



CONTROL SCIENCES

**Научно-технический
журнал**

6 номеров в год

ISSN 1819-3161

УЧРЕДИТЕЛЬ

**Институт проблем управления
им. В.А. Трапезникова РАН**

Главный редактор

Д.А. Новиков

**Заместители главного
редактора**

Л.П. Боровских, Ф.Ф. Пашенко

Редактор

Т.А. Гладкова

Выпускающий редактор

Л.В. Петракова

Издатель

ООО «СенСиДат-Контрол»

Адрес редакции
117997, ГСП-7, Москва,
ул. Профсоюзная, д. 65, к. 104.
Тел./факс (495) 330-42-66,
тел. (495) 334-92-00

E-mail: datchik@ipu.ru
www.ipu.ru/period/pu

Оригинал-макет
и электронная версия
подготовлены
ООО «ЭЛЕКТРОНИНФОРМ»

Отпечатано с готовых диапозитивов
в типографии ГКС

Подписано в печать
10.11.2006 г.

Заказ № PB606

Журнал зарегистрирован
в Министерстве
Российской Федерации
по делам печати,
телерадиовещания
и средств массовых
коммуникаций

Свидетельство о регистрации
ПИ №77-11963
от 06 марта 2002 г.

Журнал входит в Перечень ВАК
(Бюл. ВАК. — 2004. — № 3)

Подписные индексы:
81708 в каталоге Роспечати
38006 в объединенном каталоге
«Пресса России»

ПРОБЛЕМЫ УПРАВЛЕНИЯ

6.2006

СОДЕРЖАНИЕ

Обзоры

Панкова Л. А., Пронина В. А. Способы создания универсального инструментария для компьютерного моделирования. 2

Информационные технологии в управлении

Елисеев В. В., Игнатушенко В. В. Проблема надежного выполнения сложных наборов задач в управляющих параллельных вычислительных системах. 6

Подлазов В. С., Соколов В. В. Однокаскадные коммутаторы большой размерности для многопроцессорных и многомашинных вычислительных систем. 19

Управление в социально-экономических системах

Нижегородцев Р. М. Проблемы управления инфляцией: современные подходы. . . 25

Ивашенко А. А. Многокритериальные системы стимулирования. 31

Мишин С. П. Оптимальность древовидной иерархии управления симметричной производственной линией. 36

Буянов Б. Б., Лубков Н. В., Поляк Г. Л. Система поддержки принятия управленческих решений с применением имитационного моделирования. 43

Управление технологическими процессами

Генкин А. Л., Куделин А. Р. Проблемы энергосберегающего управления листопрокатным комплексом. Ч. 1. 50

Мандель А. С. Экспертно-статистические методы обработки информации в интегрированных системах управления производством и технологическими процессами. 55

Искусственный интеллект

Абрамова Н. А., Коврига С. В. О рисках, связанных с ошибками экспертов и аналитиков. 60

Гаврилова Т. Л., Клещев А. С. Внутренняя модель математической практики для систем автоматизированного конструирования доказательств теорем. Ч. 3. Модель доказательства. 68

Управление подвижными объектами

Сомов Е. И. Аналитический синтез программного гиросилового управления свободнолетающим космическим роботом в режиме транспортировки полезного груза. 72

Рудько И. М. Определение дистанции до цели, находящейся в зоне Френеля антенны. 79

Краткие сообщения

Эпштейн В. Л. Как увеличить индекс цитирования научной публикации. 83

Лебедев В. В., Гансвинд И. Н., Горохова И. Н. и др. Модульная распределенная система решения типовых региональных задач (на примере гидрологических задач). 85

Шкуратов С. Е. Подход к стратегическому управлению малым инновационным предприятием. 88

Лисицын Н. В., Викторов В. К. Определение оптимального времени функционирования информационной системы. 91

Сирина Н. Ф., Цыганов В. В. Адаптивные механизмы оценки и классификации дальновидных активных систем. 93

* * *

Указатель статей, опубликованных в 2006 г. 96

Index of papers published in 2006 98

Contents & abstracts 99

СПОСОБЫ СОЗДАНИЯ УНИВЕРСАЛЬНОГО ИНСТРУМЕНТАРИЯ ДЛЯ КОМПЬЮТЕРНОГО МОДЕЛИРОВАНИЯ

Л. А. Панкова, В. А. Пронина

Институт проблем управления РАН им. В. А. Трапезникова, г. Москва

Рассмотрены современные технологии компьютерного моделирования и классификация компьютерных моделей. Выделены подходы к созданию универсального инструментария для компьютерного моделирования на основе современных технологий.

ВВЕДЕНИЕ

Компьютерное моделирование — мощное аналитическое средство, вобравшее в себя весь арсенал новейших информационных технологий, включая развитые графические оболочки для конструирования моделей и интерпретации выходных результатов моделирования, мультимедийные средства, поддерживающие анимацию в реальном масштабе времени, объектно-ориентированное программирование, Интернет-решения и др.

В настоящей работе описывается существующая классификация компьютерных моделей по степени дискретности переменных модели и по способу описания функционирования модели (через события, действия, процессы, агентов). Рассматриваются подходы к созданию универсального инструментария для компьютерного моделирования на основе современных технологий.

1. ТЕХНОЛОГИИ КОМПЬЮТЕРНОГО МОДЕЛИРОВАНИЯ

Начиная с середины 1960-х гг. широкое распространение получила структурно-функциональная [1–3] методология анализа, проектирования и разработки моделей — SADT (Structured Analyses and Design Technique — технология структурного анализа и проектирования) [4]. В основе SADT лежит принцип иерархической декомпозиции сверху вниз (системы AUTOIDEF0, Design/IDEF). Особенность моделирования на основе SADT состоит в возможности одновременно со структурированием проблемы разрабатывать структуру базы данных. Применение методологии SADT позволяет унифицировать различные блоки модели сложной системы, распараллелить процесс составления модели и объединить отдельные модули в единую иерархическую динамическую модель. Еще одним широко известным инструментальным сред-

ством структурно-функционального моделирования, основанным на SADT, является пакет Bpwin (Macro-Project). Он предназначен для моделирования и оптимизации бизнес-процессов. Его достоинство заключается в возможности связи с известным инструментальным средством разработки баз данных Erwin (Logic Works). Отечественное инструментальное средство для структурно-функционального анализа — CASE-Аналитик.

Однако структурно-функциональная технология моделирования характеризуется слабой взаимосвязью процессов и данных, присутствующих в модели. Многочисленные наблюдения за жизнедеятельностью систем любого рода позволяют сделать вывод, что данные модели на порядок более стабильны, чем процессы, которые происходят в ней. В отличие от структурных методологий, ориентированных на функциональность системы, объектные технологии [5], ориентированные на данные, на тесную взаимосвязь данных и процессов системы, позволяют программным системам быть более надежными, легко реализуемыми и устойчивыми к изменениям. Кроме того, такая система наиболее соответствует общим концепциям поведения систем реального мира.

Объектно-ориентированный подход в последнее время стал так прочно ассоциироваться с программированием, что многие забывают о его прямой связи с моделированием: первоначально он был применен в языке моделирования SIMULA-67 [6]. В дальнейшем объектно-ориентированный подход развивался почти исключительно программистами. И только в последние годы объектно-ориентированный подход стал востребованным в своей «родной области» — моделировании. Этот подход поддерживают современные системы моделирования (Omola, Dymola, Model Vision, AnyLogic). Кандидатом на роль объекта в объектно-ориентированном моделировании (OOM) выступает компонент (блок) — совокупность переменных и поведения, взаимодейст-



вующий с внешним миром через внешние переменные. Компонент является экземпляром некоторого класса.

В итоге тридцатилетнего развития объектно-ориентированного подхода в первой половине 1990-х гг. был предложен разработанный под эгидой «Rational Software Corporation» на основе наиболее популярных объектных методов — Object Modeling Technique (OMT, James Rumbaugh), Booch (Grady Booch) и Object Oriented Software Engineering (OOSE, Ivar Jacobson) — унифицированный язык объектного моделирования (Unified Modeling Language — UML) [5], предназначенный для создания объектно-ориентированных спецификаций моделей систем.

Уже сегодня существуют множество CASE-средств, автоматизирующих процесс анализа и проектирования в UML: Rational Rose, Paradigm Plus, Select Enterprise, Microsoft Visual Modeler for Visual Basic и др. Современные CASE-средства поддерживают множество языков программирования (C++, Java, Delphi, Power Builder, Visual Basic, Centura, Forte, Ada, Smalltalk и др.), осуществляя автоматическую кодогенерацию и обратное проектирование, т. е. построение модели по программному коду, а также позволяют генерировать базу данных для большинства из существующих SQL-серверов.

В 1999 г. разработан профиль UML для моделирования систем реального времени под названием UML-RT, который позволяет иерархически строить системы из инкапсулированных модулей с четко определенными интерфейсами (портами) и явными каналами связи (соединителями).

Структурная и объектная методологии моделирования сформировались значительно позже, чем появилась возможность имитировать реальность с помощью компьютерных моделей. В русском языке появились термины «имитационная модель» и «имитационная моделирование», а в английском — «simulation modeling». Если в английском языке термин имеет вполне четкий смысл, ибо симуляция и моделирование не являются синонимами, то по-русски имитационная модель — это нонсенс. Любая модель, в принципе, имитационная, ибо она имитирует реальность [4]. Указывая, что данная модель имитационная, подчеркивается, что в отличие от других типов абстрактных моделей в этой модели сохранены и легко узнаваемы такие черты моделируемого объекта, как структура, связи между компонентами, способ передачи информации. Традиционно под компьютерным моделированием понимается лишь имитационное моделирование. В дальнейшем термин «имитационное» будем опускать.

Инструментальные средства компьютерного моделирования, или языки моделирования, появились довольно давно, почти одновременно с языками Algol и Fortran, и прошли путь от бурного развития в 1970-х гг., когда они ежегодно рождались десятками, до современного стабильного состояния, когда доминирует лишь несколько языков [7].

Все существующие средства моделирования представляют собой оболочку над каким-либо языком программирования (Fortran, Algol, C, Pascal, Java и др.), ориентированную на определенный тип модели или на конкретную область применения. Тип модели определяется такими критериями, как способ отсчета времени, степень дискретности зависимых от времени переменных, задание способа функционирования модели, сте-

пень детерминированности модели, степень изолированности модели (активности проводимого эксперимента). Кроме того, специализация системы моделирования может быть связана с проблемной ориентацией (например, банковская система, угольная промышленность, автомобилестроение и т. д.).

Специализацией систем моделирования достигается максимальное упрощение описания задач и подготовки исходных данных для пользователя, не являющегося специалистом в области программирования и численного моделирования. Чем уже проблемная ориентация, тем меньше работы по программированию. В пределе можно добиться полной параметризации системы, и тогда задание конкретной модели будет состоять лишь в указании числовых значений параметров.

В современных системах моделирования компьютерная модель создается с помощью графического языка, потом автоматически транслируется на язык моделирования и далее — на машинный язык. Трансляция на машинный язык производится в два этапа: сначала — трансляция с языка моделирования на «родной» язык программирования, затем — трансляция с языка программирования на машинный компилятором соответствующего языка программирования.

2. КЛАССИФИКАЦИЯ КОМПЬЮТЕРНЫХ МОДЕЛЕЙ

В компьютерной модели время является основной независимой величиной. Другие переменные модели зависят от времени. Модельное время может быть непрерывным или дискретным в зависимости от того, могут ли дискретные изменения зависимых переменных происходить в любые моменты времени или только в определенные [8]. Традиционно модели делятся на дискретные, непрерывные и комбинированные в зависимости от поведения зависимых переменных. В дискретной модели зависимые переменные изменяются дискретно в определенные моменты модельного времени — моменты свершения событий. В непрерывной модели зависимые переменные изменяются непрерывно в течение модельного времени. В комбинированной (гетерогенной) модели зависимые переменные изменяются дискретно, непрерывно или непрерывно на отрезках времени.

Функционирование дискретной модели (Arena, Extend, SimProcess, AutoMod, PROMODEL, Enterprise Dynamics, FlexSim, eMPlant, GPSS, SIMULA, Q-GERT, SLAM, GASP, Witness, Taylor, QUEST, SIMFACTORY II.5, SIMPLE++, ARIS и др.) задается:

- определением изменений состояния модели (совокупности значений переменных), происходящих в моменты свершения событий;
- описанием действий, в которых принимают участие элементы модели;
- описанием процессов, через которые проходят элементы модели.

Событие определяет начало или окончание действия. Процесс — это последовательность событий, которая может состоять из нескольких действий [8].

Способ описания функционирования дискретной модели определяет три альтернативные методологические подхода к построению дискретной модели: событийный, сканирования активностей и процессно-ориентированный. Каждый подход предлагает некоторую схему описания модели и каждый обладает определенны-

ми достоинствами и недостатками. Процессно-ориентированный (SIMULA, SLAM, SIMAN) подход краток и прост в изучении схемы модели, но не обладает достаточной гибкостью. Событийный подход (SIMSCRIPT) — более сложен, но обеспечивает гибкую схему модели. Подход сканирования активностей (CSL) очень эффективен, когда продолжительность действия определяется в зависимости от того, насколько состояние модели удовлетворяет заданным условиям. Но из-за необходимости сканировать условия для каждого действия он менее эффективен по сравнению с событийным подходом.

В непрерывной модели состояние системы определяется непрерывно изменяющимися зависимыми переменными (переменными состояниями). Непрерывная модель задается уравнениями для этих переменных (GSSL, Dynamo [9], Экспресс-радиус [10] и др.).

В комбинированных моделях различают два вида событий: временные события, которые совершаются в определенные моменты времени, и события состояния, которые происходят, когда модель достигает определенного состояния.

Частным случаем комбинированных (или расширением непрерывных) моделей являются так называемые гибридные модели — непрерывные модели, имеющие различное поведение в различных областях фазового пространства (Matlab-Simulink-Stateflow, Model Vision System, Modelica). Их фазовая траектория в зависимости от происходящих событий состояния, приводящих к смене поведения, оказывается то в одной области, то в другой. Таким образом, к гибридным можно отнести классические непрерывные модели, чье фазовое пространство разбивается на области с различным поведением, модели с разрывными правыми частями и модели, у переменных которых меняется размерность в различных областях фазового пространства. Каждой области можно поставить в соответствие вершину некоторого графа, а его направленные дуги трактовать как возможные пути смены текущего локального поведения. Границы областей обычно задают с помощью предикатов, которые приписываются соответствующим дугам графа. Таким образом, гибридная система может быть представлена в виде графа, вершинам которого поставлены в соответствие классические непрерывные модели и одна из вершин помечена как начальная, а дугам — условия смены поведения и мгновенные действия, выполняемые при смене поведения. Такая формализация называется гибридным автоматом. Это наиболее наглядная и удобная форма описания поведения гибридных моделей, являющаяся расширением карты состояний Харела, принятой в UML. По мере движения модельного времени фазовая траектория пересекает границы областей и меняется вид решаемых уравнений.

В работе [11] введены два типа гибридных автоматов:

- обобщенная карта состояний, применяемая для описания гибридных систем в Matlab-Simulink-Stateflow;
- карта поведения, применяемая для описания гибридных систем в Model Vision (в частном случае чисто дискретной модели всем узлам карты поведения следует приписать пустые локальные поведения и тогда карта поведения превращается в карту состояний Харела).

Традиционные модели (дискретные, непрерывные и комбинированные) характеризуются тем, что в них централизованно определяется поведение (динамика) модели в целом (моделирование сверху вниз). Разработка мо-

дели невозможна без знания о глобальных зависимостях исследуемого объекта.

Современный альтернативный подход к разработке модели называется агентным моделированием (AnyLogic [12], ЭКОМОД [13]). Общеизвестного определения «агента» не существует; до сих пор спорят о том, какими качествами должен обладать объект, чтобы «заслужить» называться агентом — инициативностью и реактивностью, ориентацией в пространстве, способностью обучаться, общаться, «интеллектом» и т. д. [12]. Определяется поведение агента на индивидуальном уровне, а глобальное поведение модели возникает как результат деятельности многих агентов, каждый из которых следует своим собственным правилам, живет в общей среде и взаимодействует со средой и с другими агентами. Поэтому агентное моделирование называют еще моделированием снизу вверх. Важное преимущество агентного моделирования в том, что разработка модели возможна в отсутствии знания о глобальных зависимостях. Агентная модель строится на основании локальных поведений участников процесса, которые и определяют глобальное поведение системы. Кроме того, агентную модель проще поддерживать: уточнения обычно делаются на локальном уровне и не требуют глобальных изменений.

В работе [12] показано, как построить агентную модель по существующей непрерывной или дискретной, и демонстрируется, как она может быть расширена для учета более сложных поведений, зависимостей и взаимодействий.

3. ПОДХОДЫ К СОЗДАНИЮ УНИВЕРСАЛЬНОГО ИНСТРУМЕНТАРИЯ ДЛЯ КОМПЬЮТЕРНОГО МОДЕЛИРОВАНИЯ

Подход первый к созданию универсального инструментария для компьютерного моделирования основан на «объединении» существующих систем моделирования [14]. В качестве компонентов универсального пакета предлагается взять уже готовые компоненты существующих пакетов, согласовав интерфейс этих компонентов.

«Идеальный» универсальный пакет моделирования должен:

- поддерживать типовые модели;
- иметь для каждой типовой модели входной язык;
- создаваться на принципах ООМ;
- представлять открытый программный продукт из типовых модулей, прототипы которых уже существуют.

Подход второй к созданию универсального инструментария для компьютерного моделирования основан на применении объектного языка программирования и современных технологий модуляризации (COM, CORBA) [15].

Модульная структура моделирующих программ практически неизменна:

- графический интерфейс — представление математической модели в виде, понятном пользователю (структурные схемы, схемы физические принципиальные, гибридные карты состояний и пр.);
- система управления базой данных (СУБД) — хранение объектов модели, а также требуемые преобразования структуры СУБД;
- математическое ядро — исполнение потоков математических функций в процессе функционирования модели;



— серверы визуализации и online-воздействий — интерфейс между функционирующим математическим ядром и пользователем.

Все перечисленные модули уже традиционно считаются независимыми программными продуктами. Разработчики моделирующих программ при создании своих продуктов, как считают сторонники этого подхода, должны придерживаться современных технологий модуляризации (COM, CORBA), а не делать все самостоятельно.

В работе [15] представлен обзор программных решений на языке C++, которые могут быть положены в основу математических ядер моделирующих программ с поточной моделью управления и могут быть выполнены в виде COM-серверов. Показана возможность интеграции математических ядер разных производителей с редактором векторной графики Visio. Сформулированы задачи модулей стыковки основных компонентов программ моделирования.

Подход третий основан на расширении унифицированного языка объектного моделирования UML.

В работе [11] для построения универсальной системы моделирования предполагается создать язык современного объектного моделирования (OOML) путем «скрещивания» языка UML и современных достижений в области непрерывного моделирования (в том числе, идей, предложенных в языке Modelica [16]). Понятия объекта, инкапсуляции, наследования, полиморфизма применительно к моделированию имеют ряд важных особенностей. В языке «физического моделирования» Modelica в настоящее время наиболее широко используются идеи объектного программирования. Однако специалисты считают, что он не может быть положен в основу гипотетического языка OOML, так как недостаточно универсален: решает проблемы компонентного моделирования с неориентированными блоками, а средства описания дискретных аспектов поведения «неудобны». Рассматривается расширение языка UML, учитывающее особенности моделирования гибридных систем.

С другой стороны, базовые механизмы UML (стереотипы, маркированные значения и ограничения) на основе существующих элементов уже позволяют создавать расширения (так называемые профили), ориентированные на определенную предметную область. Так, UML-профиль исполнения — UML Profile for Schedulability, Performance and Time Specification [17] — позволяет аннотированные UML-диаграммы проектируемого приложения использовать для создания дискретно-событийных моделей проекта на ранней стадии [18]. А разработанный гибридный UML-профиль, основанный на языке CHARON [19], может служить для построения гибридных моделей [20].

ЗАКЛЮЧЕНИЕ

В работе сделан обзор существующих средств компьютерного моделирования: описана устоявшаяся классификация моделей, традиционные и современные технологии создания средств моделирования. Выделены наметившиеся подходы к созданию универсального инструментария для компьютерного моделирования на основе современных технологий:

- «объединение» существующих систем моделирования;

- использование объектного языка программирования и современных технологий модуляризации COM, CORBA;
- расширение унифицированного языка объектного моделирования UML.

Очевидно, что последний из этих подходов в наибольшей степени соответствует уровню требований, предъявляемых к современным инструментам для разработки систем имитационного моделирования.

ЛИТЕРАТУРА

1. Вендров А. М. CASE-технологии. Современные методы и средства проектирования информационных систем. — М.: Финансы и статистика, 2003.
2. Гэйн К., Сарсон Т. Структурный системный анализ: средства и методы. — М.: Эйтекс, 1993.
3. Калашин А. Н., Калянов Г. Н. Структурные модели бизнеса: DFD-технологии. — М.: Финансы и статистика, 2003.
4. Моисеев Н. Н. Математические задачи системного анализа. — М.: Наука, 1981.
5. Фаулер М., Скотт К. UML в кратком изложении. Применение стандартного языка объектного моделирования. — М.: Мир, 1999.
6. Дал У. И., Мюрхауг Б., Ньюгорт К. Универсальный язык моделирования. — М.: Мир, 1969.
7. Бахвалов Л. А., Микулич Л. И. Компьютерное моделирование: основные тенденции развития инструментальных средств // Труды ИПУ. — М., 1999. — Т. 2. — С. 5—21.
8. Прицкер А. Введение в имитационное моделирование и язык SLAM II. — М.: Мир, 1987. — 644 с.
9. Дж. Форрестер. Мировая динамика. — М.: Наука, 1978.
10. Дорри М. Х., Рошин А. А. Инструментальные средства «Экспресс-Радиус» для автоматизации динамических расчетов систем управления // Приборы и системы управления. — 1996. — № 3.
11. Бенькович Е. С., Колесов Ю. Б., Сениченков Ю. Б. Практическое моделирование сложных динамических систем. — СПб.: БХВ, 2002. — 441 с.
12. Борцев А. От системной динамики и традиционного ИМ — к практическим агентным моделям: причины, технология, инструменты. <<http://www.xjtek.com/files/papers/fromsdto-abmru.pdf>>.
13. Экомод — интеллектуальный инструмент разработки и исследования динамических моделей экономики / А. А. Петров, И. Г. Поспелов, Л. Я. Поспелова, М. А. Хохлов // Вторая всероссийская научно-практическая конф. по вопросам применения имитационного моделирования в промышленности «Имитационное моделирование. Теория и практика». — Москва, 2005.
14. Колесов Ю. Б., Сениченков Ю. Б. Компьютерное моделирование. <<http://cnews.ru>>.
15. Клиначев Н. В. Обзор архитектурного построения программ математического моделирования динамических систем. <<http://www.vissim.nm.ru/simkernel.html>>.
16. <<http://www.modelica.org>>.
17. UML profile for schedulability, performance and time specification. <<http://www.omg.org/docs/ptc/03-03-02.pdf>>.
18. Balsamo S., Grosso M., and Marzolla M. Towards simulation-based performance modeling of UML specifications. <<http://www.dsi.unive.it/~perf/publications.php>>.
19. CHARON toolkit and information. <<http://www.cis.upenn.edu/mobies/charon>>.
20. HybridUML Profile for UML 2.0 / Kirsten Berkenk, Stefan Bisanz, Ulrich Hannemann, Jan Peleska. <<http://www-verimag.imag.fr/EVENTS/2003/SVERTS/PAPERS-WEB/23-Hannemann-hybridUMLRT.pdf>>.

☎ (495) 334-92-49;

e-mail: pron@ipu.ru



ПРОБЛЕМА НАДЕЖНОГО ВЫПОЛНЕНИЯ СЛОЖНЫХ НАБОРОВ ЗАДАЧ В УПРАВЛЯЮЩИХ ПАРАЛЛЕЛЬНЫХ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМАХ

В. В. Елисеев⁽¹⁾, В. В. Игнатущенко⁽²⁾

⁽¹⁾ НПО «Импульс», г. Северодонецк, Украина;

⁽²⁾ Институт проблем управления им. В. А. Трапезникова РАН, г. Москва

Для обеспечения надежного выполнения сложных наборов взаимосвязанных задач, со случайными временами их реализации, в управляющих параллельных вычислительных системах, предложена новая компьютерная технология, базирующаяся на статическом прогнозировании времен выполнения задач. Под надежным выполнением задач понимается их реализация за время, не превышающее заданное директивное время с требуемой вероятностью. Предложен принципиально новый подход к интеллектуальному динамическому управлению параллельными вычислительными процессами.

ВВЕДЕНИЕ. ОСНОВНЫЕ ПОНЯТИЯ

Существуют особо ответственные применения вычислительных систем, и в первую очередь, — в контурах управления реального времени, где время реакции, «отклика» ВС жестко лимитируется требованиями и особенностями объекта управления, зачастую — в условиях случайных изменений поведения объекта или его параметров. Вот примеры: верхние уровни управления сложными объектами типа атомных реакторов, управление полетами летательных аппаратов, включая космические, управление разветвленными быстропротекающими физическими и химическими технологическими процессами, и пр.

Поэтому наиболее важным критерием эффективности ВС, функционирующих в таких контурах, является способность ВС решить требуемый (заданный пользователем) набор задач управления за время, не большее заданного директивного времени. В связи с этим, для параллельных управляющих ВС реального времени в работе [1] впервые было введено понятие *надежного выполнения* требуемых наборов задач.

Под *надежным выполнением* конкретного набора задач, задаваемого пользователем, в параллельной управляющей ВС понимается выполнение их за время, не превышающее заданное (пользователем) директивное

время $T_{\max}$ с *требуемой* (удовлетворяющей пользователя) вероятностью D [1].

В понятии надежного выполнения задач, *отказ* управляющей ВС трактуется не только как потеря ее работоспособности, но и как невыполнение требуемых задач за время $T_{\max}$ [1]. Такой отказ может иметь место даже при безукоризненном, безотказном функционировании аппаратуры и программного обеспечения ВС. Его причинами могут быть внезапные изменения параметров объекта управления из-за случайных динамических изменений условий его функционирования или параметров окружающей среды, случайные изменения числа заявок к управляющей ВС, логические ветвления в программах параллельных задач и между задачами, случайные изменения объемов данных и объемов вычислений по программам задач, конфликты на общих ресурсах параллельных ВС, и пр. Разумеется, время выполнения задач может существенно увеличиваться и превышать директивное время из-за случайных неисправностей — сбоев или отказов — компонентов ВС.

Отметим также следующее чрезвычайно важное обстоятельство: в современных системах управления сложными объектами и процессами на основе ВС обычно решаются и задачи *оптимизации*, имеющие, как правило, поисковый и (или) итерационный характер, с заранее непредсказуемым числом шагов или итераций.



Вследствие случайного характера указанных явлений и эффектов, времена выполнения задач и их фрагментов (программных модулей), а также наборов задач в целом рассматриваются нами, в общем случае, как *случайные* величины, а процесс выполнения наборов задач в *параллельных* управляющих ВС (на которые ориентировано наше исследование) — как *случайный процесс*. В связи с этим возникают принципиальные, нетривиальные в решениях вопросы: можно ли оценить *вероятность* успешного завершения требуемого набора задач (в условиях упомянутых неопределенностей) за заданное директивное время на ВС с заданной или предполагаемой производительностью вычислительных ресурсов? Или наоборот: какова вероятность невыполнения этих задач, т. е. отказа ВС (в оговоренном выше смысле), за заданное директивное время? Можно ли вычислить «угрозу» такого отказа непосредственно при выполнении задач в ВС? Как управлять вычислительными процессами, чтобы предотвратить или уменьшить эту угрозу? Или даже такая проблема: как заранее «подсказать» директивное время, при котором реально обеспечивалось бы надежное выполнение конкретного набора задач в каждом сеансе управления?

Применительно к управляющим параллельным ВС, включая параллельные управляющие ПТК [2], проблема оценок такого рода получила название *математического прогнозирования* времени выполнения сложных наборов задач (программных комплексов), [1, 3–6]. Принципиально, что эти наборы рассматриваются в наиболее трудоемких версиях — с произвольными информационными и логическими связями между задачами — и потому названы *комплексами взаимосвязанных работ* — задач и (или) их параллельно-последовательных фрагментов (подзадач, процессов), а в общем случае — программных модулей.

Принятые сокращения:

ВС — вычислительная система;
КВР — комплекс взаимосвязанных работ;
МВС — многопроцессорная вычислительная система;
МКМД — множество команд и множество данных;
ОМП — обрывающийся марковский процесс;
ОП — обслуживающий прибор;
ПТК — программно-технические комплексы;
СМО — система массового обслуживания.

Формально под математическим (статическим) прогнозированием времени выполнения конкретного, заданного пользователем КВР понимается определение в статике (т. е. до реализации задач в ВС) стохастических оценок времени T реализации КВР (среднего значения, дисперсии, а главное — функции $F(t)$ распределения времени T) и определение вероятности D завершения КВР за время, не превышающее заданное $T_{\max}$, на параллельной ВС с заданной конфигурацией и производительностью ее вычислительных ресурсов [1].

1. СОСТОЯНИЕ ПРОБЛЕМЫ

Наибольшее число результатов по прогнозированию времени выполнения наборов взаимосвязанных задач в параллельных ВС получено для «детерминистских» пос-

тановок — в предположении, что время решения каждой задачи заранее известно (а не является случайной величиной, как в нашем подходе). Зачастую, например, время выполнения набора взаимосвязанных задач оценивается как сумма времен (либо заранее известных, либо средних, либо предполагаемых значений) выполнения задач вдоль некоторых путей в графе задач. Для нас же наибольший интерес представляют исследования, в которых рассматриваются *случайные* времена выполнения задач. Поэтому «идеологически» к рассматриваемой проблематике — прогнозированию времени выполнения задач управления — близки вероятностные модели, оценивающие способность ВС к выполнению заданного набора работ за ограниченное время, а также вероятностные подходы к прогнозированию производительности ВС на ранних этапах их проектирования (см., например, обзор [7]). Наиболее представительными, и «близкими по духу» к нашей проблематике являются, по мнению авторов, работы [8, 9]. В первой из них приводится вероятностная модель для оценки математического ожидания и дисперсии времени выполнения набора задач, а в другой — вычисляется среднее время отклика программы, представленной в виде графа задач, с использованием «уровня узла или задачи» в графе. Вообще, в указанных и других известных авторам исследованиях, всюду оцениваются *не распределения* времени выполнения наборов взаимосвязанных задач, а *границы* значений параметров распределений (прежде всего, первого момента).

В целом, в указанных выше и других известных авторам исследованиях по прогнозированию времени выполнения наборов взаимосвязанных задач со *случайными* временами их реализации, прогнозирование этого времени осуществляется *только в статике* (т. е. до выполнения задач в ВС) и ограничивается оценками либо средних времен наступления некоторых событий, либо границ интервалов значений искомых времен (включая оценки «наихудшего случая»), т. е. «точечными» оценками, — совершенно недостаточными для прогнозирования и анализа поведения процессов *в динамике* и оперативного управления ими в ходе выполнения задач в управляющих ВС реального времени, на которые ориентировано наше исследование.

В отличие от известных подходов, в данной работе исследуется и обобщается проблема *оперативного, динамического управления* параллельно-последовательными вычислительными процессами для обеспечения *надежного выполнения* сложных наборов задач — КВР — на основе математического (статического) прогнозирования времени их реализации в параллельных управляющих ВС, а также на основе оперативного анализа и *динамического уточнения* статических прогнозов непосредственно в процессе выполнения наборов взаимосвязанных работ (т. е. в реальном времени).

В Институте проблем управления РАН (г. Москва) с участием НПО «Импульс» (г. Северодонецк, Украина) сформулирована и развивается *общая методология* решения этой проблемы. «Общая» — поскольку методология применима для анализа выполнения сложных наборов задач и управления процессами как в однородных [1–5], так и в неоднородных [6] управляющих МВС класса

МКМД (со многими потоками команд и данных), а также в управляющих ВС с распределенной структурой, в частности — в ПТК [2, 10]. Принципиальные особенности этой методологии, помимо ее применимости для параллельных ВС различных классов, заключаются в следующем:

— в основу ее положен анализ выполнения наборов *взаимосвязанных* задач — КВР — и (или) их фрагментов с произвольными информационными и логическими связями между ними, со *случайными* временами реализации работ КВР;

— управление процессами базируется на вычислениях (в статике) и использовании (в динамике) математических прогнозов, включая *функции распределения* $F(t)$ времен выполнения как КВР в целом, так и его фрагментов;

— восстановление функций распределения $F(t)$ позволяет определять *вероятности* успешного завершения (или невыполнения) каждого заданного пользователем КВР и (или) его фрагментов за требуемое директивное время $T_{\max}$ в данной параллельной управляющей ВС, вычислять «угрозу» отказа (в оговоренном выше смысле) непосредственно при выполнении КВР в ВС и пр.

В такой комплексной постановке проблема впервые была сформулирована в работе [1].

2. ИСХОДНЫЕ ДАННЫЕ И БАЗОВАЯ МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

Для решения задачи математического (статического) прогнозирования в данной работе используются следующие исходные данные.

В качестве параллельных ВС здесь для определенности будут иметься в виду МВС класса МКМД, состоящие из k однотипных процессоров с общей оперативной памятью — например, как в работе [11].

Пользователем задается (по аналогии с работами [1, 3–5]) КВР в виде простого ориентированного графа $G(A, H)$ с конечным числом N вершин, где вершина $a_j \in A$ соответствует j -й работе ($j = 1, \dots, N$), а множество дуг H отображает информационно-логические связи между работами. На рис. 1 в качестве примера представлен граф КВР, состоящий из $N = 23$ вершин; этот пример мы и будем использовать во всем последующем изложении для иллюстрации нашего подхода. Для простоты пояснений выбран граф КВР без логических ветвлений и циклов *между работами*, причем граф G имеет только одну входную и только одну выходную висячие дуги, соответствующие начальной и конечной вершинам (работам) графа КВР.

Работа a_j считается предшественником по отношению к работе a_s , если имеется дуга $(a_j, a_s) \in H$; в этом случае работа a_s является преемником по отношению к работе a_j . Каждая работа считается готовой к выполнению, если выполнены все ее предшественники (в нашем примере). В каждый момент времени на одном процессоре МВС может выполняться только одна работа (один программный модуль). При этом в программном модуле каждой работы допускается произвольное число логи-

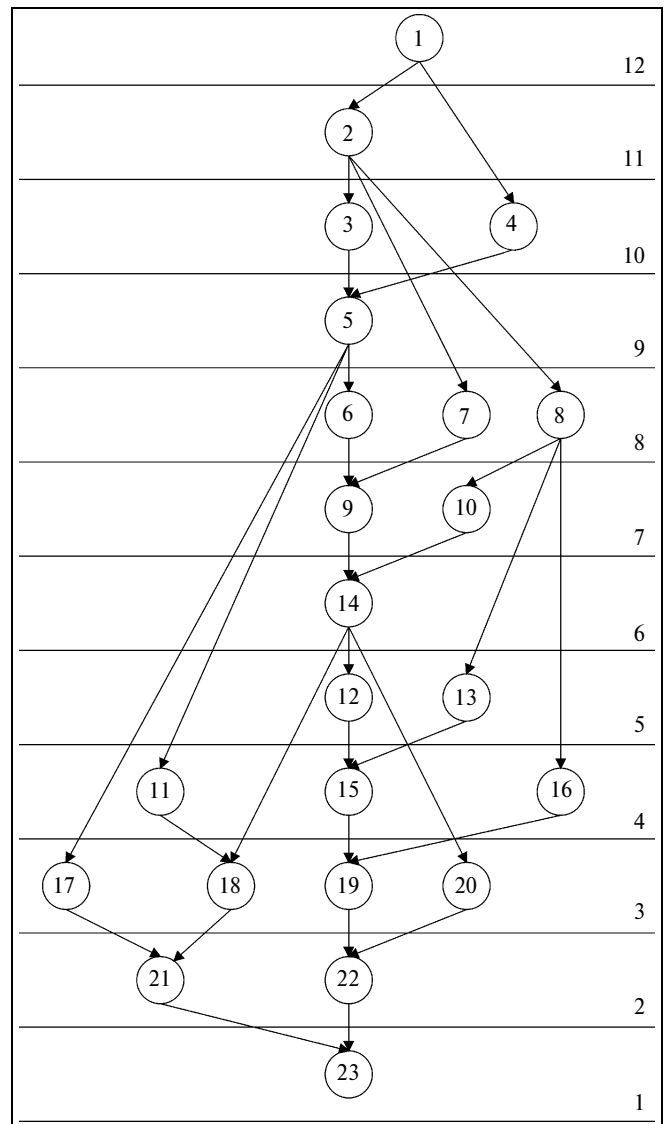


Рис. 1. Пример графа КВР

ческих ветвлений и циклов; из-за этого, в частности, время выполнения каждой работы рассматривается в нашем подходе как случайная величина, — наряду с другими причинами, указанными во Введении.

Каждая работа $a_j, j \in (1, \dots, N)$, характеризуется *случайным* временем обслуживания t_j , распределенным по *экспоненциальному* закону с параметром $\mu_j = 1/M[t_j]$, соответствующим интенсивности обслуживания конкретной работы a_j .

Конечно же, реальный закон распределения времени t_i в общем случае может быть неизвестен; обычно оказывается известным (или задается, или предполагается пользователем) лишь среднее значение случайной величины t_i . В нашем подходе выбор именно экспоненциального распределения для t_i обосновывается и мотивируется следующим образом. Получаемые при таком распределении значения прогнозов (например, среднего



времени $\bar{T}$ выполнения КВР) являются худшими — наиболее «пессимистическими» (т. е. *верхними*) — в том смысле, что при других законах распределения (с коэффициентами вариации $c < 1$, [12]) прогнозируемые значения времени выполнения КВР и его фрагментов (в частности — $\bar{T}$) были бы меньше (т. е. лучше), а вероятность успешного завершения КВР за заданное директивное время была бы большей, чем при экспоненциальном законе распределения времени t_j . Заметим, однако, что при законах распределения случайных величин t_j с коэффициентами вариации $c > 1$ (например, при гиперэкспоненциальном распределении [12]) времена выполнения КВР и его фрагментов могут оказаться еще хуже (т. е. больше), чем при экспоненциальном распределении времени t_j , принятом для расчета прогнозов.

Таблица 1

Таблица связности вершин (работ) КВР

Номер работы a_j	Пред- шест- венни- ки	Преем- ники	$M[t_j],$ τ	μ_j	Ранг r_j	Связ- ность b_j
1	2	3	4	5	6	7
1	—	2, 4	150	0,0067	12	2
2	1	3, 7, 8	30	0,0333	11	3
3	2	5	160	0,0063	10	1
4	1		80	0,0125		
5	3, 4	6, 11, 17	70	0,0143	9	3
6	5	9	120	0,0083	8	1
7	2		170	0,0059		
8		10, 13, 16	50	0,0200		3
9	6, 7	14	130	0,0077	7	1
10	8		150	0,0067		
11	5	18	100	0,0100	4	
12	14	15	160	0,0063	5	
13	8		100	0,0100		
14	9, 10	12, 18, 20	40	0,0250	6	3
15	12, 13	19	30	0,0333	4	1
16	8		170	0,0059		
17	5	21	100	0,0100	3	
18	11, 14		70	0,0143		
19	15, 16	22	130	0,0077		
20	14		80	0,0125		
21	17, 18	23	120	0,0083	2	
22	19, 20		50	0,0200		
23	21, 22	—	40	0,0250	1	0

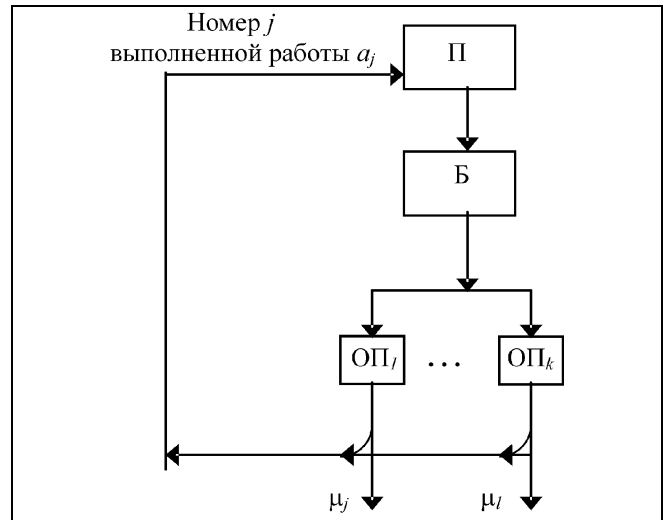


Рис. 2. Структурная схема математической модели

Граф G заданного пользователем КВР описывается таблицей связности его вершин (см., например, работу [11]). Таблица содержит N строк (по числу вершин графа), в каждой из которых указываются номер j работы a_j , номера работ (вершин) — предшественников и преемников данной работы, а также среднее время ее выполнения $M[t_j]$ в некоторых условных тактах τ и/или интенсивность ее обслуживания μ_j (табл. 1, столбцы 1—5). Значение и смысл двух последних столбцов в табл. 1 будут пояснены далее.

Среднее время выполнения каждой работы t_j задается *пользователем*. Если же это время лишь априорно предполагается, то пользователь может задавать и апробировать спектр средних значений t_j — это приводит к увеличению объема программного моделирования в статике, но не приводит к изменению рассматриваемой методологии.

Для каждого КВР *пользователем* задается директивное время его выполнения $T_{\max}$.

Процесс выполнения КВР в МВС представляется математической моделью в виде однофазной СМО, рис. 2, содержащей $k \geq 2$ обслуживающих приборов — ОП (соответствующих процессорам МВС), буфер B для готовых к выполнению работ, которые поступают из пула P , содержащего в исходном состоянии N работ. Модель функционирует в непрерывном времени.

Эта математическая модель названа базовой [1, 3—5], поскольку является основой для всех ее модификаций, используемых для прогнозирования выполнения как КВР в целом, так и его фрагментов, в параллельных ВС различных типов (упоминавшихся в § 1) и вообще в нашем подходе.

Модель функционирует следующим образом. В начальный момент времени в систему ($B, ОП$) поступает из пула P одна заявка — «начальная» работа a_1 , которая немедленно начинает обслуживаться на одном из ОП с интенсивностью μ_1 . По завершению обслуживания

работы a_1 (в общем случае — a_j) в ОП она покидает систему, «передавая» свой номер в пул Π , из которого выбираются в буфер B те приемники работы a_j , которые оказались готовы к выполнению (т. е. выполнены все их предшественники); номера этих работ однозначно устанавливаются по таблице связности вершин графа G (табл. 1, столбец 3). Из буфера B заявки выбираются — по некоторой заданной пользователем дисциплине (критерию диспетчеризации) — на обслуживание в незанятые ОП. Система массового обслуживания функционирует до тех пор, пока в пуле Π и в системе не иссякнут заявки.

Пусть T — время жизни системы (время выполнения КВР из N работ). Функционирование такой СМО можно описать ОМП $X(t)$, $t \in [0, T)$ над следующим множеством состояний:

$$X = \{(m; \vec{i}_w; \vec{i}_n) : m = \overline{0, N-1}; w = \overline{0, N-2}; n = \overline{1, k}\},$$

где m — число работ, находящихся в пуле Π в момент t ; $\vec{i}_w = i_1, \dots, i_w$; $\vec{i}_n = j_1, \dots, j_n$; причем $i_1, \dots, i_w$ — номера работ, ожидающих в буфере B , w — число таких работ; $j_1, \dots, j_n$ — номера работ, находящихся на обслуживании в ОП, n — число этих работ; k — число ОП.

В каждый момент t в системе (B , ОП) находятся $w + n$ работ. Начальным состоянием процесса $X(t)$ является состояние $X(0) = (N-1; 0, \dots, 0; 1, 0, \dots, 0)$, при котором в пуле Π находится $N-1$ работ, буфер B пуст, а на обслуживании в ОП находится работа с номером 1 (работа a_1). Процесс обрывается, т. е. переходит в поглощающее состояние $X(T) = (0; 0; 0)$ после завершения обслуживания «конечной» работы a_N (после выполнения всех N работ заданного КВР).

3. РАСЧЕТЫ БАЗОВОЙ МАТЕМАТИЧЕСКОЙ МОДЕЛИ — ПРОГНОЗИРОВАНИЕ ВРЕМЕНИ ВЫПОЛНЕНИЯ КВР И ЕГО ФРАГМЕНТОВ

Реализация принципа математического (статического) прогнозирования времени выполнения как КВР в целом, так и его фрагментов, разделяется на следующие процедуры.

Процедура 1. По алгоритму, описанному в работе [3], для базовой математической модели и заданного КВР строится *граф переходов* P (при заданном k) и матрица интенсивностей Q процесса $X(t)$. Алгоритм [3] включает в себя процедуру определения работ, готовых к выполнению, а также работ, выбираемых на обслуживание в ОП по самым различным дисциплинам диспетчеризации.

Фрагмент графа переходов P процесса $X(t)$ при выполнении КВР, граф которого изображен на рис. 1, при числе ОП $k = 2$ приведен на рис. 3 (этот фрагмент мы будем использовать в изложении и последующих процедурах). Здесь для каждого состояния указаны: m — число работ, остающихся в пуле Π ; $i_1, \dots, i_w$ — номера работ, ожидающих в буфере B (над чертой в обозначении состояния); $j_1, \dots, j_n$ — номера работ, находящихся на обслуживании в ОП (под чертой). При дугах, соот-

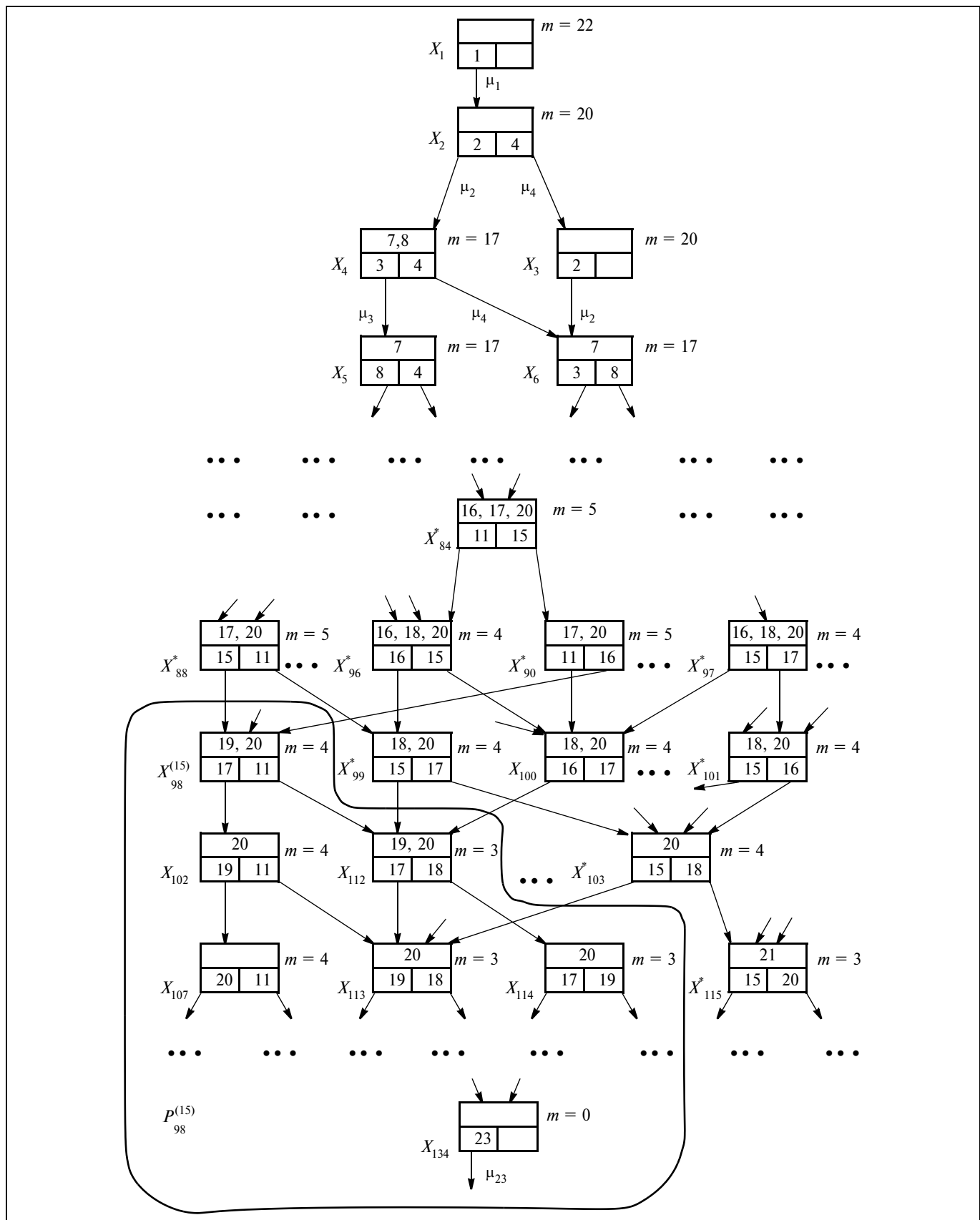
ветствующих переходам процесса из каждого состояния X_p , приписаны интенсивности переходов μ_j , где j — номер работы, завершение которой приводит к изменению состояния X_p . Граф P построен для случая, когда заявки выбираются из буфера B на обслуживание в незанятые ОП по известному критерию диспетчеризации «выбор работы с максимальным рангом r_j /степень связности b_j » вершины a_j [3, 11]; эти параметры для каждой из работ рассматриваемого КВР приведены в табл. 1, столбцы 6 и 7. Под связностью b_j вершины a_j здесь имеется в виду число приемников последней.

Состояния ОМП $X(t)$ нумеруются по правилам упорядочения состояний множества X [3], а именно: состояния нумеруются и упорядочиваются по убыванию или невозроанию значений параметра m (текущее число работ в пуле Π); состояния с одинаковыми значениями m упорядочиваются по убыванию или невозроанию значений $w + n$ (число работ в системе), см. рис. 3. При таком упорядочении состояний множества X матрица Q интенсивностей переходов ОМП $X(t)$ оказывается *нижней треугольной матрицей*, что является необходимым условием для определения, по методу [13], функции распределения $F(t)$ времени выполнения КВР в целом.

Процедура 2. Расчет базовой математической модели позволяет определить среднее время $\bar{T}$ выполнения каждого заданного КВР и, используя метод [13], восстановить функцию распределения $F(t)$, которая позволяет, в свою очередь, определить вероятность D завершения каждого конкретного КВР за заданное директивное время $T_{\max}$. Важно отметить, что расчет функции $F(t)$ не зависит от задаваемых значений $T_{\max}$.

Вообще, *вероятность* D завершения КВР за время $T_{\max}$ играет ключевую роль в нашем подходе. С одной стороны, значение D может задаваться пользователем в качестве параметра *надежного выполнения* КВР: например, заданное значение $D = 0,9$ означает, что вероятность завершения КВР за директивное время $T_{\max}$ должна быть не менее 0,9. С другой стороны, реально достигаемое значение D при заданном $T_{\max}$ может быть определено по функции распределения $F(t)$, рассчитываемой по методу [13]. Так, для КВР (см. рис. 1) с параметрами, описанными в табл. 1, функция распределения $F(t)$ представлена кривой 1 на рис. 4 (для случая $k = 2$). Если для этого КВР задано, например, директивное время $T_{\max} = 2300\tau$ (условных тактов), то вероятность D выполнения КВР за это время оказывается равной 0,96. Таким образом, если вероятность D не задана, то она может быть рассчитана в статике и предъявлена потребителю как один из параметров эффективности, «пригодности» управляющей ВС для выполнения требуемых наборов задач за заданное директивное время $T_{\max}$.

Напомним, что при любом другом законе распределения времен t_j выполнения работ КВР с коэффициентом вариации $c < 1$ прогнозируемая вероятность D завершения КВР за заданное время $T_{\max}$ была бы большей

Рис. 3. Фрагмент графа переходов P процесса $X(t)$ выполнения КВР

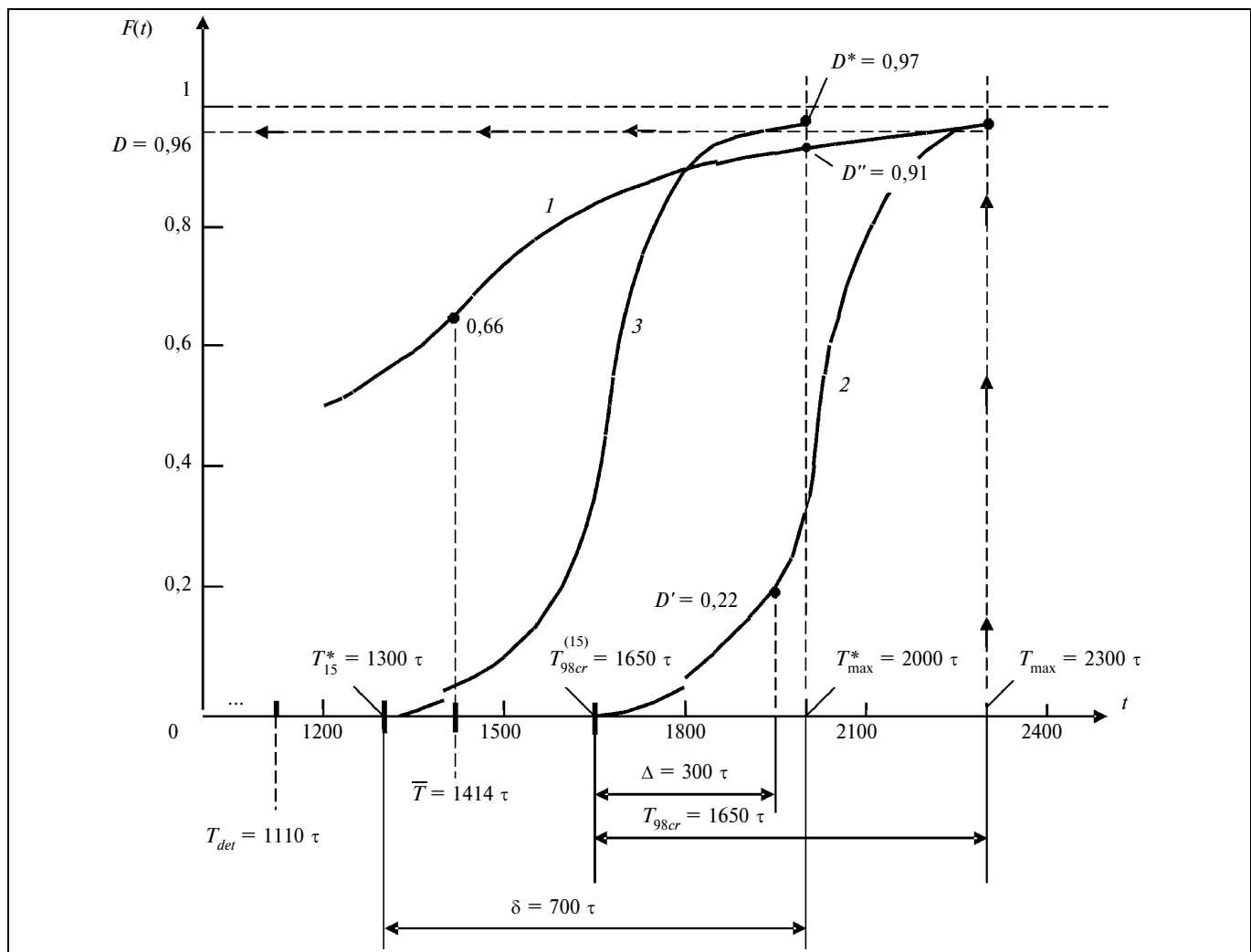


Рис. 4. Функции распределения времени выполнения КВР и его фрагментов

(т. е. более «оптимистической»), чем при используемом здесь экспоненциальном законе распределения времен t_j .

Остановимся на еще одном важном аспекте использования вероятности D завершения КВР за требуемое время. Как следует из расчетов для нашего примера КВР, среднее время его выполнения (при $k = 2$) $\bar{T} = 1414\tau$. Но тогда вероятность выполнения этого КВР за среднее время $\bar{T}$ (определяемая по функции $F(t)$, см. рис. 4, кривая 1) равна всего лишь 0,66. Это означает следующее: если при обосновании времени $T_{\max}$ руководствоваться расчетами лишь среднего времени $\bar{T}$ (например, задавать $T_{\max} = \bar{T}$), то в каждом сеансе управления КВР не будет выполняться за директивное время $T_{\max}$ с вероятностью 0,34 (в нашем примере). Такая вероятность отказа (в используемом нами понятии отказа ВС) совершенно неприемлема для многих практических систем управления реальным временем.

Сказанное свидетельствует о необходимости определения функций распределения времен выполнения рас-

сматриваемых процессов, о необходимости использования этих функций для обоснования директивных времен $T_{\max}$ для оценки «пригодности» ВС к выполнению требуемых наборов задач за заданное время (обусловленное особенностями конкретного объекта управления), а также для обоснования и выбора наборов задач и даже алгоритмов их решения.

Еще один важный тип параметров используется в нашем подходе — *контрольные события* Z_j — события завершения некоторых работ КВР, — например, отнесенных к «критическим процессам» КВР, или части работ критического пути в графе КВР, или работ, по завершению которых выдаются управляющие сигналы во внешнюю среду. Работу a_j , завершение которой рассматривается как контрольное событие Z_j , будем называть контрольной работой. Контрольные события и соответствующие им контрольные работы обычно назначаются в КВР пользователем. Если же таковые не заданы, то может быть использован формализованный аппарат назначения контрольных событий и выбора контрольных работ, изложенный в работе [14]. Для простоты, в



нашем примере мы будем рассматривать только одно контрольное событие Z_{15} , которое соответствует завершению контрольной работы a_{15} , принадлежащей критическому пути в графе КВР (см. рис. 1).

