальной инфраструктуры территории // Известия Волгоградского государственного технического университета: межвуз. сб. научн. ст. 2007. № 1 (27). С. 17—22.

УДК 004.023

А. С. Филиппова, д-р техн. наук, доц., Д. П. Брусиловский, соискатель,
Уфимский государственный авиационный технический университет,
e-mail: annamuh@mail.ru

Прогнозирование в ситуации непостоянного спроса для оптимального управления складскими запасами производственно-сборочных компаний

Рассматривается проблема оптимального управления складскими запасами малых и средних производственно-сборочных компаний (ПСК). Представлена система для управления складскими запасами ПСК и математическая модель, на базе которой формируется механизм принятия решений. Работа посвящена ситуации, когда спрос измеряется с помощью интервальной шкалы и не является постоянным, а данных прошлых периодов мало (всего 4—8 измерений). В такой ситуации классические подходы не работают. Далее формулируется и иллюстрируется новый подход, основанный на устойчивой регрессии.

Ключевые слова: управления складскими запасами; непостоянный спрос; модели Кростона; устойчивая регрессия; системы принятия решений планирования спроса.

Введение

Прогнозирование спроса на товар жизненно важно для любого предприятия — при поставках, на производстве, в розничной торговле. Предсказание будущего спроса определяет количество товаров, которое нужно закупить, произвести, отгрузить. Понимание потребительского спроса жизненно важно для любого производителя това-

ра. Он может выпустить столько товара, сколько нужно, чтобы удовлетворить потребительский спрос. Точные прогнозы являются движущей силой всей системы поставок. Они обеспечивают вводные данные для планирования спроса, производственного планирования и управления запасами на складе.

Планирование спроса для производственных и сборочных компаний, строящих свою работу на

основе спроса, должно решать для каждой сборочной единицы две основные задачи:

- определить время размещения заказа на пополнение складских запасов (задача расписания);
- определить объем заказа.

Прогнозирование спроса на период времени, для которого рассчитываются резервные запасы, представляет собой сложную, неструктурированную проблему.

На практике часто встречаются складские запасы, потребность в которых непостоянна. Данные для таких складских единиц моделируются временными рядами неотрицательных целых чисел, причем некоторые из них равны нулю.

Численные подходы к прогнозированию спроса, разработанные за последние десять лет, обсуждаются в [1]. Выделены типичные ситуации прогнозирования спроса на основе доступности и точности информации о спросе. Эти ситуации соответствуют различным шкалам измерения спроса: дихотомия (бинарное прогнозирование спроса, реализованное с помощью регрессионных моделей двоичных зависимых переменных), ординальная шкала, счетная шкала (на основе регрессионной модели Пуассона) и интервальная шкала. Новые подходы — дихотомия, ординальная шкала, счетная шкала — рассматриваются в [1].

Данная работа посвящена ситуации, когда спрос измеряется с помощью интервальной шкалы и не является постоянным, а исторических данных мало — всего 4—8 наблюдений.

1. Модель Кростона для прогнозирования непостоянного спроса

Для прогнозирования непостоянного спроса наиболее часто используется метод Кростона [2]. Он основан на прогнозах простого экспоненциального сглаживания отдельно для объема спроса и отдельно для интервалов между событиями спроса.

Обозначим ряды спроса прошлых периодов Y_1, Y_2, \dots, Y_n и предположим, что они принимают целые неотрицательные значения.

Пусть Y_t — спрос за временной период t , а X_t — переменная-индикатор для периодов ненулевого спроса, т. е. $X_t = 1$, если за временной период t наблюдался спрос, и $X_t = 0$, если спроса за этот период не было. Далее, пусть j_t — число периодов с ненулевым спросом за интервал $[0, t]$, такой что $j_t = X_i, i = 1$, т. е. j_t — индекс ненулевого спроса. Для простоты обозначений опустим букву t . Далее, пусть Y_j^* представляет размер j -го ненулевого спроса, а Q_j — время между возникновениями событий спроса Y_{j-1}^* и Y_j^* . Используя эти обозначения, можно написать, что $Y_t = X_t Y_{j_t}^*$.

Метод Кростона прогнозирует отдельно размер ненулевого спроса, отдельно интервал между

Таблица 1

Исходные данные

Изменение	Число элементов	Дата заказа	Идентификатор детали	Номер заказа	Интервал (в днях)
1	10	03 ОКТЯБРЯ 2007	34615	1	
2	20	10 ОКТЯБРЯ 2007	34615	2	7
3	0	16 ОКТЯБРЯ 2007	34615	3	6
4	10	24 ОКТЯБРЯ 2007	34615	4	8
5	20	31 ОКТЯБРЯ 2007	34615	5	7
6	10	06 НОЯБРЯ 2007	34615	6	6
7	20	13 НОЯБРЯ 2007	34615	7	7
8	20	20 НОЯБРЯ 2007	34615	8	7
9	10	27 НОЯБРЯ 2007	34615	9	7
10	20	05 ДЕКАБРЯ 2007	34615	10	8
11	10	19 ДЕКАБРЯ 2007	34615	11	14
12	10	30 ЯНВАРЯ 2008	34615	12	42
13	20	06 СЕНТЯБРЯ 2006	40959	13	1
14	10	07 НОЯБРЯ 2006	40959	14	62
15	10	03 АПРЕЛЯ 2007	40959	15	147
16	10	06 ИЮНЯ 2007	40959	4	64
17	20	03 ИЮЛЯ 2007	40959	5	27
18	40	10 АВГУСТА 2007	40959	6	38
19	60	13 СЕНТЯБРЯ 2007	40959	7	34
20	50	28 ФЕВРАЛЯ 2008	40959	8	168
21	20	16 ОКТЯБРЯ 2006	59413	1	
22	50	08 НОЯБРЯ 2006	59413	2	23
23	50	22 ЯНВАРЯ 2007	59413	3	75
24	100	21 ФЕВРАЛЯ 2007	59413	4	30
25	30	12 МАРТА 2007	59413	5	19
26	50	10 АПРЕЛЯ 2007	59413	6	29
27	100	27 АПРЕЛЯ 2007	59413	7	17
28	50	14 МАЯ 2007	59413	8	17
29	50	25 МАЯ 2007	59413	9	11
30	50	07 ИЮНЯ 2007	59413	10	13
31	100	22 ИЮНЯ 2007	59413	11	15
32	200	18 ИЮЛЯ 2007	59413	12	26

последовательными событиями спроса с помощью простого экспоненциального сглаживания. Прогнозы уточняются только после возникновения событий спроса. Пусть Z_j и P_j — прогнозируемые размер $(j + 1)$ -го спроса и время между событиями спроса соответственно, основанные на данных спроса, предшествующих событию спроса с номером j . Тогда по методу Кростона получаем:

$$Z_j = (1 - \alpha)Z_{j-1} + \alpha Y_j; \quad (1)$$

$$P_j = (1 - \alpha)P_{j-1} + \alpha Q_j. \quad (2)$$

Коэффициент сглаживания α принимает значения от 0 до 1. Предполагается, что он один и тот же для Y_j^* и Q_j .

Первоначально этот метод основывался на следующих допущениях:

- ненулевые значения размера спроса Y_j^* являются одинаково, независимо и нормально распределенными;
- временные интервалы между событиями спроса Q_j являются одинаково, независимо и геометрически распределенными;
- объемы спроса Y_j^* и временные интервалы между событиями спроса Q_j взаимно независимы.

Мы реализовали модель Кростона и провели с ее помощью анализ данных о заказах сборочно-производственной компании за прошедшие периоды (табл. 1). Результаты анализа представлены в табл. 2 и 3.

Таблица 2

Однократное экспоненциальное сглаживание: двухступенчатый прогноз с линейным трендом

Измерение	Идентификатор детали	Номер заказа	Тип	Шаг	Спрогнозированный размер заказа	Спрогнозированный интервал
1	34615	13	ПРОГНОЗ	1	17,393	15,780
2	34615	13	L70	1	9,684	2,739
3	34615	13	U70	1	25,102	28,822
4	34615	14	ПРОГНОЗ	2	18,065	16,252
5	34615	14	L70	2	10,274	3,072
6	34615	14	U70	2	25,856	29,433
7	40959	9	ПРОГНОЗ	1	57,902	88,017
8	40959	9	L70	1	42,587	9,211
9	40959	9	U70	1	73,217	166,823
10	40959	10	ПРОГНОЗ	2	64,600	90,487
11	40959	10	L70	2	49,121	10,841
12	40959	10	U70	2	80,078	170,134
13	59413	13	ПРОГНОЗ	1	112,734	1,331
14	59413	13	L70	1	59,245	-20,222
15	59413	13	U70	1	166,224	22,884
16	59413	14	ПРОГНОЗ	2	118,334	-2,971
17	59413	14	L70	2	64,274	-24,754
18	59413	14	U70	2	172,394	18,811

Примечание: L70 и U70 — соответственно нижняя и верхняя границы доверительного интервала с коэффициентом доверия 70 %.

Таблица 3

Однократное экспоненциальное сглаживание: двухступенчатый прогноз без тренда

Измерение	Идентификатор детали	Номер заказа	Тип	Шаг	Спрогнозированный размер заказа	Спрогнозированный интервал
1	34615	13	ПРОГНОЗ	1	13,732	15,192
2	34615	13	L95	1	1,538	-8,083
3	34615	13	U95	1	25,927	38,466
4	34615	14	ПРОГНОЗ	2	13,732	15,192
5	34615	14	L95	2	1,538	-8,083
6	34615	14	U95	2	25,927	38,466
7	40959	9	ПРОГНОЗ	1	34,219	80,974
8	40959	9	L95	1	-8,828	-53,059
9	40959	9	U95	1	77,267	215,008
10	40959	10	ПРОГНОЗ	2	34,219	80,974
11	40959	10	L95	2	-8,828	-53,059
12	40959	10	U95	2	77,267	215,008
13	59413	13	ПРОГНОЗ	1	92,030	21,101
14	59413	13	L95	1	-15,163	-18,509
15	59413	13	U95	1	199,222	60,711
16	59413	14	ПРОГНОЗ	2	92,030	21,101
17	59413	14	L95	2	-15,163	-18,509
18	59413	14	U95	2	199,222	60,711

Примечание: L95 и U95 — соответственно нижняя и верхняя границы доверительного интервала с коэффициентом доверия 95 %.

В модели Кростона обнаружено несколько проблем.

1. При небольшом числе измерений (небольшом объеме данных за прошлые периоды) возникают сложности в оценке параметра экспоненциального

Таблица 4

Совокупный спрос

Изменение	Идентификатор детали	Размер заказа	Дата заказа	Номер заказа	Совокупный размер заказа
1	34615	10	03 ОКТЯБРЯ 2007	1	10
2	34615	20	10 ОКТЯБРЯ 2007	2	30
3	34615	10	16 ОКТЯБРЯ 2007	3	40
4	34615	10	24 ОКТЯБРЯ 2007	4	50
5	34615	20	31 ОКТЯБРЯ 2007	5	70
6	34615	10	06 НОЯБРЯ 2007	6	80
7	34615	20	13 НОЯБРЯ 2007	7	100
8	34615	20	20 НОЯБРЯ 2007	8	120
9	34615	10	27 НОЯБРЯ 2007	9	130
10	34615	20	05 ДЕКАБРЯ 2007	10	150
11	34615	10	19 ДЕКАБРЯ 2007	11	160
12	34615	10	30 ЯНВАРЯ 2008	12	170
13	40959	20	06 СЕНТЯБРЯ 2006	1	20
14	40959	10	07 НОЯБРЯ 2006	2	30
15	40959	10	03 АПРЕЛЯ 2007	3	40
16	40959	10	06 ИЮНЯ 2007	4	50
17	40959	20	03 ИЮЛЯ 2007	5	70
18	40959	40	10 АВГУСТА 2007	6	110
19	40959	60	13 СЕНТЯБРЯ 2007	7	170
20	40959	50	28 ФЕВРАЛЯ 2008	8	220
21	59413	20	16 ОКТЯБРЯ 2006	1	20
22	59413	50	08 НОЯБРЯ 2006	2	70
23	59413	50	22 ЯНВАРЯ 2007	3	120
24	59413	100	21 ФЕВРАЛЯ 2007	4	220
25	59413	30	12 МАРТА 2007	5	250
26	59413	50	10 АПРЕЛЯ 2007	6	300
27	59413	100	27 АПРЕЛЯ 2007	7	400
28	59413	50	14 МАЯ 2007	8	450
29	59413	50	25 МАЯ 2007	9	500
30	59413	50	07 ИЮНЯ 2007	10	550
31	59413	100	22 ИЮНЯ 2007	11	650
32	59413	200	18 ИЮЛЯ 2007	12	850

сглаживания. В частности, эксперименты показали, что для восьми точек данных и прогноза, основанного на экспоненциальном сглаживании, реализованного в пакете статистических программ R, согласие данных более чем на 17 % определялось начальными данными. В результате границы доверительного интервала были неточными.

2. Нижняя граница интервала доверия для спрогнозированного спроса вполне может оказаться отрицательной.

3. Предположения о нормальности ненулевого спроса практически всегда нарушаются.

2. Новый подход к прогнозированию непостоянного спроса — модификация метода Кростона на базе устойчивой регрессии

Нами был разработан подход, преодолевающий описанные выше недостатки модели Кростона. Этот подход включает в себя следующие шаги.

1. Вместо текущего спроса используется совокупный спрос — он представляет собой неубывающую функцию по времени, и границы интервала доверия для спрогнозированного спроса не могут быть отрицательными.

2. Вместо простого экспоненциального сглаживания двух временных рядов — спроса и временного интервала между событиями спроса — используется одна регрессионная модель: совокупный спрос как функция времени спроса.

3. Для обработки выбросов и распределений, не являющихся нормальными, к данным спроса применяется метод устойчивой регрессии (M-оценивание, S-оценивание и MM-оценивание) приносят достаточно хорошие результаты, см. [3]). Минимальное число событий спроса (точек данных) для метода линейной устойчивой регрессии равно 5.

Следовательно, этот метод может начать работу с пяти точек данных, независимо от распределения спроса.

Для реализации этого подхода было разработано ядро системы принятия решений планирования спроса для производственных и сборочных компаний, строящих свою работу на основе спроса. Разработка велась на языках программирования R и C++.

Результаты применения этого метода представлены в табл. 4, 5 и 6.

Заключение

Был рассмотрен традиционный подход к прогнозированию непостоянного спроса на основе модели Кростона и описаны недостатки этого метода, когда спрос измеряется с помощью интервальной шкалы и не является постоянным, а исторических данных мало (всего 4—8 наблюдений).

Оценка углового коэффициента устойчивости регрессии и соответствующие показатели согласия

Идентификатор детали	Оценка	Среднеквадратичное отклонение	Нижняя граница интервала доверия	Верхняя граница интервала доверия	Значение χ -квадрата	Вероятность χ -квадрата	Коэффициент детерминации
34615	14,9241	0,3590	14,2205	15,6277	1728,36	<0,0001	0,8787
40959	27,6497	4,4766	18,8757	36,4236	38,15	<0,0001	0,8167
59413	62,0082	1,5304	59,0087	65,0076	1641,76	<0,0001	0,8559

Таблица 6

Прогноз устойчивой регрессии и соответствующие границы 95 %-ного интервала доверия

Измерение	Идентификатор детали	Оценка	Usage6week	Usage4week	Usage2week
1	34615	14,92	22,38	14,92	7,46
2	40959	27,64	41,47	27,64	13,82
3	59413	62,00	93,01	62,00	31,00
Usage6week_U	Usage6week_L	Usage4week_U	Usage4week_L	Usage2week_U	Usage2week_L
26,56	24,17	18,75	17,06	10,93	9,95
61,92	32,08	43,70	22,65	25,49	13,21
110,51	100,31	78,00	70,81	45,50	41,30

Примечание: Usage6week — прогнозирование спроса на 6 недель вперед, Usage6week_U и Usage6week_L — соответственно верхняя и нижняя границы доверительного интервала с коэффициентом доверия 95 %.

Был разработан и реализован новый подход к прогнозированию непостоянного спроса, не обладающий недостатками модели Кростона. Приводятся результаты применения этого метода к набору данных.

Список литературы

1. **Brusilovsky D.** A decision-making engine for optimal inventory management of the manufacturing assembly companies // The

Fifth International Conference on Mathematical Modeling and Computer Simulations of Materials Technologies. Israel Materials & Processes Society. 2008. P. 195—201.

2. **Shenstone L. and Hyndman R.** Stochastic models underlying Croston's method for intermittent demand forecasting. <http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/2003/wpl-03.pdf>.

3. **Rousseeuw P. and Leroy A.** Robust Regression and Outlier Detection // Wiley Series in Probability and Statistics. 1987. P. 14—58.

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ОБРАЗОВАНИИ

УДК 004.5

А. А. Штрик, д-р техн. наук, проф., гл. науч. сотр.,

Российский НИИ информационных технологий и систем автоматизированного проектирования,
e-mail: shtrik@land.ru

Индикаторы информационной грамотности*

Рассматриваются вопросы, связанные с концептуальной структурой построения системы индикаторов информационной грамотности (of information literacy-IL). Представлены результаты исследования таких вопросов, как определение информационной грамотности; модель, которая связывает информационную грамотность с существующими индикаторами, в частности, с уровнем квалификации и опыта в области информационно-коммуникационных технологий (ИКТ), а также описание стандартов информационной грамотности в сфере образования.