По контрольным событиям будет осуществляться анализ состояний процесса выполнения КВР в параллельной ВС и последующее оперативное управление этим процессом на основе сравнения прогнозируемого (в статике) времени выполнения фрагментов КВР с временными параметрами реальных событий, происходящих в системе в динамике выполнения КВР.

Процедура 3. Определение и анализ опорных состояний системы.

Опорными состояниями $X_s^{(j)}$ системы названы состояния, в одно из которых система может перейти при контрольном событии Z_j завершения работы a_j в некоторый момент t_j .

Событие Z_j может произойти при одном из нескольких состояний X_l^* системы, в которых контрольная работа a_j находится на обслуживании в ОП. Как уже упоминалось, в нашем примере контрольным событием является событие Z_{15} завершения работы a_{15} , принадлежащей критическому пути в графе КВР (см. рис. 1). Это событие может произойти при любом из следующих состояний системы (см. фрагмент графа P на рис. 5): X_{84}^* , X_{88}^* , X_{96}^* , X_{97}^* , X_{99}^* , X_{101}^* , X_{103}^* , X_{115}^* , X_{120}^* и X_{123}^* (для большей наглядности часть из них изображена также на рис. 3).

Очевидно, что из каждого состояния X_l^* система может перейти при завершении работы a_j только в одно

из опорных состояний $X_s^{(j)}$. Так, при завершении работы a_{15} система переходит из состояния X_{84}^* только в опорное состояние $X_{90}^{(15)}$, из состояния X_{88}^* — только в опорное состояние $X_{98}^{(15)}$, из состояний X_{96}^* и X_{97}^* — в опорное состояние $X_{100}^{(15)}$, и т. д. (рис. 3 и 5); для наглядности графические обозначения опорных состояний на рис. 5 заштрихованы, а не обозначены верхним индексом (15).

Каждому опорному состоянию $X_s^{(j)}$ соответствует начальная вершина подграфа переходов $P_s^{(j)}$ процесса выполнения работ некоторого фрагмента КВР — работ, оставшихся для обслуживания после завершения работы a_j . К множеству вершин этого подграфа $P_s^{(j)}$ относятся вершины общего графа переходов P , через которые проходят все пути от начальной вершины (опорного состояния) $X_s^{(j)}$ подграфа $P_s^{(j)}$ к конечной вершине графа P . Для иллюстрации на рис. 3 приведен подграф $P_{98}^{(15)}$ с начальной вершиной (опорным состоянием) $X_{98}^{(15)}$, который обведен сплошной линией; на рис. 5 все вершины этого же подграфа обведены штриховыми линиями.

Таким образом, число подграфов $P_s^{(j)}$ соответствует числу опорных состояний $X_s^{(j)}$, в которые может перейти система в результате контрольного события Z_j .

Таблица 2

Фрагмент таблицы прогнозов выполнения КВР

Контрольные события Z_j	Возможные состояния X_l^* процесса	Опорные состояния $X_s^{(j)}$	«Критические точки» $T_{scr}^{(j)}$ для дообслуживания КВР			
			при «штатном» выполнении КВР	при увеличении числа процессоров ($k = 3$)	при реализации упрощенных версий работ	при выполнении «критического процесса»
1	2	3	4	5	6	7
...	...	...	...	...	...	...
Z_{15} — завершение работы a_{15}	X_{84}^*	$X_{90}^{(15)}$	—	—	—	—
	X_{88}^*	$X_{98}^{(15)}$	$T_{98cr}^{(15)} = 1650\tau$	$T_{cr}^{(15)}(1) = 1760\tau$	$T_{cr}^{(15)}(2) = 1927\tau$	$T_{cr}^{(15)}(3) = 1900\tau$
	X_{96}^*, X_{97}^*	$X_{100}^{(15)}$	—	—	—	—
	X_{99}^*	$X_{112}^{(15)}$				
	X_{101}^*	$X_{105}^{(15)}$				
	X_{103}^*	$X_{113}^{(15)}$				
	X_{115}^*	$X_{125}^{(15)}$				
	X_{120}^*	$X_{128}^{(15)}$				
	X_{123}^*	$X_{132}^{(15)}$				
...	...	...	...	...	...	...

Контрольные события Z_j , соответствующие каждому из них состояния X_l^* и опорные состояния $X_s^{(j)}$ заносятся в таблицу прогнозов выполнения КВР (табл. 2, столбцы 1—3).

Производится перенумерация вершин состояний каждого подграфа $P_s^{(j)}$ в соответствии с алгоритмом упорядочения состояний процесса выполнения КВР [3, 13] и строится матрица интенсивностей переходов $Q_s^{(j)}$ для

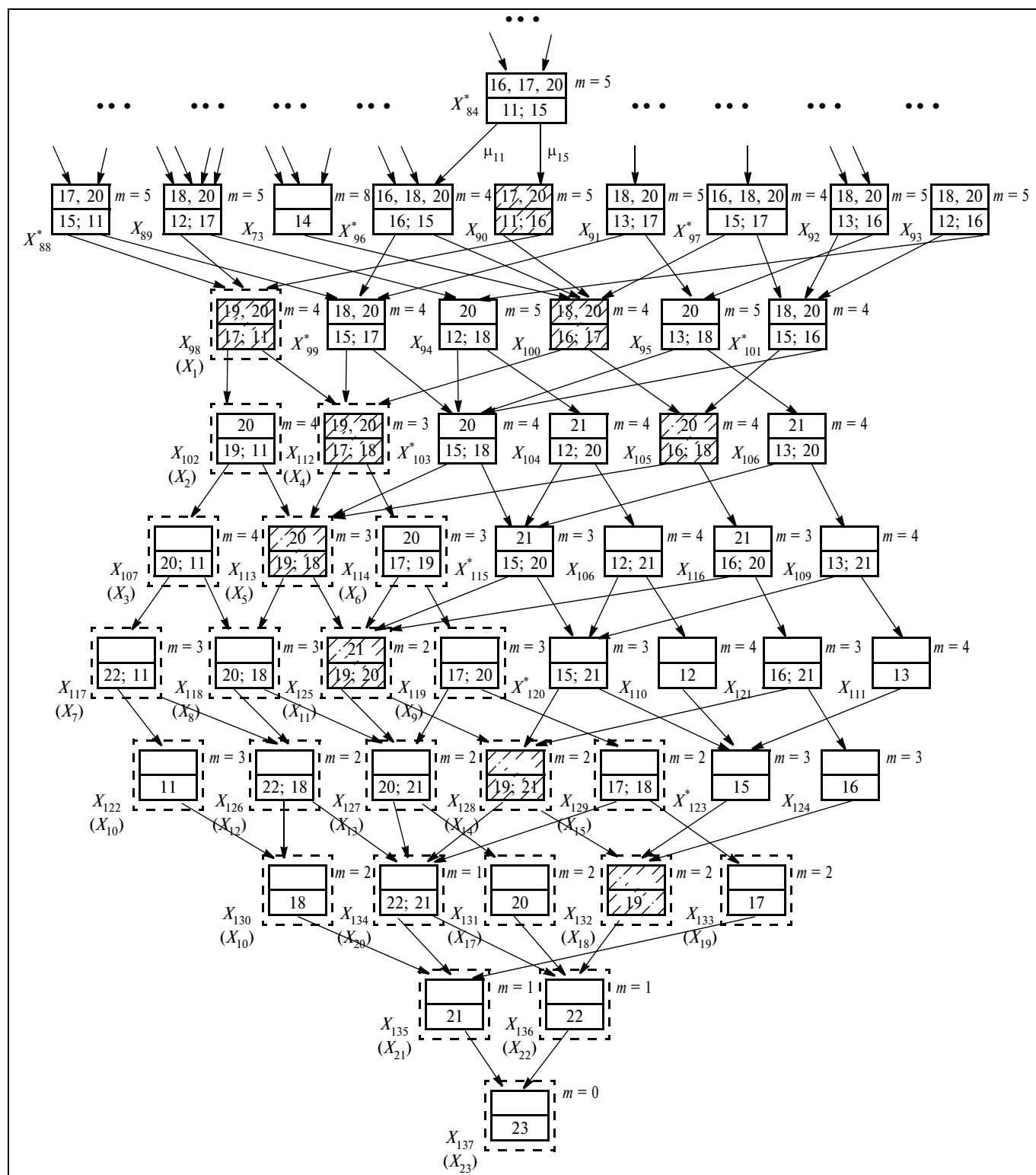


Рис. 5. Фрагмент графа переходов обрывающегося марковского процесса



этого подграфа. На рис. 5 для подграфа $P_{98}^{(15)}$ новые номера состояний указаны рядом с прежними номерами в круглых скобках.

Далее методика излагается для расчета стохастических характеристик одного (любого) ОМП, описываемого подграфом переходов $P_s^{(j)}$ и соответствующего процессу дообслуживания КВР после контрольного события Z_j , начиная с опорного состояния $X_s^{(j)}$,

Под *дообслуживанием КВР* здесь понимается выполнение его *фрагмента*, который состоит из работ, находящихся в пуле Π , буфере B (см. рис. 2) и на выполнении в ОП в момент, когда ОМП переходит в опорное состояние $X_s^{(j)}$ (в нашем примере — в состояние $X_{98}^{(15)}$).

Процедура 4. Восстановление (по методу [13]) функции распределения $F_s^{(j)}(t)$ (в нашем примере — функции $F_{98}^{(15)}(t)$) времени дообслуживания КВР после контрольного события $Z_j(Z_{15})$ с начальным состоянием $X_s^{(j)}$ ($X_{98}^{(15)}$) процесса дообслуживания.

Особенность определения этой функции (по сравнению с определением функции $F(t)$ по Процедуре 2) состоит лишь в том, что в качестве начального состояния ОМП здесь принимается не исходное состояние $X(0)$, описанное ранее, а опорное состояние $X_s^{(j)}$, соответствующее начальной вершине подграфа $P_s^{(j)}$. Так, для ОМП с подграфом переходов $P_{98}^{(15)}$ начальным состоянием является опорное состояние $X_{98}^{(15)} = (4; 19; 20; 17; 11)$ (см. рис. 3), обозначенное на рис. 5 после перенумерации состояний этого ОМП как (X_1) .

Функция распределения $F_{98}^{(15)}(t)$ представлена кривой 2 на рис. 4 (начиная не с $t = 0$, а с некоторого другого значения t по причинам, объясняемым далее).

Параметры функции $F_s^{(j)}(t)$ (в рассматриваемом случае — $F_{98}^{(15)}(t)$) — например, ее значения через интервалы 10τ или 100τ , начиная с $t = 0$, заносятся в таблицу прогнозов (для упрощения, в табл. 2 значения функций $F_s^{(j)}(t)$, соответствующие кривой 2 на рис. 4, не приведены).

Процедура 5. Определение необходимого директивного времени $T_{s\max}^{(j)}$ ($T_{98\max}^{(15)}$) дообслуживания КВР с *требуемой вероятностью D* после контрольного события $Z_j(Z_{15})$ с начальным состоянием $X_s^{(j)}$ ($X_{98}^{(15)}$) процесса дообслуживания.

Будем принимать требуемую вероятность D дообслуживания КВР такой же, как и вероятность D выполнения КВР в целом за заданное директивное время $T_{\max}$.

Тогда директивное время $T_{s\max}^{(j)}$ дообслуживания КВР с

заданной вероятностью D определяется по функции распределения $F_s^{(j)}$ — эта процедура соответствует решению «обратной задачи» по методу [13]. В нашем примере необходимое директивное время $T_{98\max}^{(15)}$ дообслуживания КВР с требуемой вероятностью $D = 0,96$ оказывается равным 650τ (см. рис. 4).

Процедура 6. Определение «критической точки» $T_{scr}^{(j)}$ ($T_{98scr}^{(15)}$) — момента наиболее позднего начала дообслуживания КВР, начинающегося с опорного состояния $X_s^{(j)}$ ($X_{98}^{(15)}$): $T_{scr}^{(j)} = T_{\max} - T_{s\max}^{(j)}$.

Смысл «критической точки» заключается в следующем: если дообслуживание КВР начинается не позже момента $T_{scr}^{(j)}$, то обеспечивается «штатное» выполнение КВР в целом за время, не большее директивного времени $T_{\max}$, с заданной вероятностью D . В нашем примере $T_{98scr}^{(15)} = T_{\max} - T_{98\max}^{(15)} = 2300\tau - 650\tau = 1650\tau$, — см. рис. 4, где рассчитанная ранее функция распределения $F_{98}^{(15)}(t)$ дообслуживания КВР (кривая 2) имеет начальную временную координату не $t = 0$, а $t = T_{98scr}^{(15)} = 1650\tau$. Рассчитанное значение $T_{scr}^{(j)}$ ($T_{98scr}^{(15)}$) заносится в таблицу прогнозов (табл. 2, столбец 4).

Процедуры 3—6, изложенные ранее, выполняются для каждого контрольного события Z_j и опорного состояния $X_s^{(j)}$, причем эти процедуры выполняются *в статике*, т. е. до реализации этого КВР на параллельной ВС, а результаты их заносятся в таблицу прогнозов (для простоты в табл. 2 результаты выполнения указанных процедур приведены лишь для контрольного события Z_{15} и опорного состояния $X_{98}^{(15)}$).

Принципиально важно, что изложенные процедуры математического (статического) прогнозирования времени выполнения задаваемых пользователем КВР и их фрагментов на основе разработанных математических моделей имеют, как было показано, «итерационный», повторяющийся характер, а вся методология прогнозирования «укладывается» в несколько типовых процедур.

4. ДИНАМИЧЕСКОЕ УПРАВЛЕНИЕ НАДЕЖНЫМ ВЫПОЛНЕНИЕМ КВР НА ОСНОВЕ СТАТИЧЕСКИХ ПРОГНОЗОВ

Использование *статических* прогнозов для *динамического* (оперативного) управления надежным выполнением КВР заключается в следующем.

При выполнении КВР в параллельной ВС в динамике фиксируется реальный физический момент t_j^* (t_{15}^*) контрольного события $Z_j(Z_{15})$ по отношению к моменту $t = 0$ начала выполнения КВР, а также состояние X_j^* (в нашем примере — X_{88}^*), в котором находился процесс выполнения КВР в этот момент. По таблице прогнозов од-

нозначно определяется опорное состояние $X_s^{(j)}$ (в нашем примере — $(X_{98}^{(15)})$), в которое переходит ОМП, и «критическая точка» $T_{scr}^{(j)}$ ($T_{98cr}^{(15)}$) для дообслуживания КВР при этом опорном состоянии (см. табл. 2, столбцы 1—4).

Сравниваются реальное значение t_j^* (t_{15}^*) и значение спрогнозированной в статике «критической точки» $T_{scr}^{(j)}$ ($T_{98cr}^{(15)}$) для дообслуживания КВР.

Если имеет место $t_j^* \leq T_{scr}^{(j)}$ (если $t_{15}^* \leq T_{98cr}^{(15)}$), т. е. контрольное событие Z_j (Z_{15}) произошло не позже, чем «момент наиболее позднего начала» дообслуживания КВР (реализация которого описывается ОМП с подграфом переходов $P_s^{(j)}$, в нашем примере — подграфом $P_{98}^{(15)}$), то обеспечивается *надежное выполнение КВР*, т. е. выполнение КВР за время, не большее директивного времени T_{max} , с вероятностью, не меньшей заданной вероятности D , и не требуется замены «штатного» алгоритма управления процессами.

Если же имеет место $t_j^* > T_{scr}^{(j)}$ (если $t_{15}^* > T_{98cr}^{(15)}$), т. е. контрольное событие Z_j (Z_{15}) произошло позже, чем это допустимо для надежного дообслуживания КВР, то вероятность завершения КВР в целом за время T_{max} оказывается меньшей, чем требуемая вероятность D , и необходимо переходить к какому-либо другому алгоритму управления дообслуживанием КВР (из заранее заданного набора алгоритмов) — с большим значением $T_{scr}^{(j)}$. Эти алгоритмы задаются пользователем.

В качестве примеров таких алгоритмов здесь выбраны:

- реконфигурация ВС — увеличение числа процессоров ВС, предоставляемых для дообслуживания КВР (числа k обслуживающих приборов в модели СМО);
- реализация упрощенных версий работ (программных модулей) фрагмента КВР, оставшегося для выполнения (как это предусматривается при многоверсионном программировании — см., например, работу [11]); это обычно приводит к уменьшению точности и времени вычислений, а следовательно — к увеличению интенсивностей μ_j обслуживания работ;
- выполнение только «критического процесса» для дообслуживания КВР. Для этого случая *пользователем* должны быть заранее (в статике) указаны работы КВР, которые *обязательно* должны быть выполнены после контрольного события Z_j (Z_{15}), — это и есть работы «критического процесса». В нашем примере КВР (см. рис. 1) «критическим процессом» является (после завершения контрольной работы a_{15}) последовательное выполнение работ a_{19} , a_{22} и a_{23} без использования данных от их предшественников — работ a_{16} , a_{20} и a_{21} .

Для каждого из алгоритмов управления процессами дообслуживания КВР, используя подграфы $P_s^{(j)}$ пере-

дов соответствующих им ОМП и матрицы $Q_s^{(j)}$ интенсивностей переходов, в статике выполняются описанные выше *Процедуры 4—6*; в результате рассчитываются значения «критических точек» $T_{cr}^{(j)}(f)$ для дообслуживания КВР по каждому из этих алгоритмов (здесь f — номер алгоритма). Для рассматриваемого примера значения $T_{cr}^{(j)}(f)$ приведены в табл. 2, столбцы 5—7, — с учетом того, что при переходе от «штатного» к другому алгоритму управления процессами требуется дополнительное время $t_{доп}$ на работу операционной системы; это время принято здесь случайным и распределенным по экспоненциальному закону с параметром $\mu_{доп} = 0,033$ (т. е. $t_{доп} = 30\tau$).

Итак, если при контрольном событии Z_j оказалось, что $t_j^* > T_{scr}^{(j)}$, то по таблице прогнозов определяются «критические точки», значения которых удовлетворяют неравенству $t_j^* \leq T_{cr}^{(j)}(f)$, и выбирается один из соответствующих им алгоритмов управления дальнейшим выполнением КВР (приоритеты назначает пользователь). Если же значение t_j^* оказалось больше любого из значений $T_{cr}^{(j)}(f)$, то выдается сообщение о том, что КВР не будет выполнен за директивное время T_{max} с требуемой вероятностью D ни при одном из предусмотренных пользователем алгоритмов управления, но может быть выполнен с вероятностью $D' < D$.

Пусть в нашем примере оказалось, что $t_{15}^* = 1910\tau$, т. е. $t_{15}^* > T_{98cr}^{(15)}$. По табл. 2 определяем, что требуемому неравенству соответствует единственное значение $T_{cr}^{(15)}(2) = 1927\tau$, и поэтому для дальнейшего управления выполнением КВР выбирается алгоритм (2) — реализация упрощенных версий работ.

Пусть в другом случае оказалось, что $t_{15}^* = 2000\tau$. Тогда для «штатного» завершения КВР остается временной интервал $\Delta = T_{max} - t_{15}^* = 2300\tau - 2000\tau = 300\tau$, вместо требуемого $T_{98max}^{(15)} = 650\tau$ (см. рис. 4). По функции распределения $F_{98}^{(15)}(t)$ (кривая 2) находим, что за оставшееся время Δ КВР может быть завершён лишь с вероятностью $D' = 0,22$.

Последний пример соответствует, в частности, резкому случайному возрастанию объемов данных, обрабатываемых в некоторых программных модулях (например, из-за внезапного радикального изменения числа и значений параметров объекта управления и окружающей среды). Для математической модели это соответствовало бы случайному уменьшению интенсивностей μ_j обслуживания всех или части работ по сравнению с предполагавшимися значениями параметров μ_j . А это может иметь место, в частности, при упоминавшемся выше гиперэкспоненциальном распределении случайных величин t_j (с коэффициентом вариации $c > 1$).



5. ДИНАМИЧЕСКОЕ УТОЧНЕНИЕ СТАТИЧЕСКИХ ПРОГНОЗОВ

Как уже указывалось, при экспоненциальном законе распределения случайных величин t_j определяются верхние, наиболее «пессимистические» оценки времени выполнения КВР и его фрагментов, а также моментов наступления контрольных событий. Эти оценки-прогнозы выражаются функциями распределения $F(t)$ и $F_s^{(j)}(t)$ (кривые 1 и 2 для нашего примера, см. рис. 4) и «критическими точками» $T_{scr}^{(j)}$. (Кстати, точная нижняя оценка времени выполнения КВР — наиболее «оптимистическая» — устанавливается при детерминированных значениях времен $t_{j\text{det}}$; если эти времена имеют значения, указанные в столбце 4 табл. 1, то КВР в нашем примере будет выполняться за время $T_{det} = 1110\tau$, см. рис. 4.) Если же реальные значения t_j подчиняются иному закону распределения (с коэффициентами вариации $c < 1$), то при выполнении КВР в ВС контрольные события Z_j могут наступать существенно раньше, чем в рассчитанные в статике моменты $T_{scr}^{(j)}$. Это может иметь место и в случае резкого уменьшения времени t_j выполнения некоторых работ, например, из-за уменьшения объемов данных, обрабатываемых по программам некоторых модулей. Уже поэтому представляется целесообразным уточнение статических прогнозов в процессе реального выполнения КВР в параллельной ВС.

Пусть для КВР (см. рис. 1 и табл. 1) задано директивное время $T_{\max}^* = 2000\tau$ (а не прежнее $T_{\max} = 2300\tau$), см. рис. 4, при той же заданной пользователем вероятности завершения КВР $D = 0,96$, а в динамике выполнения КВР оказалось, что контрольное событие Z_{15} (завершение контрольной работы a_{15}) наступило в момент $t_{15}^* = 1300\tau$, причем процесс переходит в упоминавшееся опорное состояние $X_{98}^{(15)}$. Это означает, что для надежного выполнения КВР (т. е. для завершения его в пределах $T_{\max}^*$) остается временной интервал $\delta = T_{\max}^* - t_{15}^* = 2000\tau - 1300\tau = 700\tau$.

Функция распределения $F_{98}^{(15)}(t)$ времени дообслуживания КВР из состояния $X_{98}^{(15)}$ остается, разумеется, той же, однако изменяется ее «точка отсчета». В графической интерпретации это означает, что кривая 2, соответствующая функции $F_{98}^{(15)}(t)$, смещается из точки $T_{98\text{cr}}^{(15)} = 1650\tau$ в точку $t_{15}^* = 1300\tau$ (кривая 3). По значениям этой функции (содержащимся в таблице прогнозов) устанавливается, что вероятность завершения КВР за время $\delta = 700\tau$ равна $D^* = 0,97$, т. е. обеспечивается надежное выполнение КВР в целом даже с большей вероятностью, чем заданная $D = 0,96$, и потому не требуется заменять «штатный» алгоритм управления процессами на какой-либо другой алгоритм.

Динамическое уточнение статических прогнозов может быть использовано с существенным эффектом даже тогда, когда расчеты базовой математической модели показывают, что требуемая пользователем вероятность D выполнения заданного КВР за время $T_{\max}$ не достигается. Так, если для нашего примера КВР задано упоминавшееся директивное время $T_{\max}^* = 2000\tau$, то по функции распределения $F(t)$ (кривая 1 на рис. 4) определяем в статике, что вероятность завершения КВР за время $T_{\max}^*$ равна лишь $D'' = 0,91$, т. е. меньше требуемой D . Однако это не обязательно означает, что «штатный» алгоритм управления выполнением КВР должен быть заменен с самого начала (что, например, на выполнение должны назначаться только «критические процессы» заданного КВР). Система может выполнять КВР в «штатном» режиме до наступления контрольного события Z_j , а при наступлении этого события система должна оценить временные параметры процессов по правилам, изложенным в § 4, и лишь после этого принять решение о последующем алгоритме управления процессами.

В целом динамическое уточнение статических прогнозов сводится к следующим процедурам: расчет значений δ по значениям заданного времени $T_{\max}$ и реальных моментов t_j^* наступления контрольных событий; определение вероятности D^* завершения КВР за заданное время $T_{\max}$ по полученным в динамике значениям δ и по рассчитанным в статике значениям функций $F_s^{(j)}(t)$.

ОБОБЩЕНИЕ И ЗАКЛЮЧЕНИЕ

В целом предлагается *принципиально новый подход* к интеллектуальному динамическому управлению параллельными вычислительными процессами при выполнении сложных наборов взаимосвязанных, периодических и независимых задач (со случайными временами их реализации) в режиме реального времени, к анализу и оценке эффективности параллельных управляющих вычислительных систем с точки зрения их способности выполнять возлагаемые на них задачи и функции за заданное время с требуемой вероятностью. Фактически разрабатывается *новая компьютерная технология*, направленная на обеспечение надежного выполнения сложных программных комплексов в управляющих параллельных вычислительных системах реального времени.

Почему это названо новой компьютерной технологией? Потому что:

- это есть комплекс новых принципов, математических моделей, алгоритмических методов и программных средств для априорного — в статике — расчета прогнозов времени выполнения задач со *случайными* временами их реализации, для оперативного отслеживания выполнения сложных наборов задач в динамике, динамического уточнения статических прогнозов, а в конечном счете — для выбора алгоритмов управления процессами в параллельных вычислительных системах;
- изложенная совокупность этапов и процедур обеспечения надежного выполнения КВР и их последовательность *остаются неизменными* для самых разно-

образных КВР, задаваемых пользователем, и различных классов параллельных вычислительных систем;

- технология обеспечивает анализ, оценку и выбор различных методов планирования (диспетчеризации) работ для каждого конкретного КВР; здесь уместно упомянуть, что для одного и того же КВР могут оказаться рациональными различные дисциплины диспетчеризации — в зависимости от конкретных временных параметров работ КВР.

Разумеется, разработка предъявленной технологии далека от завершения. В настоящее время исследуются следующие аспекты ее развития:

— интеллектуальное динамическое управление надежным выполнением сложных наборов и взаимосвязанных, и периодических, и независимых задач в режиме реального времени — на основе анализа и постоянного уточнения статических прогнозов, с учетом реальных временных параметров вычислительных процессов и поведения вычислительных систем, в частности, в условиях сбоев и отказов вычислительных ресурсов;

— априорная оценка «пригодности» параллельной вычислительной системы для выполнения конкретного набора задач и функций, задаваемого пользователем, за заданное время с требуемой вероятностью, даже на ранних этапах проектирования программно-аппаратурных комплексов;

— выбор такой конфигурации вычислительных ресурсов параллельной вычислительной системы, которая достаточна для надежного выполнения требуемого набора задач управления;

— адаптивное — перенастраиваемое в динамике по реальным событиям в системе — резервирование вычислительных структур и процессов.

ЛИТЕРАТУРА

1. Ignatushchenko V. V. A principle of dynamic control of parallel computing processes on the basis of static forecasting // Proc. of the 10-th Int. Conf. on Parallel and Distributed Computing Systems (PDCS-97). New Orleans, USA. — 1997. — P. 593—597.
2. Елисеев В. В., Ларгин В. А., Пивоваров Г. Ю. Программно-технические комплексы АСУ ТП. — Киев: Киевский университет, 2003. — 429 с.
3. Игнатушенко В. В., Клушин Ю. С. Прогнозирование выполнения сложных программных комплексов на параллельных компьютерах: прямое стохастическое моделирование // Автоматика и телемеханика. — 1994. — № 11. — С. 142—157.
4. Игнатушенко В. В., Подшивалова И. Ю. Динамическое управление параллельными вычислительными процессами на основе статического прогнозирования их выполнения // Там же. — 1997. — № 5. — С. 160—173.
5. Игнатушенко В. В., Подшивалова И. Ю. Динамическое управление надежным выполнением параллельных вычислительных процессов для систем реального времени // Там же. — 1999. — № 6. — С. 142—157.
6. Елисеев В. В., Барбан А. П., Подшивалова И. Ю. Об эффективности методов диспетчеризации сложных наборов задач в неоднородных многопроцессорных вычислительных системах // Тр. II междунар. конф. «Параллельные вычисления и задачи управления» РАСО'2004 / Ин-т пробл. упр. — М., 2004. — С. 395—412.
7. Balsamo S., et al. Model-Based Performance Prediction in Software Development: A Survey // IEEE Trans. on Software Eng. — 2004. — Vol. 30. — N 5. — P. 295—310.
8. Arjan J. C. Symbolic Performance Modeling of Parallel Systems // IEEE Trans. Parallel Distrib. Syst. — 2003. — N 14 (2). — P. 154—165.
9. Gelenbe E. A QMIPS Significant Example: Multiprocessor Performance with Task-Graph Models // Centre de Recherche de l'EHEI, UFR de Mathematiques et d'Informatique, Université de Paris V, 2001. <<http://www-sop.inria.fr/mistral/infos/QMIPS/Deliverables>>.
10. Методы математического прогнозирования времени выполнения сложных наборов задач в параллельных вычислительных системах с распределенной структурой / А. М. Еналиев, В. В. Игнатушенко, Е. В. Помазов, Е. А. Случанко // Автоматика и телемеханика. — 2002. — № 10. — С. 154—176.
11. Игнатушенко В. В. Организация структур управляющих многопроцессорных вычислительных систем. — М.: Энергоатомиздат, 1984. — 184 с.
12. Феррари Д. Оценка производительности вычислительных систем. — М.: Мир, 1981. — 576 с.
13. Михайлов А. Ю. Применение аппарата марковских процессов для прогнозирования времени выполнения сложных наборов задач в параллельных вычислительных системах // Тр. XXX Юбил. междунар. конф. «Информационные технологии в науке, образовании, телекоммуникации, бизнесе» (IT+SE'2003). — Украина, Ялта—Гурзуф, май 2003. — С. 39—41.
14. Случанко Е. А. Анализ правил назначения контрольных событий для прогнозирования времени выполнения взаимосвязанных задач в параллельных вычислительных системах // Тр. XXIX междунар. конф. «Информационные технологии в науке, образовании, телекоммуникации, бизнесе» (IT+SE'2002). — Украина, Ялта—Гурзуф, сентябрь 2002. — С. 96—100.

☎ (495) 334-89-29;

e-mail: ignatu@ipu.ru



Вниманию подписчиков!

В регионах страны на журнал "ПРОБЛЕМЫ УПРАВЛЕНИЯ" можно подписаться через филиалы альтернативного агентства ООО "Урал-Пресс".

В городах Екатеринбург, Березовском, В. Пышме, Первоуральске — ЗАО "Урал-Пресс":
620095, Екатеринбург, ул. Малышева, 101-335
☎ (343) 375-80-71, 375-84-93, 375-84-39
☎ (343) 375-62-74, 375-79-20
e-mail: info@ural-press.ru

В г. Нижнем Тагиле — ЗАО "Урал-Пресс",
ул. Красноармейская, 55
☎ (3435) 41-14-48, 41-77-09
e-mail: ntagil@ural-press.ru

В г. Челябинске — ООО "Южно-уральская почта":
454091, Челябинск, ул. Воровского, 63 а
☎ (351) 262-90-03, 262-90-05
e-mail: pochta@chel.surnet.ru

В г. Перми — ООО "Парма-пресс",
ул. Уральская, 116
☎ (3422) 60-24-40, 60-22-95, 60-35-42
e-mail: parma-press@permonline.ru





ОДНОКАСКАДНЫЕ КОММУТАТОРЫ БОЛЬШОЙ РАЗМЕРНОСТИ ДЛЯ МНОГОПРОЦЕССОРНЫХ И МНОГОМАШИННЫХ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ

В. С. Подлазов, В. В. Соколов

Институт проблем управления им В. А. Трапезникова РАН, г. Москва

Рассмотрен метод регулярного построения однокаскадных коммутаторов (плоских коммутируемых сетей) большой размерности для современных суперкомпьютеров.

ВВЕДЕНИЕ

Полные однокаскадные коммутаторы обеспечивают прямое соединение любого входа с любым выходом, поэтому они имеют максимальную пропускную способность и минимальные задержки. Долгое время использование однокаскадных коммутаторов большой размерности (тысячи входов/выходов — В/В) считалось неприемлемым из-за их высокой схемной сложности и большого числа В/В контактов. Развитие микроэлектроники снимает проблему схемной сложности. Остается проблема большого числа В/В контактов. Для ее решения можно сделать однокаскадный коммутатор составным, состоящим из однокаскадных коммутаторов малой размерности, но сохраняющим однокаскадную структуру. Последнее означает, что путь между любыми входом и выходом проходит только через один коммутатор малой размерности. Для подсоединения входов/выходов (В/В) к разным коммутаторам применяются мультиплексоры/демультиплексоры.

Подобный подход был предложен в работах [1, 2] для построения дешевых суперкомпьютеров из персональных компьютеров (ПК) с использованием плоских коммутируемых сетей FastEthernet. В качестве однокаскадных коммутаторов малой размерности применяются типовые сетевые коммутаторы, а каждый ПК для подсоединения к коммутаторам использует одинаковое небольшое число (3–6) сетевых карт. Для выбора размера и числа коммутаторов, числа сетевых карт и для построения схемы подсоединения станций к коммутаторам применялся метод стохастической оптимизации в виде генетических алгоритмов [1, 2]. Недостатки этого подхода — неоптимальность получаемых решений и плохая управляемость параметрами получаемых сетей.

Для преодоления указанного недостатка был предложен регулярный способ построения плоских сетей (ПС) [3]. Он основан на построении простейших ПС (ППС) — ПС минимального размера. Они используются как стро-

ительные блоки для построения ПС большого размера. Так удалось построить ППС для случая 3–4-х сетевых карт и не удалось для случая 5–6-ти сетевых карт [3]. Обсуждение этого вопроса с д-ром техн. наук М. Ф. Караваем показало, что ППС аналогичны сетям SUN (single-unit network) многопроцессорных систем, в которых любая пара процессоров соединяется только через один из нескольких блоков памяти [4]. Параметрами таких сетей выступают числа портов у процессоров и блоков памяти, числа процессоров и блоков памяти. Блоки памяти выполняют роль однокаскадных коммутаторов. Многопортовый процессор является аналогом ПК с несколькими сетевыми картами, а многопортовый блок памяти — аналогом полного коммутатора. Имеется математическая теория, описывающая параметры таких сетей. Это теория неполных уравновешенных блок-схем (УБС), рассматривающая возможности равномерного распределения элементов по блокам [5].

Будем трактовать эту теорию в терминах ППС [3]. Так, ПК является аналогом элемента, а сетевой коммутатор — аналогом блока. Число N ПК — это число различных элементов, а число коммутаторов M — это число блоков. Число портов K в каждом коммутаторе — это размер блока (число размещаемых в нем элементов), а число сетевых карт m — это число блоков, в которые входит каждый элемент. Наконец, число соединений любых пар ПК через один коммутатор λ — это число одинаковых пар элементов во всех блоках.

Теория УБС показывает, что возможны не любые сочетания значений числа портов K и числа сетевых карт m и, как следствие, возможны не любые значения числа N ПК. Кроме того, теория УБС развита для случая, когда число ПК не превосходит числа коммутаторов, т. е. для случая $N \leq M$, что ограничивает наращиваемость ПС. Для преодоления этого недостатка и было предложено собирать большие ПС из ППС [3].

В данной статье предлагаются основы теории построения ПС или однокаскадных коммутаторов большой размерности.

1. ОСНОВНЫЕ СВОЙСТВА ПРОСТЕЙШИХ ПЛОСКИХ СЕТЕЙ

В книге [5] дается следующее определение УБС: «Уравновешенной неполной блок-схемой $B(N, M, K, m, \lambda)$ называется такое размещение N различных элементов по M блокам, что каждый блок содержит точно K различных элементов, каждый элемент появляется точно в m различных блоках и каждая пара различных элементов появляется точно в λ блоках».

Теория УБС [5] устанавливает, что необходимым условием их существования является выполнение следующих простейших условий:

$$Nm = KM, \quad (1)$$

$$m(K - 1) = \lambda(N - 1). \quad (2)$$

Эта теория развита для случая, когда соблюдается условие Фишера [5]: $N \leq M$.

Нас в первую очередь интересуют УБС с $\lambda = 1$, так как они имеют естественную интерпретацию в качестве нерезервированных ППС. Будем трактовать вхождение элемента с номером $I \in [0, N - 1]$ в блок с номером $S \in [0, M - 1]$ как подключение ПК с номером I к коммутатору с номером S через одну из m сетевых карт. При такой трактовке любой ПК имеет соединение с любым другим ПК всегда через один коммутатор, используя m разных коммутаторов. Это означает, что все N ПК объединены плоской сетью, которая названа простейшей (ППС) [3]. Будем задавать ее парой K и m , т. е. как ППС (K, m) .

Симметричные УБС с $K = m$ и соответствующие им ППС (m, m) имеют наибольшее значение $N = M$ при минимальном значении M . Для существования любой

ППС (m, m) необходимо и достаточно, чтобы величина $(m - 1)$ была целой положительной степенью простого числа и выполнялись условия (1) и (2) [4, 5].

Возможно существование и несимметричных УБС и ППС с $K < m$ и $N < M$. Различные ППС представлены в табл. 1 [4, 5]. В клетках таблицы дается значение N , а значение M вычисляется по формуле (1). В табл. 1 входят и полные коммутаторы ППС $(N, 1)$ — 1-я строка, и звездные сети ППС $(1, N)$ — 1-й столбец, и симметричные ППС (m, m) — на главной диагонали. Выделим также остаточные ППС $(m - 1, m)$ — под главной диагональю, которые автоматически получаются из ППС (m, m) [5].

Обратим внимание, что все ППС в табл. 1 расположены слева от главной диагонали. Возникает вопрос о существовании ППС, расположенных справа от главной диагонали. В них путем увеличения размера коммутатора K можно было бы увеличить и число N подключенных ПК, т. е. обеспечить наращиваемость ПС.

Очевидно, что аналогом таких ППС могли бы быть блок-схемы, в которых условие Фишера нарушено и которые не являются УБС. Их можно получить из УБС с $\lambda > 1$. Пусть имеется УБС $B_1(N, M, K, m, 1)$. Образует из нее УБС $B_\lambda(N, M, \lambda K, \lambda m, \lambda)$ путем тривиального объединения λ блок-схем B_1 . При таком объединении каждый блок для B_λ получается объединением λ одноименных блоков из B_1 .

Аналогично, можно объединить λ блок-схем B_1 , имеющих *разные* множества элементов, которые совпадают в классе вычетов по $\text{mod } N$. В результате получим некоторую блок-схему с $\tilde{B}_\lambda = (\lambda N, M, \lambda K, m)$. В случае симметричной блок-схемы B_1 , представленной в разностной форме, блок-схемы B_λ и $\tilde{B}_\lambda$ изоморфны, так как элементы в них представлены вычетами по $\text{mod } N$ [5].

Определим ППС λ -го ранга — ППС (K, m, λ) , ППС $(K, m, 1)$ совпадает с ППС (K, m) . Пусть в ней к коммутатору с произвольным номером $0 \leq S < M$ подключены ПК с номерами $u_{S,1}, u_{S,2}, \dots, u_{S,K}$. Тогда в ППС $(2K, m, 2)$ к коммутатору с тем же номером дополнительно подключены ПК с номерами $u_{S,1} + N, u_{S,2} + N, \dots, u_{S,K} + N$. Вследствие различия номеров всех дополнительных ПК и постоянства числа коммутаторов построенная таким образом сеть наследует простоту и плоскость от ППС (K, m) .

Аналогично, в ППС $(\lambda K, m, \lambda)$ к коммутатору с номером S дополнительно по сравнению с ППС $((\lambda - 1)K, m, \lambda - 1)$ подключены ПК с номерами $u_{S,1} + (\lambda - 1)N, u_{S,2} + (\lambda - 1)N, \dots, u_{S,K} + (\lambda - 1)N$. Аналогично, построенная таким образом сеть любого ранга $\lambda > 1$ наследует простоту и плоскость от ППС $(\lambda - 1)$ -го ранга.

Простейшая ПС λ -го ранга является ПС, что прямо следует из изоморфности B_λ и $\tilde{B}_\lambda$. Однако следует сказать, что само это свойство было обнаружено и проверено экспериментально — с помощью компьютерной программы, которая по таблицам соединений подсчитывает число и состав соединенных ПК [3]. С помощью

Таблица 1

Начальные члены различных семейств ППС (K, m)

m	K												
	2	3	4	5	6	7	8	9	10	11	12	13	
1	2	3	4	5	6	7	8	9	10	11	12	13	
2	3	—	—	—	—	—	—	—	—	—	—	—	
3	4	7											
4	5	9	13										
5	6	—	16	21									
6	7	13	—	25	31								
7	8	15											
8	9	—	25	—	—	49	57					—	
9	10	19	28				64	73					
10	11	21	—	41				81	91				
11	12	—		45			—						
12	13	25	37	—							121	133	



Таблица 2

Начальные члены семейств ППС $(\lambda m, m, \lambda)$ и ППС $(\lambda(m-1), m, \lambda)$

m	K																		
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16				
3	4	7	8	—	14, 12	—	16	21	20	—	28, 24	—	28	35	32				
4	—	9	13		18		26	27	—		39, 36		45	52					
5		—	16	21	—		32	—	42		48		63	64					
6			—	—	25		31		—		50		62	75	—				
7					—		—		—		—		—	—	—	—			
8								49									57	94	104
9								64									73	128	
10								—		—	81		91						
11											—		—						
12												121	133						

этой программы исследовались и все другие ПС, рассматриваемые в данной статье.

По построению ППС $(\lambda K, m, \lambda)$ объединяет в λ раз больше ПК, чем сеть ППС (K, m) . Именно симметричные ППС $(\lambda m, m, \lambda)$ были предложены для построения ПС большого размера [3].

В результате проведенного построения ППС $(\lambda K, m, \lambda)$ табл. 1 может быть существенно пополнена. Пусть ячейка на пересечении I -й строки и J -го столбца содержит значение X . Тогда в этой строке на пересечении с λJ -м столбцом появится значение λX . В результате образуется многослойная таблица, задающая число ПК в ППС любого ранга $\lambda \geq 1$. Табл. 2 дает пример такого заполнения для симметричных ППС и остаточных ППС, выделенных жирным шрифтом.

2. ОСНОВНЫЕ ПРОСТЕЙШИЕ ПЛОСКИЕ СЕТИ

Тривиальная симметричная ППС (2, 2) представлена в табл. 3. В ней 0-й столбец задает номер коммутатора, а 0-я строка — номер порта в коммутаторе. В остальных ячейках таблицы задаются номера ПК, подсоединенных к соответствующим портам.

Табл. 3 задает ППС в разностной форме [5], в которой i -й столбец представляет собой циклически сдвинутый вниз 0-й столбец, задаваемый последовательностью 0, 1, ..., $N-1$. Поэтому для описания ППС достаточно задавать положение только 0-го ПК в каждом столбце. В табл. 4 в таком кратком виде представлены симметричные ППС (m, m) с $3 \leq m \leq 6$, содержащие $N = m(m-1) + 1$ ПК.

При $m > 2$ ППС (m, m) можно представить еще одним удобным способом — на базе m квадратов $(m-1) \times (m-1)$ с различными строками (нет строк с одинаковыми парами номеров ПК). Они представлены в табл. 5. Верхние квадраты выделены жирным шрифтом.

3. ПЛОСКИЕ СЕТИ БОЛЬШИХ РАЗМЕРОВ

Построение ПС больших размеров основано на использовании симметричных ППС больших рангов. Однако здесь появляется практическая задача размещения таких ПС в стандартных сетевых коммутаторах FastEthernet с фиксированным числом портов. На рынке имеются сетевые коммутаторы со следующим рядом числа портов: 16, 24, 32, 48, 64 и 96. На их основе можно построить целый спектр многограновых ППС с числом ПК от нескольких десятков до нескольких сотен.

Однако большинство из этих ПС все еще не подходят для создания ПС FastEthernet. Последние не могут функционировать без использования широкополосных

Таблица 3

ППС (2,2) для 3 ПК

S	P	
	0	1
0	0	2
1	1	0
2	2	1

Таблица 4

Краткий вид ППС (m, m) в разностной форме

m	N	ППС (m, m)					
2	3	0	1	—			
3	7			3	—		
4	13				9	—	
5	21				4	14	16
6	31			8	11	13	17

Таблица 5

ППС (m, m) в форме квадратов для $3 \leq m \leq 5$

m	3			4				5				
S	P			P				P				
	0	1	3	0	1	3	4	0	1	3	4	5
0	4	5	6	9	10	11	12	16	17	18	19	20
1		0	1		0	1	2		0	1	2	3
2		2	3		3	4	5		4	5	6	7
3	5	0	2	10	6	7	8	17	8	9	10	11
4		1	3		0	3	6		12	13	14	15
5	6	0	3	11	1	4	7	18	0	4	8	12
6		1			2	5	8		1	5	9	13
7	—	—	—	12	0	4	6	19	2	6	10	14
8					1	5	6		3	7	11	15
9					2	3	7		0	5	11	14
10				13	0	5	8	20	1	4	10	15
11					1	3	8		2	7	9	12
12					2	4	6		3	6	8	13
14				14	—	—	—	21	0	7	10	12
15									1	6	11	12
16									2	5	8	15
17									3	4	9	14
18									0	6	11	15
19									1	7	8	14
20									2	4	11	13
21									3	5	10	12

ных кадров. Для передачи последних необходим хотя бы один сетевой коммутатор 2-го уровня, объединяющий все коммутаторы 1-го уровня, составляющие ППС. Это означает, что в каждом коммутаторе 1-го уровня должен быть свободен хотя бы один порт для соединения с коммутаторами 2-го уровня.

Пусть имеется только один коммутатор 2-го уровня. В этом случае нам надо строить ПС, имеющие один свободный порт в каждом коммутаторе 1-го уровня. Такие ПС можно строить путем освобождения портов в коммутаторах в результате рациональной процедуры уменьшения числа ПК. Одна такая процедура была предложена для ППС в разностной форме [3]. Можно предложить и более простую процедуру.

Она основывается на редукции ППС (m, m) в форме квадратов к остаточной ППС ($m - 1, m$), при которой освобождается один порт во всех коммутаторах кроме 0-го, а в нем освобождаются все порты. Эта редукция представлена в табл. 6 для $m = 4$, где удаляемые ПК выделены жирным шрифтом.

Пусть в некоторой ППС λ -го ранга необходимо освободить $1 \leq l < \lambda$ портов в каждом коммутаторе. Тогда достаточно взять ППС ($\lambda - l$)-го ранга и нарастить ее до λ -го ранга, используя при каждом повышении ранга ППС ($m - 1, m$) вместо ППС (m, m). При этом будет накоплено $L = lm$ незадействованных портов в 0-м коммутаторе. Их нельзя развести по разным коммутаторам благодаря циклическим сдвигам добавляемых ППС ($m - 1, m$), так как при этом могут исчезнуть соединения некоторых ПК. Построенная таким образом ПС объединяет $N = M(\lambda - l) + (M - m)l = M\lambda - L$ ПК. Табл. 7 дает параметры ПС, построенных по указанной процедуре.

Таблица 6

ППС (4,4) (слева) и ППС (3,4) (справа)

S	P				S	P		
	0	1	3	4		0	1	3
0	9	10	11	12	—			
1		0	1	2	0	0	1	2
2		3	4	5	1	3	4	5
3		6	7	8	2	6	7	8
4	10	0	3	6	3	0	3	6
5		1	4	7	4	1	4	7
6		2	5	8	5	2	5	8
7	11	0	4	7	6	0	4	7
8		1	5	6	7	1	5	6
9		2	3	7	8	2	3	7
10	12	0	5	8	9	0	5	8
11		1	3	8	10	1	3	8
12		2	4	6	11	2	4	6

Таблица 7

Реальные плоские сети

m	λ	K	N	M	L	m	λ	K	N	M	L
3	5	15	35	7	0	5	3	15	63	21	0
	8	23	53		3		5	23	95		10
	11	31	71		6		7	31	127		20
	16	47	109		3		10	48	200		10
	21	63	147		0		13	63	283		10
	32	95	221		3		19	95	399		0
4	4	15	48	13	4	6	3	15	75	31	18
	6	23	74				4	23	118		12
	8	31	100				6	31	156		30
	12	47	152				8	47	242		6
	16	63	204				11	63	323		18
	24	95	308				16	95	490		6

Дальнейшее увеличение числа ПК, объединяемых ПС, можно получить путем увеличения числа сетевых карт m и числа портов в коммутаторах K . На этом пути возникают две трудности. Первая трудность является принципиальной — не для всех m существуют ППС (см. табл. 1). Вторая трудность носит технологический характер — в настоящее время отсутствуют ПК и сетевые коммутаторы с большими значениями m и K . Обойти эти трудности можно созданием ПС из нестандартных сетевых коммутаторов и/или благодаря разработке специализированных однокаскадных коммутаторов

4. РАСПРЕДЕЛЕННЫЕ ОДНОКАСКАДНЫЕ КОММУТАТОРЫ

Распределенный однокаскадный коммутатор был предложен в работе [6] как мультикольцевой коммутатор — однокаскадный мультикольцо.

Мультикольцом мы называем набор из $m \geq 2$ колец различной топологии (последовательности соединения узлов), задаваемой следующим образом. Предположим, что узлы перенумерованы целыми числами из $[0, N-1]$. Пусть номера соседних узлов вдоль направления передачи в кольце задаются периодической последовательностью $X_i \in [0, N-1]$ ($i = 0, 1, \dots$). Рассматриваются кольца с последовательностями вида $X_{i+1} = (X_i + S) \bmod N$, где S называется шагом кольца и $1 \leq S \leq N-1$, $0 \leq X_0 \leq S-1$.

Кольцо с положительным шагом $S \geq N/2$ будем также называть встречным кольцом с отрицательным шагом $-(N-S)$. Здесь используется сравнимость по $\bmod N$: $S \equiv -(N-S)$. При $X_0 = 0$ и $S = 1$ получаем традиционное кольцо (с шагом 1), а при $X_0 = 0$ и $S = N-1$ — встречное кольцо (с шагом -1). В дальнейшем кольцо с шагом S будем называть кольцом S , а его дугу — дугой S .

Различаются два вида колец — обычные и расщепленные. Обычное кольцо имеет взаимно простые N и S , а последовательность X_i имеет период N и содержит по одному разу все номера из $[0, N-1]$. В расщепленном кольце N и S имеют наибольший общий делитель d , а последовательность X_i разделяется на d непересекающихся последовательностей jX_i с периодом N/d ($0 \leq j \leq (d-1)$, $^jX_0 = j$), которые в совокупности содержат все номера из $[0, N-1]$. Физически расщепленное кольцо состоит из d миниколец по N/d узлов в каждом.

Мультикольцом $\{R_n\} = (S_1, S_2, \dots, S_n)$ называется набор из $2 \leq n \leq (N-1)$ различных колец $S_1 = 1, S_2, \dots, S_n$, где $S_{i+1} > S_i$. Рис. 1 дает пример мультикольца с 9-ю узлами и 4-мя кольцами — два кольца с шагами 1 и $8 \equiv -1 \pmod{9}$ и два расщепленных кольца с шагами 3 и $6 \equiv -3 \pmod{9}$.

Мультикольцо $\{R_n\}$ называется полным p -ичным мультикольцом, если $N = p^k$ и $n = k(p-1)$, а $S_{jp+i} = ip^j$ ($0 \leq j \leq k-1$, $1 \leq i \leq p-1$). Набор колец $\{R_n\}$ задается всеми цифрами разрядов в p -ичной системе счисления.

Однокаскадный мультикольцо (ОМ) — это полное p -ичное мультикольцо с $p = \sqrt{N}$. Оно имеет $n = 2(p-1)$

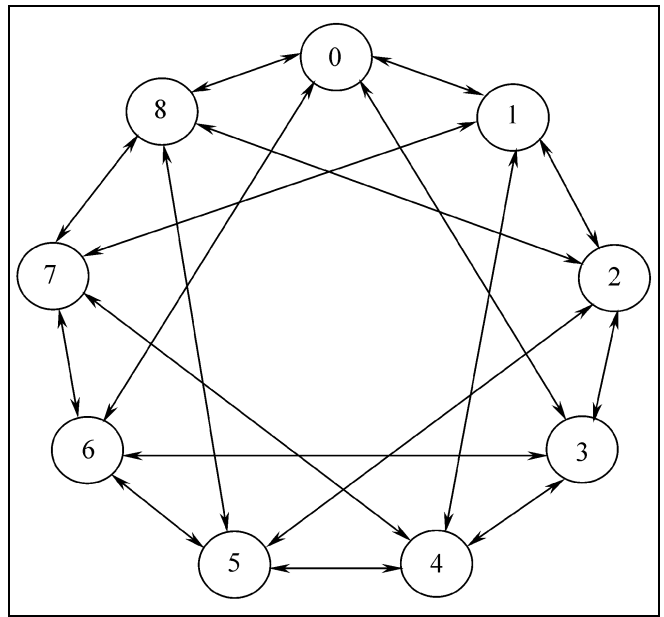


Рис. 1. Мультикольцо с $N = 9$ и $\{R_4\} = (1, 3, 6, 8) \equiv (1, 3, -3, -1) = (\pm 1, \pm 3)$

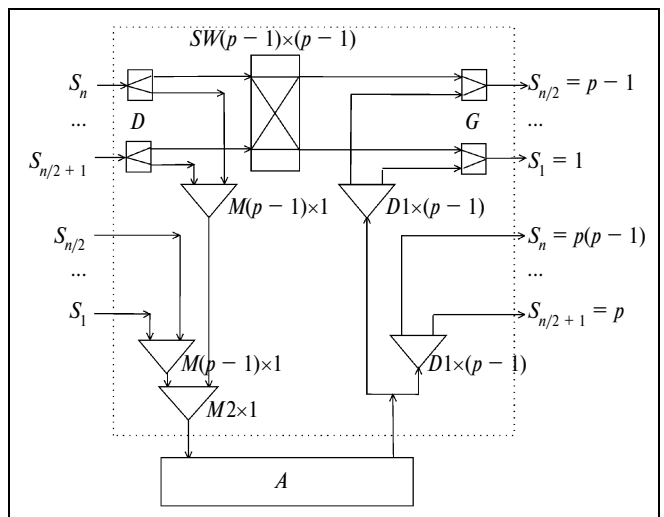


Рис. 2. Функциональная схема узла для канальной коммутации

колец с шагами $\{R_n\} = (1, 2, \dots, p-1, p, 2p, \dots, (p-1)p)$. В однокаскадном мультикольце путь между любыми двумя абонентами проходит не более чем через один промежуточный узел. С точки зрения теории коммутации ОМ является распределенным самомаршрутизируемым однокаскадным неблокируемым коммутатором и позволяет реализовать без конфликтов на дугах любую перестановку данных между абонентами.

Рассмотрим структуру узла ОМ. Она представлена на рис. 2. В ней SW означает однокаскадный коммутатор, M — входной мультиплексор, D — выходной демультиплексор и G — выходной клапан, который выполняет функцию разрешения конфликта по передаче между

коммутатором SW и абонентом A . В качестве абонента может выступать ПК, процессор или модуль памяти.

Схема на рис. 2 предназначена для работы в режиме коммутации каналов или по методу «червоточины», которые в данном случае совпадают [10]. Сначала выполняется канальное соединение абонентов отправления и назначения при передаче адресов последних. Адреса передаются в виде двух номеров выходных линий. При этом первыми передаются большие номера. В процессе соединения клапан G отдает приоритет соединению с коммутатором SW . В перестановочной модели передачи, когда каждый пакет адресуется только одному приемнику, конфликтов на клапанах G не возникает. В сетевой модели, когда имеет место произвольная адресация приемников, конфликты возможны, но они приводят к блокировкам только тех пакетов, который находятся в выходных буферах абонентов. Поэтому коммутаторы SW не имеют внутренних буферов.

Таким образом, ОМ в описанном виде представляет собой специализированную сеть, аналогичную сети Myrinet [8, 9], применяемой во многих многопроцессорных вычислительных системах. Обе эти сети имеют специализированную сетевую карту и специализированные коммутационные схемы (схемы узлов для ОМ). При этом ОМ является неблокируемой сетью, тогда как Myrinet — только перестраиваемой сетью, маршрутизируемую по методу «червоточины». Конструктивно Myrinet набирается из микросхем, представляющих собой однокаскадные коммутаторы 16×16 . Она представляет собой 3-х или 5-каскадную перестраиваемую сеть Клоза на 128 и 512 абонентов, соответственно, ОМ при его реализации на микросхемах 16×16 и 32×32 (по две на узел) объединяет 256 и 1024 абонентов со значительно меньшими задержками, чем в Myrinet.

Однотактное мультикольцо является распределенным однокаскадным коммутатором. Его можно использовать вместо сетевого коммутатора с большим числом портов при создании распределенных плоских сетей. Рассмотрим это на примере ОМ, объединяющего N ПК (пусть для определенности $N = 256$). Это означает, что каждый распределенный коммутатор имеет $K = 256$ портов. Пусть каждый ПК имеет m сетевых карт ($m = 5$). По одной карте ПК подключается к одному ОМ. Теперь создадим ППС λ -го ранга $\lambda = (K - 1)/m$ ($\lambda = 51$), используя $M = m(m - 1) + 1$ ОМ ($M = 21$). В результате такая плоская сеть соединит $\tilde{N} = M\lambda$ ПК ($\tilde{N} = 1071$). При $N = 1024$ получим уже $\lambda = 205$ и $\tilde{N} = 205 \cdot 21 - 5 = 4300$. Это в несколько раз больше, чем в сосредоточенных плоских сетях, построенных из обычных коммутаторов.

Таким образом, применение ОМ открывает возможность построения распределенных плоских сетей больших размеров, причем последние являются сетями канальной коммутации с очень малыми задержками.