Ключевые слова: базовая грамотность, индикаторы, информационная грамотность, информационные технологии, социально-экономическое развитие, развитие индикаторов, стандарты информационной грамотности.

Определение информационной грамотности

В самом общем смысле информационная грамотность является средством, позволяющим людям определять свои информационные потребности,

локализовать и оценивать качество информации, накапливать и восстанавливать информацию, проводить эффективное и этически корректное использование информации, а также создавать другую информацию и обмениваться

* Настоящая статья подготовлена по материалам [1], опубликованным в рамках программы ЮНЕСКО "Информация для всех" (UNESCO's information for All Programme — IFAP).

знаниями. Информационную грамотность для каждого отдельного гражданина можно рассматривать как базовый фундамент для достижения ими персональных, социальных, профессиональных и образовательных целей. Навыки, определяемые информационной грамотностью, позволяют людям эффективно применять технологии пожизненного обучения, а также вносить свой вклад в развитие общественно полезных знаний.

Информационная грамотность играет важную роль в процессах выработки стратегии и тактики *экономического и социального развития* как на международном, так и национальных уровнях. Более того, практика показывает, что информационную грамотность следует рассматривать не только с позиций уровня образования людей, но и в более широком контексте, охватывающем такие сферы жизни, как работа, гражданское общество, здравоохранение, уровень благосостояния и многие другие. В свою очередь, такой широкий диапазон контекстов содержит вызов как для выработки политики развития, так и для разработки индикаторов информационной грамотности.

Информационная грамотность является критически важным инструментом для развития *здравоохранения и условий жизни людей*. В сфере здравоохранения следует рассматривать две различные области, где использование информации и информационных технологий является важным. Первая область — это специалисты, работающие в области медицины и здравоохранения, когда на уровне профессиональной практики использование онлайн-доступа к современным достижениям медицинской практики и возможность оперативного доступа к конкретным данным о пациентах позволяют повысить уровень информированности специалистов и соответственно качественный уровень здравоохранения. Вторая область — это обеспечение права людей на получение информации о здоровье и условиях жизни. Особенно важно это для бедных и слаборазвитых стран и слоев населения, где уровень базовой грамотности остается низким, а доступ к соответствующей информации является ограниченным, особенно в сельских местностях, а также среди женщин.

Информационная грамотность является весьма существенным элементом для деятельности *гражданского общества*, в которой могут принимать участие практически все население страны. В первую очередь здесь идет речь о доступе граждан к нужной им информации, об их демократическом участии в политической деятельности, о возможности доступа к органам власти государственного и местных уровней и т. д. В то же время развитие демократических институтов тесно сопряжено с проблемами обеспечения общественной и национальной безопасности. Эти аспекты

также должны быть учтены при формировании целей и критериев информационной грамотности путем соответствующего планирования, разъяснения целей и поддержания сотрудничества между странами.

Стандарты информационной грамотности в сфере образования были созданы как средство для организации соответствующих работ и демонстрации важности индикаторов информационной грамотности для этой сферы. Все опубликованные стандарты имеют единую базовую основу. В качестве родовой конструкции эти стандарты ориентированы на приложения экономического сектора и пожизненного обучения, которые, по мнению разработчиков стандартов, играют ключевую роль в процессах подготовки людей к участию в гражданском обществе и в развитии или поддержании возможностей для трудовой занятости. В то же время практика показала, что необходимо, чтобы эти стандарты были распространены на эксплуатационные переменные (*operational variables*) в различных контекстах. Ранние формулировки стандартов по информационной грамотности были разработаны применительно к обучению в школах. Более поздние стандарты ориентированы на специфику и требования учреждений высшего образования. Во всех этих стандартах с различными акцентами устанавливается необходимость определения информационных потребностей, а также необходимость использования средств локализации информации, ее оценки, хранения, восстановления и применения, а также передачи новых знаний.

Информационная грамотность *в трудовых процессах и в экономической деятельности* является центральным понятием для организации обучения и для выявления уровня конкурентоспособности отдельных фирм и стран пределах глобальной экономики, ориентированной на знание. Как показывает практика, зачастую принципы управления знаниями в фирмах и в правительствах не связываются с повышением конкурентоспособности фирм. Чтобы ясно понимать важность информационной грамотности для экономического роста, необходимо различать обычное распределение информации и использование информации для создания новых знаний. Традиционно распределение информации между людьми зависит от соответствующей инфраструктуры и в основном может распределяться без введения дополнительной стоимости. Однако когда информация используется, чтобы вводить инновации, создавать новые изделия или процессы, особенно если они защищены патентами, тогда новые знания могут быть дорогостоящими для его использования конкурентами.

Все элементы информационной грамотности описываются совокупностью процессов, которые можно объединить в следующие обобщенные группы:

- выявление информационных потребностей;
- локализация и оценка качества информации;
- накопление и восстановление информации;
- эффективное и этически корректное использование информации;
- использование информации для создания и распространения знаний.

Информационная грамотность и социально-экономическое развитие

В настоящее время исследователи рассматривают в основном два вида экономик: традиционную индустриальную и развивающуюся экономику знаний или экономику, ориентированную на знание. В таблице даны сравнительные характеристики этих видов экономик.

Приведенные в таблице результаты сопоставления свидетельствуют, что в основе экономики знаний лежит информация, поэтому информационная грамотность общества является важнейшим фактором ее становления и развития. Соот-

ветственно, ключевая роль отводится методам и средствам ее измерения, т. е. индикаторам информационной грамотности.

Информационная цепочка образования знаний включает следующие четыре стадии: создание, пакетирование, распределение и использование информации. Информационная грамотность охватывает навыки и квалификацию, необходимые при использовании этой информации, и для их оценки требуются соответствующие индикаторы. Информационная грамотность является составным компонентом создания знания в цикле его производства. Этот цикл включает авторов, изобретателей, исследователей и других категорий работников, которые производят новое знание в форме статей, книг, текстов, патентов или совместно их используют с соответствующим сообществом. Эти формы знания пакетированы издателями, создателями баз данных, взб-мастерами и компаниями электронных СМИ. Эта произведенная и обработанная информация в свою очередь накапливается и распространяется книжными магазинами, библиотеками и другими информационными поставщиками, чтобы удовлетворить потребности научных работников, преподавателей, студентов, компаний и иных слоев общества.

Все существующие стандарты основаны на концепции того, что достижение информационной грамотности являются непрерывным процессом (континуумом) как с позиции средств или методов, требуемых на различных уровнях обучения человека, так и в смысле того, что применяемые средства используются интегральными способами, а не как строго линейные последовательные процессы. Так, в ходе получения образования развитие информационной грамотности начинается на самой ранней стадии в школе и продолжается при получении высшего образования и последующего пожизненного обучения. В более развитых странах развитие информационной грамотности объединяется с получением навыков по использованию ИКТ, поскольку в настоящее время цифровые технологии и электронные базы данных являются первичным источником получения информации. Таким образом, требуется объединение познавательных навыков и навыков технических, необходимых в процессе доступа и использования информации. Непрерывность процесса достижения людьми информационной грамотности, определяемая существующими стандартами, может быть представлена в виде структуры, приведенной на рис. 1.

Информационная грамотность рассматривается как центральный элемент во многих моделях, описывающих компетентность взрослых людей. Например, как показано на рис. 2, умение писать, читать и считать являются базовыми навыками;

Характеристика	Индустриальная экономика	Экономика, ориентированная на знание
Ресурсы	Материалы трудно доступны и дорогостоящие	Информация в большинстве случаев широко доступна по приемлемой стоимости
Использование	Материалы исчерпываются в процессе использования и требуют затрат при их пополнении	Информация доступна для распространения и может нарастать при использовании
Долговечность	Материалы обычно не меняются и не портятся со временем	Информационные потребности постоянно изменяются, а ее качество быстро ухудшается
Распределение	Материалы и продукция требуют транспортировки и стоимость этого определяется местом нахождения продукции и доступом к ней	ИКТ-инфраструктура распространения существует, а информация (материал) и знания (продукция) могут распределяться в широких пределах
Собственники и законы	Продукты могут быть запатентованы, доступ контролирован, а аналоги получены	Знания могут создавать трудности для правового регулирования и налогообложения
Ценообразование	Цена обычно определяется исходными затратами (накладные расходы), трудоемкостью и транспортными расходами	Цена в основном определяется затратами на Интернет и может сильно зависеть от контекста
Примечание. ИКТ — информационно-коммуникационные технологии.		