ЗАКЛЮЧЕНИЕ

Развитие микроэлектроники сделало возможной разработку однокаскадных коммутаторов на большое число

входов/выходов. В настоящее время развитие однокаскадных коммутаторов происходит, в частности, в виде плоских сетей пакетной коммутации, собираемых из стандартных сетевых коммутаторов FastEthernet. Эти сети позволяют объединить сотни персональных компьютеров. В настоящей работе развит регулярный метод построения больших плоских сетей из простейших плоских сетей. Последние сети оказались изоморфны уравновешенным симметричным блок-схемам, хорошо исследованным в комбинаторике. Рассмотрены также распределенные однокаскадные коммутаторы в форме однотактного мультикольца, позволяющего создавать плоские сети, которые объединяют несколько тысяч персональных компьютеров.

Применение коммутаторов FastEthernet является простейшим способом построения однокаскадных коммутаторов для сборки дешевых суперкомпьютеров из персональных компьютеров. При этом таблицы продвижения и фильтрации кадров в сетевых коммутаторах требуют специальной статической настройки [1, 2]. Это принципиально отличает режим использования плоских сетей FastEthernet от режима функционирования обычных древовидных сетей FastEthernet, в котором упомянутые таблицы строятся автоматически. Поэтому для больших плоских сетей желательно иметь процедуру автоматического или автоматизированного построения и занесения этих таблиц в сетевые коммутаторы.

ЛИТЕРАТУРА

1. Deitz H. G. and Mattox T. I. KLAT2's flat neighborhood network <<http://aggregate.org/FNN/>>.
2. <<http://aggregate.org/KASY0/>>.
3. Подлазов В. С. Регулярные плоские сети для суперкомпьютеров // Проблемы управления. — 2006. — № 1. — С. 26—31.
4. Finn A. M., Decker R. O. A network architecture for radar signal processing // AIAA/IEEE 8-th digital avionic systems conference. — San Jose, Calif., Oct. 1988. — P. 614—621.
5. Холл М. Комбинаторика. М.: Мир, — 1970. Гл. 10.
6. Алленов А. В., Подлазов В. С. Пропускная способность набора кольцевых каналов П. Кольцевые коммутаторы // Автоматика и телемеханика. — 1996. — № 4. — С. 162—172.
7. Подлазов В. С. Распределенные коммутаторы со статическими расписаниями для многопроцессорных вычислительных систем: Дисс. ... д-ра техн. наук Институт проблем управления РАН. — 2004. — Гл. 3.1.
8. Guide to myrinet-2000 switches and switch networks. <<http://www.myl.com/myrinet/m3switch/guide/>>.
9. Корнеев В. В. Вычислительные системы. — М.: Гелиос АРВ, 2004. — 512 с.
10. Ni L. M., McKinley P. K. A survey of wormhole routing techniques in direct networks // IEEE Computers. — 1993. — Vol. 26. — N 2. — P. 62—73.

☎ (495) 334-78-31;
e-mail: podlazov@ipu.ru



ПРОБЛЕМЫ УПРАВЛЕНИЯ ИНФЛЯЦИЕЙ: СОВРЕМЕННЫЕ ПОДХОДЫ

Р. М. Нижегородцев

Институт проблем управления РАН им. В. А. Трапезникова, г. Москва

Изложены ключевые проблемы управления современными инфляционными процессами. Обсуждены основные направления обуздания инфляции. Показана несостоятельность монетаристских методов преодоления инфляции, развеивается миф об инфляционном перегреве современной российской экономики. Даны рекомендации по снижению инфляции в экономике России, основанные на технологической модернизации.

ВВЕДЕНИЕ

Банальная и часто повторяемая мысль о том, что с инфляцией нужно бороться, на самом деле не выдерживает научной критики. Умеренная, контролируемая правительством инфляция часто *сопровождает* экономический рост, который, тем не менее, не выступает ее следствием. Однако неумелые или несвоевременные шаги правительства по обузданию инфляционных процессов способны, подавляя совокупный спрос, существенно затормозить развитие экономики и снизить темпы экономического роста, а иногда и ввергнуть макроэкономическую систему в состояние стагнации.

Необходимо также учитывать, что в любой экономической системе действуют факторы инфляции, поэтому ее никогда невозможно до конца победить. Реальной целью может быть лишь снижение темпов инфляции. Поэтому прежде чем бороться с инфляцией, необходимо задаться вопросом, для чего это необходимо, достижению каких макроэкономических целей препятствует наличие инфляции или ее слишком высокие темпы.

Инфляция — это стихия, и идея борьбы с инфляцией столь же абстрактна, как идея борьбы с водой или с огнем. Бороться нужно с наводнением и пожаром, т. е. со стихией, вышедшей из-под контроля, и лишь постольку, поскольку их последствия разрушительны и опасны. Хотя не вовремя зажженная спичка может стать источником серьезного пожара, рецепт противодействия заключается не в том, чтобы запретить продажу и применение спичек, а в соблюдении правил пожарной безопасности. Инфляцией, как и любой стихией, нужно *управлять*, т. е. удерживать ее в определенных допустимых пределах, количественные значения которых определяются спецификой экономической конъюнктуры данной макросистемы, и направлять ее энергию в конструктивное

русло, умело противодействуя ей в тех случаях, когда эта энергия препятствует достижению важнейших макроэкономических целей.

1. МИФЫ И ЗАБЛУЖДЕНИЯ АНТИИНФЛЯЦИОННОЙ ПОЛИТИКИ

Один из дискуссионных вопросов современной макроэкономической политики государства состоит в разработке комплекса мер, совокупность которых позволила бы обуздать инфляционные процессы. Трудность данной задачи заключается в том, что инфляция порождает множество экономических процессов с положительной обратной связью, когда результат какого-либо процесса выступает в то же время предпосылкой его повторения на новом уровне. В качестве примеров укажем на три закономерности:

— спираль «зарплата—цены», рост которых взаимно обусловлен;

— взаимная связь между ростом цен и ростом объемов денежной массы;

— прямая взаимозависимость между спадом физических объемов производства и развертыванием инфляционных процессов, последовательно усугубляющих друг друга (так называемая инфляционная ловушка).

В макроэкономической системе, подверженной инфляции, возникает множество проблем, порождающих друг друга подобно замкнутому кругу, и задача правительства заключается в том, чтобы обнаружить «узкие» места, в которых этот круг может быть разорван.

Рекомендации Международного валютного фонда и других международных финансовых организаций по обеспечению финансовой стабилизации сводятся к следующему комплексу мер:

— жесткий контроль над объемом денежной массы, ее сжатие до минимально возможных пределов;

— обеспечение (насколько это возможно) стабильности валютного курса и привязка эмиссии любых долговых обязательств государства к объему валютных резервов, находящихся в распоряжении правительства (эта логика валютного регулирования известна под названием *currency board*);

— последовательное снижение, а затем ликвидация бюджетного дефицита.

Заметим, что данная логика преодоления инфляции исходит из предположения, будто инфляционные процессы порождены преимущественно монетарными эффектами, и их можно нейтрализовать, приведя в действие главным образом монетарные рычаги влияния на макроэкономическую систему. Эта логика губительна и абсурдна, и ее применение ни в одной стране мира не позволило добиться желаемых результатов.

В первую очередь вызывает возражения требование сжатия объема денежной массы с целью обуздания инфляционных процессов. Если резкое увеличение денежной массы провоцирует инфляцию, то из этого не следует, что сжатие денежной массы является рецептом эффективной борьбы с инфляцией. Если открытие шлюзов на корабле приводит к его погружению под воду, то отсюда не вытекает, что, закрыв шлюзы, можно обеспечить его всплытие. Для подъема на поверхность необходимы иные алгоритмы, иная логика действий, для реализации которых положение шлюзов уже не имеет решающего значения.

Единственным объяснением того, почему сжатие объема денежной массы должно приводить к подавлению инфляции, является формула Ирвинга Фишера: $MV = PY$, где M — объем денежной массы, V — средняя скорость ее обращения, P — общий уровень цен, Y — валовой выпуск. Тем не менее из нее не вытекает, почему, сокращая объем денежной массы M , можно снизить величину P . Практика показывает, что вследствие данной логики действий будет падать лишь валовой выпуск Y . Наоборот, искусственное сокращение объема денежной массы вызовет рост коммерческой процентной ставки, что приведет к росту издержек предприятий. Поэтому вместо борьбы с инфляцией правительство спровоцирует стандартную инфляцию издержек.

В результате сжатия объема денежной массы и неконтролируемого роста коммерческой процентной ставки правительство фактически переходит к политике дорогих денег, не объявляя об этом прямо. При этом в экономике растет исходная коммерческая процентная ставка (*prime rate*), а ставка рефинансирования перестает быть регулятором экономики и плетется в хвосте: вместо того, чтобы «поджимать» ее снизу, она вынуждена ее догонять запоздалыми и совершенно нелогичными прыжками, как это было, в частности, в России в середине 1990-х гг. Таким образом, правительство отнюдь не добивается объявленных целей подавления инфляции, а только утрачивает важнейший рычаг контроля над монетарной системой страны. Статистический анализ кризисной и депрессивной российской экономики блестяще подтверждает сделанный вывод [1].

Статистические исследования временных рядов, выражающих годовые макроэкономические показатели 28 стран с переходной экономикой в 1990-х гг., показали, что положительная корреляция между темпами инфляции и совокупным объемом денежной массы $M2$ наблюдается только на Украине и в Румынии, а отрица-

тельная — в Венгрии, Словении, Чехии, Словакии, Молдавии, Азербайджане, Казахстане, Эстонии, Литве, Китае и Вьетнаме, а также в Латвии и России. При этом устойчивую отрицательную корреляцию между темпами инфляции и отношением денежной массы $M2$ к ВВП демонстрируют Польша, Словения и Китай (в этих странах она составляет более 80 %), а также Румыния и Хорватия [2].

Исследования, проведенные экспертами Всемирного Банка, также показали наличие вполне определенной *обратной* корреляции между степенью монетизации экономики (денежная масса $M2$ в процентном отношении к ВВП) и темпами инфляции: чем меньше денег в обращении, тем выше темпы инфляции [3]. Этот вывод, противоречащий монетаристской логике, неудивителен и, более того, легко объясним, если принять во внимание упомянутый выше эффект процентной ставки.

Другое объяснение этого факта (и серьезная проблема для макроэкономической политики) заключается в том, что правительство реально может контролировать (разумеется, до известных пределов) лишь объем денежного агрегата $M2$, тогда как обращение менее ликвидных агрегатов с гораздо большим трудом поддается управлению. Поэтому жесткое ограничение массы $M2$ реально вызывает не снижение объема денежной массы, а лишь ухудшение ее качества: взамен относительно ликвидных и достаточно легко учитываемых денежных ресурсов начинается обращение разного рода финансовых неликвидов (векселей, сертификатов, иных долговых обязательств).

Это лишь усугубляет проблему наполнения бюджета, поскольку «плохие» деньги, с которых можно не платить налоги, вытесняют из обращения «хорошие» деньги: экономические агенты стремятся использовать в обороте финансовые суррогаты именно потому, что их обращение непосредственно не контролируется государством. Очевидно, по иронии эксперты МВФ называют предлагаемый ими комплекс мер обретением *контроля* над объемом денежной массы, тогда как на самом деле ее реальный объем в результате применения их рекомендаций практически полностью выходит из-под контроля правительства.

Еще одним последствием жесткой монетарной политики становится бартеризация экономики — обращение товаров без участия каких бы то ни было денежных средств. Искусственно спровоцированный правительством кризис неплатежей впоследствии требует проведения так называемых взаимозачетов, когда правительство время от времени «объявляет» ликвидными деньгами разные составляющие агрегата D , и экономические агенты, которые должны друг другу определенные суммы денег, продают посредникам (банкам или финансовым фирмам, осуществляющим «расшивку» неплатежей) свои безликие долговые обязательства.

Было бы глубокой ошибкой полагать, будто кризис неплатежей является исключительно расстройством финансовой сферы. Его воздействие на развитие реального сектора ничуть не менее губительно. Подобно тому, как в здоровой, растущей экономике имеет место так называемый мультипликатор спроса (один рубль, выплаченный из бюджета государства, оборачивается несколькими рублями инвестиций, вложенных в разные отрасли хозяйства, по цепочке предъявляющие спрос на продукцию друг друга), точно так же в кризисной экономике



включается мультипликатор неплатежей. В середине 1990-х гг. в России 1 руб. внутреннего государственного долга оборачивался потерей 5–6 руб. в совокупном объеме ВВП [4].

2. НАПРАВЛЕНИЯ АНТИИНФЛЯЦИОННОЙ ПОЛИТИКИ

Поскольку инфляционные процессы в современных макроэкономических системах, как правило, имеют много причин, для их преодоления недостаточно действовать в каком-либо одном направлении. Перечислим ряд важнейших мер, которые необходимо принимать в первую очередь.

Обеспечение роста товарного покрытия денежной единицы. Некоторые учебники экономики настаивают на том, что в период высокой инфляции правительство должно принимать меры по организации так называемых общественных работ. В основе этого мнения лежит наивное убеждение, будто основная проблема заключается в том, как бы увеличить товарное покрытие, но не заплатить за это денег. На самом деле возможности правительства здесь крайне невелики, поскольку в период инфляции и без того обостряются социальные проблемы и падает уровень жизни.

Рост товарного покрытия, разумеется, следует обеспечивать с помощью включения мультипликатора инвестиций. От государства требуется избирательная поддержка отраслей, которые создадут максимальный совокупный спрос и тем самым выступят в роли «локомотива», вытягивающего из трясины всю макроэкономическую систему. Сюда же относится предпринимаемое правительством избирательное финансирование ряда долгосрочных и масштабных проектов, поддержка которых в период инфляции неминуемо выпадает из поля зрения частных экономических интересов и потому становится преимущественно заботой государства.

Ужесточение принципов кредитования и финансирования. В период инфляции следует, насколько возможно, снизить объем финансирования заведомо убыточных проектов и отраслей, поддерживая их лишь в той мере, в какой этого требуют интересы социальной и экономической безопасности страны. Проблема «плохих» (невозвращаемых) кредитов, которая всегда обостряется в период инфляции, требует от государства наведения порядка в сфере лицензирования финансовых и банковских операций и создания адекватной системы взаимного страхования и взаимного гарантирования инвестиций. Правительство в этот период не должно допускать строительства краткосрочных, спекулятивных финансовых пирамид — как частных, так и государственных.

Девальвация национальной валюты. Контролируемое правительством снижение официального курса национальной валюты, сближающее его с реальным соотношением спроса и предложения, в известной мере уменьшает некоторые негативные последствия высокой инфляции, например, утечку капитала из страны. Логика проведения девальвации национальной валюты во многом зависит от того, хочет ли правительство экономически поддержать экспортно-ориентированные производства или, напротив, импортозамещающие отрасли, развитие которых ослабляет зависимость страны от импорта. Заниженный курс национальной валюты выгоден фирмам-экспортерам, которые покупают факторы про-

изводства на внутреннем рынке, а готовую продукцию продают за рубежом; покупатели импортных товаров, напротив, в этой ситуации проигрывают. Снижение курса национальной валюты также уменьшает возможности погашения кредитов резидентами, имеющими долговые обязательства в иностранной валюте.

Одна из возможностей грамотного проведения девальвации национальной валюты в эпоху «золотого стандарта» заключалась во введении «параллельной» валюты, обеспеченной золотовалютными резервами страны. Так, в СССР, в 1922–1924 гг. была успешно проведена денежная реформа, сопровождавшаяся введением «золотого» червонца. Тем самым государство перестает поддерживать неликвидный денежный агрегат, предпринимая бегство от собственной валюты в другую более ликвидную. В принципе возможно введение параллельной валюты, «привязанной» к любому ликвидному активу (не обязательно к золоту), но источник этого актива должен находиться в руках государства.

Снижение бюджетного дефицита. В период инфляции правительство вынуждено сокращать финансирование практически всех программ, требующих затрат совокупного прибавочного продукта общества, — науки, образования, здравоохранения, транспорта, жилищно-коммунального хозяйства. Осуществляется переход к предостановлению соответствующих услуг на полноплатной основе, сокращающей до минимума необходимость бюджетных выплат на соответствующие цели. К мерам по сокращению бюджетного дефицита также относятся эмиссия долгосрочных государственных обязательств (внутренний заем), конверсия военного производства, уменьшение социальных льгот. Однако возможности «наступления» правительства на социальную сферу не слишком велики ввиду обострения социальных проблем, которое всегда имеет место в период инфляции.

С целью пополнения бюджета не следует также повышать общий уровень налогов на производственную деятельность — эта мера лишь углубит спад производства и приведет к более быстрому затуханию деловой активности и, следовательно, разрыванию новых витков инфляции. Зато можно ввести прогрессивную шкалу налогообложения *личного* имущества и личных доходов. Иначе говоря, государство должно (насколько это возможно) работать в направлении сглаживания социальной дифференциации, искать пути перераспределения национального дохода в пользу беднейших слоев населения.

Государственный контроль над ведущими отраслями хозяйства. В период роста темпов инфляции ответственность правительства за состояние экономики возрастает. Если в растущей, «здоровой» экономике многие параметры можно лишь ненавязчиво регулировать, представляя событиям разворачиваться своим чередом благодаря действию частных экономических интересов, то в экономике, подверженной инфляции и спаду производства, чаще бывает необходимо жесткое вмешательство государства. Основными целями, как правило, служат обеспечение надлежащего технологического уровня производства в ведущих отраслях хозяйства, социальная защита населения, поддержание должного уровня устойчивости макроэкономической системы и национальной безопасности страны.

Помимо политики сдерживания доходов, которая предполагает ожесточенную борьбу с «инфляционным

навесом», правительство может расширить сферу действия прямого государственного контроля над ценами, прибегая к временному их замораживанию, а иногда к прямому установлению цен на продукцию базовых отраслей и стратегически важные товары. Как правило, в период инфляции имеют место широкое применение лимитных цен, расширение сферы госзаказа. Следует заметить, что либерализация ценообразования и финансовой сферы никогда в истории не приводила к подавлению инфляционных процессов.

Обретая контроль над ведущими отраслями хозяйства, правительство должно преследовать не фискальные, а структурные цели. Необходимо проведение эффективной антимонопольной политики, поскольку интересы фирм-квазимонополистов заключаются в необоснованном поднятии цен на свою продукцию и получении дополнительной прибыли без наращивания объемов производства, что провоцирует инфляционные процессы.

Как правило, для осуществления преобразований в данном направлении достаточно провести ужесточение институциональных ограничений. Проблема приоритетного развития тех или иных форм собственности, хотя и имеет значение, все же не является важнейшей. Обретение государственного контроля вовсе не подразумевает какой бы то ни было национализации производства. Иногда более эффективно оставить основные активы в частных руках, но в то же время усилить сращивание частного бизнеса с государством. Хорошим примером такой логики действий стала гитлеровская Германия, где благодаря жесткой и грамотной инвестиционной политике правительства удалось достаточно быстро остановить инфляционные процессы. В распоряжении частных инвесторов оставалось 6 % от прибыли, остальную часть правительство перераспределяло через бюджеты всех уровней и направляло в экономику в виде автономных инвестиций и социальных расходов, направленных на стимулирование совокупного спроса. Аналогичным образом, можно создать противозатратный хозяйственный механизм без широкомасштабной приватизации, реформируя принципы управления ведущими отраслями хозяйства, как это было сделано, например, в 1980-е гг. в Китае.

Одним из направлений обретения контроля над ведущими отраслями хозяйства является введение или усиление *государственной монополии внешней торговли* по важнейшим группам товаров. Кстати, наведение порядка в данной сфере значительно затрудняет бегство капитала из страны. В России монополия внешней торговли на государственном уровне впервые была введена Иваном Грозным. Одно из возможных обоснований перехода к макроэкономической политике такого рода было дано в свое время В. И. Лениным [5]. С введением нэпа источники накопления капитала были отданы в частные руки, но воспользоваться этими источниками государство позволит лишь в той мере, в которой они будут работать на насыщение емкости *внутреннего рынка*. Если освободить эту идею от идеологической словесной шелухи, в которую она облечена, и вдуматься в экономический смысл предлагаемых мер, то станет ясно, что ничего собственно социалистического эта мера в себе не содержит: это стандартная логика действий го-

сударственной власти, оказавшейся в условиях дефицита внутренних источников накопления.

В этой же работе Ленин говорит о том, что никакая тарифная политика не может быть эффективной в условиях чудовищной разницы между странами бедными и невероятно богатыми: например, введение правительством бедной страны даже очень больших импортных пошлин будет без труда преодолеваться богатыми странами, правительства которых станут субсидировать своих производителей-экспортеров. Практика современной международной торговли блестяще подтверждает пророческий вывод Ленина. В частности, правительство США поддерживает своих фермеров экспортными субсидиями, обеспечивая экспорт сельскохозяйственной продукции, себестоимость которой является самой высокой в мире, и тем самым разоряет производителей в Африке и Латинской Америке, выпускающих намного более дешевые товары, например, хлопок и кофе. Совет Всемирной торговой организации, уже выносивший частное определение в отношении американского правительства, до сих пор не может добиться того, чтобы эта страна перестала прибегать к практике так называемой недобросовестной конкуренции и тем самым грубо нарушать устав ВТО.

Требования ВТО к правительствам беднейших стран мира по отмене всех нетарифных ограничений на их внутренних рынках направлены на обеспечение надлежащей степени открытости экономики этих стран для межстрановых переливов капитала и, следовательно, на подрыв экономического суверенитета этих стран. Только меры прямого государственного регулирования, утверждал Ленин, помогут более бедной и отсталой стране эффективно защитить свои экономические интересы, в том числе и на ее собственном внутреннем рынке, в частности, обеспечить суверенитет страны над ее собственными природными ресурсами.

Один из недавних примеров такой защиты — так называемая национализация нефтегазодобывающей отрасли в Боливии, объявленная президентом этой страны Эво Моралесом. На самом деле меры, проводимые в жизнь правительством, не представляют собой национализации в собственном смысле этого слова, хотя и предусматривают увеличение доли государства в акционерном капитале нефтегазодобывающих компаний. Основным пунктом преобразований является введение монополии внешней торговли: отныне все нефтегазодобывающие компании, действующие на территории Боливии, обязаны продавать 100 % добытых природных ресурсов государственной нефтегазовой компании, централизованно осуществляющей их экспорт. Это обретение государством контроля над ведущими отраслями хозяйства должно стимулировать рост производства, снизить темпы инфляции и повысить уровень жизни беднейших слоев населения страны.

Заметим, что для быстрого и эффективного обуздания инфляционных процессов, как правило, необходимо действовать по всем направлениям одновременно. Логика сложившейся в стране экономической ситуации предопределяет очередность принимаемых мер и степень важности каждого из указанных направлений.

Еще раз подчеркнем, что макроэкономическая стабилизация не сводится к финансовой стабилизации.



Необходимо создание предпосылок экономического подъема, предполагающее концентрацию необходимых ресурсов (как в денежной, так и в материально-вещественной форме) на важнейших направлениях экономического развития. Искаженная отраслевая и технологическая структура экономики также способна существенно затруднить выход из инфляционной ловушки, а в известной степени и порождать инфляционные процессы.

3. МИФ ОБ ИНФЛЯЦИОННОМ ПЕРЕГРЕВЕ РОССИЙСКОЙ ЭКОНОМИКИ

Инфляционный перегрев экономики — это краткосрочный эффект, выражающийся в том, что быстрый прирост инвестиций не вызывает адекватного увеличения текущего объема ВВП в силу его приближения к уровню потенциального ВВП. Желаящие могут сопоставить это определение с нестрогим и совершенно неверным (я бы сказал, непрофессиональным) описанием инфляционного перегрева, которое содержится в ряде экономических словарей: «чрезмерно высокие темпы экономического роста, приводящие к риску возникновения и развития инфляционных процессов из-за необоснованных государственных расходов» [6].

Как известно, потенциальный объем ВВП определяется последней единицей *предельного* (наиболее дефицитного) ресурса. В экономике Китая, например, в течение последних двух десятилетий таким предельным ресурсом служит электроэнергия. В трудодефицитных регионах достижение потенциального ВВП связано с состоянием так называемой полной занятости.

В период инфляционного перегрева стимулирование инвестиционного процесса наталкивается на ограничения в виде дефицита ресурсов, количество которых не может быть в надлежащей степени увеличено, вследствие чего прирост инвестиций вызывает подъем общего уровня цен без адекватного приращения ВВП.

Эта ситуация отражена на рис. 1, изображающем модель $AD-AS$ в координатах «объем производства Y — общий уровень цен P ». Стимулирование инвестиций выражается сдвигом вправо кривой AD из положения AD_0 в положение AD_1 . Непосредственным следствием этого сдвига является в первую очередь рост общего уровня цен с P_0 до P_1 , тогда как рост объемов производства наталкивается на ограничение в виде потенциального уровня ВВП, равного Y^* .

Подобная ситуация выражается прежде всего резким взлетом ресурсного индекса цен (предприятия усиливают борьбу за ресурсы), от которого заметно отстают рост индекса цен производителей, а последним реагирует на перегрев, естественно, индекс потребительских цен. Приблизительно такая ситуация сложилась на Украине летом 2004 г., что дало основания руководству Национального банка объявить об опасности инфляционного перегрева и предпринять усилия по сдерживанию темпов роста инвестиций.

В России дела обстоят существенно иначе. В 2004 г. лидером по росту цен среди непродовольственных товаров оказался бензин — 31,3 % за год. В то же время, хотя индекс промышленных цен, разумеется, в среднем обгонял индекс потребительских цен, однако цены на говядину и свинину выросли на 26,3 и 34 %, соответ-

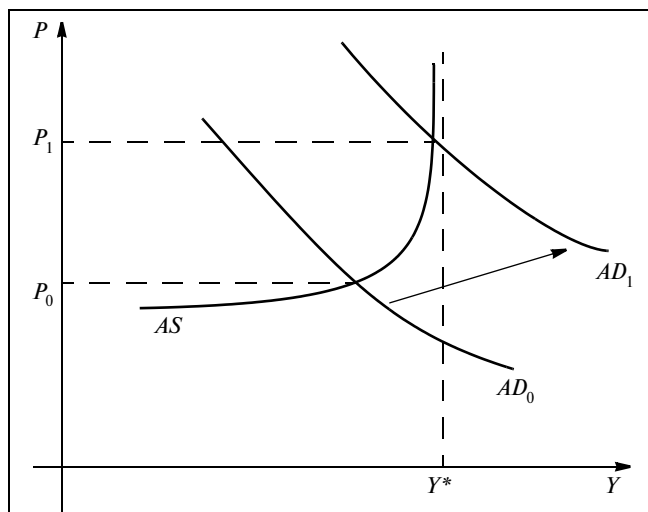


Рис. 1. Инфляционный перегрев экономики

ственно. Тарифы на жилищно-коммунальные услуги повысились за год на 23,5 %, что в 2 раза выше общего показателя инфляции 2004 г. (дефлятора ВВП), а за один лишь январь 2005 г. в некоторых регионах страны выросли еще более чем на четверть. В январе 2006 г. цены на услуги ЖКХ в некоторых регионах России взлетели на 15–25 %. Тем самым правительство в известном смысле посеяло семена инфляции, которые неизбежно через 2–3 мес. дают всходы: это естественный процесс, к перегреву не имеющий никакого отношения. Аналогичная ситуация складывалась в начале 1990-х гг., когда правительство поднимало цены на электроэнергию до уровня «мировых» (сколь бы странно ни звучал этот термин в условиях более чем двукратной разницы оптовых цен на электроэнергию в разных частях планеты).

Инфляционные тенденции в России имеют объективные причины, никак не связанные с перегревом экономики. Важнейшей из них следует считать технологическую деградацию производства, на что непосредственно указывает, в частности, министр экономического развития Герман Греф [7]. В самом деле, инвестиции в стареющие технологические уклады вызывают неизбежное нарастание инфляции издержек (подробнее см. в работе [8]). Каждая технология развивается по закону S-образной (логистической) кривой: фаза латентного (скрытого) развития, фаза резкого взлета, приносящая значительную прибыль при сравнительно небольших затратах, и, наконец, фаза отмирания, когда технология становится неэффективным объектом инвестирования. Отмирающие технологии, находящиеся в последней фазе своего жизненного цикла и приближающиеся к своим технологическим пределам, подобны «черным дырам»: они способны «проглотить» любой объем инвестиций, от которых ни страна, ни отдельное предприятие никогда не получают адекватной отдачи (рис. 2).

По этому поводу можно вспомнить прозвучавшие в разгар экономического кризиса предупреждения ряда отечественных экономистов о том, что выход из него, вероятно, осложнится отсутствием подходящих объектов для инвестирования, поддержание которых на кризисном этапе правительство сочло нерентабельным [9].

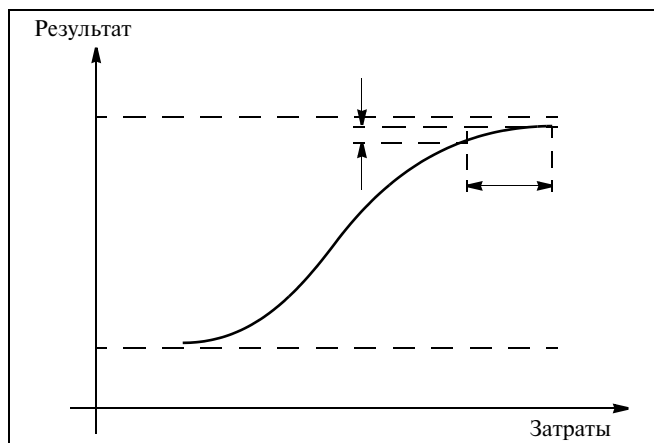


Рис. 2. Инвестиции в стареющие технологические уклады: рост затрат, снижение эффекта

Формирование «точек роста», включение инвестиционного мультипликатора, определение приоритетов технико-экономического развития и концентрация ресурсов (как технологических, так и финансовых) на избранных направлениях — это задачи *правительства*, их не под силу решить частным экономическим агентам.

Как обычно, неверная постановка диагноза плоха прежде всего тем, что из нее вытекают неверные рецепты лечения. Миф об инфляционном перегреве российской экономики опасен в первую очередь вытекающими из него рецептами антиинфляционной политики, последствия применения которых болезненно скажутся на экономической динамике. Если правительство вновь (как это было в 1990-х гг.) попытается сжать объем денежной массы (под предлогом «связывания» инвестиционных ресурсов в стабилизационном фонде или по иным причинам), то нашу страну, как нетрудно понять, ожидает очередной кризис неплатежей, в результате которого:

- реальный объем денежной массы практически не меняется, но заметно ухудшается ее качество, так как вместо полноценного ликвидного агрегата $M2$ каналы денежного обращения заполняются финансовыми суррогатами;

- «плохие» деньги, с которых можно не платить налоги, вытесняют из обращения «хорошие» деньги, а государственный бюджет недосчитывается доходов;

- затухает деловая активность — до тех пор, пока правительство не проведет взаимозачеты и не объявит в конце концов полноценными деньгами низколиквидный агрегат D (депозиты на счетах предприятий);

- возникает опасность спада производства в реальном секторе, что может лишь подхлестнуть инфляционные процессы;

- включается «мультипликатор неплатежей», аналогичный стандартному инвестиционному мультипликатору, только с обратным знаком: один рубль внутреннего государственного долга оборачивается потерей нескольких рублей в текущем объеме ВВП;

- углубляется социальная дифференциация, падает уровень жизни основной массы населения страны.

ЗАКЛЮЧЕНИЕ

Поскольку основным источником инфляции в современной экономике России является технологическая деградация производства, то наиболее эффективным способом подавления инфляционных процессов должна стать разработка национальной стратегии инновационного развития, предполагающей:

- формирование, а затем реализацию конкурентных преимуществ в высокотехнологичных отраслях хозяйства;
- создание «полюсов роста» и включение инвестиционного мультипликатора;
- концентрацию инвестиционных и технологических ресурсов на ведущих направлениях технико-экономического развития, способных обеспечить стране достойное место в мировом разделении труда.

Не случайно необходимости обеспечения технологического прорыва было посвящено одно из недавних выступлений Президента [10]. Знаковым моментом является тот факт, что оно прозвучало на заседании Совета Безопасности, поскольку от своевременной и грамотной разработки и реализации стратегии инновационного развития решающим образом зависит дальнейшая траектория экономической динамики, а в итоге — национальная безопасность страны: превращение России в сырьевой придаток более развитых стран с отчетливой перспективой распада либо выход на траекторию устойчивого экономического роста и социального развития.

ЛИТЕРАТУРА

1. Гурова Т., Полунин Ю. Цена монополии на ликвидность // Эксперт. — 2006. — № 11.
2. Страны с переходной экономикой: корреляционный анализ экономической динамики / Р. М. Нижегородцев, Б. М. Болотин, А. А. Дорофеев, А. Л. Чернявский // Информационная экономика и динамика переходных процессов: Сб. науч. тр. / Под ред. Е. Ю. Иванова, Р. М. Нижегородцева. — Москва — Барнаул: Бизнес-Юнитек, 2003.
3. Бузгалин А. В., Колганов А. И. Введение в компаративистику: Исследование и сравнительный анализ социально-экономических систем: методология, теория, применение к переходным экономикам. — М., 1997. — С. 203.
4. Эффективный экономический рост: теория и практика. — М., 2001. — С. 154.
5. Ленин В. И. О монополии внешней торговли // Полн. собр. соч., 5-е изд., М.; 1970. — Т. 45. — С. 333—337.
6. <http://www.glossary.ru/cgi-bin/gl_sch2.cgi?R0pdqutuso,lxqollz,lto9;http://www.businessvoc.ru/bv/Term.asp?theme=12&word_id=13588>.
7. Плетнев С. В чем прав Герман Греф <<http://www.rambler.ru/db/news/msg.html?mid=5418764>>.
8. Нижегородцев Р. М. Информационная экономика. Кн. 3. Взгляд в Зазеркалье: Техничко-экономическая динамика кризисной экономики России. — Москва; Кострома, 2002.
9. Нижегородцев Р. М. Технологическое будущее России: что впереди? // Мировая экономика и международные отношения. — 1995. — № 1.
10. Стенографический отчет о совещании Совета Безопасности, посвященном мерам по реализации Послания Федеральному Собранию <http://www.kremlin.ru/appears/2006/06/20/1708_type63378type63381_107461.shtml>.

☎ (495) 334-90-51

e-mail: bell44@rambler.ru





МНОГОКРИТЕРИАЛЬНЫЕ СИСТЕМЫ СТИМУЛИРОВАНИЯ

А. А. Иващенко

Институт проблем управления им. В. А. Трапезникова РАН, г. Москва

Рассмотрены теоретико-игровые модели многокритериальных систем стимулирования, в которых деятельность каждого управляемого субъекта описывается несколькими показателями, значения которых определяют размер вознаграждения, выплачиваемого ему управляющим органом.

ВВЕДЕНИЕ

Стимулированием называется побуждение (осуществляемое посредством воздействия управляющего органа — центра — на предпочтения управляемого субъекта — агента) к совершению определенных действий [1]. В настоящей работе изучаются системы многокритериального стимулирования, в рамках которых центр компенсирует агентам затраты только при условии выполнения плана. Основной акцент делается на том, что деятельность агентов описывается вектором показателей [2], от которого зависит размер вознаграждения. Сначала рассматриваются одноэлементные, а затем — многоэлементные организационные системы.

1. ОБЩАЯ ПОСТАНОВКА ЗАДАЧИ

Рассмотрим двухуровневую организационную систему (ОС), состоящую из одного центра на верхнем уровне иерархии и n агентов на нижнем. Стратегией i -го агента является выбор действия $y_i \in A_i$, $i \in N = \{1, 2, \dots, n\}$ — множеству агентов; стратегией центра — выбор системы стимулирования $\{\sigma_i(z)\}_{i \in N}$, где $z_i = Q_i(y) \in B_i$ — наблюдаемый центром результат деятельности i -го агента, $y = (y_1, y_2, \dots, y_n)$ — вектор действий агентов, $z = (z_1, z_2, \dots, z_n)$ — вектор результатов деятельности агентов, $Q_i: A \rightarrow B_i$ — оператор агрегирования, $\sigma_i: B \rightarrow \mathbb{R}^1$, $i \in N$, $A = \prod_{i \in N} A_i$, $B = \prod_{i \in N} B_i$.

Предпочтения центра отражены его целевой функцией $\Phi(z, \sigma(z)) = H(z) - \sum_{i \in N} \sigma_i(z)$, где $H(\cdot): B \rightarrow \mathbb{R}^1$ — функция дохода центра.

Предпочтения i -го агента отражены его целевой функцией $f_i(y, \sigma_i(z)) = \sigma_i(z) - c_i(y)$, где $c_i(\cdot): A \rightarrow \mathbb{R}^1$ — функция затрат i -го агента, $i \in N$.

Последовательность функционирования ОС такова: центр выбирает и сообщает агентам систему стимулиро-

вания (зависимость вознаграждения, выплачиваемого каждому из агентов, от вектора результатов их деятельности), затем агенты однократно, одновременно и независимо выбирают свои действия, которые приводят к соответствующим результатам деятельности. Целевые функции и допустимые множества, а также операторы агрегирования являются общим знанием среди всех участников ОС (центра и агентов); агенты на момент принятия решений знают выбранную центром систему стимулирования; центр наблюдает результаты деятельности агентов, но может не знать их действий.

Обозначим:

$P(\sigma(\cdot)) \subseteq A$ — множество действий, выбираемых агентами при системе стимулирования $\sigma(\cdot)$; обычно считается, что агенты выбирают действия, являющиеся равновесием их игры;

$Q(P) = \bigcup_{y \in P} \{Q_1(y_1), Q_2(y_2), \dots, Q_n(y_n)\}$ — множество

результатов деятельности агентов, которые могут реализоваться при выборе ими действий из множества P .

Эффективность стимулирования $K(\sigma)$ определяется как гарантированное значение целевой функции центра:

$$K(\sigma) = \min_{z \in Q(P(\sigma))} \left[H(z) - \sum_{i \in N} \sigma_i(z) \right].$$

В общем виде задача стимулирования формулируется следующим образом — найти допустимую систему стимулирования, обладающую максимальной эффективностью:

$$K(\sigma) \rightarrow \max_{\sigma}. \quad (1)$$

2. БАЗОВАЯ МОДЕЛЬ СТИМУЛИРОВАНИЯ

Основным аппаратом моделирования задач стимулирования в теории управления служит аппарат теории игр — раздела прикладной математики, исследующего модели принятия решений в условиях несовпадения интересов сторон (игроков), когда каждая сторона стремится воздействовать на развитие ситуации в собственных интересах [3, 4]. Простейшей игровой моделью является

взаимодействие двух игроков — центра и подчиненного ему агента, т. е. $n = 1$.

Стратегией агента является выбор действия $y \in A \subseteq \mathbb{R}^k$, $k \geq 2$, принадлежащего компактному множеству допустимых действий A . Содержательно действием агента может быть количество обрабатываемых часов, объем произведенной продукции, ее качество и иные характеристики. Пока будем считать, что агрегирование отсутствует, т. е. $z \equiv y$, а $Q(\cdot)$ — тождественное отображение, $m = k$, $B = A$.

Стратегией центра является выбор функции стимулирования $\sigma(\cdot)$, ставящей в соответствие действию агента некоторое неотрицательное вознаграждение, выплачиваемое ему центром, т. е. $\sigma : A \rightarrow \mathbb{R}_+^1$. Выбор действия $y \in A$ требует от агента затрат $c(y)$, $c : A \rightarrow \mathbb{R}^1$, и приносит центру доход $H(y)$, $H : A \rightarrow \mathbb{R}^1$.

Интересы участников организационной системы (центра и агента) отражены их скалярными целевыми функциями (функциями выигрыша, полезности и т. д., в записи которых зависимость от стратегии центра будет опускаться), которые обозначим, соответственно: $\Phi(y)$ и $f(y)$.

Целевые функции представляют собой: для агента — разность между стимулированием и затратами:

$$f(y) = \sigma(y) - c(y), \quad (2)$$

а для центра — разность между доходом и затратами центра на стимулирование — вознаграждением, выплачиваемым агенту:

$$\Phi(y) = H(y) - \sigma(y). \quad (3)$$

Определим $y_{LCA} = \arg \min_{y \in A} c(y)$ — действие агента, минимизирующее его затраты.

Введем следующие предположения, которых будем придерживаться, если не оговорено особо, в ходе дальнейшего изложения. Относительно функции затрат предположим, что она непрерывна, а затраты от выбора действия y_{LCA} равны нулю. Также допустим, что сумма вознаграждения, выплачиваемого центром агенту, неотрицательна и что функция дохода центра непрерывна и достигает максимума при действии агента, отличном от y_{LCA} .

Так как значение целевой функции агента (2) зависит как от его собственной стратегии — действия, так и от функции стимулирования, то в рамках гипотезы рационального поведения агент будет выбирать действия, которые при заданной системе стимулирования максимизируют его целевую функцию. Понятно, что множество таких действий, называемое множеством реализуемых действий, зависит от используемой центром системы стимулирования. Основная идея стимулирования как раз и заключается в том, что, варьируя систему стимулирования, центр может побуждать агента выбирать те или иные действия.

Так как целевая функция центра зависит от действия, выбираемого агентом, то эффективностью системы стимулирования называют гарантированное значение целевой функции центра на множестве действий агента, реализуемых данной системой стимулирования. Следовательно, задача стимулирования (см. выражение (1)) заключается в том, чтобы выбрать оптимальную систему стимулирования, т. е. систему стимулирования, имеющую максимальную эффективность.

Множество действий агента, доставляющих максимум его целевой функции (и, естественно, зависящих от функции стимулирования), называется множеством решений игры или множеством действий, реализуемых данной системой стимулирования:

$$P(\sigma) = \operatorname{Arg} \max_{y \in A} \{\sigma(y) - c(y)\}. \quad (4)$$

Зная, что агент выбирает действия из множества (4), центр должен найти систему стимулирования, которая максимизировала бы его собственную целевую функцию (3). Следовательно, эффективность системы стимулирования $\sigma \in M$

$$K(\sigma) = \min_{y \in P(\sigma)} \Phi(y).$$

Задача синтеза оптимальной системы стимулирования заключается в выборе допустимой системы стимулирования, имеющей максимальную эффективность (см. также (1)):

$$K(\sigma) \rightarrow \max_{\sigma}. \quad (5)$$

Перейдем к решению задачи стимулирования, практически дословно повторяя решение, описанное в работе [1], для рассматриваемого случая многокритериальной системы стимулирования. Предположим, что применялась система стимулирования $\sigma(\cdot)$, при которой агент выбирал действие $x \in P(\sigma(\cdot))$. Легко показать, что если взять другую систему стимулирования $\tilde{\sigma}(\cdot)$, которая будет равна нулю всюду, кроме точки x , и будет равна старой системе стимулирования в точке x :

$$\tilde{\sigma}(y) = \begin{cases} \sigma(x), & y = x \\ 0, & y \neq x \end{cases}, \text{ то и при новой системе стимули-}$$

рования это же действие агента будет доставлять максимум его целевой функции.

Так как центр стремится минимизировать выплаты агенту при условии, что последний выбирает требуемое действие (план, назначенный центром), то вознаграждение в случае выполнения плана должно равняться затратам агента (точнее, превосходить их на сколь угодно малую положительную величину δ — для того, чтобы целевая функция агента имела единственный максимум — точку плана). Этот важный вывод для скалярных систем стимулирования получил название «принцип компенсации затрат» [1]. Он справедлив для рассматриваемой модели и в случае многокритериального стимулирования.

Следовательно, параметрическим (с параметром — планом $x \in A$) решением задачи (5) является система стимулирования, которая называется компенсаторной:

$$\sigma_K(x, y) = \begin{cases} c(x) + \delta, & y = x, \\ 0, & y \neq x. \end{cases} \quad (6)$$

Сколь угодно малая строго положительная величина δ , фигурирующая в оптимальной системе стимулирования, получила название мотивационной надбавки, так как именно ее размер определяет значение целевой функции агента [1]. Размер мотивационной надбавки выбирается, исходя из различных соображений (см. далее).

Оптимальное реализуемое действие может быть найдено из решения следующей стандартной оптимизационной задачи:

$$y^* = \arg \max_{x \in A} [H(x) - c(x)]. \quad (7)$$



Утверждение 1. При $n = 1$, $k \geq 2$ и отсутствии агрегирования, система стимулирования (6) и (7) δ -оптимальна.

Ранее мы рассматривали случай отсутствия агрегирования информации. Теперь предположим, что агрегирование информации имеет место, т. е. доход центра $h(z)$ зависит от наблюдаемого им результата деятельности агента $z = Q(y) \in B \subseteq \mathcal{R}^m$, причем $m \leq k$, где $Q(\cdot): A \rightarrow B$ — однозначное непрерывное отображение, такое что $\bigcup_{y \in A} Q(y) = B$. При этом предполагается, что оператор агрегирования и функция затрат агента центру известны, а действия не наблюдаются.

Зафиксируем произвольный результат деятельности агента $z \in B$ и вычислим множество его действий, приводящих к данному результату:

$$Y(z) = \{y \in A | Q(y) = z\},$$

и минимальные затраты агента по достижению данного результата: (13) $C(z) = \min_{y \in Y(z)} c(y)$.

Рассмотрим систему стимулирования:

$$\sigma_K(x, z) = \begin{cases} C(x), & z = x \\ 0, & z \neq x \end{cases}, \quad x, z \in B. \quad (8)$$

Видно, что система стимулирования (8) в рамках гипотезы благожелательности (при прочих равных агент выберет действия, наиболее благоприятные с точки зрения центра) побуждает агента выбрать действия, приводящие к «плановому результату» $x \in B$, причем затраты центра на стимулирование при этом минимальны.

Оптимальный реализуемый результат деятельности может быть найден из решения следующей стандартной оптимизационной задачи:

$$z^* = \arg \max_{x \in B} [h(x) - C(x)]. \quad (9)$$

Утверждение 2. При $n = 1$, $k \geq 2$ и наличии агрегирования в рамках гипотезы благожелательности система стимулирования (8) и (9) оптимальна.

Таким образом, получено решение задачи синтеза оптимальной многокритериальной системы стимулирования в одноэлементной ОС как для случая отсутствия агрегирования информации (утверждение 1), так и для случая агрегирования информации (утверждение 2). Завершив рассмотрение механизмов стимулирования в одноэлементных ОС, перейдем к описанию механизмов многокритериального стимулирования в многоэлементных ОС.

3. СТИМУЛИРОВАНИЕ ПО ИНДИВИДУАЛЬНЫМ РЕЗУЛЬТАТАМ

Пусть $N = \{1, 2, \dots, n\}$ — множество агентов, $y_i \in A_i$ — действие i -го агента, $c_i(y)$ — скалярные затраты i -го агента, $\sigma_i(y)$ — скалярное стимулирование этого агента со стороны центра, $i \in N$, $y = (y_1, y_2, \dots, y_n)$ — вектор действий агентов, $y \in A = \prod_{i \in N} A_i$. Предположим, что центр получает доход $H(y)$ от деятельности агентов.

Обозначим $y_{-i} = (y_1, y_2, \dots, y_{i-1}, y_{i+1}, \dots, y_n) \in A_{-i} = \prod_{j \neq i} A_j$ — обстановка игры для i -го агента. Интересы и предпочтения участников ОС — центра и агентов — выражены их целевыми функциями. Целевая функция

центра $\Phi(\sigma, y)$ представляет собой разность между его доходом $H(y)$ и суммарным вознаграждением $v(y)$,

выплачиваемым агентам: $v(y) = \sum_{i=1}^n \sigma_i(y)$, где $\sigma_i(y)$ — сти-

мулирование i -го агента, $\sigma(y) = (\sigma_1(y), \sigma_2(y), \dots, \sigma_n(y))$. Целевая функция i -го агента $f_i(\sigma_i, y)$ — разность между стимулированием, получаемым от центра, и затратами $c_i(y)$

$$\Phi(\sigma, y) = H(y) - \sum_{i=1}^n \sigma_i(y), \quad f_i(\sigma_i, y) = \sigma_i(y) - c_i(y), \quad i \in N.$$

Отметим, что и индивидуальное вознаграждение, и индивидуальные затраты i -го агента по выбору действия y_i в общем случае зависят от действий всех агентов (случай сильно связанных агентов с несепарабельными затратами [1]).

Примем следующий порядок функционирования ОС. Центру и агентам на момент принятия решения о выбираемых стратегиях (соответственно — функциях стимулирования и действиях) известны целевые функции и допустимые множества всех участников ОС. Центр, обладая правом первого хода, выбирает функции стимулирования и сообщает их агентам, после чего агенты при известных функциях стимулирования одновременно и независимо выбирают действия, максимизирующие их целевые функции.

Обобщая предложенную в работе [5] модель, относительно параметров ОС введем следующие предположения:

— множество A_i допустимых действий i -го агента

является компактом в $\mathcal{R}^{k_i}$;

— функции затрат агентов непрерывны, неотрицательны и $\exists y_{LCA\ i} \in A_i$ такое, что $\forall y_{-i} \in A_{-i} \arg \min_{y_i \in A_i} c_i(y_{-i}, y_i) = y_{LCA\ i}$, причем $\forall y_{-i} \in A_{-i} c_i(y_{LCA\ i}, y_{-i}) = 0$;

— функция дохода центра непрерывна по всем переменным и достигает максимума при векторе действий агентов, отличном от $y_{LCA} = (y_{LCA\ 1}, y_{LCA\ 2}, \dots, y_{LCA\ n})$.

Так как и затраты, и стимулирование каждого агента в рассматриваемой модели зависят в общем случае от действий всех агентов, то последние оказываются вовлеченными в игру, в которой выигрыш каждого зависит от действий всех. Обозначим $P(\sigma)$ — множество равновесных при системе стимулирования σ стратегий агентов — множество решений игры (тип равновесия пока не оговаривается; единственно предположим, что агенты выбирают свои стратегии однократно, одновременно и независимо друг от друга, не имея возможности обмениваться дополнительной информацией и полезностью).

Как и в одноэлементной ОС, гарантированной эффективностью (далее просто «эффективностью») стимулирования является минимальное значение целевой функции центра на соответствующем множестве решений игры агентов: $K(\sigma) = \min_{y \in P(\sigma)} \Phi(\sigma, y)$.

Задача синтеза оптимальной функции стимулирования заключается в поиске допустимой системы стимулирования σ^* , имеющей максимальную эффективность: $\sigma^* = \arg \max_{\sigma} K(\sigma)$.

В частном случае, когда агенты независимы (вознаграждение и затраты каждого из них зависят только от его собственных действий), то оптимальной (точнее — δ -оптимальной, где $\delta = \sum_{i \in N} \delta_i$) является компенсаторная система стимулирования:

$$\sigma_{iK}(y_i) = \begin{cases} c_i(y_i^*) + \delta_i y_i = y_i^*, & i \in N, \\ 0, & y_i \neq y_i^* \end{cases} \quad (10)$$

где $\{\delta_i\}_{i \in N}$ — сколь угодно малые строго положительные константы (мотивирующие надбавки), а оптимальное действие y^* , реализуемое системой стимулирования (10) как равновесие в доминантных стратегиях (РДС) [3], является решением следующей задачи оптимального согласованного планирования: $y^* =$

$$= \arg \max_{y \in A} \left\{ H(y) - \sum_{i \in N} c_i(y_i) \right\}.$$

Если стимулирование каждого агента зависит от действий всех агентов (рассматриваемый в настоящем подразделе случай коллективного стимулирования) и затраты несепабельны (затраты каждого агента зависят в общем случае от действий всех агентов, что отражает взаимосвязь и взаимозависимость агентов), то множества равновесий Нэша [3] $E_N(\sigma) \subseteq A$ и РДС $y_d \in A$ имеют вид:

$$E_N(\sigma) = \{y^N \in A | \forall i \in N \forall y_i \in A_i \sigma_i(y^N) - c_i(y_i^N) \geq \sigma_i(y_i, y_{-i}^N) - c_i(y_i, y_{-i}^N)\};$$

$y_{i_d} \in A_i$ — доминантная стратегия i -го агента тогда и только тогда, когда

$$\forall y_i \in A_i, \forall y_{-i} \in A_{-i} \sigma_i(y_{i_d}, y_{-i}) - c_i(y_{i_d}, y_{-i}) \geq \sigma_i(y_i, y_{-i}) - c_i(y_i, y_{-i}).$$

Если при заданной системе стимулирования у всех агентов имеется доминантная стратегия, то говорят, что данная система стимулирования реализует соответствующий вектор действий как РДС.

Фиксируем произвольный вектор действий агентов $y^* \in A$ и рассмотрим следующую систему стимулирования:

$$\sigma_i(y^*, y) = \begin{cases} c_i(y_i^*, y_{-i}) + \delta_i y_i = y_i^*, & \delta_i \geq 0, i \in N, \\ 0, & y_i \neq y_i^* \end{cases} \quad (11)$$

Полученный в работе [5] результат остается в силе и для рассматриваемой модели: при использовании центром системы стимулирования (11) y^* — РДС. Более того, если $\delta_i > 0, i \in N$, то y^* — единственное РДС.

Вектор оптимальных реализуемых действий агентов y^* , фигурирующий в качестве параметра в выражении (11), определяется в результате решения следующей задачи оптимального согласованного планирования:

$$y^* = \arg \max_{y \in A} \left\{ H(y) - \sum_{i \in N} c_i(y_i) \right\}, \quad (12)$$

а эффективность системы стимулирования (11) и (12) равна следующей величине:

$$K^* = H(y^*) - \sum_{i \in N} c_i(y_i^*) - \delta.$$

Утверждение 3. При $n \geq 2, k \geq 2$ и отсутствии агрегирования система стимулирования (11) и (12) δ -оптимальна.

Таким образом, посредством обобщения результатов, полученных в работе [5], решена задача синтеза оптимальной многокритериальной системы стимулирования в многоэлементных ОС без агрегирования информации (утверждение 3). Перейдем к описанию случая агрегирования информации.

4. СТИМУЛИРОВАНИЕ ПО КОЛЛЕКТИВНЫМ РЕЗУЛЬТАТАМ

Пусть в рамках модели, рассмотренной в § 3, имеет место агрегирование информации, т. е. результат деятельности $z \in B$ организационной системы, состоящей из n агентов, является функцией их действий: $z_i = Q_i(y)$, $i \in N$. Интересы и предпочтения участников организационной системы — центра и агентов — выражены их целевыми функциями. Целевая функция центра представляет собой разность между его доходом $h(z)$ и суммарным вознаграждением, выплачиваемым агентам:

$$\Phi(\sigma(\cdot), z) = h(z) - \sum_{i \in N} \sigma_i(z_i),$$

где $\sigma_i(z_i)$ — стимулирование i -го агента, $\sigma(z) = (\sigma_1(z_1), \sigma_2(z_2), \dots, \sigma_n(z_n))$.

Целевая функция i -го агента представляет собой разность между стимулированием, получаемым им от центра, и затратами $c_i(y)$: $f_i(\sigma_i(\cdot), y) = \sigma_i(z_i) - c_i(y)$, $i \in N$.

Примем следующий порядок функционирования ОС. Центру и агентам на момент принятия решений о выбираемых стратегиях (соответственно — функциях стимулирования и действиях) известны целевые функции и допустимые множества всех участников ОС, а также функции агрегирования. Центр, обладая правом первого хода, выбирает функции стимулирования и сообщает их агентам, после чего агенты при известных функциях стимулирования выбирают действия, максимизирующие их целевые функции.

Рассмотрим случай, когда центр наблюдает только результаты деятельности агентов, от которых зависит его доход $h(z)$, но не знает и не может восстановить индивидуальных действий агентов (в противном случае мы оказались бы в рамках модели, рассмотренной ранее), т. е., имеет место агрегирование информации — центр имеет не всю информацию о векторе $y \in A$ действий агентов, а ему известен лишь некоторый их агрегат $z \in B$ — параметр, характеризующий результаты совместных действий агентов.

В рассмотренной в § 3 задаче стимулирования декомпозиция игры агентов основывалась на возможности центра поощрять агентов за выбор определенного (и наблюдаемого центром) действия. Если действия агентов не наблюдаемы, то непосредственное применение идеи декомпозиции невозможно, поэтому при решении задач стимулирования, в которых вознаграждение агентам зависит от агрегированных результатов деятельности,



следует воспользоваться подходом, описанным для одноэлементной системы ранее — найти множество действий, приводящих к заданным результатам деятельности, выделить среди них подмножество, характеризующее минимальными суммарными затратами агентов (и, следовательно, минимальными затратами центра на стимулирование при использовании компенсаторных функций стимулирования), построить систему стимулирования, реализующую это подмножество действий, а затем определить, реализация какого из результатов деятельности наиболее выгодна для центра.

Будем считать, что отображения $\{Q_i(\cdot)\}$ непрерывны и однозначны, причем $\bigcup_{y \in A} (Q_1(y), Q_2(y), \dots, Q_n(y)) = B$.

Определим множество векторов действий агентов, приводящих к заданному вектору результатов деятельности $z \in B$:

$$Y(z) = \{y \in A \mid Q_i(y) = z_i, i \in N\} \subseteq A.$$

Вычислим минимальные суммарные затраты агентов по достижению результата деятельности $z \in B$:

$$C(z) = \min_{y \in Y(z)} \sum_{i \in N} c_i(y),$$

а также множество действий $Y^*(z) = \text{Arg} \min_{y \in Y(z)} \sum_{i \in N} c_i(y)$, на котором этот минимум достигается.

Фиксируем произвольный результат деятельности $x \in B$ и произвольный вектор $y^*(x) \in Y^*(x) \subseteq Y(x)$.

Пусть выполнено одно из следующих предположений.

А. 1. $\forall x \in B$ множество $Y^*(x)$ состоит из одной точки.

А. 2. Затраты агентов сепарабельны, т. е. $c_i = c_i(y_i), i \in N$.

А. 3. $\forall i \in N, \forall x \in B, \forall y^*(x) \in Y^*(x), \forall \hat{y}_i \in A_p$ такого, что $Q(\hat{y}_i, y_{-i}^*(x)) = x$, выполнено $c_i(\hat{y}_i, y_{-i}^*(x)) > c_i(y^*(x))$.

По аналогии с тем, как это делается в работе [5], можно доказать, что:

1) при использовании центром системы стимулирования

$$\sigma_{ix}^*(z) = \begin{cases} c_i(y^*(x)) + \delta_p z = x, & i \in N, \\ 0, & z \neq x \end{cases} \quad (13)$$

вектор действий агентов $y^*(x)$ реализуется как единственное равновесие Нэша игры агентов с минимальными затратами центра на стимулирование равными $C(x) + \delta$, где $\delta = \sum_{i \in N} \delta_p$;

2) система стимулирования (13) является δ -оптимальной.

Найдем наиболее выгодный для центра результат деятельности ОС $x^* \in B$ как решение задачи оптимального согласованного планирования:

$$x^* = \arg \max_{x \in B} [h(x) - C(x)]. \quad (14)$$

Утверждение 4. Если при $n \geq 2, k \geq 2$ и наличии агрегирования выполнено одно из предположений А. 1 или А. 2, или А. 3, то система стимулирования (13) и (14) δ -оптимальна.

Таким образом, выражения (13) и (14) дают решение задачи синтеза оптимальной многокритериальной сис-

Оптимальные ($k \geq 2$) системы многокритериального стимулирования

n	Агрегирование отсутствует	Агрегирование присутствует
1 Не менее 2	Утверждение 1 Утверждение 2	Утверждение 3 Утверждение 4

темы стимулирования по агрегированным результатам совместной деятельности.

Отметим, что выше рассматривались постановки задач стимулирования, в которых вознаграждение, выплачиваемое агентам, вычиталось из дохода центра. Более простым случаем является наличие фиксированного фонда заработной платы (ФЗП), который необходимо распределить так, чтобы выбираемые (в рамках назначенной центром системы стимулирования) агентами действия максимизировали доход центра. Решение задачи стимулирования при этом останется, в основном, без изменений и центру следует по-прежнему использовать соответствующие компенсаторные системы стимулирования. Отличие будет заключаться в том, что задача согласованного планирования сведется к максимизации функции дохода центра на множестве таких действий (или результатов деятельности) агентов, что их суммарные затраты (компенсируемые центром) не превосходят имеющегося ФЗП.

ЗАКЛЮЧЕНИЕ

Итак, в настоящей работе сформулированы и решены задачи синтеза оптимальных систем многокритериального стимулирования ($k \geq 2$) для одноэлементных ($n = 1$) и многоэлементных ($n \geq 2$) организационных систем — см. таблицу.

Перспективным направлением дальнейших исследований представляется постановка и решение задач синтеза оптимальных систем многокритериального стимулирования для ситуаций, в которых класс допустимых систем стимулирования ограничен и не включает в себя компенсаторные системы стимулирования. Типовыми примерами широко распространенных на практике таких классов систем стимулирования служат линейные системы стимулирования, механизмы «бригадной» оплаты труда и ранговые системы стимулирования.

ЛИТЕРАТУРА

- Новиков Д. А. Стимулирование в организационных системах. — М.: СИНТЕГ, 2003. — 312 с.
- Модели и механизмы многокритериального стимулирования в организационных системах / А. А. Ивашенко, Д. А. Новиков, М. И. Сапико, М. А. Щепкина. — М.: ИПУ РАН, 2006.
- Губко М. В., Новиков Д. А. Теория игр в управлении организационными системами. — М.: СИНТЕГ, 2002. — 148 с.
- Myerson R. B. Game theory: analysis of conflict. — London: Harvard Univ. Press, 1991. — 568 p.
- Новиков Д. А., Цветков А. В. Механизмы стимулирования в многоэлементных организационных системах. — М.: Апостроф, 2000. — 184 с.

☎ (495) 334-90-51;

e-mail: ai@ipu.ru



ОПТИМАЛЬНОСТЬ ДРЕВОВИДНОЙ ИЕРАРХИИ УПРАВЛЕНИЯ СИММЕТРИЧНОЙ ПРОИЗВОДСТВЕННОЙ ЛИНИЕЙ

С. П. Мишин

Институт проблем управления им. В. А. Трапезникова РАН, г. Москва

Рассмотрена модель многоуровневой иерархической структуры, управляющей «производственной линией» (последовательным бизнес-процессом с одинаковой интенсивностью взаимодействия исполнителей). Затраты менеджера определены степенной функцией, зависящей от интенсивности управляемого менеджером подпроцесса. Доказано, что оптимальной будет древовидная иерархия, задача поиска которой сведена к задаче условной оптимизации. Для выпуклой функции затрат оптимальное дерево найдено в явном виде.

ВВЕДЕНИЕ

Любая экономическая система состоит из множества организованных некоторым образом сотрудников. Благодаря организации сотрудники действуют на основе определенных процедур и правил (механизмов), что позволяет достичь цели системы [1].

Сотрудники организации специализированы, что повышает их эффективность по сравнению с множеством одиночных (неорганизованных) агентов. Однако взаимодействие сотрудников различной специализации должно быть скоординировано для достижения общей цели системы. Это фундаментальная проблема любой организации, поскольку координация требует усилий, направленных на планирование совместной работы, контроль результатов, согласование целей сотрудников и др. Для реализации управленческих функций в организации создается иерархия¹ [2].

Один из ключевых факторов эффективности экономической системы — оптимальность иерархии, выполняющей управленческие функции с минимальными затратами. Для небольшой системы оптимальной может быть двухуровневая иерархия, в которой на первом (нижнем) уровне находятся рядовые исполнители, а на втором — единственный менеджер. Однако при росте системы один менеджер уже не способен управлять всеми взаимодействиями исполнителей, и необходимо на-

нять нескольких менеджеров на второй уровень иерархии, возложив на каждого из них ответственность за управление взаимодействиями внутри подчиненной группы исполнителей. Взаимодействие между группами, которые подчинены менеджерам второго уровня, порождает взаимодействие между менеджерами. Этим взаимодействием также необходимо управлять с помощью менеджеров третьего уровня и т. д. Подобным образом возникает многоуровневая иерархия с единственным менеджером на высшем уровне. Вышестоящий менеджер (начальник) в иерархии уполномочен управлять подчиненными сотрудниками (менеджерами или исполнителями), а подчиненные обеспечивают начальников информацией и выполняют их распоряжения [3].

Задача об оптимальной иерархии весьма сложна для математического исследования. По этой причине на данный момент имеется небольшое число работ по данной тематике. Среди «классических» можно отметить работы [4, 5], в которых проблема освещается достаточно общо, однако модель оптимальной иерархии в общем виде не формализована. В ряде более поздних работ (см., например, книгу [6]) исследуются строгие математические постановки указанной задачи, при этом налагается ряд весьма жестких ограничений (как правило, рассматриваются только древовидные иерархии с подчинением только между соседними уровнями и затратами, зависящими только от числа непосредственных подчиненных). В ряде недавних работ рассматривается более общая постановка, которая, несмотря на свою сложность, допускает математическое исследование [7–12]. Далее будет использоваться аппарат и результаты указанных работ.

¹ Сотрудники на более высоких уровнях иерархии обладают большими правами, чем сотрудники нижних уровней. Это позволяет системе достичь цели даже в случае конфликтов между сотрудниками.



Рассматриваемая в настоящей работе модель иерархии управления основана на заданном технологическом взаимодействии исполнителей². Интенсивность взаимодействия определяется *технологической сетью*, связи которой соответствуют *потокам* (материальным, информационным и т. п.) между исполнителями. Одним из примеров технологической сети является *производственная линия*, в которой первый исполнитель обеспечивает поставку сырья, второй и последующие производят ряд технологических операций, последний отгружает продукцию потребителю. В *симметричной производственной линии* интенсивность всех потоков одинакова; т. е. рассматривается «цепочка» исполнителей — «производственная линия» с одинаковой интенсивностью взаимодействия между соседними исполнителями. Затраты менеджера определяются монотонной функцией, зависящей от интенсивности управляемого менеджером подпроцесса (участка производственной линии). Данная функция затрат является частным случаем рассматриваемой в работах [9, 12] *секционной функции*. Для этого случая настоящая статья исчерпывающим образом решает задачу об оптимальной иерархии управления.

1. ИСПОЛНИТЕЛИ И МЕНЕДЖЕРЫ, ВИДЫ ИЕРАРХИЙ

Пусть $N = \{w_1, \dots, w_n\}$ — множество *исполнителей*, которые могут взаимодействовать друг с другом. Через w_{env} будем обозначать *внешнюю среду*, взаимодействующую с исполнителями. Иногда исполнители обозначаются $w, w', w'' \in N$. *Функцией потока* назовем следующую функцию:

$$f: (N \cup \{w_{env}\}) \times (N \cup \{w_{env}\}) \rightarrow R_+^p,$$

где для каждой пары исполнителей $w', w'' \in N$ вектор $f(w', w'')$ определяет *интенсивность потоков* между w' и w'' . Этот вектор содержит p неотрицательных компонент. Каждая компонента определяет интенсивность одного типа взаимодействия исполнителей (материальный, информационный или прочий тип потока). Таким образом, технология взаимодействия исполнителей определяет функцию потока f или взвешенную *технологическую сеть* f^3 . Вектор $f(w_{env}, w)$ соответствует интенсивности потоков между исполнителем w и внешней средой. Потоки между исполнителями назовем *потоками внутри технологической сети*. Будем говорить, что между w' и w'' отсутствует *связь* тогда и только тогда, когда поток между исполнителями нулевой.

Рассмотрим пример. Пусть $N = \{w_1, w_2, w_3\}$ и $p = 1$, т. е. имеются трое исполнителей и потоки одного типа. Будем считать, что в технологической сети имеются четыре связи $f(w_{env}, w_1) = \lambda, f(w_1, w_1) = \lambda, f(w_2, w_3) = \lambda,$

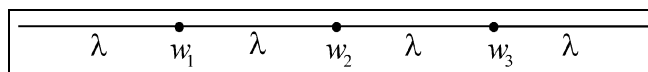


Рис. 1. Симметричная производственная линия

$f(w_3, w_{env}) = \lambda$, где λ — вектор интенсивности потоков. Эта технологическая сеть изображена на рис. 1. На рисунке не показана внешняя среда. Поэтому связи, соединяющие исполнителей w_1 и w_3 с внешней средой, имеют «висячий» конец. Данный пример может соответствовать производственной линии («бизнес-процессу»). Исполнитель w_1 принимает сырье от поставщика и передает результат исполнителю w_2 . Тот выполняет очередную технологическую операцию и передает результат далее. Последний исполнитель (в примере w_3), выполнив последнюю технологическую операцию, отгружает продукцию потребителю.

Сеть с исполнителями $N = \{w_1, \dots, w_n\}$ и потоками $f(w_{env}, w_1) = \lambda, f(w_{i-1}, w_i) = \lambda$ для всех $2 \leq i \leq n, f(w_n, w_{env}) = \lambda$ будем ниже называть *симметричной производственной линией*⁴, а λ — *интенсивностью линии*. В *несимметричной производственной линии* возможны потоки различной интенсивности. Изменение интенсивности может быть вызвано особенностями взаимодействия между исполнителями на различных этапах производства.

Обозначим через M конечное множество *менеджеров*, управляющих взаимодействием исполнителей. Менеджеры обычно будут обозначаться через $m, m', m'', m_1, m_2, \dots \in M$. Пусть $V = N \cup M$ — все множество *сотрудников организации* (исполнителей и менеджеров). Рассмотрим множество *ребер подчиненности* $E \subseteq V \times M$. Ребро $(v, m) \in E$ означает, что сотрудник $v \in E$ является *непосредственным подчиненным* менеджера $m \in M$, а m — *непосредственным начальником* сотрудника v , т. е., ребро направлено от непосредственного подчиненного к его непосредственному начальнику. Сотрудник $v \in V$ является *подчиненным* менеджера $m \in M$ (менеджер m является *начальником* сотрудника v), если существует цепочка ребер подчиненности из v в m . Будем также говорить, что начальник *управляет* подчиненным, или подчиненный *управляется* начальником. Теперь можно дать строгое определение иерархии.

Определение 1. Ориентированный граф $H = (N \cup M, E)$ с множеством менеджеров M и множеством ребер $E \subseteq (N \cup M) \times M$ назовем *иерархией*, управляющей множеством исполнителей N , если граф H ациклический, любой менеджер имеет подчиненных и найдется менеджер, которому подчинены все исполнители. Через $\Omega(N)$ обозначим множество всех иерархий. ♦

Ациклическость означает, что не существует «порочного круга» подчиненности, в котором каждый менеджер подчинен всем остальным. Определение также исключает ситуации, в которых имеются «менеджеры» без подчиненных. Существование менеджера, которому подчинены все исполнители, означает, что у любого

² Подобная модель представляется интересной, поскольку оптимальная иерархия во многом определяется именно взаимодействием исполнителей (см., например, работу [13]).

³ В работе технологическая сеть считается *неориентированной*, поскольку в рассматриваемой модели направление потока не играет роли. Также предполагаем, что сеть не содержит петель.

⁴ Все остальные потоки подразумеваются равными нулю.

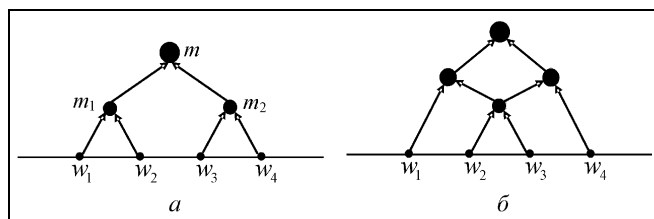


Рис. 2. Примеры иерархий над производственной линией: «классический» вид (а) и с множественным подчинением (б)

множества исполнителей найдется общий начальник, т. е. иерархия способна управлять взаимодействием всех исполнителей.

На рис. 2 приведены примеры двух иерархий, надстроенных над производственной линией из четырех исполнителей. Как видно из рис. 2, а, иерархия имеет «классический» вид — у каждого сотрудника ровно один непосредственный начальник (за исключением начальника верхнего уровня). В иерархии (см. рис. 2, б) присутствует множественное подчинение, кроме того некоторым начальникам непосредственно подчинены и менеджеры, и исполнители.

Группой исполнителей $s \subseteq N$ назовем любое непустое подмножество множества исполнителей. По определению 1 в любой иерархии H каждый менеджер имеет, по крайней мере, одного непосредственного подчиненного. Начав с любого менеджера m , мы можем двигаться по иерархии «сверху вниз» к подчиненным менеджера m . В итоге можно определить множество исполнителей, подчиненных менеджеру m . Будем называть это множество *подчиненной группой исполнителей* и обозначать $s_H(m) \subseteq N$. Будем также говорить, что менеджер m управляет группой исполнителей $s_H(m)$. В силу ацикличности каждому менеджеру подчинен хотя бы один исполнитель. Поэтому каждый менеджер управляет непустой группой исполнителей. Далее в обозначении группы $s_H(m)$ будем опускать нижний индекс, если ясно, о какой иерархии идет речь. Для удобства дальнейшего изложения будем считать, что в любой иерархии $H \in \Omega(N)$ любому исполнителю $w \in N$ «подчинена» простейшая группа $s_H(w) = \{w\}$, состоящая из самого исполнителя.

Очевидно, что для любой иерархии H и любого менеджера $m \in M$ выполнено $s_H(m) = s_H(v_1) \cup \dots \cup s_H(v_k)$, где $v_1, \dots, v_k$ — все непосредственные подчиненные менеджера m . Для любого подчиненного v менеджера m выполнено $s_H(v) \cup s_H(m)$. Например, на рис. 2, а менеджеру m непосредственно подчинены менеджеры m_1 и m_2 . Менеджеру m подчинена группа $s(m) = \{w_1, w_2, w_3, w_4\}$. Менеджерам m_1 и m_2 подчинены группы $s(m_1) = \{w_1, w_2\}$ и $s(m_2) = \{w_3, w_4\}$ соответственно. Таким образом, группа $s(m)$ разбивается на две подгруппы $s(m_1)$ и $s(m_2)$: $\{w_1, w_2, w_3, w_4\} = \{w_1, w_2\} \cup \{w_3, w_4\}$. В данном примере подгруппы не пересекаются. В общем случае, как показано на рис. 2, б, пересечения могут иметь место.

Определение 2. Иерархию назовем деревом, если в ней только один менеджер m не имеет начальников, а все остальные сотрудники имеют ровно одного непо-

средственного начальника. Менеджера m будем называть корнем дерева. ♦

На рис. 2, а изображен пример дерева. Напротив, иерархия на рис. 2, б деревом не является, так как в ней один менеджер имеет двух непосредственных начальников. Легко показать, что в дереве непосредственные подчиненные любого менеджера управляют непересекающимися группами исполнителей (не «дублируют» обязанности друг друга [12]).

Определение 3. Иерархию назовем r -иерархией, если у каждого ее менеджера не более r непосредственных подчиненных, где $r > 1$ — целое число; r -иерархию, которая является деревом, назовем r -деревом. ♦

В литературе по менеджменту часто используется термин *норма управляемости* — максимальное число непосредственных подчиненных, которыми может управлять один менеджер. Определение r -иерархии соответствует норме управляемости, равной r . Максимальную среди всех деревьев норму управляемости имеет *двух-уровневая иерархия*, в которой одному менеджеру непосредственно подчинены все исполнители.

2. ФУНКЦИЯ ЗАТРАТ МЕНЕДЖЕРА И ОПТИМАЛЬНАЯ ИЕРАРХИЯ

Чтобы исполнители взаимодействовали друг с другом в соответствии с технологической сетью необходимо управление их взаимодействием. Каждый менеджер управляет потоками между подчиненными исполнителями. Одна из интерпретаций работы менеджера — управление реализацией планов. Топ-менеджеры формулируют оперативный план, который необходимо реализовать. План может включать в себя дневной или недельный объем продаж и закупок — потоки между исполнителями и внешней средой. В процессе уточнения плана менеджеры на каждом уровне детализируют те его части, за которые они ответственны. Например, для обеспечения объема продаж, запланированного топ-менеджером, директор по производству может планировать соответствующие производственные потоки. После уточнения на всех уровнях детализированный план реализуется исполнителями. При этом каждый менеджер отслеживает реализацию своих планов. Приведем пример. Предположим, что в иерархии, изображенной на рис. 2, а, в результате конфликта между исполнителями w_2 и w_3 фактический поток между ними меньше необходимого значения потока $f(w_2, w_3)$. Исполнитель w_2 сообщает своему непосредственному начальнику m_1 , что у него возникли проблемы. Менеджер m_1 не в состоянии разрешить конфликт, так как исполнитель w_3 ему не подчинен. Аналогично, менеджер m_2 не в состоянии самостоятельно справиться с конфликтом, о котором ему сообщил исполнитель w_3 . В итоге менеджеры m_1 и m_2 сообщают о конфликте своему непосредственному начальнику m , который и примет решение, ликвидирующее конфликт. Это решение менеджеры m_1 и m_2 передадут соответственно исполнителям w_2 и w_3 . Аналогично можно рассмотреть планирование потока $f(w_2, w_3)$. Менеджер m передает план потока $f(w_2, w_3)$ менеджерам m_1 и m_2 , которые доводят план до исполнителей w_2 и w_3 , соответ-



ственно. Факт выполнения плана доводится до менеджера m в обратном порядке.

В управлении потоком $f(w_2, w_3)$ задействованы менеджеры m_1, m_2 и m . В управлении потоком $f(w_1, w_2)$ задействован только менеджер m_1 , так как он самостоятельно принимает все решения, связанные с потоком $f(w_1, w_2)$. Аналогично, в управлении потоком $f(w_3, w_4)$ задействован только менеджер m_2 . В управлении внешним потоком $f(w_{env}, w_1)$ участвуют менеджеры m_1 и m . Например, план закупок определяется менеджером m , уточняется менеджером m_1 и передается исполнителю w_1 . Аналогично, в управлении внешним потоком $f(w_4, w_{env})$ участвуют менеджеры m_2 и m . Введем формальное определение обязанностей менеджера.

Определение 4. В иерархии H менеджер m выполняет обязанности двух типов:

1) управляет потоками $f(w', w'')$ между подчиненными исполнителями $w', w'' \in s_H(m)$, которые не управляются ни одним подчиненным менеджера m ; сумму таких потоков назовем внутренним потоком менеджера m и обозначим $F_H^{int}(m)$;

2) участвует в управлении потоками $f(w', w'')$ между подчиненным исполнителем $w' \in s_H(m)$ и неподчиненным исполнителем $w'' \in N \setminus s_H(m)$ или внешней средой $w'' = w_{env}$; сумму таких потоков назовем внешним потоком менеджера m и обозначим $F_H^{ext}(m)$. ♦

Таким образом, менеджер управляет внутренним потоком и участвует в управлении внешним. *Потоком менеджера* назовем сумму его внутренних и внешних потоков. Из определения следует, что внешний и внутренний потоки менеджера m равны:

$$F_H^{ext}(m) = \sum_{\substack{w' \in s_H(m), \\ w'' \in (N \setminus s_H(m)) \cup \{w_{env}\}}} f(w', w''),$$

$$F_H^{int}(m) = \sum_{\substack{(w', w'') \subseteq s_H(m), \\ (w', w'') \not\subseteq s_H(v_j) \text{ для всех } 1 \leq j \leq k}} f(w', w''). \quad (1)$$

При подсчете внутреннего потока необходимо суммировать потоки внутри группы $s_H(m)$, которые не управляются непосредственными подчиненными. При заданных N и f внутренний и внешний поток менеджера m зависит только от $s_H(v_1), \dots, s_H(v_k)$, т. е. от групп исполнителей, которыми управляют непосредственные подчиненные менеджера m .

Определение 5. Затратами менеджера $m \in M$ в иерархии $H \in \Omega(N)$ назовем величину

$$c(s_H(v_1), \dots, s_H(v_k)) = \varphi(F_H^{int}(m) + F_H^{ext}(m)), \quad (2)$$

где $v_1, \dots, v_k$ — все непосредственные подчиненные менеджера m , $s_H(v_1), \dots, s_H(v_k)$ — управляемые ими группы, $\varphi: R_+^p \rightarrow R_+$ — монотонно неубывающая по всем переменным функция, ставящая в соответствие вектору $F_H^{int}(m) + F_H^{ext}(m)$ потока неотрицательное действительное число. ♦

Затраты менеджера определяются функцией $\varphi(\cdot)$, зависящей от потоков менеджера. Неубывание функции $\varphi(\cdot)$ означает, что при увеличении одной или нескольких компонент потока, т. е. при увеличении «объема» управленческой работы, затраты на управление не могут снизиться.

Определение 6. Затратами иерархии $H = (N \cup M, E) \in \Omega(N)$ назовем сумму затрат всех ее менеджеров

$$c(H) = \sum_{m \in M} c(s_H(v_1), \dots, s_H(v_k)) =$$

$$= \sum_{m \in M} \varphi(F_H^{int}(m) + F_H^{ext}(m)),$$

где $v_1, \dots, v_k$ — все непосредственные подчиненные менеджера m . Оптимальной иерархией назовем иерархию H^* , затраты которой минимальны: $H^* \in \text{Arg min}_{H \in \Omega} c(H)$. ♦

Оптимальных иерархий может быть несколько. Далее мы будем решать задачу поиска одной из оптимальных иерархий. При этом множество исполнителей N предполагается известным. Необходимо найти иерархию (т. е. определить число менеджеров и их подчиненность) из $\Omega(N)$, минимизирующую затраты на управление исполнителями. Согласно формулам (1) функция затрат менеджера (2) зависит только от групп $s_H(v_1), \dots, s_H(v_k)$, управляемых непосредственными подчиненными $v_1, \dots, v_k$. Функции такого вида названы в работе [12] секционными. В настоящей работе рассматривается частный случай секционной функции затрат.

3. ОПТИМАЛЬНАЯ ИЕРАРХИЯ УПРАВЛЕНИЯ СИММЕТРИЧНОЙ ПРОИЗВОДСТВЕННОЙ ЛИНИЕЙ

Рассмотрим задачу надстройки оптимальной иерархии над симметричной производственной линией (см. рис. 1), которая является простейшей сетью. Вдоль линии движется некоторый поток. Например, первый исполнитель получает сырье и передает полуфабрикат второму исполнителю. Аналогичным образом материальный поток движется далее вплоть до последнего исполнителя, который отгружает готовую продукцию. Сопутствующие информационные и прочие типы потоков также можно учитывать, рассматривая многокомпонентный поток ($p > 1$). Симметричная производственная линия накладывает на технологическую сеть два важных ограничения:

- потоки обрабатываются последовательно — каждый исполнитель взаимодействует только с предыдущим и со следующим исполнителем в линии;
- значение потока между всеми исполнителями одинаково, т. е. интенсивность взаимодействия не меняется на протяжении всей линии.

Очевидно, что на практике эти ограничения могут нарушаться. Могут существовать различные технологические маршруты, некачественная продукция может возвращаться на доработку, интенсивность потоков может возрастать и снижаться (например, учесть заготовку одного типа в начале линии может быть гораздо проще, чем десятки различных деталей, образующихся в результате обработки). Однако в некоторых случаях технологическая сеть может быть близка к симметричной линии. Такой вид сети позволяет найти оптимальную

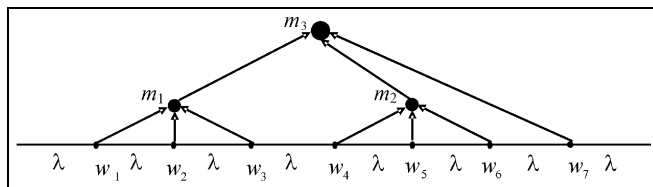


Рис. 3. Пример иерархии над симметричной производственной линией

иерархию. Если технологическая сеть значительно сложнее симметричной линии, то для поиска оптимальной иерархии могут быть применены общие методы, описанные в работе [12]. Далее считаем, что исполнители $N = \{w_1, \dots, w_n\}$ связаны технологической сетью $f(w_{env}, w_1) = \lambda, f(w_{i-1}, w_i) = \lambda$ для всех $2 \leq i \leq n, f(w_n, w_{env}) = \lambda$.

Утверждение. Существует оптимальное дерево H , управляющее симметричной производственной линией, и обладающее следующими свойствами:

- 1) в дереве H каждый менеджер управляет группой исполнителей, идущих в линии последовательно;
- 2) если функция затрат выпуклая, то у различных менеджеров дерева H число непосредственных подчиненных отличается не более чем на единицу. ♦

Доказательство приведено в Приложении.

В соответствии с этим утверждением, при поиске оптимальной иерархии можно не рассматривать недревовидные иерархии, ограничиваясь классом деревьев. В то же время в работе [12] приведен пример, показывающий, что это утверждение не верно для несимметричной линии. В силу свойства 1 утверждения можно рассматривать только такие деревья, в которых каждому менеджеру подчинена группа из исполнителей, идущих подряд, и не рассматривать иерархии, в которых некоторому менеджеру подчинена, например, группа $\{w_1, w_2, w_4\}$. Если на рис. 3 менеджеру m_1 вместо исполнителя w_3 подчинить исполнителя w_4 , а менеджеру m_2 вместо исполнителя w_4 подчинить исполнителя w_3 , то получим следующие результаты.

- Затраты менеджеров m_1 и m_2 возрастут.
- Менеджер m_1 перестанет управлять потоком $f(w_2, w_3)$, а будет лишь участвовать в управлении⁵. Менеджер m_2 перестанет управлять потоком $f(w_4, w_5)$, а будет лишь участвовать в управлении. В результате поток, управляемый вышестоящим менеджером m_3 , возрастет на $f(w_2, w_3) + f(w_4, w_5)$. Следовательно, возрастут затраты менеджера m_3 .

Пример иллюстрирует результат утверждения — подчинение менеджеру группы исполнителей, которые идут в линии не подряд, увеличивает затраты всей иерархии. Содержательная интерпретация этого свойства очевидна. Каждый менеджер должен управлять одним участком производственной линии (последовательно идущими исполнителями). Попытка подчинить менеджеру несвязанные части производства увеличит затраты иерархии и приведет к неоптимальности.

⁵ То есть этот поток станет для менеджера внешним.

Рассмотрим рис. 3. Менеджеру m_1 подчинена группа $\{w_1, w_2, w_3\}$ из трех исполнителей. В группу $\{w_1, w_2, w_3\}$ входит поток $f(w_{env}, w_1)$ из внешней среды и выходит поток $f(w_3, w_4)$. После назначения менеджера m_1 он с точки зрения затрат вышестоящих менеджеров ничем не отличается от исполнителя. Фактически, подчинив менеджеру m_1 трех исполнителей, мы «сократим» длину производственной линии на два, поскольку три исполнителя заменятся на одного менеджера. Менеджер m_2 снова «сократит» число звеньев производственной линии на два. При этом менеджеру m_2 можно было бы подчинить менеджера m_1 и двух исполнителей, что привело бы к тем же затратам. Однако подобное подчинение приведет к росту числа уровней иерархии, поэтому предпочтительнее вариант, изображенный на рис. 3.

Если в дереве H у менеджера m имеется k непосредственных подчиненных, то группа $s_H(m)$ разбивается на k подгрупп. Таким образом, некоторый участок производственной линии разбивается на k подучастков. Тогда менеджер m управляет $k - 1$ внутренними потоками и участвует в управлении двумя внешними потоками. Следовательно, если в дереве любой менеджер управляет исполнителями, идущими в линии подряд, то затраты менеджера с k непосредственными подчиненными определяются формулой:

$$\varphi((k + 1)\lambda). \quad (3)$$

Подчинив менеджеру m_1 некоторое число r_1 исполнителей, мы фактически сократим число элементов производственной линии на $r_1 - 1$ (r_1 исполнителей заменятся на одного менеджера). Аналогично, можно назначить менеджера m_2 , подчинив ему r_2 еще не подчиненных исполнителей или менеджеров, и т. д. В итоге должен остаться единственный неподчиненный менеджер m_q (напомним, что q — общее число менеджеров), т. е. $n - (r_1 - 1) - (r_2 - 1) - \dots - (r_q - 1) = 1$. Переписывая данное равенство, получим следующее ограничение на число непосредственных подчиненных всех менеджеров дерева:

$$r_1 + \dots + r_q = n + q - 1. \quad (4)$$

По формуле (3) затраты менеджеров равны $\varphi((r_1 + 1)\lambda), \dots, \varphi((r_q + 1)\lambda)$. Для решения задачи об оптимальной иерархии осталось решить следующую задачу оптимизации:

$$\varphi((r_1 + 1)\lambda) + \dots + \varphi((r_q + 1)\lambda) \rightarrow \min, \quad (5)$$

при условиях (4), $r_1, \dots, r_q \geq 2, 1 \leq q \leq n - 1$.

Итак, для симметричной производственной линии задача об оптимальной иерархии сводится к задаче (5) условной оптимизации функции, зависящей от q целочисленных переменных (такую задачу необходимо решить при каждом q). Для решения задачи (5) можно применить классические методы дискретной оптимизации или предложенные в работе [9] алгоритмы поиска оптимальных деревьев для произвольной секционной функции.

В соответствии с утверждением для выпуклых функций задача (5) решается аналитически. В оптимальном дереве у различных менеджеров число непосредственных подчиненных отличается не более чем на единицу, т. е. величины $r_1, \dots, r_q$ отличаются не более чем на еди-



ницу⁶. Пусть r — минимальное число непосредственных подчиненных менеджера. Тогда у любого менеджера либо r , либо $r + 1$ непосредственных подчиненных. Пусть $q_1 > 0$ — число менеджеров первого типа (с r непосредственными подчиненными). Тогда $q_2 = q - q_1$ — число менеджеров второго типа (с $r + 1$ непосредственными подчиненными). Левая часть выражения (4) имеет вид $q_1 r + q_2(r + 1) = qr + q_2$, т. е. выполнено условие:

$$qr + q_2 = n + q - 1, r = \lfloor (n + q - 1)/q \rfloor^7. \quad (6)$$

Если $n - 1$ делится на q нацело, то $q_2 = 0$, и у всех менеджеров одинаковое число $r = (n + q - 1)/q$ непосредственных подчиненных. Иначе согласно равенству (6) r — ближайшее снизу целое число, а q_2 — остаток от деления $n - 1$ на q .

В случае выпуклой функции при фиксированном q из формулы (6) определяются параметры $r_1, \dots, r_q$ и по формуле (5) вычисляются затраты дерева. Поэтому для решения задачи об оптимальной иерархии остается лишь найти оптимальное число менеджеров $1 \leq q \leq n - 1$. Это можно сделать с помощью сравнения $n - 1$ вариантов. Иначе говоря, для любой выпуклой функции можно найти оптимальную иерархию над симметричной производственной линией путем сравнения затрат $n - 1$ дерева. Вариант $q = 1$ соответствует двухуровневой иерархии с максимальным числом непосредственных подчиненных $r = n$. Вариант $q = n - 1$ соответствует двухуровневой иерархии с минимальным числом непосредственных подчиненных $r = 2$. Для многих выпуклых функций $\varphi(\cdot)$ задачу можно решить аналитически, не сравнивая численно затрат $n - 1$ дерева. Например, в работе [12] задача решена для важного частного случая — степенной функции затрат (для нее существует оптимальная норма управляемости, которая не зависит от n и λ).

ЗАКЛЮЧЕНИЕ

Доказано, что при поиске оптимальной иерархии управления симметричной производственной линией можно не рассматривать недревовидные иерархии, ограничиваясь классом деревьев, в которых каждый менеджер управляет одним участком производственной линии, т. е. последовательно идущими исполнителями. Таким образом, задача значительно упрощается, поскольку для ее решения достаточно найти среди указанных деревьев наилучшее — дерево с минимальными затратами. Эта задача сведена к задаче условной оптимизации функции, зависящей от q целочисленных переменных. В случае выпуклой функции затрат число непосредственных подчиненных менеджеров отличается не более, чем на единицу. Для решения задачи об оптимальной иерархии достаточно найти оптимальное число менеджеров, сравнив затраты $n - 1$ деревьев, т. е. задача оптимизации становится тривиальной.

⁶ Если значения отличаются на два и более, то в соответствии с доказательством утверждения одно можно уменьшить, а другое увеличить, без увеличения затрат дерева.

⁷ Символ $\lfloor (n + q - 1) \rfloor$ обозначает нижнее целое — максимальное целое число, не превышающее $(n + q - 1)/q$.

Доказательство утверждения. Пусть $M = \{m_1, \dots, m_q\}$ — множество менеджеров, которые управляют всеми потоками симметричной производственной линии с минимальными суммарными затратами. Обозначим через k_i число внутренних потоков, которыми управляет менеджер m_i , $1 \leq i \leq q$. Обозначим через l_i число внешних потоков, в управлении которыми участвует менеджер m_i . Тогда суммарный внутренний поток менеджера m_i равен $F^{int}(m_i) = \lambda k_i$, внешний поток равен $F^{ext}(m_i) = \lambda l_i$. Затраты менеджера m_i равны $\varphi((k_i + l_i)\lambda)$. Затраты всех менеджеров из M составят $\varphi((k_1 + l_1)\lambda) + \dots + \varphi((k_q + l_q)\lambda)$.

Любой менеджер m_i участвует в управлении, по меньшей мере, двумя внешними потоками. Действительно, пусть $w_k \in N$ — исполнитель с наименьшим номером, подчиненный менеджеру m_i . Тогда поток $f(w_{k-1}, w_k)$ (или $f(w_{env}, w_1)$ при $k = 1$) будет внешним для менеджера m_i . Аналогично можно рассмотреть исполнителя с наибольшим номером, т. е. $l_i \geq 2$. Рассмотрим $n - 1$ потоков $f(w_{i-1}, w_i) = \lambda$ для всех $2 \leq i \leq n$. Каждый такой поток управляется одним из менеджеров $m_1, \dots, m_q$ и будет внутренним, по крайней мере, для одного из менеджеров. Таким образом, выполнено неравенство $k_1 + \dots + k_q \geq n - 1$. Кроме того, $k_i \leq n - 1$.

Построим дерево $H = (N \cup M', E')$ следующим образом. В начале менеджеров нет и каждый из n исполнителей должен быть непосредственно подчинен ровно одному менеджеру. Менеджеру m'_1 непосредственно подчиним $k_1 + 1$ исполнителей, идущих в цепи подряд, начиная с первого, т. е. $s_H(m'_1) = \{w_1, \dots, w_{k_1+1}\}$. Менеджер m'_1 и каждый из исполнителей $w_{k_1+2}, \dots, w_n$ должны быть непосредственно подчинены ровно одному менеджеру. После первого шага остались неподчиненными $n - k_1$ сотрудников. Менеджеру m'_2 непосредственно подчиним менеджера m'_1 и k_2 исполнителей $w_{k_1+2}, \dots, w_{k_1+k_2+1}$. После второго шага остались неподчиненными $n - k_1 - k_2$ сотрудников. Продолжая подобные действия, можно придти к двум результатам:

1) при $k_1 + \dots + k_q = n - 1$ будут назначены $q' = q$ менеджеров; причем менеджеру m'_q будет непосредственно подчинен менеджер m'_{q-1} и k_q исполнителей $w_{k_1+\dots+k_{q-1}+2}, \dots, w_n$, еще оставшихся неподчиненными;

2) при $k_1 + \dots + k_q > n - 1$ будут назначены $q' \leq q$ менеджеров, причем последнему менеджеру $m'_{q'}$ будет непосредственно подчинен менеджер $m'_{q'-1}$ и не более $k_{q'}$ исполнителей, еще оставшихся неподчиненными.

В обоих случаях менеджер $m'_{q'}$ будет управлять всеми исполнителями, т. е. получили дерево $H \in \Omega(N)$. По построению каждый менеджер дерева управляет груп-

пой исполнителей, которые идут в цепи последовательно. Рассмотрим менеджера m'_i , $1 \leq i \leq q'$. Обозначим всех его непосредственных подчиненных $v_1, \dots, v_j$. Сотрудники $v_1, \dots, v_j$ управляют попарно непересекающимися группами, причем $s_H(m'_i) = s_H(v_1) \cup \dots \cup s_H(v_j)$, т. е. управляемый менеджером m'_i участок цепи $s_H(m'_i)$ разбивается на части $s_H(v_1), \dots, s_H(v_j)$. Тогда менеджер m'_i управляет $j - 1$ внутренними потоками и участвует в управлении двумя внешними потоками. Затраты менеджера m'_i равны $\varphi((j + 1)\lambda)$. Менеджер m'_i имеет не более $k_i + 1$ непосредственного подчиненного. Поэтому затраты менеджера m'_i не превосходят величины $\varphi((k_i + 2)\lambda)$. Таким образом,

$$c(H) \leq \varphi((k_1 + 2)\lambda) + \dots + \varphi((k_q + 2)\lambda) \leq \varphi((k_1 + 1)\lambda) + \dots + \varphi((k_q + 1)\lambda).$$

Дерево H может содержать меньше q менеджеров. В силу неотрицательности функции затрат дополнительные слагаемые не могут уменьшить затраты. Учитывая оценку затрат менеджера m'_i , докажем первое неравенство. Второе неравенство справедливо в силу монотонности функции затрат и условия $l_i \geq 2$.

В построенном дереве H не больше менеджеров, чем в M , причем затраты менеджеров дерева не превосходят затраты соответствующих менеджеров M , т. е. построено дерево, затраты которого не превышают суммарных затрат любых менеджеров, управляющих всеми потоками симметричной производственной линии, следовательно H — оптимальное дерево.

Пусть $\varphi(\cdot)$ — выпуклая функция. Для каждого менеджера m'_i дерева H через k'_i обозначим число непосредственных подчиненных, $1 \leq i \leq q'$. Если найдутся два менеджера, у которых число непосредственных подчиненных отличается более, чем на единицу, то для некоторых $1 \leq i, j \leq q' - 1$ выполнено $k'_i + 1 < k'_j$. При описанном ранее построении дерева мы можем назначать менеджеров $m'_1, \dots, m'_{q'}$ в любом порядке. Их затраты не зависят от порядка назначения. Первым можно назначить менеджера m'_i с k'_i непосредственными подчиненными, а вторым можно назначить менеджера m'_j с k'_j непосредственными подчиненными. Всех остальных менеджеров назначим в произвольном порядке. Таким образом, в новом дереве числа $k'_1, \dots, k'_{q'}$ будут переставлены, т. е. в новом дереве (после перестановки) выполнено неравенство $k'_1 + 1 < k'_2$. По построению дерева менеджеру m'_2 непосредственно подчинен исполнитель, соседний с группой $s_H(m'_1)$. Тогда можно переподчинить этого исполнителя менеджеру m'_1 . Это по-прежнему приведет к дереву, где каждый менеджер управляет группой исполнителей, которые идут в цепи последовательно. Изменились только затраты менеджеров m'_1 и m'_2 , у которых теперь $k'_1 + 1$ и $k'_2 - 1$ непосредственных подчиненных. Далее доказано, что затраты не увеличились. Можно продолжить аналогичные

переподчинения. В результате получим дерево, в котором число непосредственных подчиненных у менеджеров m'_1 и m'_2 одинаково или отличается на единицу. Кроме того, уменьшился «разброс» величин $k'_1, \dots, k'_{q'}$ (например, среднее квадратическое отклонение). Продолжая подобные действия, построим в итоге дерево, в котором у различных менеджеров число непосредственных подчиненных отличается не более чем на единицу. Затраты полученного дерева не превышают затрат исходного дерева, т. е. выполнено условие утверждения.

Для доказательства утверждения осталось показать, что выполнено неравенство $\varphi((k'_1 + 2)\lambda) + \varphi(k'_2\lambda) \leq \varphi((k'_1 + 1)\lambda) + \varphi((k'_2 + 1)\lambda)$. Обозначим $z_1 = k'_1 + 1$, $z_2 = k'_2 + 1$. По определению выпуклой функции для любых $z_1, z_2 \in R_+$ и для любого $\gamma \in [0, 1]$ выполнено неравенство $\varphi(\gamma z_1 + (1 - \gamma)z_2) \leq \gamma \varphi(z_1) + (1 - \gamma)\varphi(z_2)$. Положим $\gamma_1 = (z_2 - z_1 - 1)/(z_2 - z_1)$, $\gamma_2 = 1/(z_2 - z_1)$. Имеем, $0 < \gamma_1, \gamma_2 < 1$, $\gamma_1 + \gamma_2 = 1$. Кроме того, $\gamma_1 z_1 + (1 - \gamma_1)z_2 = z_1 + 1$, $\gamma_2 z_1 + (1 - \gamma_2)z_2 = z_2 - 1$. Подставляя значения γ_1 и γ_2 в неравенство, получим: $\varphi((z_1 + 1)\lambda) \leq \gamma_1 \varphi(z_1\lambda) + (1 - \gamma_1)\varphi(z_2\lambda)$, $\varphi((z_1 - 1)\lambda) \leq \gamma_2 \varphi(z_1\lambda) + (1 - \gamma_2)\varphi(z_2\lambda)$. Сложив неравенства, будем иметь $\varphi((z_1 + 1)\lambda) + \varphi((z_2 - 1)\lambda) \leq \varphi(z_1\lambda) + \varphi(z_2\lambda)$, что и доказывает требуемое неравенство. ♦

ЛИТЕРАТУРА

- Бурков В. Н., Новиков Д. А. Теория активных систем: состояние и перспективы. — М.: СИНТЕГ, 1999.
- Новиков Д. А. Сетевые структуры и организационные системы. — М.: ИПУ РАН, 2003.
- Новиков Д. А. Механизмы функционирования многоуровневых организационных систем. — М.: Фонд «Проблемы управления», 1999.
- Овсиевич Б. И. Модели формирования организационных структур. — Л.: Наука, 1979.
- Цвиркун А. Д. Основы синтеза структуры сложных систем. — М.: Наука, 1982.
- Задачи оптимизации иерархических структур / В. Т. Деметьев, А. И. Ерзин, Р. М. Ларин и др. — Новосибирск: Изд-во Новосиб. ун-та, 1996.
- Воронин А. А., Мишин С. П. Модель оптимального управления структурными изменениями организационной системы // Автоматика и телемеханика. — 2002. — № 8. — С. 136—150.
- Воронин А. А., Мишин С. П. Алгоритмы поиска оптимальной структуры организационной системы // Автоматика и телемеханика. — 2002. — № 5. — С. 120—132.
- Воронин А. А., Мишин С. П. Оптимальные иерархические структуры. — М.: ИПУ РАН, 2003.
- Губко М. В. Структура оптимальной организации континуума исполнителей // Автоматика и телемеханика. — 2002. — № 12. — С. 116—130.
- Мишин С. П. Оптимальное стимулирование в многоуровневых иерархических структурах // Автоматика и телемеханика. — 2004. — № 5. — С. 96—119.
- Мишин С. П. Оптимальные иерархии управления в экономических системах. — М.: ИПУ РАН, 2004.
- Mintzberg H. The structuring of organizations. — Englewood Cliffs: Prentice-Hall, 1979.

☎ (495) 377-84-19;

e-mail: smishin@newmail.ru





СИСТЕМА ПОДДЕРЖКИ ПРИНЯТИЯ УПРАВЛЕНЧЕСКИХ РЕШЕНИЙ С ПРИМЕНЕНИЕМ ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ

Б. Б. Буянов, Н. В. Лубков, Г. Л. Поляк

Институт проблем управления РАН им. В. А. Трапезникова, г. Москва

Разработаны функциональная и алгоритмическая структуры системы поддержки принятия решений, включающей в себя математическую модель сценариев управленческой деятельности и многокритериальную модель сравнения альтернатив. Описан комплекс программ, реализующий разработанные процедуры и предоставляющий пользователю интерфейс, позволяющий в интерактивном режиме формировать сценарии управленческой деятельности и систему предпочтений для выбора наилучшего решения.

ВВЕДЕНИЕ

При рассмотрении процесса управления в различных областях деятельности — экономике, технике, военном деле и в чрезвычайных ситуациях — обнаруживается, что едва ли не самую большую трудность представляет проблема неопределенности, обусловленная неполнотой и искажением информации. При выборе стратегии, например, военных действий неопределенность возрастает из-за противодействия противника, при выборе политики экономического развития или технической политики необходимо учитывать возможное противодействие в конкурентной борьбе. Все это объективно приводит к необходимости рассмотрения различных сценариев действия в зависимости от предполагаемого характера неопределенности.

В работе [1] отмечено, что при принятии решений руководитель вынужден исходить из своих субъективных представлений об эффективности возможных альтернатив и важности различных критериев. Однако при принятии субъективных решений существует и объективная составляющая, которая выражается в предопределенности действий при заданном сценарии действий. Результативность (эффективность) каждого сценария действий определяется в некоторой системе показателей, оценки по которым служат основанием для выбора. Оценки показателей носят прогнозный характер, часть из них может быть получена экспертным путем, но наиболее надежный способ получения оценок заключается в применении математических моделей. Вид модели, а зачастую это комплекс моделей, зависит от предметной области, т. е. вида деятельности, в рамках которой возникает рассматриваемая проблема. Прогнозирующие модели позволяют получать также оценки развития

процесса по заданному сценарию во времени, что может представлять дополнительную информацию, влияющую на результат выбора.

Итак, лицу, принимающему решение (ЛПР), предъявляются для выбора варианты, представленные своими векторными оценками, отдельные координаты которых (частные показатели) должны соответствовать его понятиям представлениям о предметной области. Только в этом случае ЛПР сможет сформулировать свои предпочтения по каждому показателю, что позволяет применять уже определенные формализованные процедуры сравнения вариантов для выбора единственного варианта или, по крайней мере, сужения множества альтернативных.

Таким образом, можно определить, что система поддержки принятия решения (СППР) должна включать в себя два основных блока:

- модели прогнозирования оценок показателей рассматриваемых вариантов;
- модели и процедуры формализации предпочтений ЛПР и упорядочивание рассматриваемых вариантов в порядке предпочтительности.

Собственно задача формирования вариантов (сценариев) действий выходит за рамки СППР, хотя очевидно, что по результатам прогнозирования исходных вариантов появляется дополнительная информация для генерирования последующих.

В настоящей статье описана реализация СППР в виде программного комплекса. Приведен иллюстративный пример выбора стратегии управления в задаче столкновения двух противоборствующих условных военных группировок. Для построения модели прогнозирования для сценариев военных действий используются игровые модели [2], а также модели динамики боя [3] в виде диф-

ференциальных уравнений для математических ожиданий потерь сторон при различных условиях информированности.

1. ФУНКЦИОНАЛЬНАЯ СТРУКТУРА СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЯ С МАТЕМАТИЧЕСКОЙ МОДЕЛЮ

Для своевременной и обоснованной выработки решения необходима достоверная информация о нормативах и ограничениях, регламентирующих управленческую деятельность, о располагаемых ресурсах, о возможных стратегиях реализации решения. Эти данные располагаются в базе данных и составляет информационную основу СППР. Функциональная структура СППР представлена на рис. 1.

Остановимся на двух особенностях процедуры выбора, важных для дальнейшего изложения.

1.1. Выявление системы предпочтений ЛПР

В СППР лицу, принимающему решение, предлагается выбрать наиболее предпочтительное решение на основе интервальной шкалы полезности показателей [4]. Для построения интервальной шкалы показателей

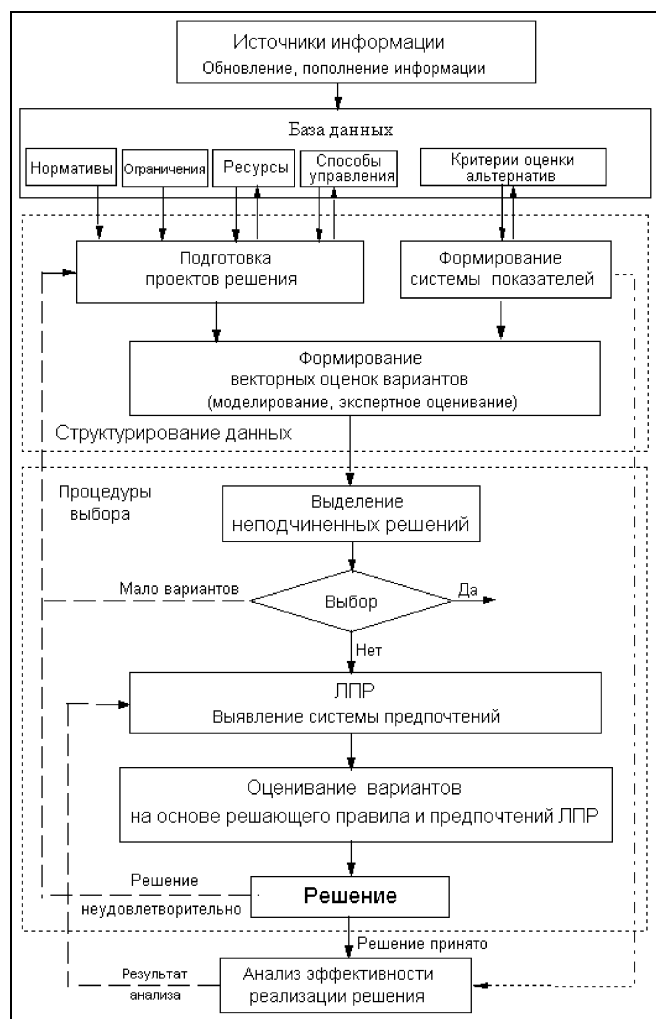


Рис. 1. Структура СППР

требуется провести парное сравнение интервалов на каждой шкале, на парах шкал и т. д. Но даже при небольшом числе вариантов и показателей (напомним, оценки вариантов образуют шкалу) возникает большое число возможных операций сравнения. Поэтому ЛПР предлагается определить относительную важность показателей и оценить характер изменения ценности свойства на шкале каждого показателя.

Относительная важность показателей определяется путем сравнения интервалов между максимальной и минимальной оценками вариантов на шкалах показателей.

Обозначим:

$s_i \in S, i = 1, 2, \dots, n, n \geq 2$, — альтернативы, варианты;

$X_j, j = 1, 2, \dots, m, m \geq 2, X_j \in X$ — система показателей;

$Y = X_1 \times X_2 \times \dots \times X_m$ — множество векторных оценок;

$\{x_{j1}, x_{j2}, \dots, x_{jk}\}$ — значения оценок вариантов $s_i \in S, i = 1, 2, \dots, n, r \leq n$ по показателю X_j ;

$(x_1^i, x_2^i, \dots, x_j^i, \dots, x_m^i)$ — векторная оценка y_i из множества Y варианта s_i , где x_j^i — оценка по j -му показателю;

$y_{\bar{j}\bar{k}}$ — элементы пространства векторных оценок $y_{\bar{j}\bar{k}} = X_1 \times \dots \times X_{j-1} \times X_{j+1} \times \dots \times X_{k-1} \times X_{k+1} \times \dots \times X_m$, не содержащего оценок по показателям X_j и X_k ;

$x_j^{\max}, x_j^{\min}$ и $x_k^{\max}, x_k^{\min}$ — максимальная и минимальная оценки вариантов по показателям X_j и X_k , соответственно.

Тогда можно дать следующие определения [4]:

— показатели X_j и X_k равноценны, если векторные оценки $(x_j^{\max}, x_k^{\min}, y_{\bar{j}\bar{k}})$ и $(x_j^{\min}, x_k^{\max}, y_{\bar{j}\bar{k}})$ одинаковы по предпочтению;

— показатель X_j важнее показателя X_k , если векторная оценка $(x_j^{\max}, x_k^{\min}, y_{\bar{j}\bar{k}})$ предпочтительнее векторной оценки $(x_j^{\min}, x_k^{\max}, y_{\bar{j}\bar{k}})$.

Если ЛПР сможет упорядочить интервалы между максимальной и минимальной оценками вариантов по шкалам показателей, то интервалам можно будет приписать числа так, что одинаковым по предпочтению интервалам будут соответствовать два равных числа, а более предпочтительному интервалу — большее число. Соответственно, числа, отражающие упорядочение интервалов и имеющие смысл важности показателя, можно сопоставить высшим оценкам по показателям, если принять за начало отсчета наименьшие оценки. Таким образом, значения, связанные с высшими оценками показателей, образуют основу шкалы ценности Z , общей для всех показателей.

Затем ЛПР выбирает вид монотонного изменения функции ценности (полезности) на шкале показателей из предлагаемого набора. Поскольку ценность высшей оценки по показателю определена в единой шкале при назначении важности показателя, выбор функции ценности позволяет СППР приписать соответствующие числа и остальным оценкам на шкале показателя и дополнить ими единую шкалу ценности Z .



Таким образом, осуществляется переход от векторных оценок u_i варианта s_i по шкалам показателей к векторным оценкам z_i по единой шкале ценности. При этом необходимо учитывать, что в интервальной шкале числа, присвоенные оценкам, только отражают отношения предпочтения между интервалами.

1.2. Выделение предпочтительных вариантов на основе решающего правила

Выделение наиболее предпочтительных вариантов производится по результатам парного сравнения векторных оценок, приведенных к единой шкале Z [5]. Каждая векторная оценка исходного множества вариантов S или множества неподчиненных решений S_π поочередно сравнивается с остальными векторными оценками. Для каждой пары векторных оценок определяется интервал между векторными оценками (z_i, z_j) по координатным вычитанием значений вектора. Численные оценки элементов вектора интервалов имеет смысл сравнивать только по отношениям «>», «=», «<», но нельзя суммировать, вычитать, умножать и т. д. Элементы вектора интервалов имеют разные знаки, в зависимости от того, какой векторной оценке — уменьшаемой или вычитаемой — принадлежит более высокая оценка по показателю.

Затем элементы вектора интервалов упорядочиваются по абсолютной величине. Равные по ценности интервалы противоположного знака можно исключить, а на упорядоченном ряду элементов вектора интервалов проверяется выполнение условий решающего правила. Если все интервалы попарно равноценны, то и варианты равноценны. Более предпочтительным может быть только тот вариант в паре, которому «принадлежит» элемент, занявший первое место в упорядоченном ряду, но при условии, что каждому элементу противоположного

знака можно поставить в соответствие элемент того же знака, что и первый, занимающий более высокое место в упорядоченном ряду. Если условие не выполняется, то пара вариантов несравнима.

Установленные отношения предпочтительности между вариантами позволяют осуществить в общем случае частичное упорядочение вариантов. Если выделено одно решение, то ЛПР остается либо согласиться с предложением СППР, либо отвергнуть его и начать всю процедуру сначала.

2. ПРОГРАММНЫЙ КОМПЛЕКС СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ С ИМИТАЦИОННОЙ МОДЕЛЬЮ

Программный комплекс позволяет автоматизировать решение основных задач, возникающих в процессе принятия решения:

- прогнозирование оценок вариантов в заданной системе показателей на основании результатов моделирования, формирование множества вариантов, подлежащих сравнению, и назначение показателей с целью выделения наиболее предпочтительного варианта;

- реализация процедур принятия решений в части получения от ЛПР в интерактивном режиме информации о предпочтениях, касающейся относительной важности показателей и формы кривой функции ценности (полезности), оценка и сравнение вариантов в соответствии с принятым решающим правилом и предоставление результатов сравнения лицу, принимающему решение, в упорядоченном виде.

Программный комплекс состоит из двух модулей: модуля моделирования функционирования системы и модуля сравнения альтернатив и принятия решения. Структурная схема программного комплекса представлена на рис. 2.