ОБРАЗОВАНИЕ	УРОВНИ	НАВЫКИ
Ступень 0 – Детский сад		
Ступень 1 – Начальная школа		Определение и озвучивание информационных потребностей
Ступень 2 – Средняя школа	Базовое	Поиск и доступ к информации
Ступень 3 – Высшая школа	Среднее	Оценка качества информации
Ступень 4 – Вторичная высшая школа	Высшее	Организация информации
Ступень 5 – Послевузовское обучение	Продвинутое	Использование информации
Ступень 6 – Научные работники		Распространение и этичное использование информации

Рис. 1. Континуум информационной грамотности человека

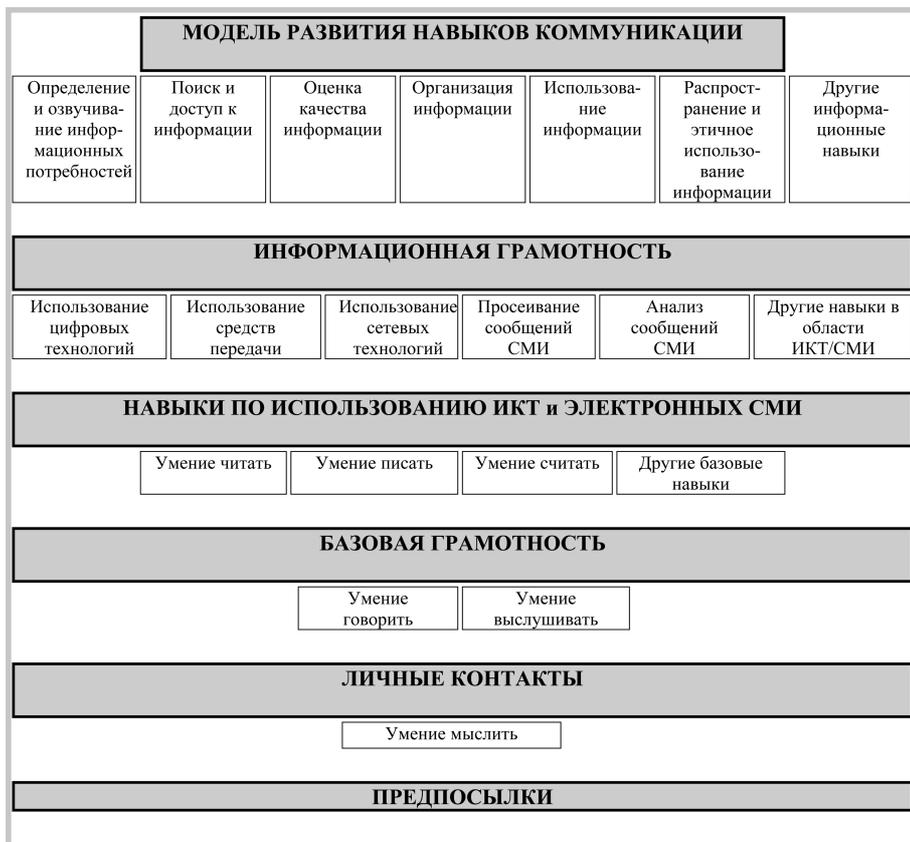


Рис. 2. Модель последовательного развития навыков (квалификации)

затем следуют навыки по использованию современных ИКТ-технологий, электронных СМИ, средств коммуникаций и сетевых технологий. Наконец, на самом высоком уровне располагается информационная грамотность, которая охватывает умение идентифицировать информационные потребности, определять местонахождение информации, восстанавливать, оценивать и использовать информацию, а также соблюдать права интеллектуальной собственности передачи информации и

знаний. Специалисты утверждают, что все эти навыки действительно в каждом обществе независимо от социально-экономического развития страны. Модель последовательного развития навыков и квалификации людей представлена на рис. 2.

На рис. 3 приведена иерархическая модель развития базовых навыков, в которой показано место высшего образования. С одной стороны, высшее образование опирается на общие базовые знания, полученные в процессе доинститутского образования, с другой — полученные навыки являются фундаментом для последующего профессионального образования в ходе пожизненного обучения.

В системе высшего образования в настоящее время известно несколько стандартизованных систем, определяющих индикаторы информационной грамотности. Ниже приведены три из них:

- в США консорциум научных библиотек разработал методику тестирования информационной грамотности, названную "Стандартизованная оценка навыков информационной грамотности" (Standardized Assessment of Information Literacy Skills — SAILS);
- несколько лет назад Служба тестирования уровня образования (Education Testing Service — ETS) опубликовала методику оценки компьютерной грамотности, названную iSkills Test, ранее известную ICT skills;

- в Австралии Совет университетских библиотек (Council of Australian University Librarians — CAUL) опубликовал свою методику определения информационных навыков (Information Skills Survey — ISS).

Три приведенные системы дают возможность провести сравнение того, каким образом проводится проверка информационной грамотности. Система **SAILS** обеспечивает проверку знаний каждого респондента об его информационной



Рис. 3. Иерархическая модель общего образования (generic skills)

грамотности, базируясь на группе стандартов, разработанных Ассоциацией американских библиотек (American Library Association — ACRL) в 2005 г. Эта система оценивает информационную грамотность людей только на общем уровне абстракции, а не на уровне фактического применения навыков работы с информацией.

Система CAUL ISS является автоматизированным средством анализа, которое обращается к отвечающим с запросом на описание того, что они делают с информацией. Система опирается на группу стандартов, разработанных в 2004 г. Австралийско-Новозеландским институтом информационной грамотности (Australian and New Zealand Institute for Information Literacy — ANZILL).

Система iSkills располагается на моделирующем компьютере и в процессе работы на этом компьютере осуществляет проверку информационной грамотности работающего.

В сфере высшего образования с широким доступом к компьютерной среде целесообразно использовать именно систему моделирования работы студентов, которая обеспечивает непосредственное измерение уровня информационной грамотности. При отсутствии доступа к компьютерной среде более целесообразным представляется использование автоматизированных средств анализа, которые позволяют проводить сравнительные оценки уровня информационной грамотности.

В сфере системы образования большую озабоченность вызывает взаимоотношение информационной грамотности и этики, и в первую очередь это связано с проблемой плагиата. С распространением сети Интернет эта проблема усугубляется. В то же время ряд исследований показывает, что внедрение принципов этического использования информации уменьшает степень плагиата. Поэтому большинство существующих стандартов в области информационной грамотности включают правила и принципы этического доступа и использования информации. Практика этического использования информации охватывает как процессы создания и распространения информации, так и ее применения. В информационно грамотном обществе люди в состоянии оценить источники поступления информации и при даль-

нейшем распространении этой информации они ссылаются на соответствующие источники. Три упомянутые выше стандартизированные системы оценки информационной грамотности в сфере высшего образования используют различные подходы к оценке выполнения этических правил и принципов использования информации. Так, в системе SAILS людям задаются вопросы, направленные на оценку их осведомленности относительно этических правил и принципов использования информации, но полученные ответы не обязательно отражают фактическое состояние. В системе CAUL ISS людей просят самостоятельно предоставить сведения относительно их навыков и осведомленности в области авторского права и плагиата, причем это делается на условиях анонимности. Наконец, в системе iSkills проверка этических аспектов информационной грамотности не проводится.

Направления развития индикаторов информационной грамотности

Необходимость создания международных индикаторов, связывающих доставку, доступ и прием информации с уровнем информационной грамотности, была сформулирована на уровне двух глобальных встреч на высшем уровне, в которых ЮНЕСКО играл ведущую роль. Первая — это

конференция " Education for All" в Дакаре в 2003 г., посвященная проблемам образования всех видов, начиная от начального и до пожизненного. Вторая — Всемирный саммит по информационному обществу 2003/2005 г. (World Summit on the Information Society 2003/2005). На нем были обозначены аспекты информационной грамотности, которые затем были включены в План действий, сформированный Партнерством для измерения ИКТ для развития (Partnership for the Measurement of ICTs for Development). В состав этого Партнерства вошли такие международные организации, как ОЭСР, ЮНКТАД, Международный союз электросвязи, Институт статистики ЮНЕСКО, Региональные комиссии ООН и ряд других. Партнерство разработало "ядро индикаторов" из 48 показателей, оценивающих использование ИКТ в бизнесе и в домашних приложениях.