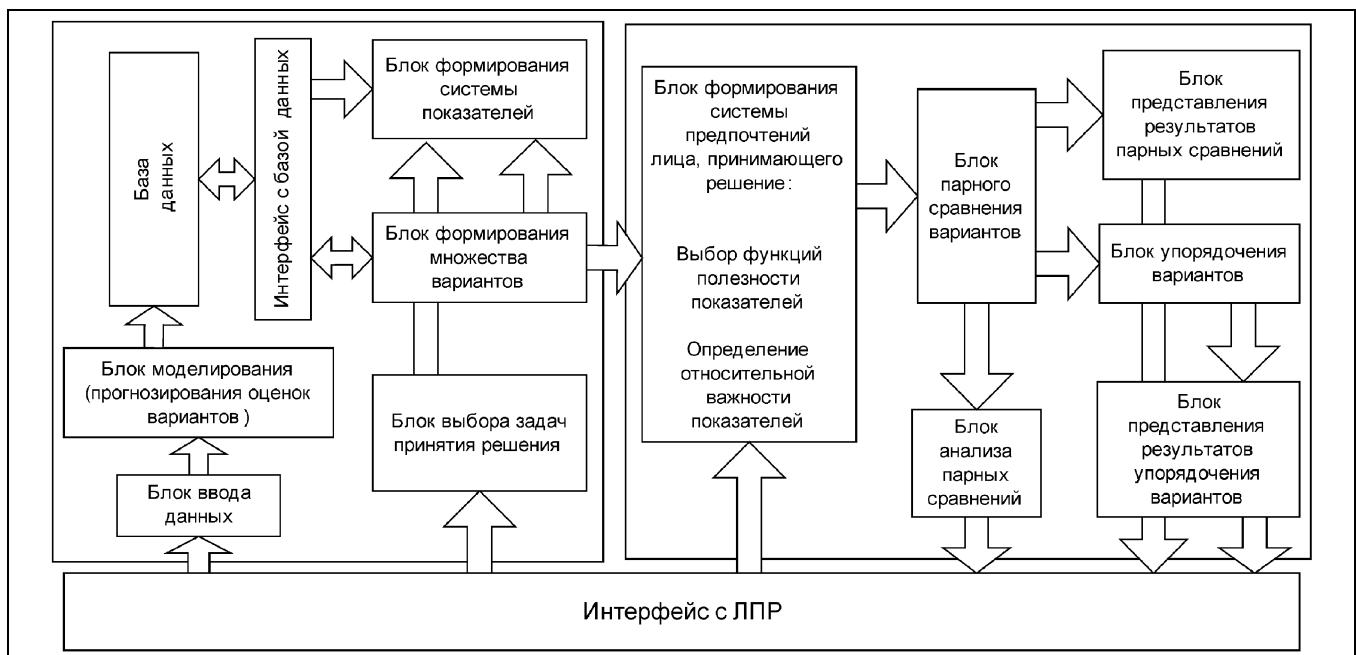


Рис. 2. Структурная схема программного комплекса

2.1. Модуль моделирования функционирования системы

Программный модуль предназначен для получения числовых оценок показателей, характеризующих варианты решения. В нем можно условно выделить следующие блоки.

Блок ввода данных для моделирования (задание характеристик исследуемого процесса).

Блок моделирования (реализует вычислительные процедуры).

Блок представления результатов моделирования (сохранения результатов в БД/файле, визуализация в табличной форме и в виде графиков).

Интерфейс с базой данных. Вся входная информация хранится в соответствующих таблицах базы данных. Интерфейс обеспечивает эффективную работу с таблицами базы данных — соединение с базой данных, формирование SQL-запросов.

База данных содержит полную информацию о результатах моделирования. Информация накапливается сегментами, содержащими результаты моделирования одной реализации, проведенной при фиксированных исходных данных решаемой задачи. В базе данных собирается информация о результатах моделирования при различных исходных данных, что позволяет использовать данные нескольких реализаций. В ней осуществляется выбор информации, необходимой для решения конкретной задачи принятия решения. Информация сохраняется и, при необходимости, преобразуется только в локальных таблицах блока. Помимо таблицы результатов, база данных содержит классификатор показателей и вспомогательные таблицы.

Блок выбора задач принятия решения позволяет ЛПР выбрать одну из нескольких постановок задач. Выбранный тип задачи предопределяет действия по формированию системы показателей и формированию множества вариантов.

Блок формирования системы показателей. Программа вызывает из базы данных показатели, участвующие в оценке вариантов, и признак свойства показателя повышать (или понижать) полезность варианта с ростом значений показателя.

Блок формирования множества вариантов. Программа извлекает из базы данных, в зависимости от задачи, либо одну реализацию, либо исходы нескольких реализаций, объединенных по определенному принципу.

2.2. Модуль сравнения альтернатив и принятия решений

Программный модуль предназначен для выявления системы предпочтений ЛПР, сравнения и упорядочения вариантов решения. В нем блоки выполняют следующие функции.

Интерфейс с ЛПР обеспечивает взаимодействие ЛПР с программным модулем в части выбора задачи принятия решения, восприятия и отображения предпочтений ЛПР, предоставление ему результатов сравнения вариантов при высказанных предпочтениях и средств анализа этих результатов.

Блок формирования предпочтений ЛПР предоставляет ЛПР графические средства для выражения своих пред-

почтений относительно важности показателей и характера изменения ценности свойства на шкале показателя. Упорядочивая показатели по относительной важности, ЛПР назначает некоторые числа высшим оценкам на шкалах показателей, соответствующие месту показателя в упорядоченном ряду, а, выбирая тот или иной вид функции полезности для каждой шкалы, обуславливает отношение порядка между всеми интервалами на шкалах показателей.

Блок парного сравнения вариантов. Здесь производится поочередно парное сравнение вариантов и регистрируются результаты сравнения.

Результаты парных сравнений представляются ЛПР для обозрения посредством *блока представления результатов парных сравнений* и для анализа.

Блок анализа парных сравнений позволяет ЛПР визуально оценить причины, по которым один вариант более или менее предпочтителен, чем другой.

Блок упорядочения вариантов. По результатам парных сравнений производится упорядочение множества неподчиненных вариантов. В общем случае упорядочение может быть частичным, так как не исключена несравнимость вариантов. Возникающее бинарное отношение на множестве вариантов характеризуется отсутствием циклов и, возможно, нетранзитивностью. Последняя обуславливается несравнимостью вариантов.

Блок представления результатов упорядочения вариантов. Полученные отношения предпочтения между вариантами отображаются на диаграмме парных отношений. Варианты распределены по уровням от более предпочтительных вариантов к менее предпочтительным.

2.3. Интерфейс пользователя СППР

Для каждого из модулей программного комплекса, реализующего процесс принятия решения, разработан интерфейс человеко-машинного диалога. Внешний вид его представлен на рис. 3—5 для рассматриваемого ниже примера.

Главное меню комплекса содержит четыре раздела: файл, модули (СППР и Модель), окна, информация о программе.

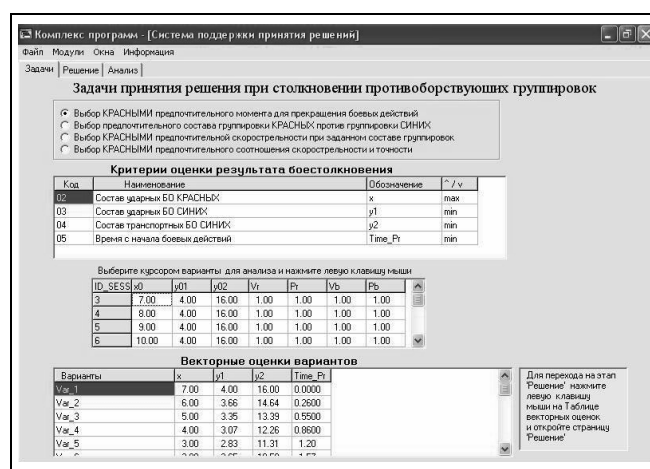


Рис. 3. Страница «Задачи» диалоговой панели СППР

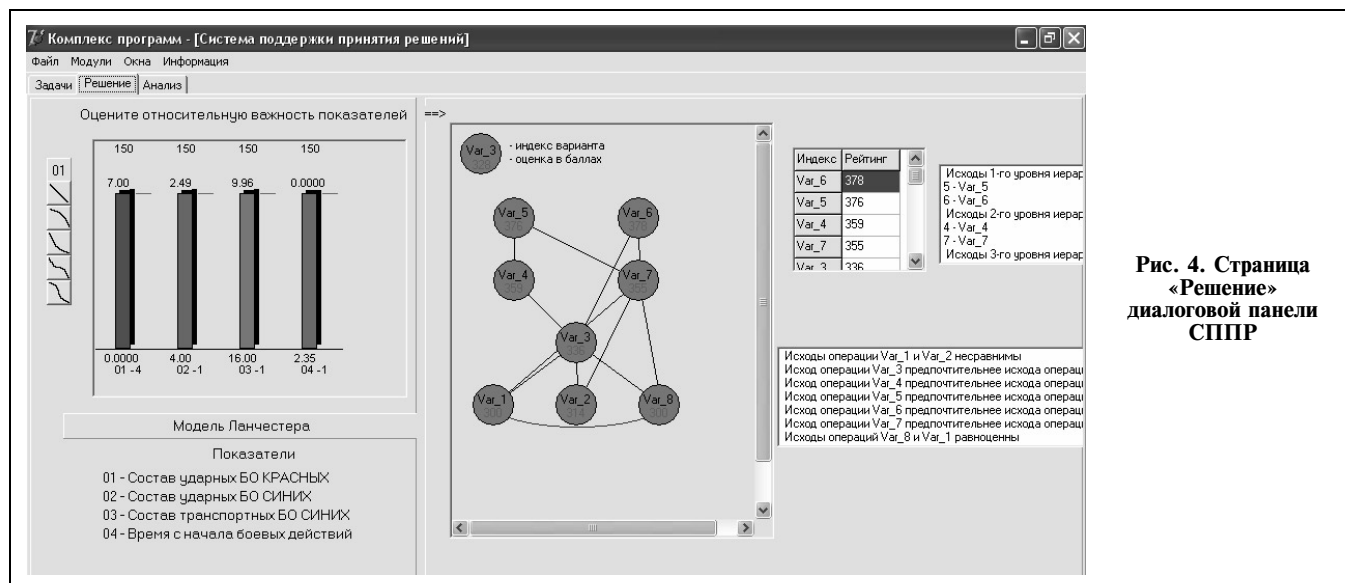


Рис. 4. Страница «Решение» диалоговой панели СППР

Доступ к программному модулю моделирования функционирования системы осуществляется через пункты главного меню программы Модули/Модель. В окне моделирования закладками представлены три страницы: исходные данные, таблица, графики.

Исходные данные. На данной закладке представлены панель выбора модели и панели задания исходных значений характеристик процесса. После выбора модели и задания данных осуществляется запуск процедур расчета показателей функционирования.

Таблицы, графики. Результаты расчетов представляются на закладках «Таблицы» и «Графики», где даются таблицы результатов моделирования и на графике показывается изменение характеристик процесса.

Доступ к программному модулю сравнения альтернатив осуществляется через пункты главного меню программы Модули/СППР. Интерфейс программного модуля сравнения альтернатив состоит из трех страниц диалоговой панели с названиями «Задачи», «Решение», «Анализ».

На первой странице — «Задачи» — расположены средства для выбора типа задачи принятия решения и ввода и редактирования необходимой для решения задачи информации (см. рис. 3). Предусмотрена возможность рассмотрения ситуации принятия решения в четырех типах задач. Общий вид страницы «Решение» приведен на рис. 4. Страница содержит две панели, связанные скользящим разделителем. На левой панели фиксируются и отображаются в графическом виде суждения ЛПР, на правой — представлены результаты обработки этих суждений. Объединение двух панелей скользящим разделителем на одной странице, с одной стороны, позволяет концентрировать больше информации на каждой панели, а с другой — предоставляет возможность ЛПР наблюдать результат своих действий.

На левой панели в центре находится поле с прямоугольниками различного цвета, число которых соответствует числу показателей, по которым сравниваются варианты. Числа над прямоугольниками соответствуют лучшей (в смысле полезности) оценке среди оценок всех

вариантов по показателю, числа под прямоугольниками — худшей оценке. Верхний ряд чисел соответствует высоте прямоугольника в условных единицах и символизирует важность показателя.

Нижний ряд цифр указывает идентификатор показателя и тип формы кривой изменения ценности (полезности) оценок показателя в зависимости от значения показателя. Сами кривые приведены в правой части панели. Цвет функции полезности совпадает с цветом прямоугольника. Выбор формы кривой производится ЛПР поочередно для каждого показателя с учетом особенностей показателя, диапазона изменения его значений и других факторов с помощью набора кнопок, расположенных в левой части панели.

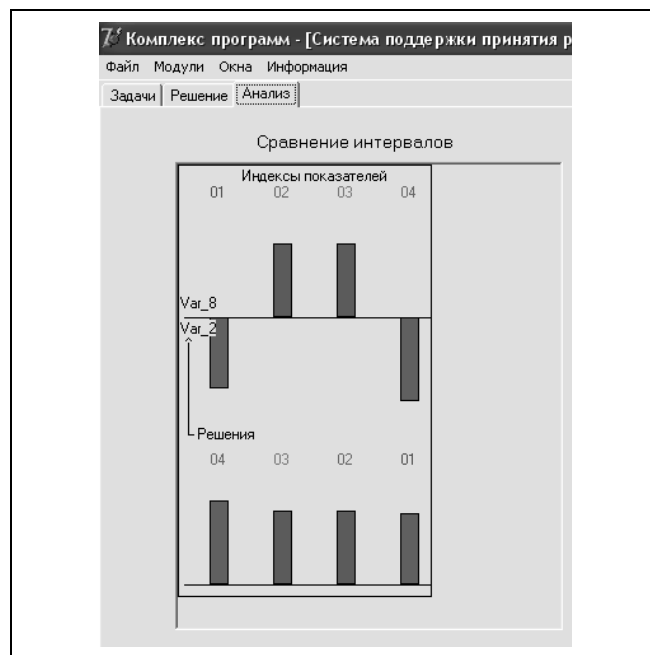


Рис. 5. Страница «Анализ» диалоговой панели

В задачу ЛПР входит установление относительной важности показателей и формы функции полезности для каждого из них. С этой целью курсором в прямоугольнике или над ним выбирается желаемая высота прямоугольника и фиксируется нажатием клавиши.

На правой панели в графической форме отображаются результаты сравнения вариантов. В левом поле изображена диаграмма парных отношений на множестве вариантов.

Вершины диаграммы отождествляются с вариантами, связи, направленные сверху вниз, — с отношениями предпочтения-безразличия. Горизонтальные дуги связывают эквивалентные варианты. В центре вершины даны название варианта и число, представляющее собой сумму числовых представлений ценности интервалов по показателям. Это число может быть использовано как дополнительный критерий упорядочения вариантов. Три других компонента панели содержат информацию, характеризующую с разных сторон результат обработки предпочтений ЛПР: результаты парного сравнения вариантов, результаты обработки диаграммы предпочтений — распределение вершин по уровням иерархии и упорядочение вариантов в соответствии с числами, записанными в вершинах.

Третья страница диалоговой панели — «Анализ» — предоставляет возможность лицу, принимающему решение, видеть основание, по которому сделано заключение о предпочтительности вариантов при парном сравнении. Фрагмент страницы приведен на рис. 5.

Левую часть страницы занимает поле, на котором дается графическое представление разности между оценками вариантов по полезности для каждого показателя — в верхней половине, а в нижней — те же интервалы, упорядоченные по их длине. Справа внизу находится компонент, содержащий результаты сравнения всевозможных пар вариантов. Процесс анализа результата инициализируется выбором нужной строки. Рассчитываются и рисуются интервалы между оценками показателей, упорядочиваются по длине. Интервалы разделены цветом на два подмножества в зависимости от того, какому из вариантов принадлежит более высокая в смысле полезности оценка показателя. На верхнем поле правой половины страницы приводятся результаты анализа.

3. ПРИМЕР ПРИМЕНЕНИЯ СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЯ

Рассматривается решение условной задачи боевого столкновения противоборствующих группировок (морской бой) при массированном использовании боевых единиц в ударе [3, с. 291]. Конвой в составе $N_B = 6$ -ти транспортов и кораблей охранения (сторона синих) атакуется $N_A = 9$ -ю самолетами (сторона красных). Каждый самолет имеет боезапас $n_A = 4$ бомбы. Вероятность попадания бомбы в транспорт $P_A = 0,45$. Среднее число попаданий бомб для потопления транспорта $\omega_A = 2,5$. На транспортах отсутствуют средства самообороны. Корабли охранения, находящиеся со стороны атаки самолетов, используют $N_{B1} = 8$ комплексов с характеристиками

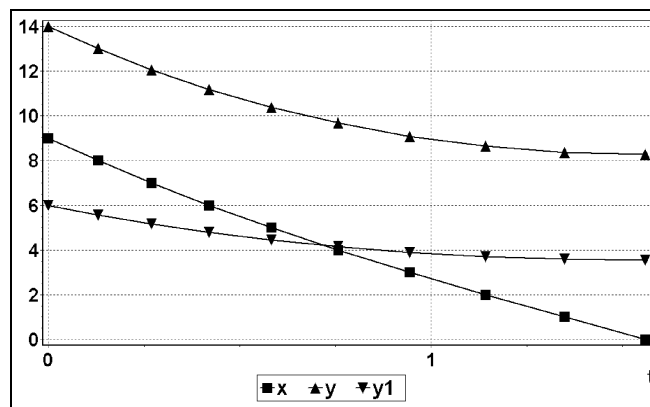


Рис. 6. Графики изменения состава сил:

x — самолеты; y — корабли охранения со стороны атаки и транспорты; y1 — транспорты

$P_{B1} = 0,6$; $n_{B1} = 2$; $\omega_{B1} = 1$, а находящиеся на линии охранения со стороны противоположной направлению атаки $N_{B2} = 6$ комплексов с характеристиками $P_{B2} = 0,4$; $n_{B2} = 1$; $\omega_{B2} = 1$. Результативность боя определяется числом пораженных транспортов и потерями самолетов. Естественно, на этапе планирования операции результативность плана операции оценивается через «ожидаемые» потери сторон.

Анализ различных стратегий боя с применением описанного программного комплекса позволяет найти наиболее предпочтительное с точки зрения ЛПР (сторона красных) решение.

Эти стратегии были исследованы на примере конкретных задач.

Задача 1. Выбор момента выхода из боя до полного уничтожения самолетов.

Задача 2. Оценка целесообразности выделения части ресурсов атакующих самолетов для поражения кораблей охранения с целью обеспечения максимального отношения числа уничтоженных транспортов к потерям самолетов.

Задача 3. Оценка эффективности увеличения числа атакующих самолетов при предположении, что все самолеты атакуют транспорты, не вступая в бой с кораблями охранения.

Комплекс программ СППР позволяет учитывать степень информированности сторон.

Задача 4. Оценка эффекта скрытного упреждающего удара самолетов по транспортам противной стороны.

В качестве конкретного примера приведем результаты решения задачи 1. В ходе моделирования реализация процесса боя длится до истощения ресурсов одной из сторон. Динамика изменения состава сил противоборствующих боевых систем (БС) приведена на рис. 6.

Определены 13 моментов окончания боя (вариантов), соответствующих времени от 0,1 до 1,12 мин. При сопоставлении своих потерь с потерями противостоящей стороны, изменяющихся в процессе боя, с учетом других характеристик, ЛПР основываясь на своих предпочтениях, может принять решение о прекращении боя.



Варианты оценивались по показателям: «потери самолетов», «потери кораблей охранения», «потери траншпортов», «длительность боя». Функции ценности для всех показателей приняты линейными. При относительной важности показателей, равной $150 : 150 : 150 : 100$, наиболее предпочтителен вариант, соответствующий времени 0,6 мин. Если цена потерь собственных сил велика по сравнению с «доходом» от потерь противостоящей стороны (относительная важность равна $135 : 100 : 100 : 110$), то рекомендуется вариант, соответствующий времени 0,4 мин. Наконец, при относительной важности показателей $120 : 170 : 170 : 100$ более предпочтительными оказываются варианты, соответствующие времени 0,7–0,8 мин.

Программный комплекс СППР, представленный в работе, позволяет рассчитать результаты боя в динамике, оценить их по векторному критерию для обоих участников боя и выбрать оптимальное решение.

ЗАКЛЮЧЕНИЕ

Концепция построения СППР, объединяющая в единый комплекс имитационную подсистему и подсистему многокритериальной оценки и выбора предпочтительного варианта действия из множества альтернативных, обеспечивает полное сопровождение процесса выработки управленческого решения от постановки задачи до принятия окончательного решения:

- формирование множества конкурирующих вариантов;
- оценку на имитационной модели задачи результативности планов действий;

— формализацию системы предпочтений ЛПР и формирование оценок полезности по отдельным показателям планируемых решений;

— выбор окончательного варианта плана действий.

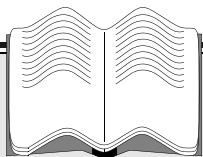
Предлагаемой технологией поддержки принятия решений предусматривается возможность проведения итерационных циклов по уточнению планов действий, системы предпочтений ЛПР (возможно при изменении условий), позволяющих повысить качество принимаемых решений.

Модульность построения СППР упрощает адаптацию системы для применения в различных областях управления, поскольку все необходимые изменения локализованы в имитационной подсистеме.

ЛИТЕРАТУРА

1. Трахтенгерц Э. А., Шершаков В. М., Камаев Д. А. Компьютерная поддержка управления ликвидацией последствий радиационного воздействия. — М.: СИНТЕГ, 2004.
2. Губко М. В., Новиков Д. А. Теория игр в управлении организационными системами. — М.: СИНТЕГ, 2002.
3. Абчук В. А., Матвейчук Ф. А., Томашевский Л. П. Справочник по исследованию операций. — М.: Воениздат, 1979.
4. Буянов Б. Б. Оценка и выбор плана операции при принятии решения в системе имитационного моделирования // Тр. Института проблем управления. — 1999. — Т. 3.
5. Методы анализа и синтеза управляющих систем / Б. Б. Буянов, Б. Г. Волик, Н. В. Лубков и др. — М.: Энергоатомиздат, 1988.

☎ (495) 334-78-01



Журнал "Банк экономических идей" —

это поток новых идей в области экономической теории и практики. Носит одновременно научно-теоретический, научно-практический и научно-популярный характер. Предлагает широким кругам ученых, бизнесменов, преподавателей экономических дисциплин, студентов и магистрантов изложенную простым языком инновационную мысль по таким кардинальным вопросам, как глобализация, грядущее общество, экономический рост, системы управления и самоуправления, налоговая политика и социальная защита населения, предмет и структура экономических наук и т. д.

Призывает читателей к сотрудничеству в генерировании, доработке и обсуждении новых идей.

Печатает также информацию о Грузии, о ее ресурсах, экономических структурах и реформах, о ее историческом прошлом и настоящем, ее литературе, искусстве и т. д.

Предполагается учреждение на базе редакции международной научно-практической организации "Банк экономических идей".

Вышел в свет первый номер журнала. До конца 2006 г. намечается выпуск № 2 и № 3. Желающие подписаться могут обратиться непосредственно в редакцию.

Стоимость 1 экз. журнала по подписке в пределах РФ — 5 долл. США (140 руб.).

О льготах при подписке пяти или более комплектов журнала по одному адресу см. сайт www.bei.ge.

Адрес: 0108, Грузия, Тбилиси, ул. Барнова, 23, ООО "Комментарии",
редакция журнала "Банк экономических идей";
e-mail: editor@bei.ge (для интерактива и вопросов к редакции);
info@bei.ge (для общей информации, подписки или переписки);
advartizing@bei.ge (по поводу размещения рекламы).

☎ (+995-32) 98-29-16; (+995-95) 50-83-46; 📠 (+995-32) 00-11-53

ПРОБЛЕМЫ ЭНЕРГОСБЕРЕГАЮЩЕГО УПРАВЛЕНИЯ ЛИСТОПРОКАТНЫМ КОМПЛЕКСОМ.

Ч. 1

А. Л. Генкин, А. Р. Куделин

Институт проблем управления им. В. А. Трапезникова РАН, г. Москва

Рассмотрены теоретические основы энергосберегающего управления листопрокатным комплексом «печи—стан», отличительная особенность которого заключается в перераспределении энергии между отдельными участками технологической линии.

ВВЕДЕНИЕ

Энергетическая сверхдержава, которой хочет стать и отчасти уже является Россия, должна не только интенсифицировать производство энергоресурсов, но и стремиться к сокращению энергопотребления в различных отраслях промышленности. Металлургическая промышленность — одна из наиболее энергоемких отраслей, в связи с чем экономия энергоресурсов является приоритетной задачей при автоматизации металлургических заводов. Актуальность решения этой задачи особенно возросла после принятия в 1996 г. Федерального Закона «Об энергосбережении», направленного на создание экономических и организационных условий для эффективного использования энергоресурсов посредством разработки энергоэффективных технологий и систем автоматизированного управления энергопотреблением.

За рубежом экономия энергетических и материальных ресурсов также является одной из важнейших задач при автоматизации металлургических заводов и, в частности, при управлении листовыми станами горячей прокатки [1, 2].

В настоящей работе из множества энергоемких объектов черной металлургии рассматривается класс объектов, общим для которых является наличие агрегата для нагрева металла и клетей для прокатки нагретого металла. При производстве горячекатаного листа — основного вида продукции металлургического завода — такой объект называется листопрокатным комплексом (ЛПК) и включает в себя печи для нагрева металла, черновую и чистовую группы клетей для прокатки листа. Энергосберегающее управление ЛПК — совокупность мер, направленных на эффективное использование энергоресурсов при нагреве и прокатке металла. Проблемы энергосберегающего управления выбранным объектом заключаются в отсутствии до настоящего времени тео-

ретического обоснования принципов энергосбережения, а также систем, позволяющих реализовать энергосберегающие принципы и технологии.

1. ТЕХНОЛОГИЧЕСКИЕ ОСОБЕННОСТИ УПРАВЛЕНИЯ ТЕМПЕРАТУРНО-ДЕФОРМАЦИОННЫМ РЕЖИМОМ ГОРЯЧЕЙ ПРОКАТКИ ПОЛОС

Цель обработки металла в ЛПК состоит в получении горячекатаной полосы определенной толщины и ширины с заданными физико-механическими свойствами. Для этого заготовки (слябы) нагревают на участке нагревательных печей до температуры, обеспечивающей возможность дальнейшей деформации металла, а затем прокатывают слябы последовательно в черновой и чистой группам клетей. Типовой печью на широкополосных станах горячей прокатки является многозонная методическая печь толкательного типа. Участок нагревательных печей содержит обычно три—пять печей. Черновая прокатка слябов осуществляется в четырех—шести последовательно расположенных клетях или за соответствующее число проходов при наличии в линии прокатки реверсивной клетки. Чистовая прокатка полос осуществляется непрерывно в шести—семи последовательно расположенных клетях.

Рассматривая листопрокатный комплекс как объект управления организационно-технологическими системами, можно выделить основные участки обработки металла в технологической линии «нагрев слябов — горячая прокатка полос»: 1) нагревательные печи; 2) черновая группа клетей; 3) промежуточный рольганг; 4) чистовая группа клетей. Параметрами, координирующими функционирование этих участков, являются температура и толщина, соответственно: $T_{сл}$ и $H_{сл}$ — сляба на выходе из печного участка; $T_{п}$ и $H_{п}$ — подката на выходе черновой группы клетей; $T_{ч}$ и $H_{ч}$ — подката на выходе проме-



жуточного рольганга (на входе в чистовую группу клетей). Эти координирующие параметры являются выходными для соответствующего участка и входными для следующего по ходу прокатки участка. Температура $T_{\text{кп}}$ и толщина h на выходе из чистовой группы клетей являются выходными параметрами, определяющими качество горячекатаной полосы. На промежуточном рольганге толщина подката не изменяется, т. е. $H_{\text{п}} = H_{\text{ч}}$.

По мере продвижения металла в рассматриваемой линии на отдельных ее участках имеют место различные энергозатраты. При нагреве металла в печи перед прокаткой расход топлива (газа) зависит, в основном, от параметров сляба и темпа выдачи слябов из печи. Прокатка раската в клетях черновой и чистовой групп сопровождается существенным расходом электроэнергии на деформацию металла. Тепловые потери металла в межклетевых промежутках и на промежуточном рольганге обусловлены излучением и конвекцией и зависят от скорости прокатки и толщины раската.

Управление каждым из рассмотренных участков осуществляется с помощью соответствующей автоматизированной системы. Цель создания энергосберегающей системы управления температурно-деформационным режимом в широкополосном стане горячей прокатки состоит в исходной настройке отдельных подсистем в соответствии с выбранным критерием оптимальности и их координации в зависимости от сложившейся производственной ситуации.

Оптимальная исходная настройка стана на прокатку заданного типоразмера полос подразумевает управление температурно-деформационным режимом обработки металла с заданными значениями $H_{\text{сл}}$, $H_{\text{п}}$, h и ширины сляба B , практически неизменной в линии прокатки. Температура металла в ключевых точках технологической линии может изменяться в определенных пределах, обеспечивая требуемые показатели качества полосы на выходе из стана и являясь наиболее важным контролируемым параметром, обеспечивающим эффективную координацию подсистем управления отдельными участками технологической линии.

Снижение температуры нагрева слябов позволяет уменьшить время нагрева, благодаря чему сокращается угар металла в печи и представляется возможным уменьшить расход топлива. Изменение температуры раската вызывает изменение энергосиловых параметров прокатки в клетях и, следовательно, влияет на потребляемую электродвигателями главного привода мощность и расход электроэнергии на прокатку.

Таким образом, основным фактором, определяющим непроизводительные затраты в производстве горячекатаного листа, являются тепловые потери на всех стадиях нагрева и прокатки металла.

Известные методы снижения тепловых потерь при производстве горячекатаного листа можно классифицировать по двум направлениям [3]:

1) методы, обеспечивающие снижение тепловых потерь благодаря дополнительному оборудованию и совершенствованию технологического процесса «нагрев слябов — горячая прокатка полос»;

2) методы, позволяющие управлять температурой металла путем оптимизации температурно-деформационного режима прокатки.

К первому направлению относятся методы снижения тепловых потерь при транспортировке металла в

линии стана, а также использование тепла предыдущих переделов, в частности, установка теплозащитных экранов, перемоточных устройств на промежуточных рольгангах, дополнительной клетки перед чистовой группой, применение прямой прокатки без промежуточного нагрева и т. п.

Характерным для методов второго направления является перераспределение энергии между отдельными участками технологической линии «нагрев слябов — горячая прокатка полос», а их реализация осуществляется в системах оптимизации технологического процесса с учетом ограничений, обусловленных качеством готовой продукции и энергосиловыми характеристиками агрегатов. Одним из наиболее эффективных методов этого направления является изменение режима обжатий в прокатных клетях.

Как показал анализ, эффективность методов снижения энергозатрат первого и второго направлений примерно одинакова, однако капитальные затраты для реализации методов второго направления существенно ниже, в связи с чем они более предпочтительны [3].

Определение критерия оптимальности исходной настройки стана осложняется многообразием требований, предъявляемых к обработке металла на различных участках технологической линии. С учетом специфики прокатного производства выбор критерия оптимизации сводится к нахождению функции, определяющей экономическое состояние производства, при соблюдении требований к качеству горячекатаной продукции (геометрические размеры и физико-механические свойства металла). В группу ограничений вводятся также требования, обусловленные предельными значениями параметров технологического процесса, конструктивными характеристиками агрегатов и их компоновкой. Учитывая многомерность решаемой задачи, основными способами ее решения являются эвристические подходы. Основные критерии оптимальности при управлении процессом прокатки на различных участках технологической линии следующие.

- Максимальная производительность стана. Реализация этого критерия требует четкой координации пропускной способности различных звеньев технологической линии (нагревательных печей, черновой и чистовой групп клетей).
- Минимальный расход топлива в печах. Обеспечивает оптимальный режим нагрева металла при заданной температуре $T_{\text{сл}}$. Минимально допустимая температура нагрева металла в печах всегда соответствует минимальным затратам на нагрев металла.
- Минимальный расход электроэнергии в главных приводах прокатных клетей. Этот критерий может быть использован при невозможности изменения температуры нагрева металла или в иных специфических ситуациях. С повышением температуры $T_{\text{сл}}$ расход электроэнергии в клетях уменьшается.
- Минимальные суммарные затраты на нагрев и прокатку металла. В критерий включаются затраты на топливо в печах при нагреве слябов перед прокаткой и на электроэнергию в главных приводах прокатных клетей с учетом дополнительных потерь металла при его нагреве.

В определенных ситуациях может возникнуть необходимость использования критериев равномерной загрузки клетей, прокатки с фиксированной температурой

на входе и выходе из группы клетей, минимизации потерь тепла и др.

Отметим, что те или иные критерии могут быть реализованы в зависимости от возможностей оборудования на соответствующих участках технологической линии. Так, например, отсутствие в нагревательных печах возможности дифференциации нагрева (изменения $T_{сл}$ в зависимости от параметров сляба) не позволяет уменьшить $T_{сл}$ и, тем самым, сократить расход топлива и угар металла при его нагреве; эффективное снижение тепловых потерь на промежуточном рольганге может быть обеспечено лишь при его оборудовании теплоизолирующими экранами.

Анализ энергетических потерь в линии широкополосных станов горячей прокатки показал, что основными статьями расходов в ЛПК «печи—стан» являются расход топлива при нагреве металла в печах и электроэнергии в главных приводах прокатных клетей, а также потери металла на окалину при его нагреве [4]. Учитывая примерно равноценную эффективность обоих методов снижения энергозатрат и существенно более низкие капитальные затраты при реализации методов второго направления, следует отдать предпочтение разработке систем управления ЛПК по экономическому критерию.

2. ОСОБЕННОСТИ МОДЕЛИРОВАНИЯ ТЕХНОЛОГИЧЕСКОЙ ЛИНИИ «НАГРЕВ СЛЯБОВ — ГОРЯЧАЯ ПРОКАТКА ПОЛОС»

Разработка энергосберегающей системы управления должна предваряться созданием адекватной математической модели объекта управления (ЛПК «печи—стан»), состоящего из нагревательных печей и прокатных клетей.

В Институте проблем управления РАН еще в первой половине 1970-х гг. профессором А. Б. Челюсткиным была поставлена задача исследования широкополосного стана как объекта управления температурным режимом горячей прокатки полос [5—8]. В результате были построены математические модели, предназначенные как для исследования изменения параметров прокатки, так и для разработки алгоритмов управления соответствующими параметрами. Точность полученных моделей позволяет адекватно оценивать изменение параметров процесса в условиях их широкого варьирования.

В качестве элементарного звена принимается прокатная клеть с прилегающим к ее входу промежуток. Для первой клетки черновой группы, ближайшей к нагревательным печам, этим промежуток является печной рольганг, для первой клетки чистовой группы — промежуточный рольганг, для остальных клетей — межклетевые рольганги.

Математическая модель элементарного звена представляет собой систему алгебраических и дифференциальных уравнений, описывающих основные процессы формоизменения и теплопередачи при изменении температуры металла на входе звена, толщины и скорости раската на входе и выходе звена [7]. Выходным параметром является температура металла на выходе звена. Система дополнена специфическими формулами, определяющими энергетическое состояние звена при изменении параметров прокатки.

Моделирование нагрева металла в печи и расчет технико-экономических показателей нагрева и прокатки металла выполнены с помощью методики, изложенной в работе [9].

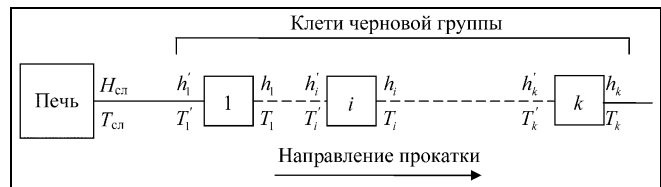


Рис. 1. Схема нагрева и черновой прокатки

В качестве объекта моделирования рассмотрим часть технологической линии, включающую в себя нагревательные печи и черновую группу клетей (рис. 1). Слябы шириной B и толщиной $H_{сл}$ нагреваются в печи до температуры $T_{сл}$ и затем прокатываются в клетях черновой группы от толщины $H_{сл} = h'_1$ на входе в черновую группу до толщины подката $H_{п} = h_k$ на выходе последней k -й клетки черновой группы, причем температура раската меняется от T'_1 на входе в черновую группу до температуры подката $T_{п} = T_k$ на выходе из нее. Отметим, что толщина раската на входе в i -ю клеть равна толщине на выходе предыдущей $(i - 1)$ -й клетки, т. е. $h'_i = h_{i-1}$, $1 < i \leq k$; скорость прокатки раската в i -й клетке V_i постоянна и зависит от номера клетки.

С учетом изложенного, в формализованном виде черновая группа клетей может быть представлена в виде последовательного соединения k звеньев, каждое из которых характеризуется фазовым вектором

$$x = (x_1, x_2, \dots, x_n) \quad (1)$$

n -мерного фазового пространства [10]. Координаты вектора фазового состояния i -го звена — геометрические, теплотехнические и энергосиловые параметры прокатки, которые должны подчиняться условию фазового ограничения $x \in \Omega$, где множество Ω — совокупность допустимых фазовых состояний i -го звена.

Управление процессом состоит из совокупности управлений для каждого i -го звена (прокатной клетки):

$$u = (u_1, \dots, u_i, \dots, u_k), \quad (2)$$

а управляющие воздействия должны принадлежать допустимой области управлений $u \in U$.

Область допустимых управлений в клетях определена ограничениями на усилия и моменты прокатки, мощность главного привода клетей, а также условиями захвата металла валками и длиной межклетевых рольгангов.

К управляющему воздействию относится также изменение температуры нагрева сляба $T_{сл}$, ограниченной сверху условиями оплавления металла в печи, а снизу — возможностью безаварийной прокатки недостаточно нагретого металла.

Цель управления состоит в расчете температуры $T_{сл}$ нагрева сляба в печи путем выбора допустимых управляющих воздействий u в каждом i -м звене с учетом ограничений, накладываемых на координаты вектора фазового состояния звена. Для того, чтобы управление было оптимальным, необходимо определить критерий оптимальности.

Из теории и технологии прокатки известно, что при заданной температуре подката $T_{п}$ минимальный расход топлива обеспечивается при наиболее низкой темпера-



туре нагрева сляба в печи $T_{\text{сл}}$. Наоборот, минимальный расход электроэнергии имеет место при максимально возможной температуре прокатываемого металла. В общем случае критерий оптимальности может быть представлен в виде суммарных удельных (на единицу массы металла) затрат на нагрев и прокатку металла:

$$Q = Z + \sum_{i=1}^k \varphi(x_i, u_i), \quad (3)$$

где $Z = f(T_{\text{сл}}, H_{\text{сл}}, B)$ — удельные затраты на нагрев металла и его потери вследствие угара в печи; $\varphi(x_i, u_i)$ — удельные затраты на электроэнергию в i -м звене, вектор x фазового состояния которого определен выражением (1), а управляющее воздействие u — выражением (2).

С учетом изложенного, оптимальными будут такие управления $T_{\text{сл}}$ и u , которые обеспечивают минимум функционала (3).

Можно видеть, что управляемый процесс естественным образом разбивается на k шагов, а критерий оптимальности Q зависит от совокупности шаговых управлений. Моделирование энергосберегающего управления осуществляется с помощью метода покоординатного спуска, представляющего собой итеративную процедуру, в которой переходят шаг за шагом от одного допустимого решения к другому так, что значение целевой функции улучшается.

Расчет состоит из k этапов, соответствующих k звеньям черновой группы, и производится против хода прокатки, т. е. от k -го звена к первому. Задаются начальное состояние процесса: $H_{\text{сл}}, H_{\text{п}}, T_{\text{п}}$, начальные значения толщин на выходе каждой i -й клетки h_i , другие параметры клеток и межклетевых рольгангов, а также область допустимых управлений. На k -м этапе рассчитываются параметры прокатки, соответствующие начальному состоянию процесса, и начальное состояние целевой функции Q_0 . На всех остальных этапах, вплоть до первого, изменяется толщина раската на выходе из клетки, номер которой соответствует номеру этапа. Толщина раската на выходе остальных клеток остается неизменной. По окончании расчета параметров прокатки во всех клетях в рассмотренном i -м этапе определяется значение функции цели, которое сравнивается с ее предыдущим значением. Если новое значение функции цели больше предыдущего, толщина раската на выходе i -й клетки изменяется в противоположную сторону, и i -й этап повторяется заново. Если новое значение функции цели меньше предыдущего, расчет этапа с изменением толщины раската в i -й клетки в том же направлении повторяется до тех пор, пока не будет достигнуто ограничение по какому-либо параметру в клетке или минимум функции цели.

Окончание расчета всех k этапов соответствует одному циклу. Число циклов повторяется до тех пор, пока не будет достигнуто равенство (с определенной точностью) значений функции цели в двух последовательных циклах. Полученная в результате моделирования последовательность решений, минимизирующая критерий оптимальности (3), является оптимальной стратегией управления для энергосберегающего режима и представляет собой распределение толщин на выходе каждой клетки $h = (h_1, \dots, h_p, \dots, h_k)$ и оптимальную по критерию (3) температуру нагрева сляба в печи $T_{\text{сл}}$ (см. рис. 1).

3. РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ

Анализ результатов моделирования показал, что для каждого типоразмера прокатываемого металла (по толщине сляба $H_{\text{сл}}$ и подката $H_{\text{п}}$, ширине сляба B) может быть определена область допустимых управлений, схематично представленная на рис. 2 [10, 11]. В координатах $(T_{\text{п}}, T_{\text{сл}})$ эта область ограничена треугольником, сторона ab которого соответствует прокатке с максимально возможной для данного типоразмера температурой нагрева сляба $T_{\text{сл}}$, сторона ac — прокатке с минимально возможной температурой $T_{\text{сл}}$, а сторона bc — максимально возможной $T_{\text{сл}}$ по условиям оплавления металла в печи $T_{\text{сл}}^{\text{max}}$. Иными словами, каждому типоразмеру соответствует некоторая область изменения координат $T_{\text{п}}$ и $T_{\text{сл}}$, внутри и на границах которой может быть осуществлена прокатка, а вне ее пределов прокатка данного типоразмера невозможна.

Так, например, для обеспечения $T_{\text{п}} = T_{\text{п}}^*$ температура сляба должна удовлетворять следующему неравенству:

$$T'_{\text{сл}} \leq T_{\text{сл}} \leq T''_{\text{сл}},$$

где $T'_{\text{сл}}$ и $T''_{\text{сл}}$ — соответственно, минимально и максимально возможные для данного типоразмера температуры нагрева слябов.

Если допустить $T_{\text{сл}} < T'_{\text{сл}}$, то для данного типоразмера невозможно обеспечить $T_{\text{п}} = T_{\text{п}}^*$. В этом случае будет иметь место $T_{\text{п}} < T_{\text{п}}^*$. И, наоборот, если $T_{\text{сл}} > T''_{\text{сл}}$, то также невозможно обеспечить $T_{\text{п}} = T_{\text{п}}^*$ ($T_{\text{п}} > T_{\text{п}}^*$).

Можно видеть (см. рис. 2), что для определенного типоразмера металла имеют место две экстремальные точки [11]. Первая из них определяется координатами $(T_{\text{п}}^{\text{min}}, T_{\text{сл}}^{\text{min}})$. При нагреве сляба до температуры $T_{\text{сл}} < T_{\text{сл}}^{\text{min}}$ прокатка в черновой группе невозможна из-за перегрузки клеток. Вторая точка с координатами $(T_{\text{п}}^{\text{max}}, T_{\text{сл}}^{\text{max}})$ определяет максимально возможную тем-

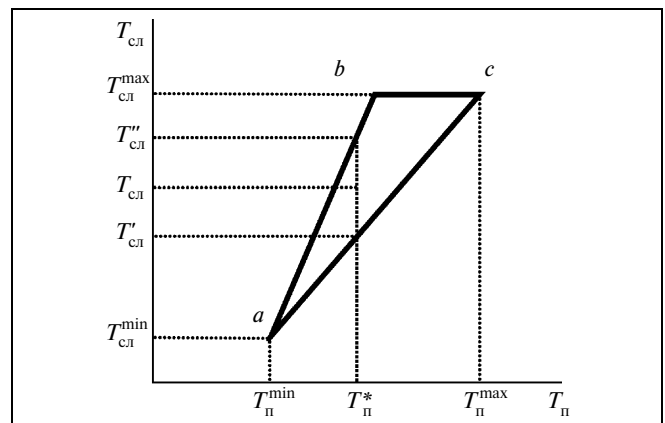


Рис. 2. Область допустимых управлений листопрокатным комплексом «печи—стан»

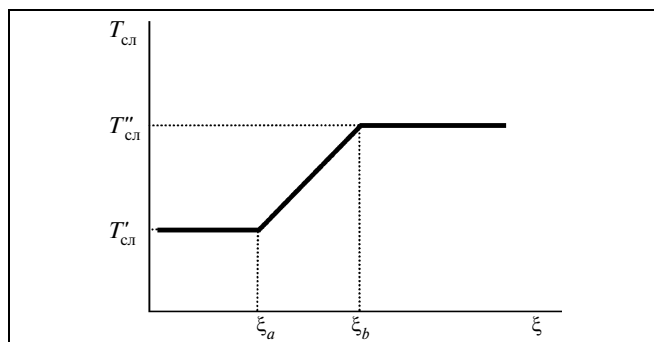


Рис. 3. Зависимость оптимальной температуры нагрева слэба $T_{сл}$ от соотношения цен на электроэнергию и топливо

температуру подката $T_{п}^{max}$, которая может быть обеспечена при нагреве слэба до температуры $T_{сл}^{max}$. Эта точка находится на рис. 2 пересечением зависимости $T_{сл} = f(T_{п})$, соответствующей прокатке с минимально возможной температурой $T_{сл}$ и максимально возможной по условиям оплавления металла температуры слэба $T_{сл}^{max}$ (точка с на рис. 2).

Результаты моделирования энергосберегающего режима позволили выявить закономерность изменения энергетических расходов в зависимости от соотношения цен на электроэнергию $C_{э}$ и топливо $C_{т}$ [12]. В соответствии с этой закономерностью для каждого типоразмера прокатываемого металла на рис. 3 представлена зависимость температуры нагрева слэба $T_{сл}^*$, обеспечивающей минимальные суммарные энергозатраты на нагрев и прокатку металла при заданной температуре подката $T_{п}^*$, от соотношения цен $\xi = C_{э}/C_{т}$ [10].

При изменении ξ в диапазоне $\xi < \xi_a < \xi_b$ минимум энергозатрат обеспечивается минимизацией функционала (3). На участках $\xi < \xi_a$ и $\xi > \xi_b$ температура слэба достигает, соответственно, минимального $T'_{сл}$ и максимального $T''_{сл}$ значений (см. рис. 2). Режим управления на участке $\xi < \xi_a$ называется топливосберегающим и обеспечивает минимум удельных затрат на топливо при нагреве металла, а на участке $\xi > \xi_b$ — обеспечивает минимум удельных затрат на электроэнергию при прокатке металла и называется электросберегающим. В настоящее время соотношение цен ξ таково, что минимум суммарных удельных затрат на нагрев и прокатку металла обеспечивает топливосберегающий режим.

ЗАКЛЮЧЕНИЕ

Рассмотрены методы энергосберегающего управления листопрокатным комплексом «печи—стан», отличительная особенность которых состоит в перераспределении энергии между его отдельными участками. Разработана экономико-математическая модель технологической линии «нагрев слэбов — горячая прокатка полос».

Сформулированы теоретические основы энергосберегающего управления ЛПК «печи—стан», в соответствии с которыми критерий оптимальности представлен

в виде суммарных удельных (на единицу массы металла) затрат на нагрев и прокатку металла. Разработана методика расчета оптимальных управляющих воздействий. Выявлена закономерность изменения энергетических расходов в зависимости от соотношения цен на электроэнергию и топливо. Определены границы возможных управлений ЛПК.

Дальнейшее развитие теоретического обоснования и разработка принципов построения энергосберегающих систем управления ЛПК «печи—стан» будут рассмотрены во второй части статьи.

ЛИТЕРАТУРА

1. Кюттнер Г. Новые задачи автоматизации металлургических заводов // Steel Times International. — 1997. — Вып. 5. — С. 30—32.
2. Новая концепция контроля процесса нагрева слэбов / А. Шутти, Э. Шиллер, А. Ортнер и др. // Черные металлы. — 2005. — № 3. — С. 31—36.
3. Системы и средства автоматизации энергосберегающей технологии в прокатном производстве / Ю. П. Божко, А. Л. Генкин, Я. С. Масальский, Л. Ф. Ромашкевич // Черная металлургия: Бюлл. ин-та «Черметинформация». — 1987. — Вып. 23. — С. 21—36.
4. Программно-технический комплекс оптимального управления широкополосным станом горячей прокатки / А. А. Меленков, А. В. Зиле, Ю. П. Божко и др. // Сталь. — 1990. — № 8. — С. 47—51.
5. Челюсткин А. Б., Койнов Т. А., Генкин А. Л. Математическая модель и алгоритмы управления непрерывным широкополосным станом горячей прокатки // Динамическое моделирование и управление технологическими процессами с помощью ЭВМ: Тр. советско-финского симпозиума (Тбилиси, 12—19 декабря 1973 г.). — М.: ВЦ АН СССР, 1974. — Кн. 1. — С. 247—255.
6. Optimal control of hot strip mills according to economic criterion / A. B. Cheliustkin, G. G. Grigorian, A. A. Logovatsky, et al // Optimization methods (applied control): Preprints of the IFAC-IFORS symposium. — Varna, Bulgaria, 1974. — P. 427—432.
7. Челюсткин А. Б., Цифринович Б. А., Генкин А. Л. Изменение температуры горячекатаных полос в условиях принудительного межклетевого охлаждения // Изв. вузов. Черная металлургия. — 1976. — № 1. — С. 100—105.
8. Койнов Т. А., Масальский Я. С., Челюсткин А. Б. Оптимальное управление производственным потоком технологического комплекса «непрерывная разливка стали — горячая прокатка полос» // Автоматика и телемеханика. — 1977. — № 1. — С. 168—178.
9. Интегрированное проектирование металлургических комплексов / С. А. Власов, С. А. Малый, В. С. Томашевская, А. И. Тропкина. — М.: Металлургия, 1983. — 144 с.
10. Генкин А. Л., Куделин А. Р., Масальский Я. С. Моделирование энергосберегающего управления листопрокатным комплексом // Тр. II Междунар. конф. SICPRO'03 / «Идентификация систем и задачи управления» / Ин-т пробл. упр. РАН. — М.: 2003. — С. 2399—2408.
11. Генкин А. Л. Принципы построения энергосберегающих систем управления листовыми станами горячей прокатки // Тр. Ин-та пробл. упр. РАН. — М., 1998. — Т. I. — С. 40—47.
12. Генкин А. Л., Куделин А. Р. Особенности моделирования энергосберегающих режимов управления горячей прокаткой полос // Тр. Ин-та пробл. упр. РАН. — М., 2000. — Т. VIII. — С. 71—76.

☎ (495) 334-87-59;

e-mail: genfone@ipu.rssi.ru





ЭКСПЕРТНО-СТАТИСТИЧЕСКИЕ МЕТОДЫ ОБРАБОТКИ ИНФОРМАЦИИ В ИНТЕГРИРОВАННЫХ СИСТЕМАХ УПРАВЛЕНИЯ ПРОИЗВОДСТВОМ И ТЕХНОЛОГИЧЕСКИМИ ПРОЦЕССАМИ¹

А. С. Мандель

Институт проблем управления им. В. А. Трапезникова РАН, г. Москва

Исследована возможность применения экспертно-статистических методов обработки информации для решения задач управления производством и технологическими процессами в рамках действующей на предприятии интегрированной системы управления.

ВВЕДЕНИЕ

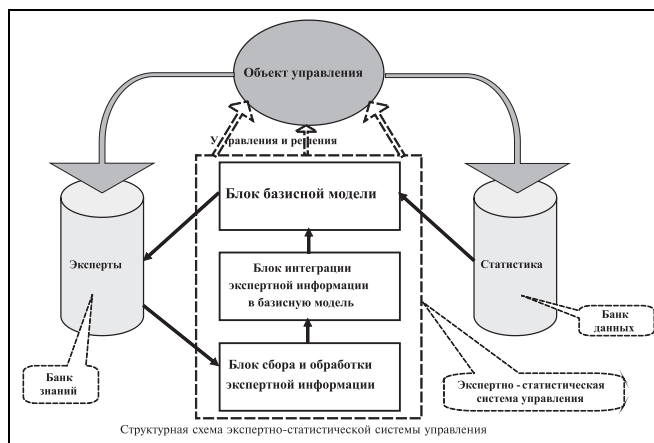
Интегрированные системы управления производством (ИСУ), охватывающие все уровни управления современным предприятием — от уровня ERP-планирования до уровня управления отдельными технологическими процессами, — находят все более широкие применения. Главное достоинство ИСУ заключается в возможности согласованного и координированного решения многочисленных управленческих проблем, при том неявном предположении, что для каждой из этих проблем наличествуют адекватные и эффективные методы их решения. При этом возникает иллюзия, что сам факт создания и внедрения ИСУ на современном промышленном предприятии является определяющим фактором триумфального успеха всей его последующей деятельности. Однако опыт автора настоящей статьи свидетельствует о том, что инструмент глобальной интеграции всех процессов управления предприятием — от проблем стратегического планирования до вопросов управления отдельными технологическими процессами — отнюдь не становится панацеей от вполне вероятных и, порою, весьма серьезных провалов, в том числе и стратегических. Особенно в тех случаях, когда для решаемых в рамках ИСУ отдельных («частных») управленческих задач отсутствуют адекватные практике способы их точной формализации (идентификации).

В качестве важнейших причин отсутствия достаточно точных математических моделей тех или иных процессов управления современным предприятием назовем: (а) чрезмерную сложность формальных описаний,

которая в процессе поиска оптимальных, робастных или, попросту, рациональных управленческих решений заставляет прибегать к их неоправданному (в смысле адекватности идентифицируемой модели) упрощению; (б) отсутствие возможности единовременного и интегрированного использования при идентификации моделей таких разнородных источников информации, как результаты объективных измерений параметров и характеристик управляемых процессов и мнения, высказываемые опытными экспертами-управленцами; (в) наличие резких и плохо предсказуемых возмущений (прежде всего, во внешней среде — например, на рынке), которые могут кардинально изменить ситуацию и потребовать отказа от ранее разработанных математических описаний в пользу новых моделей, более адекватных вновь возникшей ситуации, и (г) возможную неполноту имеющейся объективной информации, частичную наблюдаемость или даже полную ненаблюдаемость важнейших характеристических переменных, а в результате нехватку доступных для регистрации массивов данных для построения статистически достоверных моделей управляемых процессов.

Первая из указанных ранее причин особенно характерна при решении проблем управления сложнейшими многосвязными технологическими процессами в таких отраслях, как нефтепереработка, нефтехимия [1] и ряд других. Наличие второго источника возможных проблем отмечается во множестве случаев, относящихся, прежде всего, к вопросам, связанным с управлением на верхнем уровне АСУТП и на высших уровнях ИСУ во многих отраслях. Третий из названных источников неприятностей чаще всего проявляется тогда, когда речь идет о взаимодействии предприятия через отделы сбыта и маркетинга с внешним рынком ими при возникновении аварийных ситуаций. Наконец, четвертая причина нередко связана со спецификой решаемых управленческих про-

¹ Работа выполнена при финансовой поддержке РФФИ, проект № 06-08-00415-а.



Структурная схема экспертно-статистической системы управления

блем (например, когда речь идет об организации производства принципиально новых видов продукции).

В середине 1990-х гг. в целях преодоления проблем, связанных с причинами, помеченными литерами (б) и (г), автором был предложен инструмент интеграции разнородной информации в рамках единой системы управления, который получил название экспертно-статистических методов обработки информации [2, 3], т. е. основным объектом применения нового инструмента стали социально-экономические системы и организационные системы управления. В настоящей статье делается попытка распространить действие экспертно-статистического подхода на ИСУ, включая уровень АСУТП.

1. ЭКСПЕРТНО-СТАТИСТИЧЕСКИЙ ПОДХОД К ОБРАБОТКЕ ИНФОРМАЦИИ

Структурная схема экспертно-статистической системы (ЭСС) управления представлена на рисунке. Собственно ЭСС управления состоит из блока базисной модели, блока сбора и обработки экспертной информации и блока интеграции экспертной информации в базисную модель.

Многочисленные примеры ЭСС для решения различных задач организационного управления, управления запасами и маркетинга можно найти в работах [2—4]. Наибольший интерес представляет блок интеграции экспертной информации в базисную модель. Именно структура и алгоритмическое наполнение этого блока определяет возможность успешного решения соответствующей задачи управления. Далее мы рассмотрим устройство данного блока применительно к задаче идентификации и управления технологическими процессами.

2. БЛОК ИНТЕГРАЦИИ ЭКСПЕРТНОЙ ИНФОРМАЦИИ В БАЗИСНУЮ МОДЕЛЬ В ЗАДАЧАХ АСУТП

Блок интеграции в блок базисной модели представляет собой совокупность процедур (методик проведения интервью, опросов, анкетирования и пр.) извлечения экспертных знаний и формируемой в результате соответствующих процедур извлечения (сбора) знаний не-

которой разновидности экспертной системы. Однако в большинстве случаев, в отличие от обычных систем, работающих со знаниями, экспертные системы, которые входят в состав ЭСС, как правило, гораздо проще. Дело в том, что ЭСС обычно разрабатываются для объектов управления в тех случаях, когда имеется достаточно высокая степень уверенности в том, что та априорная и апостериорная информация, которой располагают разработчики и заказчики, по поводу способов функционирования и характеристик подобных систем достаточно для формирования «почти» полной модели описываемой системы. Критерием такой «почти достаточности» объективных данных (вполне возможно и иллюзорной) является малость остаточной дисперсии для предварительной «эскизной» модели системы, которая формируется на этапе предварительного обследования по априорным данным. Приведенную логику перепромулируем в виде следующего утверждения, достоверность которого неоднократно подтверждена практикой внедрения ЭСС.

Утверждение 1. *Априорные данные, пополненные, быть может, той информацией, которая поступит в интервале от момента окончания предварительного обследования до момента создания эскизной модели, образуют обучающую выборку. Если остаточная дисперсия (или какая-либо другая аналогичная характеристика качества модели) окажется на этой обучающей выборке мала, то искомого качества управления можно будет достичь на пути создания ЭСС.*

К сожалению, многочисленные попытки теоретически получить точные оценки уровня остаточной дисперсии или аналогичных ей характеристик, которые позволяли бы строго подтвердить, что мы находимся в области применения экспертно-статистического подхода, оказались — и не случайно! — безуспешными. Очевидно только, что при малых значениях остаточной дисперсии можно, скорее всего, говорить о выборе между классическими методами статистической идентификации и экспертно-статистическими методами. При этом с дальнейшим увеличением остаточной дисперсии следует определиться с границей, отделяющей экспертно-статистический подход от чисто экспертных методов, которые, в конце концов, приводят к созданию экспертных систем.

Условность любых «строгих» выводов обусловлена тем, что даже при малых значениях остаточной дисперсии (полученных на обучающей выборке!) нет никаких гарантий, что лицо, принимающее решения (ЛПР), примет рекомендации, полученные по идентифицированной модели, к исполнению. За подобным отказом может скрываться как ошибка или нерешительность ЛПР, так и его уверенность в том, что обучающая выборка не является статистически представительной (не содержит информации обо всех «режимах» функционирования описываемого объекта).

Яркой иллюстрацией к этим рассуждениям служит картина, которая наблюдается при автоматизации сложных технологических процессов в таких областях, как химия, нефтехимия, нефтепереработка и многие другие (см., например, работы [5—7]). В этом случае классические, успешные попытки автоматизации и предшествующие им этапы идентификации сводятся к применению двух основных подходов.



Первый из них заключается в формировании комплекса сложных нелинейных статических моделей, которые строятся методами кусочной аппроксимации [6, 7]. Этот подход позволяет без труда промоделировать многорежимные объекты, однако ни в малейшей степени не отображает их динамики.

Второй подход (который применяется, в частности, и для создания тренажеров [5]) опирается на описание технологического процесса с помощью совокупности локальных физико-химических моделей с последующим объединением их в глобальную модель с учетом геометрии и физико-механических свойств соединяющих локальные процессы транспортных устройств. Данный подход, по сравнению с первым, кажется поначалу более адекватным реальному процессу, однако гораздо хуже справляется с многорежимностью. В целях повышения его адекватности проводятся эксперименты по моделированию рассматриваемого технологического процесса с последующим предъявлением результатов моделирования экспертам-технологам. Изучив результаты моделирования, технологи формулируют свои замечания, сообщая, как по их мнению, должен был бы вести себя процесс в предъявленных ситуациях. Их пожелания отображаются в некую разновидность экспертной системы, которая подменяет созданный комплекс физико-химических и физико-механических моделей в отмеченных экспертами-технологами критических ситуациях.

В качестве альтернативы (или дополнения) к рассматриваемым подходам можно воспользоваться экспертно-статистическим подходом, очертив, хотя бы качественно, границы его применимости. По опыту автора, границы применимости разных подходов можно охарактеризовать значениями отношения χ корня квадратного из остаточной дисперсии к математическому ожиданию наблюдаемых выходных переменных, которые приведены в таблице. Взаимное наложение областей, указанных в таблице, подтверждает высказанную ранее мысль о том, что возможность построения точных дискриминантных границ достаточно условна. Правда, из таблицы следует вывод, что в интервале $20\% \leq \chi \leq 30\%$ можно достаточно уверенно говорить о целесообразности применения именно экспертно-статистического подхода.

К тезису, содержащемуся в утверждении 1, непосредственно примыкает и еще один вывод относительно ЭСС, который составляет суть следующего утверждения.

Утверждение 2. Системы экспертно-статистической обработки дополняют традиционные системы управления и экспертные системы и отличаются от последних тем, что дешевле и проще (в части применения методологии искусственного интеллекта), так как не требуют специально организованных процедур извлечения знаний и формирования слишком сложных правил вывода.

Границы (нечеткие) применимости различных подходов

Тип подхода	Статистическая идентификация	Экспертно-статистические методы	Чисто экспертные методы
χ , %	Менее 20	10—60	Более 30

3. ERP-ПЛАНИРОВАНИЕ В УСЛОВИЯХ НЕОПРЕДЕЛЕННОСТИ

Успешность функционирования предприятия, охваченного действием ИСУ, в значительной степени определяется рациональностью выбранной им маркетинговой политики. Суть этой политики прежде всего связана с тщательным изучением и попытками воздействия на потребительские предпочтения, формирующие, в конечном счете, рыночный спрос. Обычной технологией решения задач оценки спроса на отдельные виды традиционных и впервые производимых товаров является построение сетевой модели формирования спроса (товаропроводящей сети), состоящей из собственно рынка, на котором реализуется товар, предприятий, осуществляющих выпуск этого товара и предприятий-поставщиков полуфабрикатов, используемых при изготовлении товара. На рынке нефти и нефтепродуктов в качестве первых выступают нефтеперерабатывающие заводы, а в качестве вторых — предприятия нефтедобычи и транспортная сеть.

В конечном счете, решением задачи прогнозирования спроса является установление пропорций (рыночных долей), в которых рынок распределяется между разными производителями, абсолютных значений спроса на товар и их поведения во времени.

Для аппроксимации поведения спроса на продукцию конкретных производителей во времени предлагается аппарат так называемых S -кривых (символом S обозначается основная форма кривой). В частности, такая S -кривая может быть найдена как решение следующего дифференциального уравнения:

$$\frac{dz}{dt} = l(1 - z(t)/p)[z(t)/p]^{-1},$$

где $z(t)$ — уровень спроса в момент времени t , p — доля продукции данного производителя в установившемся режиме, l — постоянная времени.

Решение этого уравнения при $t \rightarrow \infty$ хорошо аппроксимируется следующей функцией

$$z(t) \approx p(1 + be^{-lt}). \quad (1)$$

При наличии многих производителей соответствующие переменные $x(t)$ должны быть снабжены индексом производителя (или вида товара) i .

Чтобы оценить параметры S -кривых в условиях неполной информации о состоянии рынка воспользуемся схемой экспертно-статистической обработки на базе метода анализа иерархий (МАИ) [8]. Однако, прежде чем сделать это, рассмотрим еще один более наглядный подход к задаче оценки спроса на товарном рынке. Проиллюстрируем подход на примере рынка персональных компьютеров (ПК). В этом случае в товаропроводящей сети можно выделить собственно рынок (потребителей), изготовителей ПК, изготовителей комплектующих узлов (микросхем) и системного программного обеспечения.

Тогда уравнение материального баланса в вершине i сети будет иметь вид:

$$\sum_{j_1} z_{i,j_1} + z_{\text{экспорт}} = \sum_{j_2} z_{j_2,i} + z_{\text{экспорт}}, \quad (2)$$

где z_{i,j_1} — объем продажи потребителю j_1 в вершине i , $z_{j_2,i}$ — объем поставок в вершину i от поставщика j_2 , а

$z_{\text{экспорт}}$ и $z_{\text{импорт}}$ — объемы экспорта и импорта в вершине i , соответственно.

Тогда задача прогнозирования рынка ПК может быть поставлена как следующая задача оптимизации: найти

$$\min_t \left\{ z_j(t) - \sum_i f_i(z_{ij}(t)) \right\} \quad (3)$$

при условиях (2) и выполнении ограничений вида

$$b_i < f_i(z_{si}) < d_i, \quad (4)$$

где f_i — производственная функция поставщика i , работающего на удовлетворение спроса потребителя j ; параметры b_i и d_i определяют максимальную и минимальную производительности, соответственно; t — время; s — номер поставщика комплектующих узлов. В любой момент времени, конечно, должно выполняться условие $f_i(z_{ij}) < c_{ij}$, где c_{ij} — предельный уровень спроса, который связан с производственной функцией соотношением типа формулы (1):

$$f_i(z_{ij}(t)) = c_{ij}(1 + b_{ij}e^{-a_{ij}t}).$$

Теперь можно вернуться к МАИ. В его рамках задача преобразуется в некоторую иерархическую структуру, на каждом уровне которой находится нескольких элементов. Каждый из этих элементов в свою очередь (на следующем уровне) также представляется некоторой совокупностью элементов. Каждому элементу любого уровня приписывается сравнительный уровень значимости, например, от 1 до 9, где «единицей» обозначается равнозначность элементов, а «девяткой» — абсолютная степень превосходства. При этом два элемента одного уровня сравниваются с соответствующим элементом более высокого уровня. Результаты сравнения после обработки заносятся в квадратную матрицу, размер которой равен числу элементов данного уровня. Способ обработки выбирается так, чтобы получилась матрица, обратная матрице значений относительных коэффициентов значимости. Нормированное значение доминантного собственного вектора этой матрицы определяет относительную значимость элементов по сравнению с вышестоящим элементом.

Если применить МАИ к прогнозированию спроса на рынке, то, используя в качестве основы создания оценочной иерархии модель товаропроводящей сети (2)–(4), необходимо сформировать МАИ для соответствующей сетевой схемы [9]. Применительно к рынку персональных компьютеров на одном из уровней МАИ могут быть размещены такие характеристики, как производственно-технологические особенности, «дружественный» интерфейс, внешняя привлекательность, цена и прочее (уровень требований). На альтернативном уровне МАИ могут быть представлены такие характеристики, как оценки конкурентоспособности фирм-производителей, политика ценообразования, технология использования. В результате введения этих уровней и соответствующей параметризации формируется набор весовых коэффициентов, которые характеризуют относительные уровни значимости и могут быть получены посредством опроса экспертов. Найденные значения параметров непосредственно связываются со значениями параметров S -кривой.

Предложенная процедура формирования оценок параметров S -кривой чисто экспертная. Одновременно можно осуществить обработку ретроспективных данных

о спросе, на основе которой будут сформированы «объективные» (или более объективные) оценки уровня насыщения p (в установившемся режиме) и постоянной времени l . Эти оценки могут быть предъявлены экспертам и стать для них дополнительным источником информации, используя который эксперты могут «объективизировать» высказываемую ими точку зрения.

Так, например, допустим, что поведение цен свидетельствует о том, что в настоящее время на рынке имеет место равновесие. Используя этот факт, можно сформировать алгоритм оценки функции спроса вида [10]:

$$\hat{x}_{t+1} = x_t - g_t(\hat{p}_t - p_0),$$

где $\hat{x}_{t+1}$ — оценка спроса на $(t+1)$ -м шаге, x_t — спрос на t -м шаге, $\{g_t\}$ — последовательность коэффициентов, удовлетворяющих известным условиям на коэффициенты алгоритмов метода стохастической аппроксимации, $\hat{p}_t$ — оценка степени удовлетворения спроса на t -м шаге (в данном случае совпадает с фактической рыночной долей), а p_0 — заданное значение (в данном случае совпадает с желаемой рыночной долей).

В этом случае эксперту могут предъявляться для корректировки значения $\hat{p}_t$ и значения $\hat{x}_{t+1}$. Последние используются для корректировки значений коэффициентов g_t .

4. УПРАВЛЕНИЕ РЕСУРСАМИ ДЛЯ РАБОТЫ В АВАРИЙНЫХ СИТУАЦИЯХ

Солидная часть предъявляемого на рынке спроса обусловлена запросами на ресурсы, которые формируются в силу хозяйственной или каких-либо других форм деятельности предприятий, действующих на рынке в роли потребителей. Поэтому изучению механизмов выбора структуры ресурсов придавалось весьма большое значение. В соответствии с указанными в § 2 границами применимости экспертно-статистического подхода ниже рассматривается случай решения проблемы выбора структуры ресурсов в условиях достаточно высокой степени неопределенности исходных характеристик.

Именно такие задачи распределения ресурсов возникают, в частности, при планировании деятельности подразделений, направляемых для работы в редко возникающих ситуациях, например, чрезвычайных или аварийных [11].

Вероятные варианты аварийных ситуаций классифицированы по типам, поэтому рассмотрение ведется в рамках некоторого фиксированного типа (класса) аварийных ситуаций, преодоление которых требует наличия некоторого набора ресурсов, характеризуемого вектором $(x_1, x_2, \dots, x_k)^T$. Здесь компоненты вектора ресурсов могут представлять собой численности групп специалистов, состав парка транспортных средств, интенсивность материальных и финансовых потоков и т. п.

Предполагается, что время возникновения, глубина и время преодоления последствий аварийных ситуаций случайны, при этом соответствующие случайные величины обладают неизвестными априори вероятностными характеристиками. Иначе говоря, считается, что неизвестны следующие функции распределения (ф. р.): $F_n(\theta_n)$ — ф. р. интервала (θ_n) от момента преодоления последствий n -й аварийной ситуации до возникновения



$(n + 1)$ -й аварийной ситуации, $G_n(\tau_n)$ — ф. р. времени τ_n преодоления последствий n -й аварийной ситуации и $H_n(\mathbf{u}_n)$ — ф. р. «глубины» n -й аварийной ситуации, характеризующейся запросом на вектор ресурсов $\mathbf{u}_n$, который связан с самим фактом возникновения n -й аварийной ситуации при условии выбора «оптимального» плана преодоления ее последствий («оптимальность» завышена, поскольку в условиях неопределенности оценка оптимальности может быть лишь субъективной, поэтому фактически следует говорить не об оптимальности, а о профессиональном, грамотном планировании работ по преодолению последствий аварийной ситуации).

Заданы следующие компоненты затрат: $A(\mathbf{u}_n)$ — единовременные потери от возникновения n -й аварийной ситуации; $B_\tau(\mathbf{x} - \mathbf{u}_n)$ — интенсивность затрат на преодоление последствий n -й аварийной ситуации в момент времени $\tau \in [0, \tau_n]$; $C_\theta(\mathbf{x})$ — интенсивность затрат на поддержание уровня ресурса $\mathbf{x}$ в момент времени $\theta \in [\theta_n + \tau_n, \theta_{n+1}]$.

Нетрудно показать, что возникновение каждой аварийной ситуации, принадлежащей к рассматриваемому классу, может быть использовано для последовательного уточнения оценок необходимого вектора ресурсов $\mathbf{x}$. А именно, оценка вектора ресурсов, который необходимо иметь после преодоления последствий n -й аварийной ситуации, может быть представлена в виде:

$$\hat{\mathbf{x}}_{n+1} = \mathbf{x}_n - \Gamma_n \left[\int_0^{\tau_n} \text{grad}_{\mathbf{x}_n} B_\tau(\mathbf{x}_n - \mathbf{u}_n) d\tau + \int_{\theta_{n-1} + \tau_{n-1}}^{\theta_n} \text{grad}_{\mathbf{x}_n} C_\theta(\mathbf{x}_n) d\theta \right], \quad (5)$$

где через Γ_n обозначена матрица коэффициентов, удовлетворяющих известным условиям на коэффициенты алгоритмов стохастической аппроксимации, а $\text{grad}_{\mathbf{x}_n} G(\mathbf{z})$ — градиент функции $G(\mathbf{x}_n)$ в точке $\mathbf{x}_n = \mathbf{z}$.

Выбор начальных условий и значений коэффициентов в алгоритме (5) должен осуществляться с учетом требования обеспечения достаточно высокой скорости сходимости (акселерации) алгоритма к некоторой окрестности на конечном интервале значений n .

ЗАКЛЮЧЕНИЕ

В 1990-х гг. для решения некоторых задач исследования операций был предложен экспертно-статистический подход к обработке информации, суть которого заключалась в том, чтобы в процессе идентификации модели, предназначенной для синтеза системы принятия решений (например, при управлении материально-техническим снабжением) воспользоваться «на равных» информацией объективного (результатами измерений) и субъективного (экспертными оценками) происхождения. В настоящей работе сделана первая попытка распространить идеологию экспертно-статистического подхода на решение различных задач управления в рамках глобальной структуры интегрированной системы управления современным промышленным предприятием.

Очерчены границы применимости классического (названного в настоящей работе статистическим), экс-

пертно-статистического и чисто экспертного подходов к решению проблем АСУТП и ERP-планирования. Эти границы нечеткие, они размыты. Однако при некоторых вариантах исходных данных — см. таблицу в § 2 — экспертно-статистические системы оказываются для разработчиков наиболее предпочтительным вариантом (в качестве альтернативы подходам, основанным на статистических методах идентификации и методах построения комплексов взаимодействующих локальных физико-химических и физико-механических моделей).

Применительно к задачам ERP-планирования также возможно применение экспертно-статистических методов обработки информации в сочетании с такими методами экспертного оценивания, как, например, метод анализа иерархий Т. Саати.

Важным компонентом экспертно-статистических систем являются адаптивные и робастные схемы управления. Они, нередко, составляют алгоритмическую основу, используемую не только для уточнения статистических параметров базовой модели, но и являющуюся инструментом интеграции в эту модель оценок, полученных чисто экспертными методами. Иллюстрацией этой идеи служит описанная в статье (см. § 4) алгоритмическая модель, которая предназначена для решения проблемы планирования оптимальных ресурсов на случай возникновения на предприятии аварийных ситуаций.

ЛИТЕРАТУРА

1. Соркин Л. Р. Современные технологии управления в нефтегазовом комплексе. — М.: МФТИ, 2003. — 104 с.
2. Мандель А. С. Экспертно-статистические системы в задачах управления и обработки информации. Часть I // Приборы и системы управления. — 1996. — № 12. — С. 34–36.
3. Мандель А. С. Экспертно-статистические системы в задачах управления и обработки информации. Часть II // Там же. — 1997. — № 2. — С. 11–13.
4. Беляков А. Г., Мандель А. С. Прогнозирование временных рядов на основе метода аналогов (элементы теории экспертно-статистических систем). — М.: ИПУ РАН, 2002. — 60 с.
5. Дозорцев В. М., Кнеллер Д. В. Типовой тренажерный комплекс для обучения операторов ТП // Автоматизация в промышленности. — 2003. — № 2. — С. 9–14.
6. Дорофеев А. А., Касавин А. Д., Торговицкий И. Ш. Контроль качества и восстановление характеристик объектов, работающих в нескольких режимах // В кн.: Доклады II Всесоюзного совещания по статистическим проблемам теории управления. Идентификация. — М.: Наука, 1970. — С. 181–182.
7. Касавин А. Д. Адаптивные алгоритмы кусочной аппроксимации в задаче идентификации // Автоматика и телемеханика. — 1972. — № 12.
8. Саати Т. Метод анализа иерархий. — М.: Сов. Радио, 1992.
9. Мандель А. С. Экспертно-статистические системы и метод анализа иерархий // В кн. Управление большими системами. Материалы Междунар. научн.-практ. конф. — М.: СИНТЕГ, 1997.
10. Лотоцкий В. А., Мандель А. С. Адаптивное управление запасами // Тез. докл. 10-го Всесоюзного совещания по проблемам управления. — М., 1986.
11. Мандель А. С. Экспертно-статистический подход к решению задачи оперативного управления работами по преодолению последствий ЧС // Проблемы управления в чрезвычайных ситуациях: Тез. докл. Четвертой междунар. конф. — М., 1997. — С. 65.

☎ (495) 334-89-69;

e-mail: manfoon@ipu.ru



О РИСКАХ, СВЯЗАННЫХ С ОШИБКАМИ ЭКСПЕРТОВ И АНАЛИТИКОВ¹

Н. А. Абрамова, С. В. Коврига

Институт проблем управления им. В. А. Трапезникова РАН, г. Москва

Представлен новый подход к анализу ситуаций, традиционно интерпретируемых как ошибки экспертов, аналитиков, лиц, принимающих решение. Подход опирается на анализ практической деятельности экспертов и теоретические модели, учитывающие как практику, так и психологические знания. Факторы риска, обусловленные неадекватной интерпретацией «ошибок», анализируются на известных примерах нарушения транзитивности предпочтений.

ВВЕДЕНИЕ

В условиях расширяющегося внедрения в практику различных компьютерных технологий, обрабатывающих качественную информацию и оценочные данные, усиливается роль человеческих факторов, и становится актуальной проблема рисков, связанных с ними, особенно когда такие технологии применяются в ответственных приложениях с высокой ценой возможного ущерба [2]. Эта проблема относится к различным типам технологий поддержки организованной интеллектуальной деятельности (называются ли эти технологии информационными, экспертными, интеллектуальными, когнитивными, поддержки принятия решений или иначе)². Ключевым признаком, объединяющим такие технологии, с точки зрения человеческих факторов и связанных с ними рисков, является формализация знаний экспертов, аналитиков, лиц, принимающих решение (ЛПР), и других носителей знаний в проблемной области. Для удобства анализа мы называем этих специалистов обобщенным понятием «эксперт-аналитик».

Наш анализ показывает, что одним из значимых аспектов проблемы рисков являются практические ситуации, в которых «носители» технологий оценивают знания экспертов-аналитиков в ходе их формализации и нередко квалифицируют их как ошибочные (см., например, работы [3, 4]). Альтернативное отношение к этим

ситуациям (не исключающее возможности ошибок) связано с понятием психологической корректности технологий и используемых в них методов и операций приобретения знаний у человека [2, 5, 6]. Интуитивный смысл этого понятия, введенного О. И. Ларичевым и его коллегами³, состоит в том, что человек может предоставлять недостаточно надежную информацию в условиях, когда разработчики тех или иных методов и технологий не учитывают психологических особенностей «человеческой подсистемы переработки информации»⁴.

Настоящая статья посвящена анализу рисков, связанных с такими ситуациями. Анализ основан на изучении практики применения экспертных методов и теоретических исследований, которые вкратце представлены далее. К ним относятся:

- модель действия стереотипов [2, 6], опирающаяся, в основном, на работу [7], а также наши исследования [8];
- обобщенная модель эксперта-аналитика;
- оригинальная концепция схем представления знаний, ориентированных на корректность [8, 9].

Статья завершается анализом рисков на примере известной ситуации нетранзитивности парных предпочтений.

¹ Статья представляет собой расширенное изложение доклада, представленного на конференции [1], отредактированное и дополненное некоторыми ссылками на публикации.

² В докладе [1] такие технологии для краткости обобщенно называются интеллектуальными.

³ «Под психологически корректными мы понимаем такие операции получения информации от ЛПР, использование которых приводит к согласованным, непротиворечивым результатам» [5].

⁴ Расширение понятия психологической корректности в работах [2, 6] рассчитано на учет факторов психологического влияния, которые действуют путем навязывания человеку не только отдельных операций, но и всей структуры знаний и деятельности, ведущей к получению конечного результата.



1. МОДЕЛЬ ДЕЙСТВИЯ СТЕРЕОТИПОВ И ФАКТОРОВ ВЛИЯНИЯ

Вкратце модель, определяющая действие стереотипов и других эвристик и факторов влияния при принятии решений [3, 4], такова.

- Всякая мыслительная операция, которая может быть интерпретирована как принятие решения⁵, осуществляется посредством (не обязательно осознаваемой) операции подбора эвристики: готового решения или способа, при помощи которого затем находится и принимается решение⁶.
 - Человек располагает эвристиками разной степени проверенности, освоенности, пригодности для конкретной ситуации. Среди них особое место занимают стереотипы, которые обладают повышенной активностью (связываемой с механизмом условного рефлекса), благодаря закреплению опытом успешного применения или влиянию, например, научных парадигм.
 - Применение немотивированных эвристик, неадекватных конкретной ситуации, является довольно типичным, и этому способствуют:
 - объективные и субъективные условия, такие как недостаток времени, информационная сложность ситуации, неопределенность и др.;
 - специфика действия стереотипов при подборе эвристик;
 - факторы веры — факторы влияния, повышающие субъективную уверенность в адекватности решений, помимо объективной адекватности, вплоть до отклонений от рационального поведения.
- Типичными факторами веры являются принципы Чалдини: принцип последовательности, срабатывающий в силу обязательства (выбранной позиции), принцип социального доказательства, принцип авторитета.

2. МОДЕЛЬ ЭКСПЕРТА-АНАЛИТИКА

Исходя из анализа практической деятельности экспертов-аналитиков, которая не всегда согласуется с нормативными моделями структуры их знаний и деятельности в различных технологиях, и, ориентируясь на анализ рисков, принята следующая модель эксперта-аналитика.

Специалист является экспертом-аналитиком в некоторой области профессиональной деятельности, если он:

- *обладает совокупностью эвристик*, рутинное применение которых в стереотипных для него ситуациях обеспечивает относительно успешное решение (общих или частных) задач, адекватных ситуации, по крайней

мере, в условиях отсутствия факторов влияния, толкающих к отказу от применения собственных профессиональных и интеллектуальных знаний и умений;

— способен с более или менее высокой степенью успеха *отличать нестереотипные (и пограничные) ситуации* из своей области деятельности и — в случае ситуаций, для которых его стереотипные эвристики неадекватны, — *творчески решать задачи* на основе собственных эвристик или привлекаемых знаний, *ориентируясь на успешность конечных результатов* этой деятельности⁷.

Предполагается, что знание, которое формирует эксперт-аналитик, действующий в области стереотипных для него ситуаций, является наиболее достоверным из-за успешности используемых эвристик, подтвержденной практикой, тогда как в случае нестереотипных ситуаций достоверность снижается из-за риска неадекватных решений. (Для простоты мы не касаемся ситуаций, когда результаты работы эксперта-аналитика недостоверны из-за недостоверной исходной информации или иных объективных факторов).

Предполагается также, что за рамками названных областей более или менее результативной (по успешности, достоверности) деятельности, в которых человек выступает как эксперт-аналитик за счет репродуктивного или творческого переноса своих профессиональных знаний и умений, лежат области с меньшей достоверностью результатов, в которых возрастает роль немотивированных эвристик. Грубо говоря, они составляют область «обыденного мышления», когда эксперт уже перестает быть экспертом⁸.

Помимо характера решаемых задач и имеющихся для этого мыслительных средств, на достоверность результатов деятельности эксперта-аналитика в рамках определенной технологии влияет *фактор навязанности способов мышления*, который обусловлен регламентацией структуры знаний и интеллектуальной деятельности эксперта-аналитика. В частности, речь идет о моделях знаний, в соответствии с которыми его знания (исходные или порождаемые в ходе анализа) должны быть формализованы для их последующей обработки формальными методами.

О влиянии фактора навязанности на достоверность результатов и на наличие в них ошибок прямо или косвенно свидетельствуют как практические наблюдения, так и теоретические исследования, в том числе, по психологической корректности интеллектуальных и экспертных технологий [2, 5, 6]; по действиям человека в условиях, когда алгоритм в него «введен извне» [11].

Как следует из экспериментальных исследований и объяснительных гипотез психолога С. И. Шапиро [11], фокусирование сознания на выполнении отдельных умственных действий угнетает в сознании «логический

⁵ Принятие решения понимается в психологическом, а не формальном теоретическом смысле; при этом, например, любая единичная операция оценивания — это принятие решения.

⁶ В составе эвристик могут быть и знания, приобретенные из опыта, не обязательно осознаваемые, и научные знания, которые проверены на практике, или, по крайней мере, истинны и значимы для практики, с точки зрения эксперта, и мыслительные средства манипулирования профессиональными знаниями.

⁷ В психологии выделенные области деятельности соответствуют репродуктивному переносу стереотипных знаний и творческому переносу [8].

⁸ Противопоставление экспертного и обыденного мышления заимствовано из работы Л. Г. Ионина [10], который различает эти виды мышления по используемым стереотипам (хотя и в несколько иной терминологии).

механизм», «генерирующий и контролирующий эти действия» и придающий им целостность. Затруднения возникают из-за недостаточного понимания роли отдельных действий в общей структуре; теряется «схватывание целого» (гештальта), которое играет решающую роль. Человек оказывается в ситуации «операторного» уровня мышления, характерного для неопытных программистов.

Изложенная модель эксперта-аналитика в сочетании с моделью действия стереотипов и факторов влияния позволяет предположить, что навязанность определенных форм (схем) мышления является значимым фактором в ряду условий, при которых те или иные методы решения задач, основанные на операциях получения знаний от человека, становятся психологически некорректными. Разумеется, степень риска снижения достоверности знаний, получаемых от эксперта-аналитика знаний при недостаточном использовании его собственных ресурсов компетентности, зависит от многих факторов, и этот вопрос требует изучения.

3. СХЕМЫ ПРЕДСТАВЛЕНИЯ ЗНАНИЙ, ОРИЕНТИРОВАННЫЕ НА КОРРЕКТНОСТЬ

Чтобы связать модель эксперта-аналитика, в рамках которой рассматривается вопрос достоверности знаний, с рисками в ситуациях, которые расцениваются как ошибки носителей знаний о проблемной области, обратимся вначале к понятию корректности (отсутствия ошибок).

Наши прежние поиски ответа на вопросы о том, какие знания считать корректными и какими путями обеспечить их корректность, привели к гипотезе, согласно которой *человеческие способности оценивать те или иные знания как некорректные связаны с представлением знаний по определенным схемам*⁹ [8, 9]. Возникла концепция *схем представления знаний, ориентированных на корректность*, т. е. схем, которые, благодаря присущей им общей структуре знаний или определенным семантическим ограничениям на наполнение этой структуры, обеспечивают определенные свойства корректности конкретных знаний¹⁰.

Поясним действие схем представления знаний, ориентированных на корректность, при формализации, т. е. представлении конкретных знаний по определенной схеме. (Предполагается, что схема является общей для соответствующего типа знаний.) При формализации или порождении новых знаний по заданной схеме воз-

можна конфликтная ситуация. Она состоит в том, что общая схема, по априорной оценке человека, применима к рассматриваемому конкретному знанию, но представить его по этой схеме, иначе говоря, вложить его в схему, не удастся: знание оказывается так или иначе не соответствующим схеме. Такой конфликт нередко выражается в форме субъективных оценок типа «не понимаю», «что-то тут не так», «неправильно» и т. п.¹¹.

Типичная реакция носителя знаний в случаях конфликта между схемой и конкретным знанием, которое в нее вкладывается, заключается в том, что *приоритет схемы над конкретным знанием признается как какая-то некорректность этого знания*, даже если в первичном представлении (т. е. в представлении по другим, не обязательно осознаваемым, схемам) некорректность не была видна. Именно в этом, в *признании приоритета общей схемы* (заимствованной из логики, опыта, определенной теории, действующих норм и т. д.) *над конкретными знаниями* при оценке их корректности и состоит *ключевое свойство схем, ориентированных на корректность*.

По-видимому, древнейшими общепризнанными схемами, ориентированными на корректность (для знаний в виде умозаключений), являются силлогизмы Аристотеля, которые выглядят присущими человеческому разуму. Характерными примерами некорректностей, выявляемых посредством подходящих схем, служат неполнота, неоднозначность, противоречивость и т. п. Оценка знания как некорректного, содержащего какую-то ошибку (или ошибки), означает признание необходимости его корректировки, например, дополнения недостающего, согласно схеме, фрагмента.

Формально субъективную реакцию носителя первичных знаний на оценку результата применения схемы, ориентированной на корректность, можно описать утверждением

$$\begin{aligned} & \text{применимо } (S, k) \rightarrow \\ & \rightarrow (\text{некорректно } (k, S) \rightarrow \text{некорректно } (k)), \end{aligned} \quad (1)$$

где S — схема (общая для однотипных, в формальном аспекте, знаний); k — конкретное знание; *применимо* (S, k) — оценка применимости схемы S к знанию k ; *некорректно* (k, S) — оценка несоответствия знания k схеме S (иначе, наличия конфликта между ними), возникшая при применении S к k , в какой бы словесной формулировке эта оценка ни выражалась; *некорректно* (k) — оценка некорректности знания k как такового, безотносительно к схеме. При этом все оценки считаются бинарными (да, нет).

Выражение (1) означает, что если человек признает применимость общей схемы к его конкретному знанию, он признает и некорректность этого знания, если она проявляется при применении к нему схемы.

Как следует из выражения (1) некорректность знания относительно схемы проявилась, единственная возможность не отнести эту некорректность на счет конкретного знания — это принять, что схема не применима к данному знанию. И действительно, наряду с типичной

⁹ Известное понятие схемы представления знаний, или просто схемы, довольно подробно анализируется в работе [8].

¹⁰ Эти вопросы, в первую очередь, исследовались в рамках проблемы надежности программного обеспечения, которая была порождена ошибками в компьютерных программах. Было известно, с одной стороны, что формализация способствует обнаружению ошибок, а с другой, — что она может быть источником новых ошибок, ошибок перевода по Г. Майерсу [12]. Вместе с тем уже тогда нам стало понятным, что проблема некорректности формализованных знаний может быть актуальной и для информационных систем и систем, основанных на знаниях [9].

¹¹ Для простоты мы не рассматриваем ситуацию, когда трудности обусловлены тем, что человек не владеет схемой.



реакцией носителя знаний на конфликт конкретного знания и схемы, на практике случается смена априорной, до применения схемы, оценки «схема применима к данному знанию» на апостериорную — «схема не применима». Тем самым признается приоритет конкретного знания — надо менять схему.

Стоит уточнить, что, как показывают наблюдения, в конкретных ситуациях, подразумевающих оценку формализованного знания, роль схем может не осознаваться, но схемы проступают более или менее явно при обосновании необходимости коррекции. Особенно это заметно в ситуациях, когда конфликт обнаружен не носителем знания, а другим человеком. То же относится и к ситуациям, когда оценка «неправильно» появляется без предварительно заданной схемы. (Человек подбирает нужную схему (норму), чтобы обосновать свою оценку «неправильно».)

В изложенной концепции схем представления знаний, ориентированных на корректность, неявно предполагается, что схема может выступать как ориентированная на корректность в рамках определенного человеческого сообщества, если она признается в этом качестве всеми членами сообщества как *общая норма*. Это является еще одним *ключевым свойством* схем представления знаний, ориентированных на корректность. Оно создает основу для достижения соглашения относительно корректности конкретных знаний или необходимости исправлений.

Позиция носителя знаний выделена потому, что именно его знания должны подвергаться коррекции в случае обнаружения конфликта. Коррекции может подвергаться не только фрагмент, который оценивается как некорректный, но и, возможно, все, что с ним связано в его внутренних представлениях.

В различных технологиях в качестве схем, по которым осуществляется представление знаний экспертов-аналитиков, выступают *общие модели знаний, лежащие в основе технологий*. В целях сокращения рисков, ведущих к недостаточной достоверности знаний, порождаемых по некоторой технологии, естественно стремиться к тому, чтобы используемые модели знаний были ориентированы на корректность. В противном случае может иметь место *эффект маскирования ошибок*. Комплекс вопросов, касающихся рисков, обусловленных таким маскированием, составляет самостоятельную проблему. Далее мы ограничиваемся рассмотрением рисков, связанных с противоположными ситуациями, когда конфликт между общей моделью знаний и формализуемым знанием проявляется, что может, но отнюдь не обязательно, свидетельствовать об ошибках.

4. РИСКИ ПРИ РАЗРЕШЕНИИ КОНФЛИКТА СХЕМЫ И КОНКРЕТНОГО ЗНАНИЯ

Рассмотрим ситуацию разрешения конфликта между общей схемой представления знаний эксперта-аналитика, лежащей в основе некоторой технологии, и конкретным знанием конкретного эксперта-аналитика. При

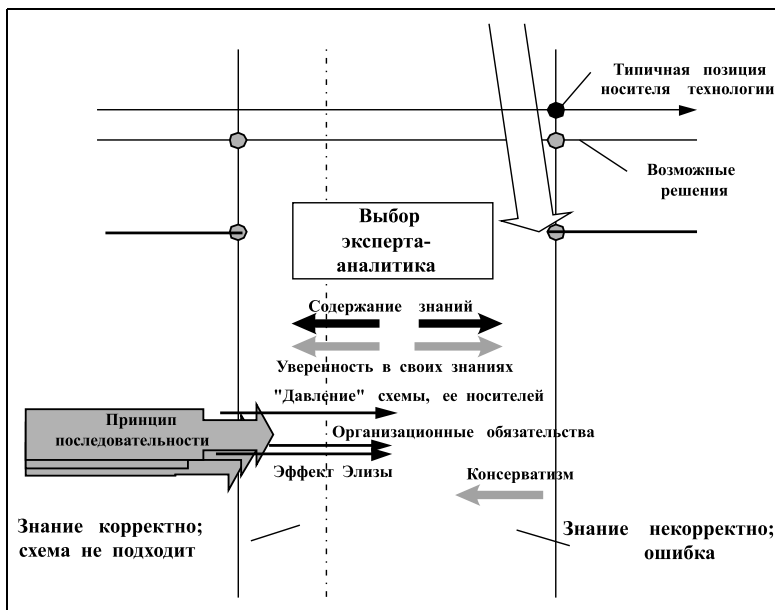


Рис. 1. Разрешение конфликта схемы и конкретного знания

этом не предполагается, что схема обязательно ориентирована на корректность. Предполагаем, что в процессе формирования формализованных знаний участвуют и эксперт-аналитик, и носитель технологии.

Имеются два возможных решения, снимающих конфликт. Одно из них соответствует оценке «знание некорректно; ошибка», другое — оценке «знание корректно; схема не подходит»¹² (рис. 1). Первое решение, в пользу схемы, означает необходимость коррекции знаний эксперта-аналитика, второе — отказ от схемы. Во втором случае при «жестких» технологиях возможен отказ от технологии или, напротив, от эксперта или его мнения.

Для носителя технологии типичная позиция состоит в решении «ошибка». Это легко объяснить моделью действия стереотипов: принятие приоритета схемы, заложенной в технологию, соответствует влиянию принципа последовательности. По той же модели, выбор эксперта-аналитика обусловлен целым рядом факторов влияния.

Вначале будем полагать, что решаемые задачи и формализуемые знания относятся к области наиболее достоверного знания эксперта-аналитика.

В сторону принятия адекватного решения, каким бы оно ни было, его толкают *объективное содержание знаний и субъективная уверенность в своих знаниях*. В сторону решения «ошибка», независимо от его адекватности, толкает, прежде всего, принцип последовательности, а также, в большей или меньшей степени — другие при-

¹² Для простоты мы исключаем из рассмотрения решения, требующие рефлексивного осознания того, что формализованное знание и первичное (внутреннее) знание эксперта-аналитика различаются по смыслу. Такое осознание требуется, в частности, для того, чтобы идентифицировать «описки» — расхождения между первичным знанием и его описанием.

нции и правила Чалдини: правило «взаимного обмена» (с носителями технологии), принципы благорасположения, авторитета, социального доказательства. Они воплощаются в ряде факторов, таких как «давление» схемы, которая уже принята и может быть более или менее признанной, организационные обязательства. Игрет роль и так называемый «эффект Элизы», известный по исследованиям психологов в области искусственного интеллекта, который состоит в преувеличенном доверии к решениям, полученным формальными методами. В своем большинстве факторы этой группы являются относительно объективными, т. е. не зависят от особенностей конкретной личности (хотя реакция на них зависит от личности). В противоположную сторону, своей правоты, независимо от адекватности решения, может толкать субъективный фактор консерватизма — неприятия навязанных схем.

Как показывает анализ выделенной совокупности факторов влияния, при разрешении конфликта *преобладающую роль играют систематические факторы влияния в сторону интерпретации конфликта как ошибки эксперта-аналитика*, прежде всего — принцип последовательности. При этом рассматриваемые факторы влияют лишь на субъективную уверенность в решении, *независимо от его объективной адекватности*.

Тем самым, *имеется риск ошибочно признанной ошибки* и, как следствие, ошибочно скорректированного знания эксперта-аналитика в ходе применения технологии: замены более или менее достоверного знания, не вкладываемого в схему, на результат его подгонки под схему. Объективная достоверность формализованных знаний при наличии такого риска снижается.

Основным противовесом от принятия экспертом-аналитиком ошибочного решения о признании ошибки служит его субъективная уверенность в своих знаниях, относящихся к области высокой достоверности его решений.

Специфика схем, ориентированных на корректность, при разрешении конфликта состоит, прежде всего, в существенном усилении влияния схемы, поскольку она признана как общая норма. Решая вопрос о рисках, связанных с признанием ошибки при использовании схем, ориентированных на корректность, естественно предположить, что такого рода схема становится нормой благодаря опыту успешного использования. Иными словами, схема, ориентированная на корректность, может рассматриваться как общепризнанная эвристика, ставшая нормой, подтвержденной опытом успешного использования. Как и всякая эвристика, она имеет свои границы успешного применения. Тогда можно ожидать, что риск мал или даже пренебрежим при использовании таких схем в своих границах. Можно ожидать увеличения риска в условиях, когда

- принятие схемы в качестве нормы не обусловлено опытом успешного использования;

- границы применимости нечетки.

Обратимся теперь к ситуациям, которые относятся к областям меньшей достоверности формализованных знаний. Обобщенно рассматривая изменения среди факторов, влияющих на принятие решения о наличии ошибки по отношению к ситуациям с высокой досто-

верностью решений, в среднем можно ожидать снижения субъективной уверенности эксперта-аналитика в достоверности формализованных знаний. (Например, это может происходить из-за осознания недостаточности своих профессиональных ресурсов для формирования знаний приемлемой достоверности.) Однако при этом нет оснований считать, что риск ошибочно признанной ошибки существенно изменится.

5. АНАЛИЗ ПРИМЕРА НЕТРАНЗИТИВНОСТИ ПРЕДПОЧТЕНИЙ

Рассмотрим ситуацию, которую обычно связывают с нарушением транзитивности строгих предпочтений, определяемых экспертом на некотором множестве объектов. Нетранзитивность отношения строгого предпочтения, «>», при ранжировании трех объектов a , b и c проявляется в том, что имеют место оценки $a > b$, $b > c$ и при этом $c > a$, а не $c < a$, как это было бы в случае транзитивности. Нетранзитивность предпочтений довольно типична в практике принятия решений [4] и нередко интерпретируется сторонниками определенных технологий и их теоретиками как ошибка экспертов или, по крайней мере, нарушение тех или иных норм рациональности. Например, Ю. В. Сидельников говорит о «нарушении основного правила логического вывода — транзитивности» и непоследовательности экспертов [3], Б. Г. Литвак — о противоречивости и непоследовательности [4].

Если следовать логике теоретиков, говорящих о противоречивости и непоследовательности, и принять во внимание понятие психологической корректности операций получения информации от человека [5], операции парных сравнений, результат которых должен выражаться в виде строгих предпочтений, приходится рассматривать как психологически некорректные. Ведь признаками этого являются несогласованность и противоречивость результатов¹³.

Точка зрения, согласно которой применение принципа транзитивности является свидетельством рациональности мышления, а «нетранзитивное поведение» говорит об отклонениях от рациональности, распространена и среди психологов. Достаточно назвать широко известные работы А. Тверского и Д. Канемана [13].

Вместе с тем на сегодня накоплено немало свидетельств в пользу того, что применимость (адекватность) принципа транзитивности предпочтений ограничена. Исследования в этом направлении, свидетельствующие о динамике в понимании проблемы транзитивности, сегодня ведутся среди психологов. Назовем лишь недавнюю публикацию А. Н. Поддьякова [14], в которой представлено сложившееся представление о принципе транзитивности предпочтений в психологии и собраны и обобщены результаты целого ряда исследований, опровергающих универсальность принципа. Опровержение, в основном, строится путем демонстрации объек-

¹³ Нельзя не отметить, что сам О. И. Ларичев относил операции попарного сравнения к числу надежных, психологически корректных операций.

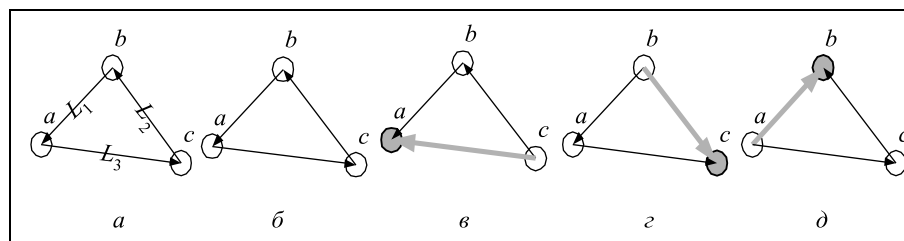


Рис. 2. Варианты сравнительных оценок:

a и b — исходные экспертные оценки согласно отношениям (2) и (3), $в$, $г$ и $д$ — альтернативные исправления «ошибки» в варианте 2 поведения эксперта

тивных отношений превосходства (как более или менее известных по публикациям, так и найденных автором), которые характеризуются нетранзитивностью.

Все больше свидетельств в пользу того, что нетранзитивные предпочтения могут иметь рациональные обоснования и объяснения, сегодня находят специалисты в области формальных методов принятия решений (см., например, работу [15]). Эти объяснения находятся в системе предпочтений экспертов, или, иначе говоря, в совокупности эвристик, применяемых при решении задачи. Характерная и, по-видимому, наиболее известная эвристика такого рода — это многокритериальная система предпочтений.

Хотя на сегодня ощутим недостаток единого методологического подхода к анализу многоаспектной проблемы транзитивности, который позволил бы согласовать различные точки зрения или, по крайней мере, объяснить их наличие и взаимосвязь, уже сейчас представляется целесообразным признать, что транзитивность предпочтений — это лишь проверенная практикой стереотипная эвристика. Она обеспечивает успешность на широком классе практических задач; однако, как и все эвристики, имеет ограниченную область применимости, о чем свидетельствует наличие многих контрпримеров.

Здесь мы остановимся на одном из аспектов нетранзитивности предпочтений: разрешении конфликтов при ее обнаружении в ходе применения экспертных методов и технологий и связанных с этим рисков. Анализ известного по публикациям примера, в котором применяются многокритериальные сравнения, проводится на основе представленных выше теоретических построений.

Пример. Пусть три объекта a , b и c сравниваются экспертом на основе признаков (критериев) L_1 , L_2 и L_3 , и в каждой паре доминирующим при оценке оказывается один из признаков, при неразличимости (примерном равенстве)¹⁴ по остальным. Результат сравнения можно представить отношением

$$(a \overset{L_1}{>} b \& a \overset{L_2}{\cong} b \& a \overset{L_3}{\cong} b) \& (b \overset{L_1}{\cong} c \& b \overset{L_2}{>} c \& b \overset{L_3}{\cong} c) \& (a \overset{L_1}{\cong} c \& a \overset{L_2}{\cong} c \& a \overset{L_3}{<} c),$$

которое, в свою очередь, исключением несущественных признаков сводится к отношению

$$a \overset{L_1}{>} b \& b \overset{L_2}{>} c \& a \overset{L_3}{<} c. \quad (2)$$

¹⁴ Отношение примерного равенства не является транзитивным.

Тогда, при необходимости вложить этот результат экспертного анализа в схему строгих парных предпочтений, его естественно редуцировать к виду

$$a > b \& b > c \& a < c, \quad (3)$$

где обнаруживается нетранзитивность.

Здесь хорошо видно, что нетранзитивность получается благодаря применению к знаниям эксперта более грубой модели знаний, которая навязана технологией; эта схема требует обобщения результатов попарных сравнений в виде оценок предпочтения. Можно привести и другие примеры эвристик сравнения, приводящих к сходным последствиям.

Рассмотрим три варианта возможной реакции эксперта на оценку отношения (3) как ошибочного, в соответствии с требованием транзитивности, выполнение которого позволяло бы определить наилучший элемент. Отношения (2) и (3) графически представлены на рис. 2, a и b .

Вариант 1. Пусть эксперт нацелен на задачу определения наиболее предпочтительного элемента. Пусть при этом он уверен в адекватности своих эвристик, использованных при сравнении (в выборе критериев, в оценке по отдельным критериям), уверен в достоверности полученных результатов, и пусть он слабо подвержен влияниям в сторону признания ошибки. Он может даже априорно признавать транзитивность предпочтений (т. е., с формальной точки зрения, признавать, что отношение предпочтений является отношением строгого порядка). Проведя свой анализ, он может прийти к выводу, что требование транзитивности неприменимо к конкретной решаемой задаче, в силу установленного им отношения (2), которое означает отсутствие наиболее предпочтительного элемента и в котором он уверен.

Непризнание экспертом требования транзитивности для конкретной задачи рационально обосновывается не только его недостаточной мотивированностью, но и в силу факта отсутствия наилучшего элемента в конкретном примере.

Вариант 2. Пусть, решая ту же задачу, эксперт больше подвержен влиянию навязанной схемы или иным факторам влияния. Тогда при помощи отношения транзитивности, он легко может подогнать результат под требование корректности, выбирая за основу любую пару предпочтений. На рис. 2, $в$, $г$, $д$ показаны три версии решения с использованием транзитивности для определения последней из трех оценок. (Серым цветом показано предположение, выводимое из более достоверных оценок, полученных по экспертной эвристике.) Недостоверность конечного результата: определения наиболее предпочтительного элемента, разного во всех трех вер-

сиях, — очевидна. Это — пример реализации риска исправления ошибочно признанной ошибки, результатом которой оказывается менее достоверное знание, чем исходное (рис. 2, а, в).

Понятно, что тем же путем эксперт мог действовать изначально, проводя детальное сравнение для каких-то двух пар элементов и к третьей применяя транзитивность. В сравнении с результатом более полного анализа всех трех пар это лишний раз свидетельствует о недостаточной мотивированности использования свойства транзитивности для решения задачи определения наиболее предпочтительного элемента при многокритериальном сравнении.

Если рассматривать отношение строгих предпочтений со свойством транзитивности в качестве схемы представления знаний эксперта при решении задачи определения наиболее предпочтительного элемента с позиций психологической корректности, даже рассмотренный частный пример позволяет считать, что эта схема обладает признаками психологической некорректности. Ведь она толкает эксперта к предоставлению знаний, не соответствующих его возможностям как эксперта (ресурсам компетентности). Правда, в данном случае некорректными оказываются не отдельные операции попарных сравнений (как это следовало бы из определения в работе [3]), а недостаточно мотивированное требование транзитивности системы предпочтений в целом (согласно определению из работы [6]). Источником риска можно считать то, что эвристика транзитивности опирается на обширный человеческий опыт успешного использования, сделавший ее стереотипом, при том, что границы ее применимости нечетки.

Вариант 3. Пусть эксперт нацелен на задачу (локальных) парных сравнений («от вас требуется только сравнение пар»), и при этом, как и в варианте 1, он уверен в адекватности своих эвристик и слабо подвержен влияниям в сторону признания ошибки. Оценка результата (3) (см. рис. 2, б) полученного редукцией адекватного отношения (2) (см. рис. 2, а) как ошибочного и здесь может быть не признана экспертом из-за неприменимости требования транзитивности, но уже по другой причине: транзитивность не требуется для определения предпочтения в любой отдельно взятой паре элементов. Последний вариант может показаться не убедительным «противной стороне». Однако следует иметь в виду, что задачи определения наиболее предпочтительного элемента и определения локальных предпочтений психологически различны и могут опираться на разные эвристики. Здесь мы вновь сталкиваемся с навязыванием психологически некорректного решения.

Основные выводы, которые можно сделать из анализа примера с формированием нетранзитивных экспертных предпочтений и гипотетических реакций экспертов на оценку «ошибка», в соответствии с описанным теоретическим подходом к анализу, таковы.

- Возможно более или менее рационально обоснованное непризнание экспертом требования транзитивности для конкретной задачи, причем — непризнание апостериорное (т. е. после применения к конкретным знаниям), опирающееся на приоритет конкретных знаний над общим требованием. Это является дополнительным свиде-

тельством того, что отношение предпочтений со свойством транзитивности не может рассматриваться в качестве общепринятой нормы или, иначе, в качестве универсальной схемы, ориентированной на корректность.

- Различные контраргументы в пользу транзитивности в случае многокритериальных сравнений, а значит, и в общем случае могут рассматриваться лишь как факторы психологического влияния на экспертов. (Анализ наиболее популярных контраргументов, таких как принцип «денежного насоса» А. Тверского или «нарушение основного правила логического вывода — транзитивности», выходит за рамки данной статьи, заслуживая отдельной публикации.)

- Можно ожидать, что некоторые ошибки, особенно относящиеся к области меньшей достоверности результатов деятельности эксперта, будут обнаруживаться посредством требования транзитивности (признаваться экспертами как ошибки); однако при этом имеется риск ошибочного признания ошибки.

- Из интуитивных соображений, в условиях таких эвристик сравнительной оценки, как многокритериальное сравнение, требование транзитивности агрегированных предпочтений представляется психологически некорректным, если оценивать его по достоверности результатов. Однако само понятие психологической корректности как теоретическое понятие требует развития и уточнений, особенно, — в контексте примеров, связанных с обнаружением «ошибок» и ошибок.

ЗАКЛЮЧЕНИЕ. ОБЩИЕ ВЫВОДЫ И РЕЗУЛЬТАТЫ

- Концепция схем представления знаний, ориентированных на корректность, и примененный подход к анализу рисков при разрешении конфликта между общей схемой и конкретным профессиональным знанием, позволяют установить, что проявлениями риска могут быть ошибочно признанные ошибки и замена относительно достоверного знания на менее достоверный результат его подгонки под схему.

В свете этого можно считать, что недостаточная психологическая корректность определенных методов и технологий, поддерживающих деятельность по принятию решений, может выражаться не только в виде типичных ошибок, но и в маскировании недостоверности.

- Представленные качественные модели поведения и анализ рисков, проведенный на их основе, позволяют предположить следующее. Явление психологической некорректности при получении формализованных знаний от человека-носителя конкретных практических знаний (ЛПР, эксперта, аналитика), впервые обозначенное и исследованное О. И. Ларичевым и его коллегами, имеет более глубокие и многообразные корни, чем они, по-видимому, предполагали, учитывая лишь «ограничения человеческой системы переработки информации».

Представляется целесообразным проводить исследования этого явления и его роли в создании технологий компьютерной поддержки интеллектуальной деятельности человека с привлечением психологических знаний и психологов.



- Схема анализа факторов, влияющих на достоверность решения о признании экспертом-аналитиком своей ошибки, примененная в данной работе для обобщенного анализа, позволяет проводить анализ рисков в конкретных технологиях и даже отдельных ситуациях вплоть до учета особенностей людей, влияющих на принятие решения. Для такого анализа целесообразно применять рефлексивный подход, когда участники взаимодействия, носители знаний о проблемной области и конкретных решаемых проблемах и носители технологии взаимно оценивают друг друга.
- Ввиду практической значимости проблемы транзитивности предпочтений для решения многих практических задач управления представляется необходимым провести комплексный анализ проблемы с учетом всех сложившихся на сегодня теоретических точек зрения и известных фактов, включая и факт уверенности многих специалистов в универсальности принципа транзитивности, вопреки имеющимся знаниям.

ЛИТЕРАТУРА

1. *Абрамова Н. А., Коверига С. В.* О рисках, связанных с ошибками экспертов и аналитиков // Тр. 4-й Междунар. конф. «Когнитивный анализ и управление развитием ситуаций». — М., 2004. — Т. 2. — С. 12—23.
2. *Абрамова Н. А.* О поиске подходов к созданию психологически корректных интеллектуальных технологий // Тр. 3-ей Междунар. конф. «Когнитивный анализ и управление развитием ситуаций» CASC'2003. — М., 2003. — Т. 2. — С. 153—158.
3. *Сидельников Ю. В.* Теория и организация экспертного прогнозирования. — М.: ИМЭМО АН СССР, 1990.
4. *Литвак Б. Г.* Экспертные оценки и принятие решений. — М.: «Патент», 1996.
5. *Асанов А. А., Ларичев О. И.* Влияние надежности человеческой информации на результаты применения методов принятия решений // Автоматика и телемеханика. — 1999. — № 5. — С. 20—31.
6. *Абрамова Н. А.* О некоторых мифах в оценке качества программного обеспечения // Надежность. — 2004. — № 1. — С. 38—63.
7. *Чалдини Р.* Психология влияния. — СПб.: ПИТЕР, 2001. — 270 с.
8. *Поиск подходов к решению проблем / И. В. Прангишвили, Н. А. Абрамова, В. Ф. Спиридонов и др.* — М.: СИНТЕГ, 1999. — 192 с.
9. *Абрамова Н. А., Сперанская И. В.* Развиваемая семантическая модель как механизм формализации знаний // Тез. докл. Всес. конф. по искусственному интеллекту. — Переславль-Залесский, 1988. — Т. 1. — С. 44—48.
10. *Ионин Л. Г.* Понимание и экспертиза // Вопросы философии. — 1991. — № 10.
11. *Шапиро С. И.* Мышление человека и переработка информации ЭВМ. — М.: Сов. радио, 1980.
12. *Майерс Г.* Надежность программного обеспечения. — М.: Мир, 1980.
13. *Kahneman D., Tversky A.* Judgment under uncertainty: heuristics and biases. — Cambridge: Cambridge University Press, 1983.
14. *Поддьяков А. Н.* Непереходность (нетранзитивность) отношений превосходства и принятия решений // Психология. Журнал высшей школы экономики. — 2006. — Т. 3. — № 3. — С. 88—111.
15. *Новиков Д. А.* Современные проблемы теории управления организационными системами // В кн.: Человеческий фактор в управлении / Под ред. Н. А. Абрамовой, К. С. Гинсберга, Д. А. Новикова. — М.: КомКнига, 2006. — С. 313—344.

☎ (495) 334-92-09;

e-mail: abramova@ipu.ru



Новая книга

Новиков Д. А. Теория управления организационными системами. — М.: МПСИ, 2005. — 584 с.

Книга посвящена описанию основ математической теории управления организационными системами. Ее цель — показать возможность и целесообразность применения математических моделей для повышения эффективности функционирования организаций (предприятий, учреждений, фирм и т. д.).

Описаны более сорока типовых механизмов — процедур принятия управленческих решений (реализующих функции планирования, организации, стимулирования и контроля): управления составом и структурой организационных систем, институционального, мотивационного и информационного управления. Их совокупность может рассматриваться как «конструктор», элементы которого позволяют создавать эффективную систему управления организацией.

Адресована студентам вузов, аспирантам и специалистам (теоретикам и практикам) в области управления организационными системами.

Более подробную информацию можно найти на сайте www.mtas.ru.