Сегодня рассматриваются несколько основных направлений создания системы индикаторов информационной грамотности. Первое — это разработка системы косвенных индикаторов информационной грамотности на базе существующих традиционных показателей. Сюда могут включаться такие показатели, как число опубликованных книг, число зарегистрированных патентов, распространенность сети Интернет и т. п. Однако многие эксперты считают, что подобные показатели определяются экономическими условиями и в большей степени являются практическим результатом применения информационной грамотности, а не собственно индикатором, характеризующим уровень информационной грамотности.

Другим направлением является разработка новой международной системы оценки навыков информационной грамотности. Эта система должна быть полной и исчерпывающей, но ее создание потребует значительных затрат, особенно на этапах разработки и последующего внедрения.

Третье, и самое предпочтительное, направление — это создание системы индикаторов информационной грамотности, опирающейся на существующие международные показатели, в том числе и

с использованием характеристик, содержащихся в рассмотренных выше стандартизованных системах индикаторов информационной грамотности.

Для развития системы индикаторов информационной грамотности важное значение имеет разработанная Институтом статистики ЮНЕСКО Программа работ стран ОЭСВ по оценкам и мониторингу грамотности (Literacy Assessment and Monitoring Programme — LAMP). Подробное описание этой и ряда других инициатив по созданию международной системы индикаторов информационной грамотности, а также существующие системы показателей для проведения их вторичного анализа приводится в [1].

* * *

В заключение следует отметить, что вне зависимости от того, по какому направлению пойдет развитие системы индикаторов, любая система должна удовлетворять определенной совокупности базовых требований и условий. В качестве основных в их число входят следующие: пригодность; современность; точность; частота сбора данных; затраты; обоснованность; надежность; взаимная непротиворечивость; прозрачность; сопоставимость, Перечень этих требований может быть легко продолжен.

Таким образом, разработка системы индикаторов информационной грамотности является необходимым условием развития глобальной информационной грамотности. При этом, как считают авторы [1], предложенный в их работе механизм создания системы индикаторов на базе вторичного анализа существующих показателей является рентабельным с точки зрения сокращения затрат и эффективным с позиции того, что такая система может быть широко доступна в ближайшем будущем.

Список литературы

1. **Towards** Information Literacy Indicators. UNESCO / Ralph Catts and Jesus Lau. Paris: UNESCO. 2008.

Д. А. Пономарев, аспирант,
Нижегородский государственный
лингвистический университет,
e-mail: dima-tensor@yandex.ru

Двухэтапный алгоритм обучения речи слабослышащих на основе информационной теории восприятия речи

Рассматривается новая методика обучения речи глухонемых и слабослышащих детей и подростков на основе информационной теории восприятия речи и критерия минимума информационного рассогласования. Показано, что главное отличие данного алгоритма обучения речи по сравнению с известными аналогами состоит в визуализации степени информационного рассогласования произношения обучаемого не с одним, а с группой эталонных реализаций. Приведен пример практического применения алгоритма в задаче обучения речи.

Ключевые слова: речь, методика обучения речи, метод "улитки", информационная теория восприятия речи, теоретико-информационный подход.

Введение

Одной из актуальных проблем в области специальной педагогики является проблема обучения речи и постановки произношения у детей и подростков с нарушениями слуховых функций (глухонемых и слабослышащих). Суть большинства разработанных на данный момент методик обучения [1, 2] сводится к поэтапному улучшению произношения звуков по отношению к некоей эталонной реализации. Между тем такой подход скрывает в себе очевидное противоречие между постановкой задачи и возможностями ее реализации, существенно снижающее эффективность данных методик: обучаемый не слышит эталонной реализации и поэтому не может максимально близко приблизиться к ней в произношении. Принципиально другим подходом в процессе обучения речи глухонемых является применение методик обучения с визуализацией степени рассогласования произношения обучаемого по отношению к определенному речевому сигналу-эталону [3]. Существующие методики основаны, главным образом, на применении временных диаграмм и частотных спектров речевых сигналов. Обучаемому предстоит последовательно во времени улучшать качество своего произношения звуков за счет приближения формы временных и частотных диаграмм к соответствующим диаграммам эталонных реали-

заций. Тем не менее, указанная методика имеет существенный недостаток: добиться близости спектров/временных диаграмм в процессе обучения на всех участках не представляется возможным.

Рассматриваемый в данной статье подход предложен В. В. Савченко [4] и в противовес известным подходам использует один, интегральный, критерий оптимальности — минимум информационного рассогласования (МИР) [5]. Его обоснованием служит информационная теория восприятия речи (ИТВР), а наиболее перспективным вариантом реализации является метод "улитки", которому и посвящена настоящая статья. Существенным отличием от существующих методик является использование в процессе обучения не одного эталона (произнесенного диктором, имеющим правильное произношение), а целой группы эталонных реализаций, состоящей из лучших образцов произнесений фонем ранее обученных слабослышащих.

Основная идея

Центральная идея ИТВР — это определение минимальной речевой единицы: фонемы как элемента звукового строя национального языка через множество $X_r = \{x_{r,j}\}, j = \overline{1, J_r}$, одноименных звуков-реализаций, объединенных друг с другом в кластер по критерию МИР в метрике Кульбака—Лейблера. Здесь $J_r < \infty$ — объем r -го множества. Каждая фонема-кластер имеет четко очерченные границы вокруг информационного центра — эталона (ИЦ-эталона) x_r^* , математически заданного вектором $a_r = \{a_{r,i}\}$ авторегрессионных (АР) коэффициентов $a_{r,i}$. При этом каждый ИЦ-эталон характеризуется минимальной суммой информационных рассогласований по Кульбаку—Лейблеру относительно всех других его звуков-реализаций $\{x_{r,j}\}, j = \overline{1, J_r}$ [4]:

$$x_r^* = \text{Arg}_k \min_{j=1}^{J_r} \rho_{x_j^{(r)}}(x_k), k = \overline{1, J_r}, \quad (1)$$

где $\rho_{x_j^{(r)}}(x_k)$ — величина информационного рассогласования (ВИР) между звуками-реализациями $x_{r,j}$ и $x_{r,k}$ r -го элемента звукового строя языка.

Данный ИЦ и выполняет роль эталонной реализации (произнесения) соответствующего r -го элемента звукового строя языка. А набор ИЦ $\{a_r\}, r = \overline{1, R}$, всех R применяемых в языке фонем определяет искомым эталон всего звукового строя.

Одной из основных проблем при обучении речи и постановке произношения у глухонемых и слабослышащих является проблема выбора эталонов: использование в качестве эталонных реализаций фонем, произнесенных дикторами с высо-

ким качеством произношения, является малоэффективным ввиду невозможности повторения их произношения людьми с ограниченными слуховыми возможностями. Преодоление указанной проблемы возможно посредством одновременного использования нескольких эталонных реализаций — лучших произнесений ранее обученных слабослышащих. Таким образом, в данной методике в роли "учителя" предлагается использовать не одного, а целую группу дикторов, причем с заведомо неидеальным произношением. Каждый отдельный диктор-учитель порождает свой собственный фонетический ряд $\{a_r\}$, $r = \overline{1, R}$, в роли эталона звукового строя национального языка. А множество всех дикторов-учителей порождает целое множество, или группу $X_r = \{x_{r,j}\}$, $j = \overline{1, J_r}$, эталонных реализаций для каждой отдельной фонемы $r = \overline{1, R}$ из звукового строя. Таким образом, процесс обучения в данной методике представляет собой последовательное приближение каждого произнесения к любому из используемого множества ИЦ-эталонов каждой отдельной фонемы.

Критерий МИР

В терминах теоретико-информационного подхода задача в общем случае формулируется как J_r -альтернативная проверка гипотез $P_{i,j}$, $j = \overline{1, J_r}$ о законе распределения вектора речевого сигнала. Следуя критерию МИР, приходим к решающему правилу вида:

$$x \in x_r, v: \rho_{x_v}^{(r)}(x) = \min_j \rho_{x_j}^{(r)}(x), j = \overline{1, J_r}, \quad (2)$$

где $\rho_{x_j}^{(r)}(x) = \iint \frac{d\hat{P}(x)}{d\hat{P}_j(x)} d\hat{P}(dx)$ — величина информационного рассогласования [6] выборочного закона распределения $\hat{P}(x)$ относительно его j -й альтернативы $P_j(x)$, $j = \overline{1, J_r}$. Решение существенно упрощается, если воспользоваться гауссовой аппроксимацией закона распределения речевого сигнала $P_j = N(\mathbf{K}_j)$, где \mathbf{K}_j — его автокорреляционная матрица (АКМ) размера $n \times n$, $n \geq 1$. Как следует из работы [7], свойство оптимальности критерия МИР в этом случае сохраняется. А соответствующий набор оптимальных решающих статистик из выражения (2) переписывается следующим образом:

Матрица ВИР для r -го элемента звукового строя языка

	$x_{r,1}$	$x_{r,2}$...	x_r^*	...	x_{r,J_r-1}	x_{r,J_r}
x_r^*	$\rho_{x_1}^{(r)}(x_1)$	$\rho_{x_2}^{(r)}(x_2)$		$\rho_{x^*}^{(r)}(x^*) = 1$		$\rho_{x_{J_r-1}}^{(r)}(x_{J_r-1})$	$\rho_{x_{J_r}}^{(r)}(x_{J_r})$

$$\rho_{x_j}^{(r)}(x) = \frac{1}{2} [\text{tr}(\hat{\mathbf{K}} \mathbf{K}_j^{-1}) - \log|\hat{\mathbf{K}} \mathbf{K}_j^{-1}| - n], j = \overline{1, J_r}. \quad (3)$$

Здесь $\hat{\mathbf{K}}$ — это выборочная оценка АКМ анализируемого сигнала x ; символами $\text{tr}(\cdot)$ и $|\cdot|$ обозначены соответственно след и определитель квадратных $(n \times n)$ -матриц. Задача сводится теперь к определению множества альтернативных АКМ $\{\mathbf{K}_j\}$ на заданном J_r — множестве сигналов-эталонов $X_r = \{x_{r,j}\}$. Это стандартная задача статистического анализа.