ВНУТРЕННЯЯ МОДЕЛЬ МАТЕМАТИЧЕСКОЙ ПРАКТИКИ ДЛЯ СИСТЕМ АВТОМАТИЗИРОВАННОГО КОНСТРУИРОВАНИЯ ДОКАЗАТЕЛЬСТВ ТЕОРЕМ.

Ч. 3. Модель доказательства¹

Т. Л. Гаврилова, А. С. Клецев

Институт автоматики и процессов управления ДВО РАН, г. Владивосток

Статья завершает цикл из трех работ, посвященных модели математической практики для систем автоматизированного доказательства теорем. Сформулированы правила рассуждения, используемые при конструировании доказательств. Определена модель полного доказательства. Приведены примеры формулировок математических утверждений на языке ММД (модели математического диалекта) и модель правильного полного доказательства одного из них.

ВВЕДЕНИЕ

В доказательствах, которые строят математики, центральную роль играют утверждения, имеющие форму импликации, — пропозициональные тавтологии, лежащие в основе логических рассуждений, математические утверждения (аксиомы и теоремы), лежащие в основе математических рассуждений, и метаматематические утверждения, позволяющие использовать в рассуждениях свойства логических и нелогических кванторов, а также логические и нелогические принципы. Правило вывода, связанное с использованием утверждений, имеющих форму импликации, известно как *Modus ponens*. Это правило вывода позволяет сводить доказательство утверждения, совпадающего с заключением импликации, к доказательству утверждений, совпадающих с ее посылками (декомпозиция), либо выводиться утверждение, совпадающее с заключением импликации, из утверждений, совпадающих с ее посылками (вывод). Совпадение обеспечивается процедурой унификации (конкретизации). Сама же импликация доказывается на основе теоремы дедукции. Кроме того, в процессе доказательства выдвигаются и используются различные

предположения. В моделях математической логики, в том числе используемых в системах интерактивного доказательства теорем, эти свойства математических доказательств представлены в крайне усеченных вариантах, что существенно затрудняет управление построением доказательств. Настоящая статья завершает цикл работ, начатый в работах [1, 2].

1. ПРАВИЛА РАССУЖДЕНИЯ

Два правила рассуждения (декомпозиция и вывод) определяют все возможные шаги, из которых могут конструироваться полные доказательства. Эти правила рассуждения формулируются на математическом диалекте. Каждое правило рассуждения позволяет получить непустую совокупность предположений.

Будем говорить, что *математическое утверждение справедливо*, если оно входит в математические знания или является конкретизацией метаматематического утверждения, входящего в математические знания. Будем говорить, что *предложение справедливо*, если оно либо получено в результате доказательства без выдвижения предположений, либо является конкретизацией пропозициональной тавтологии, либо является конкретизацией справедливого математического утверждения, сделанной без предположений (в последнем случае предвзвешенно должны быть доказаны все утверждения о допустимости синтаксической подстановки, связанные с этой конкретизацией).

Будем говорить, что *предложение справедливо при совокупности предположений*, если оно либо получено в ре-

¹ Работа выполнена при финансовом содействии программы № 16 Президиума РАН, проект «Теоретические основы интеллектуальных систем, основанных на онтологиях, для интеллектуальной поддержки научных исследований» и программы № 16 ОЭММПУ РАН проект «Синтез интеллектуальных систем управления базами знаний и базами данных».



зультате доказательства с использованием этой совокупности предположений, либо является конкретизацией справедливого математического утверждения, с которой связана эта совокупность предположений.

Внутренняя модель математической практики содержит два правила рассуждения.

Декомпозиция. Если справедливо (при совокупности предположений) предложение $f_1 \& \dots \& f_k \Rightarrow f$ и требуется доказать предложение f , то для этого достаточно доказать (при тех же предположениях) предложения $f_1, \dots, f_k$.

Вывод. Если справедливо (при совокупности предположений) предложение $f_1 \& \dots \& f_k \Rightarrow f$ и справедливы (при некоторых совокупностях предположений) предложения $f_1, \dots, f_k$, то справедливо (при объединении всех этих предположений) предложение f .

2. МОДЕЛЬ ПОЛНОГО ДОКАЗАТЕЛЬСТВА

2.1. Вспомогательные понятия

С каждым предложением f , входящим в полное доказательство, свяжем множество (возможно, пустое) $t^+(f) = t_1^+(f) \cup t_2^+(f)$ предположений, при которых предложение f истинно, причем $t_1^+(f)$ есть множество предположений о свойствах произвольных объектов, а $t_2^+(f)$ — множество предположений о справедливости предложений, при которых истинно f . Если « $a \in t$ » — предположение о свойствах произвольного объекта, то « $a \in t$ » $\in t_1^+(\text{«}a \in t\text{»})$; если же f — предположение о справедливости предложения, то $t^+(f) = \{f\}$. Кроме того, с каждым предложением f , требующим доказательства, свяжем множество (возможно, пустое) $t^-(f) = t_1^-(f) \cup t_2^-(f)$ предположений, при которых предложение f должно быть доказано, причем $t_1^-(f)$ есть множество предположений о свойствах произвольных объектов, а $t_2^-(f)$ — множество предположений об истинности предложений, при которых f должно быть доказано. Если $t^+(f) = \emptyset$, то f справедливо безусловно; если $t^-(f) = \emptyset$, то f должно быть доказано без всяких предположений.

2.2. Определение модели полного доказательства

Модель полного доказательства математического утверждения $(v_1: t_1) \dots (v_m: t_m)f$ есть доказательство предложений $t'_1 \neq \emptyset, \dots, t'_m \neq \emptyset$ и предложения f' , являющегося конкретизацией формулы f , сделанной при полном наборе предположений P , связанных с этой конкретизацией; здесь $t'_1, \dots, t'_m$ — результаты применения подстановки, при которой сделана эта конкретизация, к термам $t_1, \dots, t_m$. Для $i = 1, \dots, m$ имеет место $t^-(t'_i \neq \emptyset) = P$; кроме того, $t^-(f') = P$.

Модель полного доказательства предложения $f_1 \& \dots \& f_k \Rightarrow f$ (имеющего форму импликации) есть доказательство предложения f в предположении, что справед-

ливо предложения $f_1, \dots, f_k$. При этом $t^-(f) = t^-(f_1 \& \dots \& f_k \Rightarrow f) \cup \{f_1, \dots, f_k\}$.

Модель полного доказательства предложения f , не имеющего форму импликации, может быть либо его декомпозицией, либо его выводом, либо пустым доказательством.

Декомпозиция предложения f есть результат применения правила декомпозиции к предложению f . Она характеризуется множеством компонент декомпозиции, каждая из которых состоит из предложения и его доказательства. Смысл декомпозиции состоит в том, что для доказательства предложения f достаточно доказать предложения всех компонент декомпозиции (задача поиска доказательства предложения сводится к множеству подзадач поиска доказательств предложений для всех компонент декомпозиции). Пусть в результате применения правила декомпозиции выдвигается множество предположений P (если предположения не выдвигаются, $P = \emptyset$). Если доказывается предложение f , и $f_1, \dots, f_m$ — предложения компонент декомпозиции, то для $i = 1, \dots, m$ имеет место $t^-(f_i) = P \cup t^-(f)$.

Вывод предложения f характеризуется процессом вывода, на каждом шаге которого применяется правило вывода. Начальным состоянием процесса вывода является множество предложений $t^-(f)$. Если на текущем шаге в состоянии процесса вывода содержится предложение f , то процесс вывода завершается (если предложение f содержится в начальном состоянии, то процесс вывода является пустым). Если при этом $t^+(f) \subseteq t^-(f)$, где $t^+(f)$ получено в результате процесса вывода, то процесс вывода завершается нормально, и предложение f считается доказанным. Заметим, что если в том же самом или в другом доказательстве было получено такое $t^+(f)$, что выполнено $t^+(f) \subseteq t^-(f)$, то доказательством предложения f может служить ссылка на процесс вывода в том же самом или в другом доказательстве, где было получено это $t^+(f)$. Применение правила вывода на очередном шаге процесса вывода состоит в следующем. Если значения посылок правила вывода — предложения $f_1, \dots, f_m$ — справедливы (входят в текущее состояние процесса вывода или получаются в результате конкретизации), то состояние процесса вывода на следующем шаге получается из состояния на текущем шаге добавлением предложения f — результата применения правила вывода. Пусть в результате применения правила вывода выдвигается множество предположений P (если предположения не выдвигаются, $P = \emptyset$). Тогда $t^+(f) = P \cup t^+(f_1) \cup \dots \cup t^+(f_m)$.

Модель доказательства справедливого предложения — пустое доказательство.

2.3. Примеры

Примеры определений терминов. Далее приводится запись на языке ММД нескольких определений из теории пределов последовательностей.

1. **Последовательности** $\equiv I[1, \infty) \rightarrow R$.

2. *Предел* $\equiv (\lambda(x: \text{последовательности})(a: R)(\forall(\varepsilon: R(0, \infty))(\exists(N: I[1, \infty))(\forall(n: I[N, \infty))(|x(n) - a| < \varepsilon)))$.

3. *Бесконечно малые величины* $\equiv \{(x: \text{последовательности}) \text{ предел}(x, 0)\}$.

Примеры определений новых способов записи. Далее приводится запись на языке ММД нескольких определений новых способов записи.

4. $(v: R)R(v, \infty) \equiv \{(w: R)w > v\}$.

5. $(v: R)|v| \equiv / (v \geq 0 \Rightarrow v)(v < 0 \Rightarrow -v)/$.

6. $(v: R) - v \equiv 0 - v$.

7. $(v_1: I)I[v_1, \infty) \equiv \{(v_2: I)v_2 \geq v_1\}$.

8. $(v_1: I)(v_2: I[v_1, \infty))I[v_1, v_2] \equiv \{(v_3: I)v_3 \geq v_1 \ \& \ v_3 \leq v_2\}$.

Примеры пропозициональных тавтологий. Доказательство равносильного утверждения:

9. $(v_1: L)(v_2: L)(v_1 \Rightarrow v_2) \ \& \ (v_2 \Rightarrow v_1) \Rightarrow (v_1 \Leftrightarrow v_2)$.

Доказательство от противного:

10. $(v: L)(\neg v \Rightarrow \text{ложь}) \Rightarrow v$.

Доказательство противоречия:

11. $(v: L)v \ \& \ \neg v \Rightarrow \text{ложь}$.

Примеры математических утверждений. Далее приводятся примеры записи на языке ММД математических утверждений.

Теорема о единственности предела числовой последовательности:

12. $(x: \text{последовательности})(a_1: R)(a_2: R) \text{ предел}(x, a_1) \ \& \ \text{предел}(x, a_2) \Rightarrow a_1 = a_2$.

Теорема о сумме членов конечного отрезка натурального ряда:

13. $(n: I[0, \infty))(\Sigma(i: I[0, n])i) = n * (n + 1)/2$.

Правило приведения к общему знаменателю дроби и числа:

14. $(x: R)(y: R \setminus \{0\})(z: R)x/y + z = (x + y * z)/y$.

Правило выноса общего множителя за скобки:

15. $(x: R)(y: R)(z: R)x * y + z * y = (x + z) * y$.

Свойство коммутативности умножения:

16. $(x: R)(y: R)x * y = y * x$.

Примеры метаматематических утверждений. Далее приводятся примеры записи на метаязыке некоторых метаматематических утверждений.

Одна из форм принципа полной математической индукции:

17. $(v_1: I[0, \infty))(v_2: I[0, v_1])f \vdash 0 \ \& \ (f \vdash v_2 \Rightarrow f \vdash v_2 + 1) \Rightarrow f \vdash v_1$.

Свойство квантора Σ — сумма с одним слагаемым:

18. $(v: I)(\Sigma(v: I[v, v])t \vdash v) = t \vdash v$.

Свойство квантора Σ — добавление слагаемого в сумму:

19. $(v_1: I)(v_2: I[v_1, \infty))(\Sigma(v: I[v_1, v_2 + 1])t \vdash v) = (\Sigma(v: I[v_1, v_2])t \vdash v) + t \vdash v_2 + 1$.

Принцип замены равных термов (замена в формуле вхождения терма равным ему):

20. $f \vdash t_1 \ \& \ t_1 = t_2 \Rightarrow f \vdash t_2$.

Пример формального доказательства. Доказывается теорема о сумме членов начального отрезка натурального ряда (см. пример 13).

Модель доказательства теоремы представлена в виде последовательности шагов, каждый из которых характеризуется правилом рассуждения, примененным на этом шаге, значениями посылок, возможно, способом кон-

кретизации при их получении, значением заключения правила, а также, возможно, выдвинутыми предположениями.

Формулировка теоремы на языке ММД:

$$(n: I[0, \infty))(\Sigma(i: I[0, n])i) = n * (n + 1)/2.$$

Модель доказательства теоремы, по определению, есть доказательство предложений

$$I[0, \infty) \neq \emptyset \quad (1)$$

и

$$(\Sigma(i: I[0, a])i) = a * (a + 1)/2 \quad (2)$$

в предположении, что

$$a \in I[0, \infty).$$

Шаги доказательства:

1) *модель доказательства предложения* (1) здесь не приводится;

2) *модель доказательства предложения* (2):

из принципа полной математической индукции в примере 17 посредством конкретизации получается аксиома:

$$\begin{aligned} (v_1: I[0, \infty))(v_2: I[0, 1])(\Sigma(i: I[0, 0])i) &= \\ = 0 \ \& \ ((\Sigma(i: I[0, v_2])i) = v_2 * (v_2 + 1)/2 \Rightarrow \\ \Rightarrow (\Sigma(i: I[0, v_2 + 1])i) = (v_2 + 1) * (v_2 + 2)/2) \Rightarrow \\ \Rightarrow (\Sigma(i: I[0, v])i) = v_1 * (v_1 + 1)/2; \end{aligned} \quad (3)$$

из нее в предположении, что $b \in I[0, \infty]$, посредством конкретизации получается предложение

$$\begin{aligned} (\Sigma(i: I[0, 0])i) = 0 \ \& \ ((\Sigma(i: I[0, b])i) = b * (b + 1)/2 \Rightarrow \\ \Rightarrow (\Sigma(i: I[0, b + 1])i) = (b + 1) * (b + 2)/2) \Rightarrow \\ \Rightarrow (\Sigma(i: I[0, a])i) = a * (a + 1)/2; \end{aligned} \quad (4)$$

применяя предложение (4) для декомпозиции (2), получаем компоненты декомпозиции — предложения:

$$(\Sigma(i: I[0, 0])i) = 0 \quad (5)$$

и

$$\begin{aligned} (\Sigma(i: I[0, b])i) = b * (b + 1)/2 \Rightarrow \\ \Rightarrow (\Sigma(i: I[0, b + 1])i) = (b + 1) * (b + 2)/2; \end{aligned} \quad (6)$$

3) *модель доказательства предложения* (5):

из свойства квантора Σ (сумма с одним слагаемым) в примере 18 посредством конкретизации получается аксиома:

$$(v: I)(\Sigma(i: I[v, v])i) = v;$$

из нее посредством конкретизации получается предложение (5) (пустое доказательство);

4) *модель доказательства предложения* (6), имеющего форму импликации, по определению, есть доказательство предложения

$$(\Sigma(i: I[0, b + 1])i) = (b + 1) * (b + 2)/2 \quad (7)$$

в предположении, что

$$(\Sigma(i: I[0, b])i) = b * (b + 1)/2; \quad (8)$$

5) *модель доказательства предложения* (7) есть вывод этого предложения.



Шаг 1 вывода предположения (7): из принципа замены равных термов в примере 20 посредством конкретизации получается предложение (аксиома):

$$\begin{aligned} & (\Sigma(i: I[0, b+1])i) = \\ & = (\Sigma(i: I[0, b])i) + b + 1 \ \& \ (\Sigma(i: I[0, b])i) = \\ & = b * (b+1)/2 \Rightarrow (\Sigma(i: I[0, b+1])i) = \\ & = b * (b+1)/2 + b + 1; \end{aligned} \quad (9)$$

из свойства квантора Σ (добавление слагаемого в сумму) в примере 19 посредством конкретизации получается аксиома:

$$\begin{aligned} & (v_1: I)(v_2: I[v_1, \infty))(\Sigma(i: I[v_1, v_2+1])i) = \\ & = (\Sigma(i: I[v_1, v_2])i) + v_2 + 1; \end{aligned}$$

из нее посредством конкретизации получается предложение

$$(\Sigma(i: I[0, b+1])i) = (\Sigma(i: I[0, b])i) + b + 1;$$

из него и предположения (8) с учетом аксиомы (9) выводится предложение

$$(\Sigma(i: I[0, b+1])i) = b * (b+1)/2 + b + 1. \quad (10)$$

Шаг 2 вывода предположения (7): из принципа замены равных термов в примере 20 посредством конкретизации получается предложение (аксиома)

$$\begin{aligned} & (\Sigma(i: I[0, b+1])i) = \\ & = b * (b+1)/2 + b + 1 \ \& \ b * (b+1)/2 + b + 1 = \\ & = (b * (b+1) + 2 * (b+1))/2 \Rightarrow (\Sigma(i: I[0, b+1])i) = \\ & = (b * (b+1) + 2 * (b+1))/2; \end{aligned} \quad (11)$$

из математического утверждения в примере 14 посредством конкретизации получается предложение

$$b * (b+1)/2 + b + 1 = (b * (b+1) + 2 * (b+1))/2; \quad (12)$$

предварительно должно быть доказано, что $b*(b+1) \in R$, $2 \in R \setminus \{0\}$, $b+1 \in R$ (эти доказательства здесь не приводятся); из предложений (10) и (12) с учетом аксиомы (11) выводится предложение

$$(\Sigma(i: I[0, b+1])i) = (b * (b+1) + 2 * (b+1))/2. \quad (13)$$

Шаг 3 вывода предположения (7): из принципа замены равных термов в примере 20 посредством конкретизации получается предложение (аксиома)

$$\begin{aligned} & (\Sigma(i: I[0, b+1])i) = \\ & = (b * (b+1) + 2 * (b+1))/2 \ \& \ b * (b+1) + 2 * (b+1) = \\ & = (b+2) * (b+1) \Rightarrow (\Sigma(i: I[0, b+1])i) = \\ & = (b+2) * (b+1)/2; \end{aligned} \quad (14)$$

из математического утверждения в примере 15 посредством конкретизации получается предложение

$$b * (b+1) + 2 * (b+1) = (b+2) * (b+1); \quad (15)$$

из предложений (13) и (15) с учетом аксиомы (14) выводится предложение

$$(\Sigma(i: I[0, b+1])i) = (b+2) * (b+1)/2. \quad (16)$$

Шаг 4 вывода предположения (7): из принципа замены равных термов в примере 20 посредством конкретизации получается предложение (аксиома)

$$\begin{aligned} & (\Sigma(i: I[0, b+1])i) = \\ & = (b+2) * (b+1)/2 \ \& \ (b+2) * (b+1) = \\ & = (b+1) * (b+2) \Rightarrow (\Sigma(i: I[0, b+1])i) = \\ & = (b+1) * (b+2)/2; \end{aligned} \quad (17)$$

из математического утверждения в примере 16 посредством конкретизации получается предложение

$$(b+2) * (b+1) = (b+1) * (b+2); \quad (18)$$

из предложений (16) и (18) с учетом аксиомы (17) выводится предложение (7).

ЗАКЛЮЧЕНИЕ

В цикле из трех статей [1, 2 и настоящая статья] предложена формальная система с расширяемым логико-математическим языком высокого порядка. Наряду с этим языком в формальной системе присутствуют язык пропозициональных утверждений для представления правил логических рассуждений и метаязык для представления свойств логических и нелогических кванторов, а также логических и нелогических принципов. Основным правилом вывода в формальной системе является *Modus ponens*. Так как набор метаматематических аксиом не фиксирован и может расширяться, традиционное понятие полноты исчисления оказывается неприменимым к этой формальной системе (что имеет место и для математической практики). Понятно, что отсутствие большинства ограничений, характерных для моделей математической логики, используемых в компьютерных системах доказательства теорем, еще не является гарантией того, что доказательства, представленные в предложенной формальной системе будут сопоставимы по сложности с математическими доказательствами. Поэтому одно из направлений дальнейших исследований заключается в экспериментальном изучении математических доказательств средствами этой формальной системы с целью выявления в них конструкций еще более высокого уровня.

ЛИТЕРАТУРА

1. Гаврилова Т. Л., Клещев А. С. Внутренняя модель математической практики для систем автоматизированного конструирования доказательств теорем. Ч. 1. Общее описание модели // Проблемы управления. — 2006. № 4. — С. 32—35.
2. Гаврилова Т. Л., Клещев А. С. Внутренняя модель математической практики для систем автоматизированного конструирования доказательств теорем. Ч. 2. Модель математического диалекта // Там же. — № 5. — С. 68—73.

☎ (4232) 31-40-01, 31-04-24

E-mail: gavrillov@iacp.dvo.ru ; kleshev@iacp.dvo.ru



АНАЛИТИЧЕСКИЙ СИНТЕЗ ПРОГРАММНОГО ГИРОСИЛОВОГО УПРАВЛЕНИЯ СВОБОДНОЛЕТАЮЩИМ КОСМИЧЕСКИМ РОБОТОМ В РЕЖИМЕ ТРАНСПОРТИРОВКИ ПОЛЕЗНОГО ГРУЗА¹

Е. И. Сомов

Самарский научный центр РАН, г. Самара

Представлен метод аналитического синтеза пространственного программного углового движения свободнолетающего космического роботизированного модуля в режиме транспортировки полезного груза. Показано, что при выполнении поворотного маневра модуля с нежестким грузом при краевых условиях общего вида обеспечивается достижение слабого возбуждения упругих колебаний переносимой конструкции благодаря аналитически рассчитанному гиросиловому управлению.

ВВЕДЕНИЕ

Среди общих проблем сборки больших космических конструкций [1] непосредственно на орбите с помощью свободнолетающих космических роботизированных модулей (КРМ) [2, 3] выделяется актуальная задача управления пространственным угловым движением КРМ в режиме транспортировки полезного груза при малых массовых и энергетических затратах. Как известно [4], при необходимости многочисленных, разнообразных и быстрых пространственных поворотных маневров любого космического аппарата, в частности КРМ, важными преимуществами по массовым и энергетическим затратам обладают системы управления движением, в состав которых входят двухстепенные силовые гироскопы — гиродины (ГД). Задача пространственного углового маневра КРМ существенно отличается от традиционной задачи управления ориентацией космических аппаратов в режимах, осуществляющих поиск внешних физических ориентиров. В настоящей статье представлен метод аналитического синтеза пространственного программного углового движения КРМ, совершающего свобод-

ный полет по баллистической траектории [5] в промежутке между участками, на которых применяется управление движением его центра масс. Модуль с транспортируемым полезным грузом, зачастую нежестким, должен выполнять поворотные маневры с краевыми условиями общего вида [6], причем требуется ограничение возбуждения упругих колебаний конструкции такой связки твердых и упругих тел в процессе маневра путем обеспечения «гладкости» воздействий со стороны аналитически рассчитанного гиросилового управления.

1. ЗАДАЧА АНАЛИТИЧЕСКОГО СИНТЕЗА ПРОГРАММЫ ПОВОРОТНОГО МАНЕВРА КРМ

Данная задача на заданном интервале времени $t \in T_p \equiv [t_0^p, t_f^p]$, $t_f^p \equiv t_0^p + T_p$ состоит в определении явных функций времени [6]: кватерниона ориентации $\Lambda(t)$ связанного с корпусом КРМ базиса $\mathbf{B}$ относительно известного инерциального базиса $\mathbf{I}$, векторов угловой скорости $\boldsymbol{\omega}(t)$, ускорения $\boldsymbol{\varepsilon}(t)$ и его производной $\dot{\boldsymbol{\varepsilon}}(t) = \boldsymbol{\varepsilon}^*(t) + \boldsymbol{\omega}(t) \times \boldsymbol{\varepsilon}(t)$. Кватернион $\Lambda = (\lambda_0, \boldsymbol{\lambda})$, $\boldsymbol{\lambda} = \{\lambda_i\}$, векторы скорости $\boldsymbol{\omega} = \{\omega_i\}$ и ускорения $\boldsymbol{\varepsilon} = \{\varepsilon_i\} = \dot{\boldsymbol{\omega}}$, а также вектор производной ускорения $\dot{\boldsymbol{\varepsilon}} = \{\dot{\varepsilon}_i\} = \boldsymbol{\varepsilon}^* + \boldsymbol{\omega} \times \boldsymbol{\varepsilon}$ должны удовлетворять следующим краевым условиям на левом

¹ Работа поддержана РФФИ (грант 05-08-18175), Отделением энергетики, механики, машиностроения и процессов управления РАН (программа 16).



($t = t_0^p$) и правом ($t = t_f^p$) концах траектории поворотного маневра:

$$\Lambda(t_0^p) = \Lambda_0; \omega(t_0^p) = \omega_0 \equiv \mathbf{e}_0^0 \omega_0; \varepsilon(t_0^p) = \varepsilon_0 \equiv \mathbf{e}_0^0 \varepsilon_0; \quad (1)$$

$$\Lambda(t_f^p) = \Lambda_f; \omega(t_f^p) = \omega_f \equiv \mathbf{e}_f^0 \omega_f; \varepsilon(t_f^p) = \varepsilon_f \equiv \mathbf{e}_f^0 \varepsilon_f; \\ \dot{\varepsilon}(t_f^p) = \dot{\varepsilon}_f \equiv \mathbf{e}_f^* \varepsilon_f^* + \omega_f \times \varepsilon_f. \quad (2)$$

Последнее из этих условий представляет собой требования к *гладкости сопряжения* правого конца траектории с последующим участком причаливания и механической стыковки КРМ с собираемой с его помощью космической конструкцией.

Подход к решению этой задачи основывается на *необходимом* и *достаточном* условии разрешимости классической задачи Дарбу — определения кватерниона $\Lambda(t)$ в аналитическом виде (в том числе в квадратурах) из уравнения $\dot{\Lambda}(t) = \frac{1}{2} \Lambda(t) \odot \omega(t)$ при известных Λ_0 и $\omega(t)$, где $\odot$ — знак умножения кватернионов.

Введем базис $\mathbf{E}_0$, фиксированный в инерциальном базисе $\mathbf{I}$ кватернионом Λ_0 (1), и подвижные базисы $\mathbf{E}_k$, $k = 1, \dots, n$, где базис $\mathbf{E}_n$ совпадает со связанным базисом $\mathbf{B}$. Необходимое и достаточное условия разрешимости задачи Дарбу состоят в возможности представления вектора угловой скорости $\omega(t)$ в виде $\omega(t) = \omega_n(t) + \omega_{n-1}(t) + \dots + \omega_1(t)$, где вектор $\omega_k(t)$ имеет неизменное направление в базисе $\mathbf{E}_{k-1}$ и является вектором угловой скорости базиса $\mathbf{E}_k$ относительно базиса $\mathbf{E}_{k-1}$, т. е. в виде

$$\omega = \omega_n^{n-1} + \tilde{\Lambda}_n \odot (\omega_{n-1}^{n-2} + \tilde{\Lambda}_{n-1} \odot (\omega_{n-2}^{n-3} + \tilde{\Lambda}_{n-2} \odot \dots + \omega_1^0 \dots \odot \Lambda_{n-2}) \odot \Lambda_{n-1}) \odot \Lambda_n,$$

где вектор-столбец $\omega_k^{k-1} \equiv \omega_k^{k-1}(t)$, $k = 1, \dots, n$, составлен из проекций вектора $\omega_k(t)$ фиксированно-

Типы элементарных движений и их краевые условия

k	Тип	Краевые условия на левом конце траектории ($t = t_0^p$)				Краевые условия на правом конце траектории ($t = t_f^p$)			
		φ	ω	ε	$\dot{\varepsilon}$	φ	ω	ε	$\dot{\varepsilon}$
1	Гашение ускорения ε_0		0	ε_0	$\dot{\varepsilon}_1^0$	φ_1^f			
2	Гашение скорости ω_0		ω_0		$\dot{\varepsilon}_2^0$	φ_2^f	0		
3	Позиционный переход				$\dot{\varepsilon}_3^0$	φ_3^f		0	
4	Разгон до скорости ω_f			0	$\dot{\varepsilon}_4^0$	φ_4^f	ω_f		
5	Разгон до ускорения ε_f				$\dot{\varepsilon}_5^0$	φ_5^f		ε_f	
6	Разгон до производной ускорения $\dot{\varepsilon}_f$				$\dot{\varepsilon}_6^0$	φ_6^f	0	0	ε_f^*

го направления в базисе $\mathbf{E}_{k-1}$, а $\Lambda_k(t)$ является кватернионом ориентации базиса $\mathbf{E}_k$ относительно базиса $\mathbf{E}_{k-1}$.

Решение поставленной задачи представляется как результат сложения в общем случае шести *одновременно* происходящих элементарных поворотов «вложенных» базисов $\mathbf{E}_k$ вокруг ортов $\mathbf{e}_k$ осей Эйлера, положение которых определяется из краевых условий (1) и (2) исходной пространственной задачи. Краевые условия всех 6-ти элементарных поворотов относительно ортов $\mathbf{e}_k$ приведены в таблице, где для первых *пяти* элементарных движений требуется обеспечить равенство нулю локальной (собственной) производной ускорения на правом конце траектории.

Кватернион $\Lambda(t)$ ориентации КРМ в базисе $\mathbf{I}$ определяется произведением

$$\Lambda(t) = \Lambda_0 \odot \Lambda_1(t) \odot \Lambda_2(t) \odot \Lambda_3(t) \odot \Lambda_4(t) \odot \Lambda_5(t) \odot \Lambda_6(t).$$

Здесь индексы 1—6 кватернионов $\Lambda_k(t)$ соответствуют их номерам в таблице, причем кватернион $\Lambda_k(t) = (\cos(\varphi_k(t)/2), \mathbf{e}_k \sin(\varphi_k(t)/2))$, где $\varphi_k(t)$ и $\mathbf{e}_k$ — текущий угол и орт оси Эйлера k -го поворота. В силу *неподвижности* орта $\mathbf{e}_k$ в базисе $\mathbf{E}_{k-1}$ имеем $\omega_k(t) = \dot{\varphi}_k(t) \mathbf{e}_k$; $\varepsilon_k(t) = \ddot{\varphi}_k(t) \mathbf{e}_k$ и $\dot{\varepsilon}_k(t) = \ddot{\varphi}_k(t) \mathbf{e}_k$. Вектор угловой скорости $\omega(t)$, векторы углового ускорения $\varepsilon(t)$ и его производной $\dot{\varepsilon}(t)$ при начальных обозначениях векторов $\omega^{(1)}(t) = \omega_1(t)$, $\varepsilon^{(1)}(t) = \varepsilon_1(t)$, $\dot{\varepsilon}^{(1)}(t) = \dot{\varepsilon}_1(t)$ определяются аналитически по рекуррентному алгоритму: для верхних индексов $k = \overline{2, 6}$ последовательно вычисляются

$$\omega_q^k(t) := \tilde{\Lambda}_k(t) \odot \omega^{(k-1)}(t) \odot \Lambda_k(t);$$

$$\omega^{(k)}(t) = \omega_k(t) + \omega_q^k(t);$$

$$\varepsilon_q^k(t) := \tilde{\Lambda}_k(t) \odot \varepsilon^{(k-1)}(t) \odot \Lambda_k(t);$$

$$\varepsilon^{(k)}(t) = \varepsilon_k(t) + \varepsilon_q^k(t) + \omega_q^k(t) \times \omega_k(t);$$

$$\dot{\varepsilon}_q^k(t) := \tilde{\Lambda}_k(t) \odot \dot{\varepsilon}^{(k-1)}(t) \odot \Lambda_k(t);$$

$$\dot{\varepsilon}^{(k)}(t) = \dot{\varepsilon}_k(t) + \dot{\varepsilon}_q^k(t) + (2\varepsilon_q^k(t) + \omega_q^k(t) \times \omega_k(t)) \times \omega_k(t) + \omega_q^k(t) \times \varepsilon_k(t), \quad (3)$$

искомые векторы получаются как $\omega(t) = \omega^{(6)}(t)$, $\varepsilon(t) = \varepsilon^{(6)}(t)$, $\dot{\varepsilon}(t) = \dot{\varepsilon}^{(6)}(t)$.

Функции $\varphi_k(t)$, представляющие в аналитическом виде углы элементарных поворотов, выбираются в классе полиномов (сплайнов) соответствующей степени.

Гашение начальных значений угловой скорости и углового ускорения. Для *первых* двух движений выберем функции $\varphi_k(t)$, $k = 1, 2$ с краевыми условиями

$$\varphi_k(t_0^p) = 0; \quad \dot{\varphi}_k(t_0^p) = \omega_0 \equiv |\omega_0|; \quad \ddot{\varphi}_k(t_0^p) = \varepsilon_0 \equiv |\varepsilon_0|;$$

$$\dot{\varphi}_k(t_f^p) = 0; \quad \ddot{\varphi}_k(t_f^p) = 0; \quad \ddot{\varphi}_k(t_f^p) = 0$$

в виде сплайнов $\varphi_k(\tau)$ 5-й степени нормированного времени $\tau = (t - t_0^p)/T_p \in [0, 1]$:

$$\ddot{\varphi}_k(t) = \dot{\varepsilon}_k(t) = -6(a_0 - a_1\tau + 2a_2\tau^2)/T_p;$$

$$\dot{\varphi}_k(t) = \varepsilon_k(t) = \varepsilon_0 - \tau(6a_0 - 3a_1\tau + 4a_2\tau^2);$$

$$\dot{\varphi}_k(t) = \omega_k(t) = \omega_0 + T_p\tau[\varepsilon_0 - \tau(3a_0 - a_1\tau + a_2\tau^2)];$$

$$\varphi_k(t) = T_p\tau\{\omega_0 + T_p\tau[\varepsilon_0 - \tau(20a_0 - 5a_1\tau + 4a_2\tau^2)/10]/2\}, \quad (4)$$

где коэффициенты a_i определяются так: $a_0 = 2\omega_0/T_p + \varepsilon_0$; $a_1 = 8\omega_0/T_p + 3\varepsilon_0$ и $a_2 = 3\omega_0/T_p + \varepsilon_0$. Здесь формально гашение начального углового ускорения получается по соотношениям (4) при $\omega_0 = 0$, а гашение начальной угловой скорости — при $\varepsilon_0 = 0$. На левом конце траектории производные ускорений $\dot{\varepsilon}_k(t_0^p) \equiv \dot{\varepsilon}_k^0 = \ddot{\varphi}_k(t_0^p) = -6a_0/T_p$, а углы φ_k на правом конце траектории принимают значения $\varphi_k(t_f^p) \equiv \varphi_k^0 = T_p^2 [8(\omega_0/T_p) + \varepsilon_0]/20$.

Разгон до заданных значений угловой скорости и ускорения. Для четвертого и пятого движений функции $\varphi_k(t)$, $k = 4, 5$ должны удовлетворять краевым условиям

$$\varphi_k(t_0^p) = 0; \quad \dot{\varphi}_k(t_0^p) = 0; \quad \ddot{\varphi}_k(t_0^p) = 0;$$

$$\dot{\varphi}_k(t_f^p) = \omega_f \equiv |\omega_f|; \quad \ddot{\varphi}_k(t_f^p) = \varepsilon_f \equiv |\varepsilon_f|; \quad \ddot{\varphi}_k(t_0^p) = 0,$$

поэтому $\varphi_k(t)$ также выбираются в виде сплайнов 5-й степени

$$\ddot{\varphi}_k(t) = \dot{\varepsilon}_k(t) = 6(a_0 + a_1\tau + 2a_2\tau^2)/T_p;$$

$$\dot{\varphi}_k(t) = \varepsilon_k(t) = \tau(6a_0 + 3a_1\tau + 4a_2\tau^2);$$

$$\dot{\varphi}_k(t) = \omega_k(t) = T_p\tau^2(3a_0 + a_1\tau + a_2\tau^2);$$

$$\varphi_k(t) = T_p^2\tau^3(20a_0 + 5a_1\tau + 4a_2\tau^2)/20,$$

где нормированное время $\tau = (t - t_0^p)/T_p \in [0, 1]$ и коэффициенты a_i вычисляются по соотношениям $a_0 = 2\omega_f/T_p - \varepsilon_f$; $a_1 = -8\omega_f/T_p + 5\varepsilon_f$; $a_2 = 3\omega_f/T_p - 2\varepsilon_f$. На левом конце траектории собственные производные ускорений $\dot{\varepsilon}_k(t_f^p) \equiv \dot{\varepsilon}_k^0 = \ddot{\varphi}_k(t_0^p) = 6a_0/T_p$, а углы φ_k принимают конечные значения $\varphi_k(t_f^p) \equiv \varphi_k^f = T_p^2 [12(\omega_f/T_p) - 3\varepsilon_f]/20$, $k = 4, 5$.

Разгон до заданного значения локальной производной углового ускорения. Для шестого элементарного движения функция $\varphi_6(t)$ должна удовлетворять краевым условиям

$$\varphi_6(t_0^p) = 0; \quad \dot{\varphi}_6(t_0^p) = 0; \quad \ddot{\varphi}_6(t_0^p) = 0; \quad \ddot{\varphi}_6(t_f^p) = 0;$$

$$\ddot{\varphi}_6(t_f^p) = 0; \quad \ddot{\varphi}_6(t_f^p) = \varepsilon_f^*,$$

она представляется сплайном 5-й степени нормированного времени $\tau \in [0, 1]$ в виде

$$\varepsilon_6^*(t) = b_6[1 - 6\tau + 6\tau^2]/T_p; \quad \varepsilon_6(t) = b_6\tau[1 - 3\tau + 2\tau^2];$$

$$\omega_6(t) = b_6T_p\tau^2[1 - 2\tau + \tau^2]/2;$$

$$\omega_6(t) = b_6T_p^2\tau^3[10 - 15\tau + 6\tau^2]/60,$$

где $b_6 = \varepsilon_f^* T_p$. На левом конце траектории $\varepsilon_6^*(t_0^p) \equiv \varepsilon_6^0 = \ddot{\varphi}_6(t_0^p) = \varepsilon_f^*$, угол φ_6 поворота относительно орта $\mathbf{e}_6$ принимает конечное значение $\varphi_6(t_f^p) \equiv \varphi_6^f = T_p^2 b_6/60 = T_p^3 \varepsilon_f^*/60$.

Позиционный переход. Функция позиционного перехода $\varphi_3(t)$ по углу поворота в третьем замыкающем элементарном движении должна удовлетворять краевым условиям

$$\varphi_3(t_0^p) = 0; \quad \dot{\varphi}_3(t_0^p) = 0; \quad \varphi_3(t_f^p) = \varphi_3^f; \quad \dot{\varphi}_3(t_f^p) = 0;$$

$$\ddot{\varphi}_3(t_f^p) = 0; \quad \ddot{\varphi}_3(t_f^p) = 0; \quad \ddot{\varphi}_3(t_0^p) = 0. \quad (5)$$

Здесь угол $\varphi_3^f = \varphi^* = 2\arccos(\lambda_0^*)$ определяется краевыми условиями (1) и (2), где λ_0^* — скалярная часть кватерниона $\mathbf{\Lambda}^* \equiv (\lambda_0^*, \boldsymbol{\lambda}^*) = \tilde{\mathbf{\Lambda}}_2(t_f^p) \odot \tilde{\mathbf{\Lambda}}_1(t_f^p) \odot \tilde{\mathbf{\Lambda}}_0 \odot \mathbf{\Lambda}_f \odot \tilde{\mathbf{\Lambda}}_6(t_f^p) \odot \tilde{\mathbf{\Lambda}}_5(t_f^p) \odot \tilde{\mathbf{\Lambda}}_4(t_f^p)$ с ортом оси Эйлера $\mathbf{e}_3 = \boldsymbol{\lambda}^*/\sin(\varphi^*/2)$, а кватернионы $\mathbf{\Lambda}_k(t_f^p)$, $k = 1, 2, 4, 5, 6$ однозначно определяются углами φ_k^f , представленными выше в явном виде, (см. таблицу) и ортами

$$\mathbf{e}_1 = \mathbf{e}_0^{\varepsilon}; \quad \mathbf{e}_2 = \mathbf{e}_0^{\omega};$$

$$\mathbf{e}_4 = \mathbf{\Lambda}_6(t_f^p) \odot \mathbf{\Lambda}_5(t_f^p) \odot \mathbf{e}_f^{\omega} \odot \tilde{\mathbf{\Lambda}}_5(t_f^p) \odot \tilde{\mathbf{\Lambda}}_6(t_f^p);$$

$$\mathbf{e}_5 = \mathbf{\Lambda}_6(t_f^p) \odot \mathbf{e}_f^{\varepsilon} \odot \tilde{\mathbf{\Lambda}}_6(t_f^p); \quad \mathbf{e}_6 = \mathbf{e}_f^{\varepsilon^*} \equiv \frac{\mathbf{e}_f^*}{|\mathbf{e}_f^*|}.$$

На модуль скорости движения в позиционном переходе может накладываться ограничение с заданной константой ω_3^* в виде неравенства

$$\omega_3^m \equiv \max_{\tau \in [0, 1]} |\dot{\varphi}_3(\tau)| \leq \omega_3^*. \quad (6)$$

Рассмотрим сначала случай позиционного перехода без ограничения (6). Весь интервал такого перехода в нормированном времени $\tau \in [0, 1]$ разделим на два участка одинаковой длительности. На каждом участке введем свое нормированное время $\tau_1 = (t - t_0)/T_1$, $T_1 \equiv T_p/2$ и $\tau_2 = (t - t_0 - T_1)/T_2$, $T_2 \equiv T_p/2$. Угловое движение на первом участке представим функцией $\varphi_3^{(1)}(\tau_1) = \omega_3^m T_1[\tau_1^2/2 + (1 - \cos(2\pi\tau_1))/2\pi]$, удовлетворяющей краевым условиям (5) на левом конце, а на втором участке — зеркально-симметричной по отношению к $\varphi_3^{(1)}(\tau_1)$ функцией



$\varphi_3^{(2)}(\tau_2) = \varphi_{31} + \omega_3^m T_2[\tau_2 - \tau_2^2/2 - (1 - \cos(2\pi\tau_2))/2\pi]$, для которой выполняются краевые условия (5) на правом конце. Здесь значение $\varphi_{31} = \varphi_3^{(1)}(1) = \omega_3^m T_1/2$. В момент абсолютного времени $t = t_0 + T_p/2$ происходит гладкое сопряжение участков элементарного движения позиционного перехода с достижением максимальной скорости $\omega_3^m = \varphi^*/T_1 = 2\varphi^*/T_p$ с выполнением равенств

$$\begin{aligned}\varphi_3^{(1)}(1) &= \varphi_3^{(2)}(0) = \varphi_{31}; & \dot{\varphi}_3^{(1)}(1) &= \dot{\varphi}_3^{(2)}(0) = \omega_3^m; \\ \ddot{\varphi}_3^{(1)}(1) &= \ddot{\varphi}_3^{(2)}(0) = 0; & \ddot{\varphi}_3^{(1)}(1) &= \ddot{\varphi}_3^{(2)}(0) = 0.\end{aligned}$$

Функции позиционного перехода $\varphi_3^{(1)}(t)$, $\varphi_3^{(2)}(t)$ и их производные принимают вид:

$$\text{участок 1: } \begin{cases} \dot{\varepsilon}_3^{(1)}(t) = \ddot{\varphi}_3^{(1)}(t) = \dot{\varepsilon}_3^{(m1)} S_{\alpha 1}; \\ \varepsilon_3^{(1)}(t) = \ddot{\varphi}_3^{(1)}(t) = \varepsilon_3^{(m1)}(1 - C_{\alpha 1}); \\ \omega_3^{(1)}(t) = \dot{\varphi}_3^{(1)}(t) = \omega_3^{(m)}(\tau_1 - S_{\alpha 1}/2\pi); \\ \varphi_3^{(1)}(t) = \omega_3^m T_1[\tau_1^2/2 - (1 - C_{\alpha 1})/4\pi^2]; \end{cases} \quad (7)$$

$$\text{участок 2: } \begin{cases} \dot{\varepsilon}_3^{(2)}(t) = \ddot{\varphi}_3^{(2)}(t) = -\dot{\varepsilon}_3^{(m2)} S_{\alpha 2}; \\ \varepsilon_3^{(2)}(t) = \ddot{\varphi}_3^{(2)}(t) = -\varepsilon_3^{(m2)}(1 - C_{\alpha 2}); \\ \omega_3^{(2)}(t) = \dot{\varphi}_3^{(2)}(t) = \omega_3^{(m)}(1 - \tau_2 + S_{\alpha 2}/2\pi); \\ \varphi_3^{(2)}(t) = \varphi_{31} + \omega_3^m T_2[\tau_2 - \tau_2^2/2 + \\ + (1 - C_{\alpha 2})/4\pi^2]. \end{cases} \quad (8)$$

Здесь $\tau_1 = (t - t_0)/T_1$, $\tau_2 = (t - t_0 - T_1)/T_2$; функции $S_{\alpha i} = \sin(\alpha^{(i)})$, $C_{\alpha i} = \cos(\alpha^{(i)})$, $\alpha^{(i)} = 2\pi\tau_i$, $i = 1, 2$; значения $\dot{\varepsilon}_3^{(mi)} = 2\pi\omega_3^m/T_i^2$, $\varepsilon_3^{(mi)} = \omega_3^m/T_i$, $i = 1, 2$, и, как указано ранее, $\varphi_{31} = \omega_3^m T_1/2$.

На левом конце траектории позиционного перехода собственная производная ускорения $\dot{\varepsilon}_3(t_0^p) \equiv \dot{\varepsilon}_3^0 = \ddot{\varphi}_3(t_0^p) = 0$ и на правом конце $\dot{\varepsilon}_3(t_f^p) \equiv \dot{\varepsilon}_3^f = \ddot{\varphi}_3(t_f^p) = 0$ в соответствии с краевым условием (5), а угол φ_3 принимает конечное значение $\varphi_3(t_0^p) = \varphi_3^f = \varphi^* = \omega_3^m[(T_1/2) + (T_2/2)] = \omega_3^m T_p/2$.

Рассмотрим теперь вариант наличия условия (6). Здесь решение заключается в том, что между первым и вторым участками позиционного перехода с уменьшенными длительностями $T_1 = T_p/(2 + q)$ и $T_2 = T_p/(2 + q)$ «гладко вставляется» участок с постоянной скоростью $\dot{\varphi}_3(t) = \omega_3^* = \text{const}$ («полка») длительностью $T_c = T_p - (T_1 + T_2) = T_p q/(2 + q)$. Нормированный параметр q , определяющий длительность T_c «полки», находится по соотношению $q = 2(a - 1)/(1 - 2a)$, где $a = \omega_3^* T_p/(2\varphi^*)$.

Если $0 < a \leq 1/2$ либо $a \geq 1$, то значение $q \leq 0$, тогда следует назначить $q = 0$, и позиционный переход при $\omega_3^m < \omega_3^*$ будет без «полки», а при $\omega_3^m > \omega_3^*$ требуемое движение неосуществимо. Если же параметр $a \in (0,5; 1)$, то $q > 0$ и позиционный переход обязательно имеет «полку» длительностью $T_c = T_p(1 - a)/a$. Явное описание позиционного перехода на участке 1 (от начала движения до времени появления «полки») и участке 2 (с момента времени схода с «полки» до завершения движения) по-прежнему дается выражениями (7) и (8), но с подстановкой в них значений ω_3^* вместо ω_3^m , уменьшенных длительностей $T_1 = T_p(2 + q)$ и $T_2 = T_p(2 + q)$, нормированного времени на втором участке $\tau_2 = (t - (t_0 + T_1 + T_c))/T_2 \in [0, 1]$ и угла $\varphi_{31} = \omega_3^*(T_c + T_1/2) = \omega_3^* T_p(q + 1/2)/(2 + q)$. На участке «полки» движение КРМ в позиционном переходе определяются формулами

$$\begin{aligned}\dot{\varepsilon}_3^{(c)}(t) &= \ddot{\varphi}_3^{(c)}(t) = 0; & \varepsilon_3^{(c)}(t) &= \ddot{\varphi}_3^{(c)}(t) = 0; \\ \omega_3^{(c)}(t) &= \dot{\varphi}_3^{(c)}(t) = \omega_3^*; & \varphi_3^{(c)} &= \omega_3^*(T_1/2 + T_c\tau_c),\end{aligned}$$

где нормированное время $\tau_c = (t - t_0 - T_1)/T_c \in [0, 1]$.

В итоге угол φ_3^f принимает значение

$$\begin{aligned}\varphi_3(t_f^p) &= \varphi_3^f = \varphi^* = \omega_3^*[(T_1/2) + T_c + (T_2/2)] = \\ &= T_p\omega_3^*(1 + q)/(2 + q).\end{aligned}$$

2. СИЛОВЫЕ ГИРОСКОПИЧЕСКИЕ КОМПЛЕКСЫ

При выборе схемы расположения гиродинов (ГД) в связанной с корпусом КРМ системе координат $Oxzy$ учитываются требования, предъявляемые к системе управления движением. Возможности силового гироскопического комплекса по созданию вектора управляющего момента определяются только расположением осей подвеса ГД в связанной системе координат. Применение групп (в простейшем случае пар) ГД с коллинеарными осями подвеса дает важное преимущество — все внутренние сингулярные состояния такой схемы являются *проходимыми*. В рамках такого подхода при использовании только четырех ГД возможна лишь классическая схема 2-SPE (система двух ножничных пар — 2 Scissored Pair Ensemble). Анализ свойств силового гироскопического комплекса удобно проводить в его ортогональном каноническом $\mathbf{E}_c^g(x_c^g, y_c^g, z_c^g)$ и косоугольном гироскопическом $\mathbf{E}^g(\mathbf{x}, \mathbf{y}, \mathbf{z})$ базисах (рис. 1).

В прецессионной теории силовых гироскопов представления нормированного вектора кинетического момента силового гироскопического комплекса $\mathbf{h}^c \equiv \{x_c^g, y_c^g, z_c^g\}$ в каноническом базисе и $\mathbf{h} = \{x, y, z\}$ в гироскопическом базисе связаны соотношением $\mathbf{h}^c = \mathbf{A}_\gamma \mathbf{h}$, а вектор нормированного управляющего момента гироскопического комплекса $\mathbf{m}^g$ в каноническом базисе

$$\mathbf{m}^g = -\dot{\mathbf{h}}^c = -\mathbf{A}_\gamma \mathbf{A}_\beta(\beta) \mathbf{u}^g, \quad \dot{\beta} = \mathbf{u}^g, \quad (9)$$

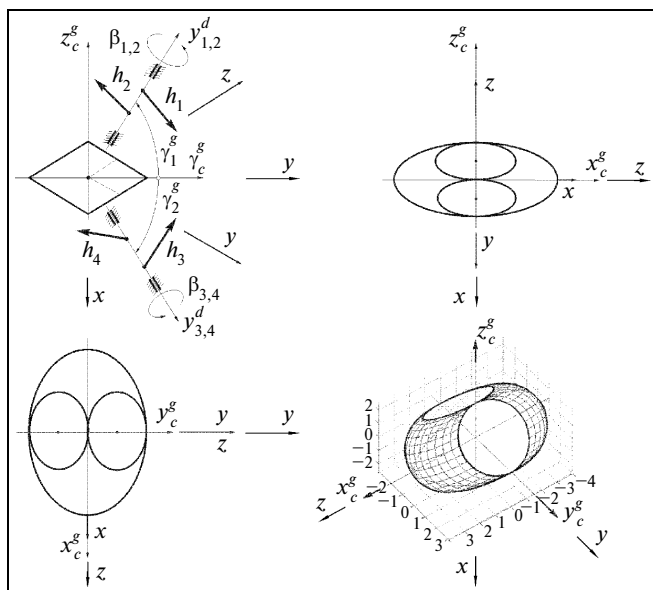


Рис. 1. Схема силового гироскопического комплекса 2-SPE и оболочка области S вариации вектора его кинетического момента

где матрицы $\mathbf{A}_\gamma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & S_{\gamma_1^g} & S_{\gamma_2^g} \\ 0 & -C_{\gamma_1^g} & C_{\gamma_2^g} \end{bmatrix}$,

$\mathbf{A}_h(\beta) = \frac{\partial \mathbf{h}}{\partial \beta} = \begin{bmatrix} -y_1 & -y_2 & -z_3 & -z_4 \\ x_1 & x_2 & 0 & 0 \\ 0 & 0 & -x_3 & -x_4 \end{bmatrix}$;

$x_p = C_{\beta_p}, p = \overline{1, 4}; y_p = S_{\beta_p}, p = 1, 2;$

$z_p = S_{\beta_p}, p = 3, 4; S_\alpha \equiv \sin \alpha; C_\alpha \equiv \cos \alpha; \beta = \{\beta_p\};$

$x = x_{12} + x_{34}; x_{12} = x_1 + x_2; x_{34} = x_3 + x_4;$
 $y = y_1 + y_2; z = -(z_1 + z_2).$

При обозначениях $\gamma = (\gamma_1^g + \gamma_2^g)/2; \alpha_1 = (\beta_1 + \beta_2)/2, \alpha_2 = (\beta_3 + \beta_4)/2, \alpha_i \in [-\pi, +\pi]; \delta_1 = (\beta_1 - \beta_2)/2, \delta_2 = (\beta_3 - \beta_4)/2, \delta_i \in [-\pi/2, \pi/2]$ и $\mathbf{A}_{\gamma h}(\beta) \equiv \mathbf{A}_\gamma \mathbf{A}_h(\beta)$ определитель матрицы Грамма $\Delta_G(\beta) = \det[\mathbf{A}_{\gamma h} \mathbf{A}_{\gamma h}^t]$ выражается в симметричном виде:

$$\Delta_G = S_{2\gamma}^2 G; G \equiv 2[(C_{\alpha_1}^2 C_{\delta_1}^2 + S_{\alpha_1}^2 S_{\delta_1}^2) S_{2\delta_2}^2 + (C_{\alpha_2}^2 C_{\delta_2}^2 + S_{\alpha_2}^2 S_{\delta_2}^2) S_{2\delta_1}^2].$$

Поэтому все естественные сингулярные состояния гироскопического комплекса, в которых $\Delta_G = 0$, описываются единым соотношением

$$(C_{\alpha_i} C_{\delta_i} S_{\delta_j} C_{\delta_j} = 0) \& (S_{\alpha_i} S_{\delta_i} S_{\delta_j} C_{\delta_j} = 0), \\ i, j \in \{1, 2\}, i \neq j, \quad (10)$$

где нет зависимости от значения угла $2\gamma \neq n\pi (n = 0, 1, 2, \dots)$ между осями подвеса пар ГД. Отсюда следует принципиальная важность синтеза закона настройки схемы гироскопического комплекса в «каноническом» варианте, когда углы $\gamma_1^g = \pi/2, \gamma_2^g = 0$ (т. е. $2\gamma = \pi/2$). Анализ соотношения (10) приводит к следующему описанию всех естественных множеств сингулярных состояний:

- $[C_{\delta_i}] \Rightarrow \delta_i = \pm\pi/2, i = 1, 2$: обе пары ГД находятся во *внутреннем* сингулярном состоянии, при этом вектор нормированного кинетического момента $\mathbf{h} = 0$ (рис. 2);
- $[(C_{\delta_i} = 0) \& (S_{\alpha_i} = 0)] \Rightarrow \alpha_i = 0 \vee (\pm\pi); \delta_i = \pm\pi/2, (i = 1) \vee (i = 2)$: одна из пар ГД находится во *внутреннем* сингулярном состоянии и *одновременно* векторы $\mathbf{h}$ ее гироскопов направлены перпендикулярно оси x гироскопического базиса, при этом вектор $\mathbf{h}$ формируется только другой парой ГД и принадлежит соответствующему кругу радиуса 2, см. рис. 2, а;
- $[(C_{\delta_i} = 0) \& (S_{\delta_j} = 0)] \Rightarrow \delta_i = \pm\pi/2, \delta_j = 0, i, j \in \{1, 2\}, i \neq j$: одна из пар ГД находится во *внутреннем* сингулярном состоянии, а другая — во *внешнем* сингулярном состоянии, при этом конец вектора $\mathbf{h}$ принадлежит окружности радиуса 2, см. рис. 2, а;
- $[(S_{\delta_i} = 0)] \Rightarrow \delta_i = 0, i = 1, 2$: обе пары ГД находятся во *внешнем* сингулярном состоянии, при этом конец вектора $\mathbf{h}$ принадлежит *трехмерной* двояковыпуклой поверхности $\mathbf{S}^* = \{\mathbf{h}(x, y, z): \varphi(x, y, z) \equiv x^2 + y^2 + z^2 \mp 2q_y q_z - 8 = 0, q_s = \sqrt{4 - s^2}, |s| \leq 2, s = y, z\}$, где верхний знак соответствует оболочке $\partial \mathbf{S}$ области вариации $\mathbf{S}$ кинетического момента силового гироскопического комплекса (см. рис. 1).