В асимптотическом случае $n \rightarrow \infty$, когда в качестве объектов статистического анализа рассматриваются стационарные гауссовские процессы, и при дополнительном и актуальном для задач автоматического распознавания речи (АРР) условии нормировки АР-модели сигналов по дисперсии их порождающего шума, выражения для решающих статистик (3) могут быть записаны в частотной области в эквивалентном виде [7]:

$$\rho_{x_j}^{(r)}(x) = \frac{1}{2F} \sum_{f=-F}^F \left| \frac{1 + \sum_{m=1}^P a_{r,j}(m) \exp(-i\pi mf/F)}{1 + \sum_{m=1}^P a_{r,x}(m) \exp(-i\pi mf/F)} \right|^2 - 1, \quad (4)$$

где $a_{r,j}(m)$ и $a_{r,x}(m)$ — АР-коэффициенты сигналов x_j и x соответственно; F — верхняя граница частотного диапазона сигнала или используемого для его передачи канала связи [Гц]; f — частота [Гц]; m — счетчик; P — порядок АР-модели; i — мнимая единица.

Синтез алгоритма

Рассмотрим в качестве примера множество эталонных реализаций r -го элемента звукового строя языка, состоящее из $J_r < \infty$ сигналов-эталонов, заданных векторами $a_i = \{a_{i,j}\}$, $i = \overline{1, J_r}$, $j = \overline{1, P}$, АР-коэффициентов порядка P . Данное множество эталонов характеризуется информационным центром-эталонном $a_r = \{a_{r,j}\}$, $j = \overline{1, P}$, определяемым по формуле (1). В качестве метрики "близости" ИЦ и остальных сигналов-эталонов множества выберем величину их информационного рассогласования (4). Тогда для определения

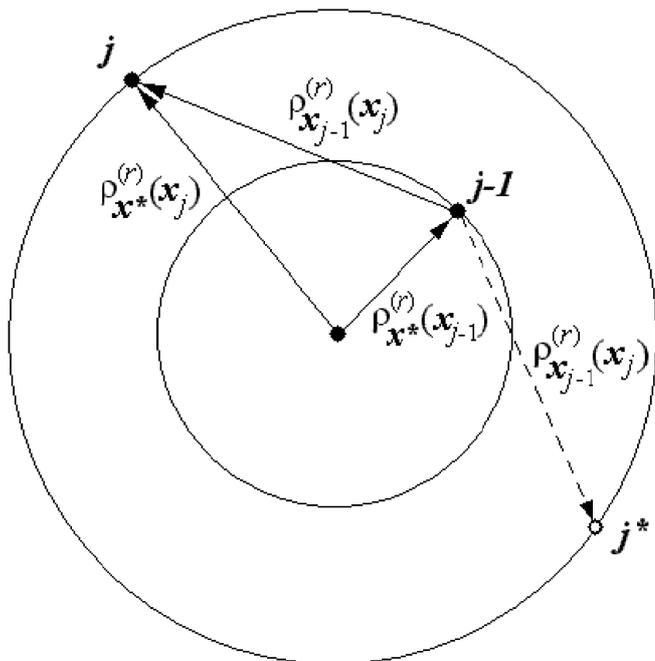


Рис. 1. Взаимное расположение эталонов на координатной плоскости

взаимного расположения сигналов-эталонов и ИЦ-эталона внутри множества необходимо рассчитать матрицу ВИР $\rho = \{\rho_{x^*}^{(r)}(x_j)\}$, $j = \overline{1, J_r}$, общий вид которой приведен выше в таблице.

Здесь $\rho_{x^*}^{(r)}(x_j)$ — ВИР между ИЦ-эталонном и j -м элементом множества эталонных реализаций. Чем меньше значение $\rho_{x^*}^{(r)}(x_j)$, тем "ближе" расположен j -й сигнал-эталон к ИЦ-эталону в теоретико-информационном смысле. При визуализации множества сигналов-эталонов будем придерживаться двух следующих утверждений.

Утверждение 1: j -я эталонная реализация сигнала расположена на концентрической окружности с радиусом $\rho_{x^*}^{(r)}(x_j)$, в центре которой размещен ИЦ-эталон.

Утверждение 2: j -я эталонная реализация сигнала расположена на расстоянии $\rho_{x_{j-1}}^{(r)}(x_j)$ от $(j-1)$ -й эталонной реализации, где $\rho_{x_{j-1}}^{(r)}(x_j)$ — ВИР между $(j-1)$ -м и j -м сигналами-эталонами множества (4).

Следует отметить, что утверждению 2 всегда удовлетворяют два решения — j и j^* (рис. 1). Выбор в пользу одного из них осуществляется по принципу раскручивания против часовой стрелки ломаной линии, соединяющей сигналы-эталоны (рис. 2, см. третью сторону обложки). По своей форме такая линия напоминает контуры улитки. В соответствии с данным принципом автор метода В. В. Савченко назвал предложенный им метод методом "улитки".

Графически реализация данных утверждений показана на рис. 1.

Из рисунка видно, что ИЦ-эталон, $(j-1)$ -й и j -й сигналы-эталон образуют на плоскости треугольник со сторонами $\rho_{x_{j-1}}^{(r)}(x_j)$, $\rho_{x^*}^{(r)}(x_j)$ и $\rho_{x_{j-1}}^{(r)}(x_j)$.

Поэтому для визуализации сигналов на плоскости необходимо выполнение условия существования треугольника [8], которое в нашем случае может быть записано в виде (5):

$$\begin{aligned} \rho_{x^*}^{(r)}(x_j) - \rho_{x_{j-1}}^{(r)}(x_{j-1}) &\leq \rho_{x_{j-1}}^{(r)}(x_j) \leq \\ &\leq \rho_{x^*}^{(r)}(x_j) + \rho_{x_{j-1}}^{(r)}(x_{j-1}). \end{aligned} \quad (5)$$

Сигналы-эталон, для которых система неравенств (5) не имеет решения, не могут быть отображены на координатной плоскости в соответствии с утверждениями 1 и 2.

Таким образом, визуализация множества сигналов-эталон состоит в следующей последовательности операций:

1) определяем ИЦ для множества эталонных реализаций (1);

2) рассчитываем матрицу ВИР $\rho = \{\rho_{x^*}^{(r)}(x_j)\}$, $j = \overline{1, J_r}$ между ИЦ-эталонном и другими эталонами множества по формуле (4);

3) сортируем ρ по возрастанию ВИР;

4) выбираем из ρ заданное число M сигналов-эталон, удовлетворяющих системе неравенств (5);

5) строим на координатной плоскости концентрические окружности вокруг ИЦ-эталон для M эталон в соответствии с утверждением 1;

6) второй сигнал-эталон отображаем на первой окружности — по горизонтали от ИЦ-эталон;

7) третий и последующие сигналы-эталон отображаем на соответствующих окружностях в соответствии с утверждением 2.

Внешний вид $M = 20$ сигналов-эталон из множества показан на рис. 2 (см. третью сторону обложки).

Отличительной особенностью предлагаемой методики обучения речи и постановки произношения является возможность сравнения произношения не с одним, а с несколькими сигналами-эталонами. Таким образом, задача, стоящая перед диктором, заключается в последовательном приближении своего произношения к одному, наиболее удобному для него эталону из группы. В данной ситуации имеет смысл говорить не об абсолютной, а об относительной близости произношения к какому-то из эталон.

Первым этапом процесса визуализации реализации-произнесения обучаемого на координатной плоскости является определение ближайшего к нему сигнала-эталон из группы. Для этого рассчитываем матрицу ВИР $\rho = \{\rho_{x_j}^{(r)}(x)\}$, $j = \overline{1, M}$.

Пример практического применения

В соответствии с критерием МИР (2) сигнал-эталон x_v , для которого значение ВИР минимально является искомым. Тогда для визуализации сигнала от обучаемого будем придерживаться двух следующих утверждений.

Утверждение 3: сигнал от обучаемого x находится на окружности, в центре которой расположен ИЦ-эталон, и с радиусом $\frac{\rho_x^{(r)}(x_v) + \rho_x^{(r)}(x_{v+1})}{2}$, где $\rho_x^{(r)}(x_v)$ — ВИР между

ИЦ-эталонном и ближайшим к x сигналом-эталонном x_v .

Утверждение 4: сигнал от обучаемого x располагается на расстоянии $\rho_{x_v}^{(r)}(x)$ от ближайшего к нему сигнала-эталона x_v .

Следует отметить, что реализации утверждения 4 удовлетворяет пара решений, выбор в пользу одного из них необходимо осуществить по принципу, аналогичному для утверждения 2.

Таким образом, визуализация процесса обучения речи и постановки произношения состоит в следующей последовательности операций:

1) рассчитать матрицу $\rho = \{\rho_{x_j}^{(r)}(x)\}$, $j = \overline{1, M}$, ВИР между сигналами-эталонами и сигналом от обучаемого по формуле (4);

2) определить ближайший к x сигнал-эталон x_v по критерию МИР (2)—(4);

3) определить координаты сигнала x в соответствии с утверждениями 3 и 4;

4) повторить пп. 1—3 для следующего сигнала от обучаемого;

5) соединить реализации-произнесения обучаемого ломаной линией, представляющей собой траекторию обучения.

Таким образом, предлагаемый алгоритм включает в себя два этапа обработки информации: этап подготовки данных и этап обучения. Первый этап состоит в последовательном определении взаимного расположения сигналов-эталонов внутри множества эталонных реализаций элемента звукового строя языка. Результатом выполнения данного этапа является структурированная область допустимых решений данного элемента звукового строя, позволяющая визуально определять степень рассогласования произношения отдельных сигналов-эталонов по отношению друг к другу. Второй этап сводится к последовательному обучению речи и постановке произношения за счет визуализации степени рассогласования произношения обучаемого по отношению к выбранному речевому сигналу-эталону.

Проиллюстрируем работу алгоритма обучения речи на следующем примере.

Для иллюстрации работы алгоритма обучения речи и постановки произношения диктором с высоким качеством произношения были проговорены в режиме информативного звучания 50 эталонных реализаций звука "а". На первом этапе работы алгоритма в автоматическом режиме из них были отобраны $M = 20$ реализаций с последующей визуализацией, приведенной на рис. 2. Далее в режиме обучения речи в качестве входных сигналов использовались произнесенные тем же диктором следующие комбинации звуков "м" и "а": "mmaaa", "maaa", "mmaaamm", "aaamm", "aaam", "aaa", "а". Указанные сигналы последовательно во времени обрабатывались по алгоритму обучения 1—5 и отображались на плоскости под порядковыми номерами от 1 до 7 (рис. 3, см. третью сторону обложки). При этом для расчета коэффициентов авторегрессии применялась высокоскоростная рекуррентная процедура Берга—Левинсона [9]:

$$\left. \begin{aligned} a_m(i) &= a_{m-1}(i) + c_m a_{m-1}(P-i), \quad i = \overline{1, P}; \\ c_m &= 2(N-P) \frac{\sum_{t=m}^{N-1} \eta_{m-1}(t) v_{m-1}(t-1)}{\sum_{t=m}^{N-1} [\eta_{m-1}^2(t) + v_{m-1}^2(t-1)]}; \\ \eta_m(t) &= \eta_{m-1}(t) - c_m v_{m-1}(t-1); \\ v_m(t) &= v_{m-1}(t-1) - c_m \eta_{m-1}(t), \\ t &= \overline{0, N-1}; \\ \sigma_m^2 &= (1 - c_m^2) \sigma_{m-1}^2, \quad \sigma_0^2 = \frac{1}{N} \sum_{t=1}^N |x(t)|^2, \\ m &= \overline{1, P}. \end{aligned} \right\} \quad (6)$$

при ее инициализации системой равенств $v_0(t) = \eta_0(t-1) = x(t)$. Здесь a_m — авторегрессионные коэффициенты порядка m ; c_m — коэффициент отражения; $\eta_m(t)$ — ошибка линейного предсказания вперед; $v_m(t)$ — ошибка линейного предсказания назад; σ_m^2 — дисперсия возбуждающего белого шума; P — порядок АР-модели; N — объем выборки.

Из рисунка видно, что данному диктору наиболее удобным для произношения, по-видимому, является сигнал-эталон № 13, и чем четче проговаривается звук "а" в процессе обучения, тем ближе располагаются реализации к эталону (реализации 5 и 7). При этом становится ясным механизм обучения: чем четче произносится звук, тем ближе он располагается к определенному сигналу-эталону, а чем ближе данный сигнал-эталон расположен к ИЦ-эталону, тем лучше произношение и качество речи. Следует отметить, что первые три

реализации соответствуют неверному произношению звука "а" и находятся вне области допустимых решений. При этом, чем дальше расположена реализация от сигнала-эталона № 19, тем хуже качество ее произнесения.

Заключение

Таким образом, рассмотренный выше метод "улитки" позволяет проводить обучение речи и постановки произношения у детей и подростков с нарушением слуховых функций в режиме самостоятельного обучения. Следует отметить, что возможности данного метода не ограничиваются обучением произношения только отдельных фонем языка. Напротив, указанный процесс является лишь первым этапом обучения. Дальнейшей ступенью обучения речи является обучение произношения аллофонов — позиционных произносительных вариантов фонем, звучание которых существенно зависит от звукового окружения. На данном этапе сигналы от обучаемого представляют собой последовательности фонем, выделение аллофонов из которых осуществляется путем автоматического сегментирования сигнала [10]. Обучение произношения аллофонов является важнейшим этапом процесса обучения речи. Полученные результаты могут найти также практическое применение и в исследовательских целях,

например, при сравнительном анализе качества произношения разных дикторов, анализе зависимости их произношений от внешних факторов, в частности от эмоционального состояния, и т. п.

Работа выполнена при поддержке гранта РФФИ 07-07-12042-офи.

Список литературы

1. Рау Ф. Ф., Слезина Н. Ф. Методика обучения произношению в школе глухих. М.: Просвещение, 1981.
2. Комаров К. В. Методика обучения русскому языку в школе для слабослышащих детей. М.: Оникс 21 век, 2005.
3. Кукушкина О. И., Королевская Т. К., Зеленская Ю. Б. Информационные технологии в обучении произношению. М.: Полиграф-Сервис, 2004.
4. Савченко В. В. Информационная теория восприятия речи // Известия вузов. Радиоэлектроника. 2007. Вып. 6.
5. Савченко В. В. Автоматическая обработка речи по критерию минимума информационного рассогласования на основе метода обеляющего фильтра // Радиотехника и электроника. 2005. Т. 50. № 3. С. 309—314.
6. Кульбак С. Теория информации и статистика. М.: Наука, 1967. 408 с.
7. Савченко В. В. Различение случайных сигналов в частотной области // Радиотехника и электроника. 1997. Т. 42. № 4. С. 426—431.
8. Выгодский М. Я. Справочник по элементарной математике. М.: Наука, 1967. 416 с.
9. Марпл С. Л. Цифровой спектральный анализ и его приложения. М.: Мир, 1990.
10. Савченко В. В., Пономарев Д. А. Автоматическое сегментирование речевых сигналов методом обеляющего фильтра // Известия вузов. Радиоэлектроника. 2008. Вып. 6.

Информация

Фирма "IC", основанная в 1991 г., специализирующаяся на разработке и поддержке компьютерных программ и баз данных, провела в июне 2009 г. телеконференцию, транслировавшуюся на территорию России, Украины и Беларуси, по отраслевому решению IC:PDM — "УПРАВЛЕНИЕ ИНЖЕНЕРНЫМИ ДАННЫМИ: информационные технологии для эффективного управления в новых экономических условиях".

В ходе телеконференции участники познакомились с возможностями системы IC:PDM в решении задач сокращения затрат на проектирование изделий, эффективного распределения производственных ресурсов, формирования свода знаний технологической подготовки производства и др.

В докладе "Применение функциональных возможностей PDM-систем в текущих экономических условиях. Решение основных задач антикризисного управления компанией с помощью производственной PDM-системы" генеральный директор компании APPIUS (www.appius.ru) А. Тимошин раскрыл важность направленности IC:PDM на решение производственных задач, наличие единой базы данных с ERB системой. Руководитель отдела разработки И. Берендеев познакомил слушателей с основными функциями системы по управлению электронной структурой изделия, нормативно-справочной информацией, электронным и бумажным архивом документации; технологической подготовкой производства, нормированием, интеграцией с CAD, web-интерфейсом. О методике внедрения IC:PDM рассказал руководитель отдела внедрения О. Бессмертный.

Содержание докладов, вопросы участников и ответы экспертов, методистов "IC", дискуссия по затронутым вопросам свидетельствуют о возросшем интересе к системе IC:PDM особенно в текущих экономических условиях, когда предприятиям нужна не просто автоматизация отдельных подразделений, а комплексная автоматизация, от конструкторско-технологических подразделений до контура планирования и расчета себестоимости, с единой базой данных, для эффективного управления предприятием как единым комплексом.

CONTENTS

Vasenin V. A. Critical Energetic Infrastructure: Cyberterrorist Threat. 2

The present publication is dedicated to problems of protecting critical energetic infrastructure from destructive computer impacts with cyberterroristic aims. It primarily focuses on identifying the infrastructure and objects composing it as critically important to the state, systematizing threats and methods of their implementation. This field is being studied using computer science techniques, approaches to information security of large systems and the publication presents results of such studies.

Keywords: energetics infrastructure, fuel-energy complex, hydro-energetics, nuclear energetics, technological process, automated control system, critically important object, destructive impact, information security, threat classification.

Ratmanova I. D., Korovkin S. D., Zheleznyak N. V. Information Model of Fuel and Energy Complex as a Basis of Energy Security Analysis. 9

The article describes approach to information model construction of fuel and energy complex for regional energy policy infotainment. The model is based on metalevel data warehouse architecture considering particular qualities of investigated data domain.

Keywords: energy security, fuel and energy complex, decision support systems (DSS), data warehouse, metadata, logical data model, multidimensional analytical data model, fuel and energy balance.

Devyatov D. Kh., Fainshtein S. I., Belyavsky A. B., Torchinsky V. E. A Heuristic Optimization Model for NP-Complete Constrained Problems Arising in Operational Planning of Plate Rolling. . . . 16

Given model is aimed at recasting a class of NP-complete constrained problem arising in particular in operational planning of Metallurgical works plate rolling to an optimization problem. This is accomplished by converting the constraints to a penalty function and giving bonuses for following external optimization criterions. Values of the penalties, the bonuses and some other additional parameters are setting constants aimed at tuning heuristic threshold SF-algorithm in to actual data and real-time changing optimization criterions.

Keywords: optimization methods, constrained problems, heuristic algorithms, operational planning, plate rolling.

Kukhareenko B. G. Independent Component Analysis of Oscillation Records in Spectral Analysis Technology Based on the Fast Prony Transform 20

Mutual preprocessing multi-dimensional time-series by the Independent component analysis is presented. It increases exactness and enlarge frequency resolution of the spectral Prony method. As example, non-stationary records of a rotor blade collective oscillation due to a flutter in turbo compressor are under study. It has been found that preprocessing the multi-dimensional oscillation records by the Independent component analysis gives an opportunity to study a synchronization process of rotor blade unstable oscillations by the Prony method.

Keywords: multi dimensional time-series, Independent component analysis, spectral analysis, Prony method, compressor blade flutter, oscillation synchronization.

Nechaev V. V. Conceptual Model Representation of a Problem as a System 26

The study discusses creation and formalized representation of a Problem Conceptual Model (PCM) as an integral system.

A PCM version focusing on system applications is proposed based on the results of a multipronged system analysis. The necessity for such a model is caused by the pressing need for automation of system problem setting, description, and solution processes in various subject areas.

Keywords: problem, metaproblem, model, metamodel, system, data, requirement, problem solver, method, adequacy, relevance, intelligence, algorithm, information, technology, system problem, conceptual model, information technology, intellectual problem solver.

Bryndin E. G. *Theoretical Aspects of Technology of Continual Processing on Virtual Storage* 33

This paper is devoted the problem of large program and data continuous processing by a computer with forestalling parallel replacement on virtual storage.

Keywords: program with determinate connection of modules, continual processing on virtual storage, forestalling rotation of common data, forestalling analyses of connections of program modules, forestalling replacement processor of program modules on virtual storage.

Grebnev S. V. *Design and Implementation of Two-Level Voice Control Approach Using Hidden Markov Models* 40

The article describes voice control system based on hidden markov models (HMM) and two-level command recognition approach. System includes author's researches as well as the art of speech recognition technology. Speed and accuracy of proposed system are good enough to use it in real time practice.

Keywords: hidden markov models, speech recognition, keyword spotting, sliding window models.

Malyhina G. F., Merkusheva A. V. *Signals Form Reconstruction Based on Measurement Information on its Mixture. Conceptions, Algorithm Modifications and Applications* 46

Possible accesses and methods are given that enable to get the form of separate signals on the base of measurement information on their mixture. As the base of algorithms and neural network structures realizing the reconstruction of signal mixture-components under registered their joint mixture, there are used different conceptions concerning the type of signal mixture forming, ability to have the noise characteristics for registered sensor signals, and also the possibility of noise cancellation for separated signals.

Keywords: source signals, reconstruction, independent components, analysis, neural network, adaptive filtration, structures, signal separation, noise cancellation, problem applications.

Asadov H. H., Chobanzadeh I. G. *Information Method for Radiometric Correction and Calibration of on-Board Spectrometers During Flights* 54

In the article the information method for carrying out of radiometric correction of on-board spectrometers during flights is suggested. It is shown, that used methods of radiometric correction may violate the authenticity of hyperspectral data. According to suggested the plot of maximal information capability should be found on the basis of given limitation conditions. Then the information of areas not included to the plot may be substituted with information derived from correlation links between elements of lines and lines itself.

Keywords: radiometric correction, calibration, information, spectrometer, atmosphere.

Bozhday A. S. *A Complex Infrastructure of a Territory: Methods and Models of an Informational Monitoring* 57

The paper is devoted to the perspective way of an informational monitoring of socio-economic systems. It's introducing the new conception — a complex infrastructure on a territory (CIT). The author considers the decomposition method for a CIT formalized representation (in relation to low- and middle-complexity social infrastructures). In this connection, the paper takes up the opportunity for a fusion of advanced informational technologies (such as OLAP, GIS, CAD, telecommunications) in the context of an unified datalogical model.

Keywords: complex infrastructure of territory (CIT), informational monitoring, decomposition, socio-economic systems, CIT mathematical model, geoinformation systems, decision support.

Filippova A. S., Brusilovsky D. P. *Forecasting of Intermittent Demand for Optimal Inventory Management of the Manufacturing Assembly Companies*. 63

In this paper, the problem of optimal inventory management for the small or medium-size manufacturing assembly companies is described. The system for optimal inventory management of the manufacturing assembly companies and the mathematical model, which the inventory management decision-making engine is based on, are presented.

This paper is dedicated to the situation when the demand is measured on interval scale, is intermittent, and when the history of demand is small (just 4—8 observations). In this situation classical Croston's approach does not work. A new approach, based on robust regression, is formulated and illustrated by an example.

Keywords: inventory management, Intermittent demand, Crostons model, robust regression, demand planning decision making engine.

Shtrik A. A. *Information Literacy Indicators* 67

In this article are considered the questions connected with a conceptual framework of information literacy and a list of potential international indicators for information supply, access and supporting skills by UNESCO Institute for Statistics. Results of research of such questions, as definition of information literacy, the elements and models of information literacy, economics and information literacy, information literacy standards, information literacy indicators in elementary and higher education is submitted.

Keywords: basic levels literacy, indicators, information literacy, information technology, economic development, development indicators, information literacy standards.

Ponomaryov D. A. *Two-Phase Algorithm Speech Ttraining for the Deaf Based on the Information Theory of Speech Perception* 73

Based on the information theory of speech perception and minimum informational error criterion the new speech training technique for deaf and deaf-mute children and teenagers is considered. It is shown that its main difference consists in visualization of information error degree of trainee's pronunciation not with an etalon but with a group of etalon realizations. The example of practical adaptation of algorithm in speech training is given.

Keywords: speech, speech training technique, information theory of speech perception, "snail" method, the information approach.

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. Пчелкина*

Сдано в набор 07.07.2009. Подписано в печать 17.08.2009. Формат 60×88 1/8. Бумага офсетная. Печать офсетная.
Усл. печ. л. 9,8. Уч.-изд. л. 11,28. Заказ 745. 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