Экстремальные значения определителя Δ_G достигаются в таких конфигурациях:

локальный максимум $\max \Delta_G = 2S_{2\gamma}^2 \Rightarrow \alpha_i, \alpha_j = \pm\pi/2; \delta_p, \delta_j = \pm\pi/4, i, j \in \{1, 2\}, i \neq j$, в четырех точках ($x = 0; y = \pm\sqrt{2}; z = \pm\sqrt{2}$) внутри области вариации $\mathbf{S}$;

глобальный максимум $\max \max \Delta_G = (64/27)S_{2\gamma}^2 \Rightarrow \alpha_i, \alpha_j = \{0, \pm\pi\}; \delta_p, \delta_j = \{\pm\delta_p, \pi \pm \delta_p\}, i, j \in \{1, 2\}, i \neq j$, где

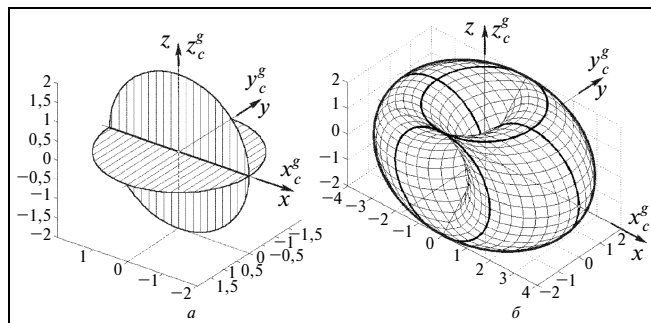


Рис. 2. Естественные сингулярные состояния схемы гироскопического комплекса для канонического варианта:

а — плоские поверхности; б — криволинейные поверхности

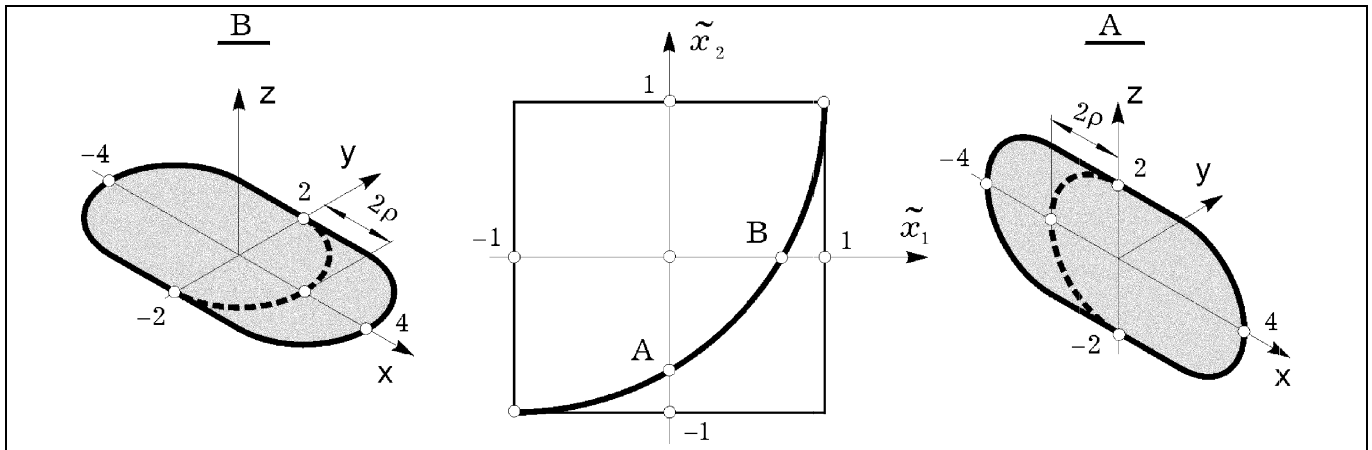


Рис. 3. Сингулярные множества силового гироскопического комплекса при законе настройки (3) — половины эллипсов

$\delta_p = \arccos(\sqrt{2/3}) = 35^\circ 25'$, достигается в трех точках ($x = \{0, \pm 4\sqrt{2/3}\}$; $y = 0$; $z = 0$).

При произвольных углах установки осей подвеса пар ГД оболочка области $\mathbf{S}$ вариации кинетического момента и множества сингулярных состояний имеют подобный вид (см. рис. 1).

Самая принципиальная проблема управления силовым гироскопическим комплексом с избыточной структурой заключается в выборе функции распределения потребного суммарного вектора кинетического момента между ГД — синтезе закона настройки гироскопического комплекса. Наиболее рациональными являются *явные* законы настройки, которые позволяют получить все характеристики движения каждого ГД по явным аналитическим соотношениям.

Введем функцию распределения вектора суммарного кинетического момента силового гироскопического комплекса между парами ГД

$$f_p(\beta) = \tilde{x}_1 - \tilde{x}_2 + \rho(\tilde{x}_1 \tilde{x}_2 - 1); \quad (11)$$

$$\tilde{x}_1 = x_{12}/q_y; \quad \tilde{x}_2 = x_{34}/q_z$$

с фиксированным параметром ρ , $0 < \rho < 1$. При условии $f_p(\beta) = 0$ внутри области $\mathbf{S}$ остаются сингулярными только два одномерных множества

$$\mathbf{S}_y = \{(x/(2\rho))^2 + (z/2)^2 = 1, x < 0; y = 0, |y_1| = |y_2| = 0\};$$

$$\mathbf{S}_z = \{(x/(2\rho))^2 + (y/2)^2 = 1, x > 0; z = 0, |z_3| = |z_4| = 0\}, \quad (12)$$

(рис. 3) с двумя точками $A(0, -\rho)$ и $B(\rho, 0)$ в плоскости переменных $\{\tilde{x}_1, \tilde{x}_2\}$.

Применяемый закон настройки силового гироскопического комплекса имеет вид

$$D^+ f_p(\beta) = \Phi_p(f_p(\beta), \mathbf{h}(\beta)) =$$

$$\equiv \begin{cases} -\text{Sat}(\phi_p, \mu_p f_p(\beta)), & \mathbf{h}(\beta) \in \mathbf{S}/\mathbf{Q}_{yz} \\ \phi_p \text{Relh}(a_s, l_p, r_s), & \mathbf{h}(\beta) \in \mathbf{Q}_s, \quad s = y, z. \end{cases} \quad (13)$$

Здесь D^+ — символ правой производной по времени, ϕ_p , μ_p и l_p — положительные параметры, множество

$\mathbf{S}_{yz} \equiv \mathbf{S}_y \cup \mathbf{S}_z$, зависимости $r_y = M_\pi(\beta_1 - \beta_2 - \pi)$ и $r_z = M_\pi(\beta_3 - \beta_4 - \pi)$, где при стандартной функции насыщения $\text{Sat}(a, x)$ используются нелинейные функции

$$\text{Relh}(a, l_p, x) \equiv \begin{cases} 1, & x > -l_p \\ -1, & x < l_p \end{cases}$$

$$\text{Relh}(a_s, l_p, r_s(\beta_0)) = a_s \in \{-1; 1\};$$

$$M_\pi(\alpha) \equiv \begin{cases} \alpha, & |\alpha| \leq \pi \\ \alpha - 2\pi \text{sign}(\alpha), & |\alpha| > \pi \end{cases}$$

При законе настройки (13) обеспечивается принадлежность конца нормированного вектора $\mathbf{h}(\beta)$ кинетического момента множеству $\mathbf{Q}_{yz}(\beta)$ внутренних сингулярных состояний только в отдельные моменты времени (меры нуль по Лебегу) и биективная связь вектора $\mathbf{m}^g(\beta)$ (9) с векторами-столбцами β и $\mathbf{u}^g = \dot{\beta}$. Анализ выражений (12) приводит к результату: в косоугольном гироскопическом базисе $\mathbf{E}^g(x, y, z)$ для любых значений нормированного вектора $\mathbf{h} = \{x, y, z\}$, принадлежащего сфере $\mathbf{S}_\rho^g \equiv \{\mathbf{x}^2 + \mathbf{y}^2 + \mathbf{z}^2 \leq (r_\rho^g)^2\}$ радиуса $r_\rho^g < 2\rho$, отсутствуют сингулярные состояния гироскопического комплекса. В каноническом базисе $\mathbf{E}_c^g(x_c^g, y_c^g, z_c^g)$ сфере $\mathbf{S}_\rho^g$ соответствует эллипсоид

$$\mathbf{S}_\rho^c = \{x_c^g, y_c^g, z_c^g: (x_c^g)^2 + \frac{(y_c^g)^2}{2S_\gamma^2} + \frac{(z_c^g)^2}{2C_\gamma^2} \leq (r_\rho^g)^2\}.$$

3. АНАЛИТИЧЕСКИЙ РАСЧЕТ ПРОГРАММНОГО ГИРОСИЛОВОГО УПРАВЛЕНИЯ

При заданном в каноническом базисе векторе кинетического момента $\mathbf{h}^c \in \mathbf{S}_\rho^c$ и при выполнении условия $f_p(\beta) = 0$ для функции распределения $f_p(\beta)$ (11) положение каждого ГД однозначно определяется по явным соотношениям. Вектор-столбец $\mathbf{h} = \{x, y, z\} = \mathbf{A}_\gamma^{-1} \mathbf{h}^c$

также известен, а разность $\Delta = x_{12} - x_{34} = d\{1 - \sqrt{1 - 4\rho[(q_y - q_z)(x/2) + \rho(q_y q_z - (x/2)^2)]/d^2}\}/\rho$, где $d \equiv q_y + q_z$. Отсюда при введении функций $a_y = (x + \Delta)/2$; $b_y = y$; $a_z = (x - \Delta)/2$; $b_z = -z$; $r_s \equiv \sqrt{a_s^2 + b_s^2}$ и $x_s = \sqrt{(2 - r_s)(2 + r_s)}/r_s$, $s = y, z$ получаются проекции вектора кинетического момента каждого ГД первой пары $x_{1,2} = a_y \mp b_y x_y/2$; $y_{1,2} = b_y \pm a_y x_y/2$ и аналогично для второй пары ГД. При введении обозначений

$$s_{1,2}^0 = \sin(\beta_1 - \beta_2) \equiv y_1 x_2 - y_2 x_1;$$

$$s_{3,4}^0 = \sin(\beta_3 - \beta_4) \equiv z_3 x_4 - z_4 x_3;$$

$$r_{yx} = x_{12} y p_{zx} q_y^2; \quad r_{zx} = x_{34} z p_{yx} q_z^2;$$

$$p_s = d_s/d_{yz}, \quad s = yx, zx,$$

$$D(\beta) = \begin{bmatrix} -x_2 p_{yx}/s_1^0 & -(y_2 - x_2 r_{yx})/s_1^0 & -x_2 r_{zx}/s_1^0 \\ x_1 p_{yx}/s_2^0 & (y_1 - x_1 r_{yx})/s_2^0 & x_1 r_{zx}/s_2^0 \\ -x_4 p_{zx}/s_3^0 & -x_4 r_{yx}/s_3^0 & (z_4 + x_4 r_{zx})/s_3^0 \\ x_3 p_{zx}/s_4^0 & x_3 r_{yx}/s_4^0 & -(z_3 + x_3 r_{zx})/s_4^0 \end{bmatrix},$$

где $d_{yx} = q_y - \rho x_{12}$; $d_{zx} = q_z - \rho x_{34}$ и $d_{yz} = d_{yx} + d_{zx}$, вектор программного управления силовым гироскопическим комплексом $u^g = \dot{\beta}$ вычисляется по формуле $u^g = \dot{\beta} = -D(\beta)A_y^{-1}m^g$. При аналитическом задании программы движения корпуса КРМ (функций $\Lambda(t)$, $\omega(t)$, $\varepsilon(t)$, $\dot{\varepsilon}(t)$) и выполнении тождества $f_p(\beta) \equiv 0$ в силу наличия векторного интеграла суммарного кинетического момента механической системы «КРМ + гиросилового комплекс» векторы-столбцы углов $\beta(t)$, угловых скоростей $\dot{\beta}(t)$, а также и ускорений $\ddot{\beta}(t)$, всех четырех ГД определяются по явным соотношениям, полученным в работах [7–10].

ЗАКЛЮЧЕНИЕ

Предложен метод аналитического синтеза программного углового движения КРМ при выполнении им поворотного маневра с краевыми условиями общего вида. В процессе такого «профилированного» поворотного маневра естественным способом достигается слабое возбуждение упругих колебаний транспортируемого с помощью КРМ нежесткого груза благодаря обеспечению

достаточной «гладкости» аналитически рассчитанного гиросилового управления. При кинетическом моменте $h^c(t) \in S_p^c$ гарантируется отсутствие сингулярных состояний силового гироскопического комплекса, расчет потребных максимальных по модулю значений скорости и ускорения каждого ГД выполняется по явным аналитическим соотношениям.

В дальнейшем представляется целесообразным оценить влияние характеристик транспортируемого упругого груза на выбор коэффициентов программного управления силовым гироскопическим комплексом $u^g = \dot{\beta}$ с позиций обеспечения требуемого качества реализации поворотных маневров КРМ.

ЛИТЕРАТУРА

1. Рутковский В. Ю., Суханов В. М. Большие космические конструкции: модели, методы исследования и принципы управления. I, II // Автоматика и телемеханика. — 1996. — № 7. — С. 52–65; № 8. — С. 55–66.
2. Putz P. Space Robotics. — Amsterdam: Kluwer Academic Publishers, 1999. — P. 547–596.
3. Yoshida K., Umetani Y. Control of space free-flying robot // Proc. of 29th IEEE Conference on Decision and Control, 1990. — P. 97–102.
4. Раушенбах Б. В., Токарь Е. Н. Управление ориентацией космических аппаратов. — М.: Наука, 1974.
5. Иванов Н. М., Лысенко Л. Н. Баллистика и навигация космических аппаратов. — М.: Дрофа, 2004.
6. Сомов Е. И., Бутырин С. А. Аналитический синтез программного движения космических аппаратов наблюдения // Изв. Самарского науч. центра РАН. — 2004. — № 1. — С. 168–179.
7. Управление силовыми гироскопическими комплексами космических аппаратов / Е. И. Сомов, С. А. Бутырин, А. В. Сорокин, В. Н. Платонов // Труды X С.-Петербургской Междунар. конф. по интегрированным навигационным системам. — СПб.: ЦНИИ «Электронприбор», 2003. — С. 278–294.
8. Somov Ye. I. Nonlinear spacecraft gyromoment attitude control // Proc. of 1st Intern. conf. on Nonlinear Problems in Aviation and Aerospace. — Daytona Beach: ERAU, USA, 1997. — P. 625–630.
9. Сомов Е. И., Герасин И. А. Оценка реализуемости поворотного маневра космического аппарата, управляемого избыточной системой гироскопов // Управление движением и навигация летательных аппаратов. — Самара: ПФ Российской академии космонавтики им. К. Э. Циолковского, 1998. — С. 138–143.
10. Ultra-precision attitude control of a large low-orbital space telescope / Ye. I. Somov, S. A. Butyrin, V. M. Matrosov, G. P. Anshakov et al. // Control Eng. Practice. — 1999. — Vol. 7, N 7. — P. 1127–1142.

☎ (846) 242-38-82;

e-mail: e_somov@mail.ru



Уважаемые читатели!

Если вы не успели или забыли подписаться на журнал «Проблемы управления», то через Редакцию Вы можете оформить льготную подписку в любое время и с любого номера (дешевле, чем через каталоги агентств) или приобрести номера журнала за прошедшие годы.

Можно также заказать электронные версии как необходимого вам номера журнала, так и отдельных статей.

Обратитесь в редакцию по тел. (495) 330-42-66 или пришлите заказ по адресу (datchik@ipu.ru) — и подписка будет оформлена за один день. Расходы по пересылке журнала редакция берет на себя. Не забудьте указать свой полный почтовый адрес!

Наш адрес: 117997, Москва, В-342, ГСП-7, Профсоюзная ул., 65, ИПУ РАН, оф. 104.

ОПРЕДЕЛЕНИЕ ДИСТАНЦИИ ДО ЦЕЛИ, НАХОДЯЩЕЙСЯ В ЗОНЕ ФРЕНЕЛЯ АНТЕННЫ

И. М. Рудько

Институт проблем управления им. В. А. Трапезникова РАН, г. Москва

Рассмотрен способ определения дальности до цели при пространственно-временной обработке сигналов в режиме пассивной локации целей, находящихся в зоне Френеля антенны, в случае, когда излучаемый целью сигнал содержит узкополосную составляющую.

ВВЕДЕНИЕ

В пассивных гидролокационных системах обнаружение объектов и оценка их координат основаны на приеме и обработке акустических сигналов, излучаемых целью. Работа пассивных гидролокационных систем лимитируется потерями на распространение акустических сигналов и уровнем аддитивных шумов приемника, а также неточностью знаний акустических параметров целей и обусловленным водной средой размытием сигналов, излучаемых целью, как по времени, так и по частоте.

При обработке сигналов от целей, расположенных в дальней зоне приемной антенны, волновой фронт акустической волны в пределах апертуры антенны является практически плоским (цель находится в зоне Фраунгофера). В этом случае путем пространственно-временной обработки сигналов можно осуществлять разрешение целей по направлению [1].

При пассивной локации полное определение местоположения цели возможно лишь тогда, когда она находится в зоне Френеля антенны (цилиндрический фронт волны). При этом единственным источником прямой информации о дальности до источника излучения является кривизна фронта волны.

В настоящей статье рассматривается способ определения дальности до цели методом максимального правдоподобия при пространственно-временной обработке сигналов в режиме пассивной локации целей, находящихся в зоне Френеля, в случае, когда излучаемый целью сигнал содержит узкополосную составляющую.

1. ПЛОТНОСТЬ ВЕРОЯТНОСТИ СИГНАЛА ПРИ НАЛИЧИИ ИЗОТРОПНОЙ ПОМЕХИ

Способы формирования диаграммы направленности антенны, пространственно-временной обработки сигнала и его обнаружения, которые могут осуществляться как во временной, так и в частотной областях, широко известны [1, 2] и в дальнейшем не рассматриваются.

В приведенном далее алгоритме обработка входного сигнала производится в частотной области, в координатах (время — t , пространство приемников — $j = 1, \dots, M$).

Пусть сигнал от цели C , коррелированный по фронту волны и содержащий узкополосную составляющую с частотой f_0 и интенсивностью A , находящейся на расстоянии r_0 под углом θ (здесь и далее углы отсчитываются от нормали к плоскости антенны в ее центре), принимается на M гидрофонов линейной эквидистантной антенны с расстоянием между гидрофонами d в узкой полосе Δf , где $\Delta f/f_0 \ll 1$, на фоне узкополосной помехи с пространственно-корреляционной матрицей R_0 размерностью M (рис. 1). Считаем, что система обнаруже-

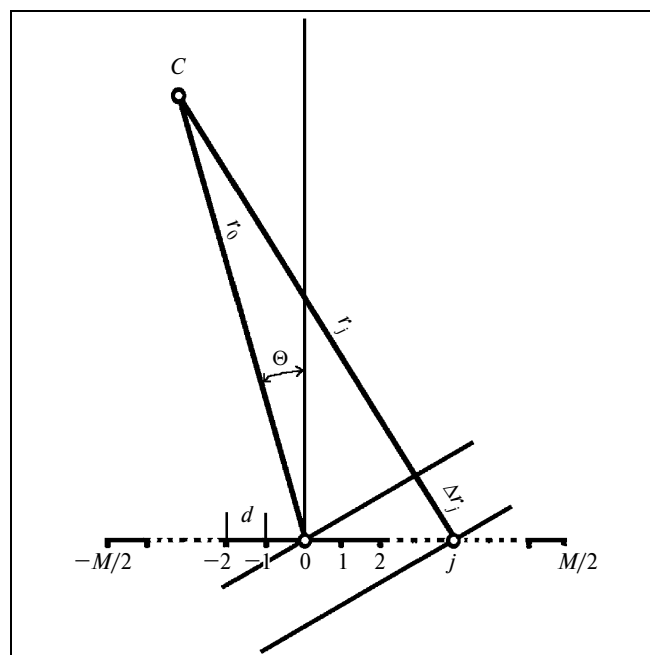


Рис. 1. Взаимное расположение цели и антенны

ния антенны обнаружила узкополосный сигнал от цели C на частоте f_0 в полосе Δf под углом θ . Далее по умолчанию считается, что сигнал обрабатывается на частоте f_0 в полосе Δf . Необходимо оценить дистанцию до цели r_0 .

Априорные сведения о начальной фазе φ сигнала отсутствуют. Узкополосный сигнал $m(t) = A \exp\{-i(2\pi f_0 t + \varphi)\} = U \cos 2\pi f_0 t + iV \sin 2\pi f_0 t$ представляет собой комплексный процесс с независимыми нормально распределенными компонентами U и V , имеющими нулевые математические ожидания и одинаковые дисперсии:

$$N(0, \sigma^2). \quad (1)$$

В момент времени t сигнал на j -м гидрофоне антенны: $z(j, t) = m(t) \exp\{-i2\pi f_0 \Delta r_j / c\} + n(j, t)$, $j = 1, \dots, M$, где c — скорость звука, $\Delta r_j = r_0 - r_j$ — разность расстояний между источником и центром антенны и источником и j -м гидрофоном, $n(j, t)$ — изотропный гауссов белый шум в полосе Δf . Под наблюдением X будем понимать M -мерный вектор-столбец комплексных коэффициентов дискретного Фурье-преобразования по времени t на частоте f_0 сигнала $z(j, t)$: $X = \{x(j_1), x(j_2), \dots, x(j_M)\}^T$, где $\{\cdot\}^T$ — операция транспонирования, $0 \leq t \leq T$, $[0, T]$ — интервал наблюдения, $T = 1/\Delta f$.

Как показано в работе [2], формула для расчета условной плотности вероятности сигнала при наличии изотропной помехи имеет следующий вид:

$$p(X|Y) = \frac{1}{\pi^M \det R_0} \exp\{-(X - Y\Phi_\theta)^T R_0^{-1} (X - Y\Phi_\theta)\}, \quad (2)$$

где $Y = U + iV$ — комплексный коэффициент дискретного Фурье-преобразования сигнала $m(t)$ в полосе Δf на частоте f_0 ,

$$\Phi_\theta = \left\{ \exp\left(-i \frac{2\pi f \Delta r_1}{c}\right), \exp\left(-i \frac{2\pi f \Delta r_2}{c}\right), \dots, \exp\left(-i \frac{2\pi f \Delta r_M}{c}\right) \right\}^T \quad (3)$$

вектор-столбец апертурных функций антенны, а черта сверху означает комплексно сопряженную величину.

Когда волновые размеры антенны соизмеримы с дальностью до цели (цель — в зоне Френеля), при построении апертурных функций Φ_θ необходимо учитывать кривизну фронта волны. Расстояние от цели до j -го приемника, $j = \{0, \pm 1, \pm 2, \dots, \pm M/2\}$, [3]

$$r_j = \sqrt{r_0^2 + j^2 d^2} - 2jdr_0 \sin \theta,$$

$$\Delta r_j = r_0 - r_j \approx jd \sin \theta - j^2 d^2 \cos^2 \theta / 2r_0. \quad (4)$$

С учетом выражений (4) элементы вектор-столбца апертурных функций (3) имеют вид:

$$\begin{aligned} \varphi_j(\theta, r) &= \exp\left(i\pi \frac{j^2 d^2}{r_0 c} f \cos^2 \theta - i2\pi f \frac{jd}{c} \sin \theta\right) = \\ &= \exp[i2\pi \alpha(r) - i2j\beta], \quad j = 1, \dots, M, \end{aligned} \quad (5)$$

где $\alpha(r) = j^2 d^2 f \cos^2 \theta / 2r_0 c$, $\beta = jdf \sin \theta / c$.

Из выражения (5) следует, что при нахождении цели в зоне Френеля апертурные функции $\varphi_j(\theta, r)$ зависят не только от угла прихода сигнала θ , но и от расстояния до цели r_0 . При этом антенна будет фокусироваться в точки кривой, описываемой уравнением $r = r_n \cos^2 \theta$, где r_n — расстояние по нормали до центра антенны.

Пусть корреляционная матрица шума в полосе Δf на частоте f_0 равна $R_0 = \delta^2 I_M$, где δ^2 — дисперсия шума, I_M — единичная матрица размерности M . Тогда для фиксированного радиуса фокусировки r и угла θ с учетом условия (1) можно преобразовать плотность вероятности (2) к виду:

$$\begin{aligned} p(X|r, \theta) &= \int p(X|r, \theta, Y) p(Y) dY = \\ &= \gamma \exp\left\{ \frac{2\rho^2 X^T \Phi \bar{\Phi}^T X}{\delta^2 (1 + 2\rho^2 M)} \right\}, \end{aligned} \quad (6)$$

где γ — нормировочный коэффициент, не зависящий от r , $\rho = \sigma/\delta$ — отношение сигнал/помеха с направления θ на частоте f_0 . Скалярное произведение

$$\begin{aligned} \bar{X}^T \Phi &= \bar{\Phi}^T X = \sum_{j=1}^M x_j \varphi_j(\theta, r) = \\ &= \sum_{j=1}^M \{x_j \exp[i2\pi \alpha(r)]\} \exp(-i2\pi j\beta) \end{aligned}$$

можно рассматривать как дискретное Фурье-преобразование входного сигнала по пространству M приемников. При этом сам сигнал на j -м приемнике $\{x_j \exp[i2\pi \alpha(r)]\}$ представляет собой входной сигнал, сдвинутый на фазовый угол, определяемый кривизной фронта волны. Таким образом [4],

$$E[\bar{X}^T \Phi \bar{\Phi}^T X] = E[(\bar{X}^T \Phi)^2] = S(r, \theta),$$

где E — оператор математического ожидания, $S(r, \theta)$ — спектральная плотность мощности (периодограмма) входного сигнала при фиксированных радиусе фокусировки r и угле θ . Вычисление оценки спектральной плотности мощности может осуществляться, например, путем последовательного вычисления дискретного Фурье-преобразования по M приемникам, возведения в квадрат и усреднения [4].

С учетом изложенного выражение (6) преобразуется к виду:

$$p(X|r, \theta) = \gamma \exp\left\{ \frac{2\rho^2 S(r, \theta)}{\delta^2 (1 + 2\rho^2 M)} \right\}. \quad (7)$$

2. ОЦЕНКА ДИСТАНЦИИ ДО ЦЕЛИ В ЗОНЕ ФРЕНЕЛЯ

При приеме узкополосного сигнала от цели, находящейся в зоне Френеля, наблюдается пространственная избирательность по дальности, причем энергия сигнала имеет максимумы на расстояниях, определяемых выражением [5]:

$$R_k = L^2 / 2k\lambda, \quad k = 1, 2, \dots, K, \quad (8)$$

где $\lambda = c/f$ — длина излучаемой волны, k — номер зоны Френеля, $L = (M - 1)d$ — база антенны.



Ширина зоны Френеля определяется выражением:

$$b_k = 2L^2/\lambda(4k - 1). \quad (9)$$

Соотношения (8) и (9) справедливы лишь для значений $r \geq \sqrt[3]{L^4/(32\lambda)}$, поскольку при их выводе использо-

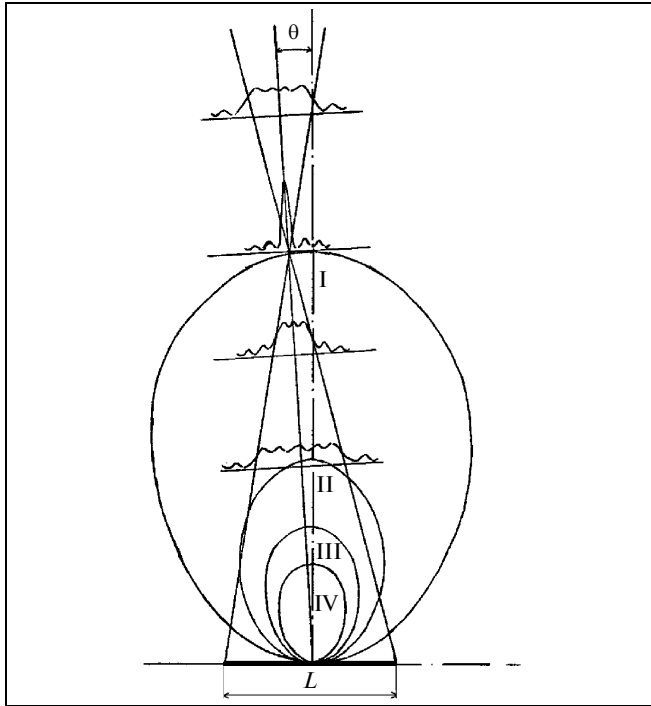


Рис. 2. Конфигурация первых четырех зон Френеля

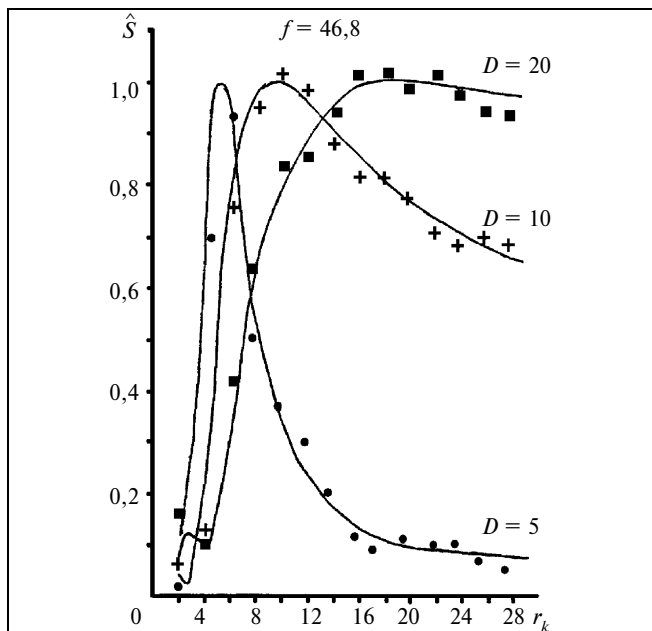


Рис. 3. Оценки энергии сигнала для различных радиусов фокусировки антенны:

■ — $D = 20$ миль (1-я зона Френеля); + — $D = 10$ миль (2-я зона Френеля); ● — $D = 5$ миль (4-я зона Френеля)

Расстояния до первой зоны Френеля $R_1(L, f)$, км

L , м	F , Гц (λ , м)			
	20(75)	50(30)	100(15)	200(7,5)
500	1,7	4,2	8,3	16,7
1000	6,7	16,7	33,3	66,7
1500	15,0	37,5	75,0	150,0
2500	41,7	104,2	208,3	416,7

валось квадратичное приближение для цилиндрического фронта волны [3]. На рис. 2 показаны первые I, II, III, IV зоны Френеля и их конфигурация в пространстве, а в таблице приведены значения $R_1(L, f)$.

Построив условную плотность вероятности (7), можно по критерию максимального правдоподобия определить дистанцию до цели. При этом аналитический поиск максимума при выполнении конкретных вычислений заменяется поиском по дискретному множеству дистанций.

Следовательно, для фиксированного вектора наблюдений X , произведя перебор по дискретному ряду радиусов фокусировки r_k при заданной частоте f_0 и вычислив оценки спектральной плотности $\hat{S}(r_k, \theta)$, отношения сигнал/помеха $\hat{\rho}$ и энергии шума $\hat{\delta}^2$ для каждого радиуса r_k , можно построить оценку плотности вероятности $p(X|r_k, \theta)$, определяемую формулой (7), в заданном диапазоне дистанций. Эти оценки используются для построения оценки плотности вероятности $\hat{p}(X|r_k, \theta)$ в дискретном ряде точек по дистанции от $r_{\min}$ до $r_{\max}$, по которой можно определить оценку дистанции из условия обеспечения $\max_r p(X|r_k, \theta)$.

3. РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ

Приведенный алгоритм оценки дистанции до цели, находящейся в зоне Френеля, был реализован в виде программы для ЭВМ и проверен в полунатурных условиях: а именно, в качестве сигнала использовался генерируемый специальной программой модельный (синусоидальный) сигнал с заданными частотой f_0 , амплитудой A , направлением θ и дистанцией до антенны r_0 , а в качестве шума — записи шума моря, полученные с буксируемой геофизической антенны ($d = 50$ м, $M = 32$). На рис. 3 и 4 приведены некоторые результаты моделирования — по формуле (7) построены оценки спектральной плотности $\hat{S}(f_0, \theta)$ при переборе по радиусам r_k для фиксированного вектора наблюдений $\{X(t)\}$ для двух случаев:

- 1) в отсутствии шума ($\delta = 0$) — сплошные линии;
- 2) при заданном отношении сигнал/помеха ($\rho = 15$ дБ) — символы ●, +, ■.

На рис. 3 для фиксированной частоты излучения цели $f = 46,8$ Гц построены оценки (относительные) энергии сигнала при различных радиусах фокусировки ан-

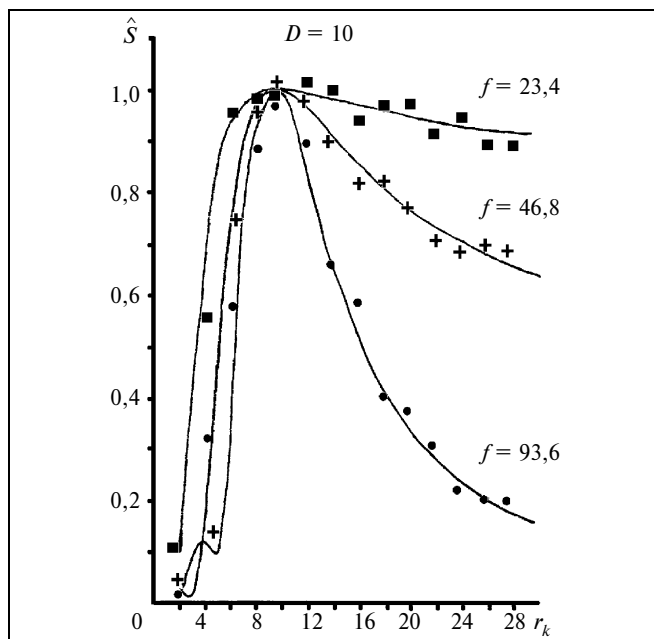


Рис. 4. Оценки энергии сигнала при различных частотах излучаемого сигнала:
 ■ — $f = 23,4$ Гц (1-я зона Френеля); + — $f = 46,8$ Гц (2-я зона Френеля); ● — $f = 93,6$ Гц (4-я зона Френеля)

тенны в диапазоне от 2 до 28 миль на различных дистанциях D до цели.

На рис. 4 для фиксированной дистанции до цели $D = 10$ миль построены оценки энергии сигнала при различных частотах сигнала.

Как видно из приведенных графиков, в отсутствии шума результаты моделирования совпадают с теоретическими, а при $\rho = 15$ дБ — близки к ним.

ЗАКЛЮЧЕНИЕ

Наличие узкополосной составляющей в спектре сигнала цели, находящейся в зоне Френеля приемной антенны, позволяет определять дистанцию до цели без применения триангуляционных методов определения дистанции. Точность определения дистанции определяется номером зоны Френеля, в которой находится цель, и частотой излучаемого узкополосного сигнала, а именно:

— при фиксированной частоте сигнала при уменьшении дистанции до цели точность определения увеличивается, так как цель попадает в зоны Френеля с более узкой шириной;

— при фиксированном радиусе фокусировки антенны, совпадающем с дистанцией до цели, с ростом частоты принимаемого сигнала точность определения увеличивается, так как цель также попадает в зоны Френеля с более узкой шириной зоны.

ЛИТЕРАТУРА

1. Найт У. Г., Придем Р. Г., Кэй С. М. Цифровая обработка сигналов в гидролокационных системах // ТИИЭР. — 1981. — Т. 69, № 11. — С. 84–154.
2. Монзинго Р. А., Миллер Т. У. Адаптивные антенные решетки. — М.: Радио и связь, 1986. — 446 с.
3. Кремер И. Я., Понькин В. А. О потенциальной точности определения местоположения цели при пространственно-временной обработке сигналов в общем случае // Радиотехника и электроника. — 1975. — Т. XX, — № 6. — С. 1186–1193.
4. Бендат Дж., Пирсол А. Применение корреляционного и спектрального анализа. — М.: Мир, 1983. — 312 с.
5. Басс Ф. Г., Фукс И. М. Рассеяние на статистически неоднородной поверхности. — М.: Наука, 1972. — 424 с.

☎ (495) 334-79-59;

e-mail: dobroid@ipu.ru



Новая книга

Левин В. И. Структурно-логические методы в теории расписаний. — Пенза: Изд-во Пенз. гос. технол. акад., 2006. — 176 с.

Дан обстоятельный обзор основных результатов в теории расписаний, полученных с помощью математического аппарата непрерывной логики и логических определителей. Подробно излагается указанный аппарат и его применение к решению следующих задач: расчет характеристик быстродействия и загрузки систем, состоящих из последовательно соединенных блоков и получающих на входе совокупность работ, выполняемых затем в этих блоках; анализ характеристик быстродействия и загрузки систем при варьировании временных параметров работ и блоков; планирование оптимального расписания выполнения работ в системе при постоянном порядке прохождения работ через все блоки; расчет характеристик быстродействия и планирование оптимального расписания выполнения работ в системе при допустимости изменения порядка прохождения работ при переходе от одного блока к другому. Большое внимание уделено возможности вычислений в случае высокоразмерных систем, содержащих большое число блоков и работ.

Предназначена для студентов вузов, аспирантов, специалистов по управлению и планированию в экономических, социальных и технических системах.

КАК УВЕЛИЧИТЬ ИНДЕКС ЦИТИРОВАНИЯ НАУЧНОЙ ПУБЛИКАЦИИ

В. Л. Эпштейн

Институт проблем управления им. В. А. Трапезникова РАН, г. Москва

Приведены сравнительные данные о цитируемости печатных и электронных публикаций. Для увеличения индекса цитирования предложено осуществлять интернет-маркетинг «экстратекстов» научных работ. Дано определение «экстратекста». Высказано предположение о том, что предлагаемый подход позволит ускорить распространение информации о новых достижениях науки; предложено ввести в научный и экономический оборот интеллектуальный капитал, хранящийся на полках наших библиотек в виде опередивших свое время или недооцененных и незаслуженно забытых публикаций; предложено создать общедоступное научно-информационное обеспечение для инновационной экономики.

С тех пор как на столе у каждого современного учебного появился компьютер, подключенный к Интернету, индекс цитирования любой научной публикации все больше зависит от того, представлена ли она во Всемирной паутине.

Как показало исследование [1]:

- с каждым годом все больше цитируются электронные статьи, открытые для бесплатного доступа;
- среднее число библиографических ссылок на печатную статью — 2,74; среднее число ссылок на бесплатную электронную — 7,3 или в 2,6 раза больше, чем на печатную;
- если взять процентное соотношение печатных и бесплатных электронных статей по годам, а затем усреднить полученные результаты на интервале 1989—2000 гг., то окажется, что электронные статьи цитируются в 4,5 раза чаще печатных, и это соотношение быстро возрастает.

Объяснение очевидно: с каждым годом ученые проводят все меньше времени в библиотеках и все больше черпают информацию из Всемирной паутины. Отмечается, что «...статья, недоступная в электронном формате, становится как бы статьей второго сорта не потому, что она худшего качества, а потому, что люди предпочитают доступность вещей, которые они могут получить прямо сейчас» [2].

В этих условиях каждый ученый должен понять, что для признания его идей, гипотез, научных результатов и адекватной их цитируемости мировым научным сообществом

недостаточно печатных публикаций и публичных выступлений, нужно позаботиться о том, чтобы представить их в Веб, привлечь к ним внимание и заинтересовать предельно широкую целевую аудиторию.

Как это сделать?

Научная публикация (препринт, статья, книга) — носитель знания, добытого в результате интеллектуального труда или, переходя на язык рыночной экономики, товар, который нужно «продать» возможно большему числу потребителей [3]. «Продать» в этом контексте означает, что после ознакомления с «рекламным сообщением» о новом товаре на рынке знаний потребитель просит прислать полный текст публикации или сообщить, где ее можно прочитать.

Конечно не факт, что после изучения публикация будет процитирована. Но это уже вопрос качества работы и ее адекватности интересам потребителя. С другой стороны, можно представить себе, что «рекламное сообщение» содержит достаточную информацию о товаре (например, об идее или о полученном результате), и это может оказаться достаточным для того, чтобы потребитель сослался на нее или обратился с контрактным предложением. Нельзя также не учитывать, что в постиндустриальную эпоху знания структурированная информация, результаты фундаментальных исследований (даже незавершенных) переходят из категории бесплатного общественного блага в категорию товара, приносящего порой миллиардную прибыль [4].

Основываясь на сказанном, зададимся теперь вопросом о том, как должно быть устроено «рекламное сообщение», чтобы пользователь мог быстро оценить «товар»: быстро понять, чему посвящена данная публикация, какие получены результаты и что предлагается. По сути дела все эти моменты, уже с начала XX в., принято, так или иначе, отражать в стандартном формате журнальной статьи, обозначаемом в англоязычной литературе как IMRAD (Introduction, Methods, Results and Discussion) [5].

Теперь для Интернет-маркетинга научных публикаций мы предлагаем использовать формат, который называем «экстратекстом». В экстратекст включаем: *Заголовки публикации, Сведения об авторах, Выходные данные, Аннотацию, Введение, Перечень разделов, Заключение (Выводы) и Список литературы.*

Введение, как известно, должно содержать информацию о том, чему посвящена работа; *Перечень разделов* позволяет понять какие рассматриваются вопросы; *Заключение* — какие получены результаты; *Список литературы* — характеризует информированность автора, имеет самостоятельную ценность. Экстратекст (за вычетом *Списка литературы*) не должен занимать более 1–2-х стр., чтобы его можно было просмотреть достаточно быстро. Это важно, так как многочисленные исследования поведения Интернет-пользователей показывают, что пользователь покидает сайт, если за 20–30 с у него не возникает заинтересованности.

Суммируя сказанное, в качестве ответа на поставленный в заголовке этой заметки вопрос можно рекомендовать следующую схему действий.

1. Отредактируйте *Аннотацию, Введение и Заключение* так, чтобы они в совокупности на 1–2-х страницах представляли квинтэссенцию вашей работы: какое вы добыли новое знание, кому и зачем оно может быть полезно.

Аннотация, Введение и Заключение в экстратексте не обязательно должны быть точной копией из ранее опубликованной работы. Разумеется, если каких-то элементов экстратекста нет в статье, то их не нужно выдумывать. Ключевая задача, которую должен решить автор, состоит в том, чтобы заинтересовать читателя своими идеями, предложениями, результатами.

2. Переведите экстратекст на английский язык. Английский язык — латынь современной науки. С этим нельзя не считаться. Проблему перевода на английский язык не всей статьи, а лишь одной — двух страниц экстратекста нужно решить, если хотите максимизировать известность вашей работы и ее цитируемость.

3. Создайте двухстраничный сайт («мини-сайт»). На первую страницу загрузите экстратекст, на второй укажите, по какому адресу нужно обратиться и на каких условиях можно получить полный текст вашей работы.

4. Оптимизируйте мини-сайт экстратекста.

Примеры экстратекстов статей, опубликованных в журнале «Проблемы управления», можно посмотреть на сайтах <http://extratext.by.ru/> и <http://epchtein.ru/>

Мини-сайт электронной версии книги В. А. Трапезникова «Управление и научно-технический прогресс» находится по адресу: <http://trap-ipu.narod.ru>. Книга была напечатана в 1982 г. и практически забыта. Менее чем за год после публикации мини-сайта поступило (и продолжают поступать) свыше 100 обращений от различных научно-исследовательских и учебных институтов, предприятий, органов территориального управления, научных работников, преподавателей, аспирантов и даже студентов с просьбой прислать электронную версию упомянутой книги.

Теперь об интересах. Интересы авторов очевидны. Еще в начале 1980-х гг. Ю. Гарфилд обратил внимание на строгую корреляцию между цитируемостью, академической карьерой и такими формами признания научных заслуг, как присуждение почетных премий, включая Нобелевскую [6].

Менее очевидно, но не менее, а, может быть, более важно другое. Интернет-маркетинг экстратекстов открывает возможность создать широко доступное научно-информационное обеспечение, необходимое для ускоренного развития инновационной экономики, в том числе за счет реанимации огромного интеллектуального капитала, хранящегося на полках наших библиотек в виде опередивших свое время или недооцененных и незаслуженно забытых публикаций.

Напомним, *инновационной* называется экономика, основанная на практическом использовании новых знаний, добываемых фундаментальной наукой.

ЛИТЕРАТУРА

1. Lawrence W. On line or invisible? <<http://www.neci.com/>>
2. Rayburn S. and Bouton E.N. "If it's not on the Web, it doesn't exist at all": Electronic information resources — Myth and reality. — 1997. <<http://www.eso.org/genfac/libraries/lisa3/stevens-rayburns.html>>.
3. Энштейн В. Л. Антропоцентрическое информационное взаимодействие (вопросы терминологии) // Проблемы управления. — 2003. — № 1. <<http://epchtein.ru/AiV.htm>>.
4. Новая постиндустриальная волна на Западе / Под ред. В. Л. Иноземцева. — М.: Academia, - 1999.
5. Day R. A. How to write & publish a scientific paper. — N.-Y.: Oryx Press, 1998.
6. Маркусова В. А. Импакт-фактор российских журналов. <<http://www.ng.ru/science>>.

☎ (495) 334-89-80;

e-mail: epstein@ipu.ru;

caim: www.epchtein.ru





МОДУЛЬНАЯ РАСПРЕДЕЛЕННАЯ СИСТЕМА РЕШЕНИЯ ТИПОВЫХ РЕГИОНАЛЬНЫХ ЗАДАЧ (на примере гидрологических задач)

В. В. Лебедев, И. Н. Гансвинд, И. Н. Горохова, Е. И. Куприянова, В. А. Харитонов

Научный геоинформационный центр РАН, г. Москва

На основе анализа общей структуры социально-экономической системы региона и подсистемы окружающей среды предложена функциональная структура автоматизированной системы решения типовых региональных гидрологических задач.

Опыт решения многих гидрологических задач (их примеры перечислены на рисунке) показывает, что системный подход к их решению реализуется с помощью модульной распределенной системы, подсистемы которой логически интегрированы и связаны единой базой данных или, возможно, ее частями, размещенными на нескольких связанных телекоммуникационных средствах программно-вычислительных комплексах [1]. Распределенные системы соответствуют классу гидрологических задач в силу территориального распределения процессов формирования речного стока, характеристик водного баланса территорий, самих гидрологических объектов и др.

Для понимания типологических региональных характеристик рассмотрим строение региональной социально-экономической системы [2]. Социально-экономические системы регионов состоят из разноуровневых функционально-структурных подсистем, объединенных системообразующими отношениями, связями и общей целью развития. В их составе выделяются следующие подсистемы: демографическая, производственная, природная, социальная, рекреационная, производственно-инфраструктурная, социально-инфраструктурная, институционально-инфраструктурная, экологическая, духовная среда, ресурсная. Все эти подсистемы в своем развитии подчинены единой цели — удовлетворению экономических, социальных и духовных потребностей населения региона при сохранении устойчивости природной среды.

Общим для всех регионов является структура самой региональной социально-экономической системы. Так, в соответствии с характером связей внутри системы и, в частности, связи производственной подсистемы с природой выделяются природно-ресурсные отрасли, основывающиеся на использовании почвенно-земельных, водных, биологических, минерально-сырьевых ресурсов. В качестве профильных форм их организации выступают, соответственно, сельскохозяйственные, лесозаготовительные, гидроэнергетические, рыбопромысловые, горнодобывающие и перерабатывающие сырье предприятия.

Цели и задачи региональной политики (равно как ее формы и методы) разных государств могут существенно

различаться. Однако существуют общие, генерализованные цели и задачи, присущие региональной политике практически всех без исключения стран. В частности, это предотвращение загрязнения окружающей среды, экологическая защита регионов, оценка динамики изменения окружающей среды и ее прогнозирование и др.

Среди перечисленных региональных подсистем авторам наиболее близка подсистема окружающей природной среды, поэтому в настоящей работе рассматривается модульная распределенная система решения гидрологических задач, относящихся к задачам природной, точнее, природно-антропогенной среды. Предполагается, что эта система пригодна и для решения других природно-антропогенных задач (гидрогеологических, климатических, почвенных и др.). Гидрологические задачи многообразны, их ставят и решают для определения количества и качества воды в водных объектах, характеристик движения воды в руслах рек, прогнозирования расходов воды в их створах, выявления роли отдельных источников в питании рек, для определения процессов формирования паводков и передвижения их волн, формирования стока рек, термического и ледового режима рек, стока растворенных веществ и др.

К разработке модульной распределенной системы решения региональных гидрологических задач авторов привели исследования, связанные с решением задачи прогнозирования объемов стока половодья [3]. Модульность оказывается необходимой по ряду причин, основной из которых является резкий рост сложности взаимосвязей в системе с ростом числа ее элементов. При проектировании модульных систем должны быть обеспечены следующие основные свойства входящих в них модулей [1]:

— функциональность (модуль должен содержать функционально законченную и максимально независимую совокупность операций по обработке данных; обращение к модулю осуществляется как к единому целому, и значение вызываемого параметра отражает специфику функций модуля);

— связность (модуль реализует совокупность взаимосвязанных функций, работающих с одними и теми же данными);

— алгоритмичность (функции модуля группируются на алгоритмической основе);

— последовательность (модуль осуществляет несколько функций, причем выходные результаты одной функции являются входными для другой);

— однородность (в модуль объединяются однородные по своему функциональному назначению процедуры).

Основное содержание настоящего этапа исследования состояло в определении структуры разрабатываемой модульной системы решения общих региональных задач (см. рисунок) и описании иерархически верхнего уровня модулей — подсистем, с учетом перечисленных свойств модулей. Система распадается на 8 подсистем, названия которых подчеркивают их функциональную суть: моделирование, реализация моделей, создание слоев информации, синтез и др. Связность модулей показана на схеме горизонтальными линиями. Процесс решения задач начинается с создания (или подбора среди разработанных) расчетных или прогнозных моделей (блок 4), которые реализуются с помощью математического аппарата (блок 5); блоки 6–8 решают подзадачи создания аналитической информации; блоки 9–10 — вырабатывают выходную информацию и таким образом осуществляют достижение цели исследования.

База данных. Входная информация для гидрологических моделей, представленная этой подсистемой, характеризует природно-антропогенные факторы, определяющие формирование гидрологических процессов и явлений, а также представляет собой данные о гидрологических процессах, объектах, явлениях: атмосферные осадки, испарение, уровни залегания грунтовых вод, характеристики подстилающей поверхности — литологический и механический состав, влажность, глубина промерзания почвогрунтов, растительный покров (его виды и состояние), уклоны земной поверхности, распаханность склонов, залесенность; уровенный, гидрохимический, температурный, ледовый режимы объектов, скорости течения.

Подсистема «*Моделирование режимов водных объектов, процессов и явлений*» содержит набор моделей, соответствующих конкретно поставленной задаче. Так, разработанный авторами, совместно с учеными Гидрометцентра РФ подход к решению задачи прогнозирования объемов паводкового стока, основывается на модели многопараметрического картографического компьютерного синтеза.

Содержание подсистемы «*Подбор математического аппарата для реализации моделей*» вытекает из анализа класса гидрологических задач и накопленного опыта их решений. Наиболее востребован аппарат математической статистики и теории вероятностей (для определения зависимости речного стока от параметров водосборной территории, определения временной изменчивости гидрологических характеристик водных объектов и формирующих их факторов и др.). В задачах гидравлического содержания (передвижение паводковой волны в русле реки, формирование уровней грунтовых вод, гидравлически связанных с водохранилищем в процессе его функционирования и др.) применяется теория дифференциальных уравнений, причем однотипные уравнения могут использоваться для разных регионов и объектов исследования. Меняются лишь значения параметров.

Подсистема «*Создание слоев дистанционной информации*» выделена по отношению к любой другой информа-

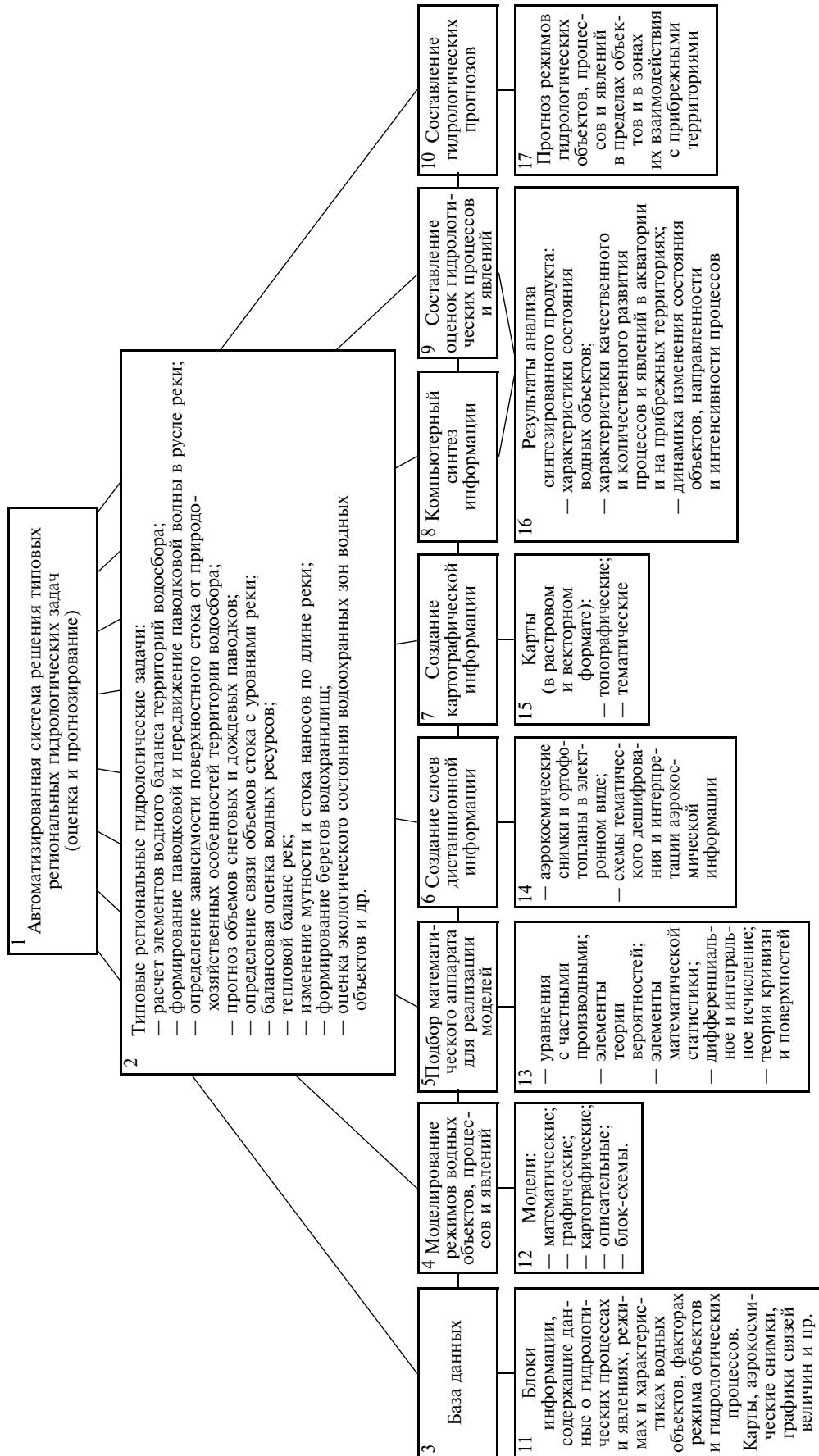
ции, содержащейся в базе данных. Ее актуальность состоит в том, что в стране практически разрушена сеть гидрологических наземных наблюдений, поэтому аэрокосмические и другие дистанционно получаемые виды информации приобретают особую ценность, так как в основном только они формируют обновляемый банк данных о состоянии водных объектов, процессах в акватории и на прибрежных территориях. Помимо этого, подсистема содержит алгоритмы, реализующие методы обработки аэрокосмических и других видов съемки с применением геоинформационных технологий.

Подсистема «*Создание картографической информации*» отражает специфику разрабатываемой системы — распределенность физических элементов решаемых задач в пространстве, заданную картографической моделью. Этот модуль реализует технику создания карт с использованием аэрокосмических снимков (создание растровых карт, сшивка частей изображения объекта в единое полотно, его привязка к топографической основе) [4]. Лишь после такого комплекса процедур тематические карты могут участвовать в компьютерном синтезе.

Подсистема «*Компьютерный синтез информации*» является ключевой, так как она обеспечивает достижение конечного результата. Анализ направленности природных и антропогенных процессов в их количественном выражении достигается построением матриц взаимных переходов классов объектов (процессов, явлений). В основу этого подхода положены теоретические разработки и реализация матричного анализа и прогноза динамики сложных экологических систем по аэро- и космическим снимкам [5].

На базе автоматизированной технологии синтеза аэрокосмической, наземной и картографической информации можно создавать выходные картографические документы различной тематической направленности:

- трансформации экосистем — карты взаимных переходов основных компонентов экосистем за два-три исследуемых срока (пашни, луга, леса, водоемы, населенные пункты и др.);
- ландшафтно-экологические (динамики загрязнения почв и вод радионуклидами и тяжелыми металлами-токсикантами на ландшафтной и (или) топографической основе);
- ландшафтно-геохимические (динамики засоления почвы токсичными солями на ландшафтной и (или) топографической основе);
- факторов риска (с отображением на единой картографической основе процессов, приводящих к аварийным разрывам магистральных и промысловых газо- и нефтепроводов и других продуктопроводов (геодинамических, гидрогеологических, гидродинамических, геокриологических и других факторов);
- социально-экологические (динамика заболеваемости населения на фоне антропогенного загрязнения окружающей среды) и многие другие тематические карты, отражающие пространственно-временные изменения экосистем и их компонентов, природных и антропогенных процессов и явлений: изменения видового состава и состояния растительности, изменения глубины залегания и минерализации грунтовых вод, затопления, подтопления и заболачивания территорий, загрязнения и зарастания акваторий, трансформации береговых линий, схода снегового и ледового покрова и др.



Структурная схема модульной распределенной автоматизированной системы решения типовых региональных задач

Синтезированные компьютерные карты стокоформирующих факторов служат геоинформационной основой получения оценок гидрологических процессов и явлений и составления научно-обоснованных гидрологических прогнозов.

Подсистема «Составление оценок гидрологических процессов и явлений» отвечает на главный вопрос задачи и на основе синтеза информации дает искомые оценки.

В подсистеме «Составление гидрологических прогнозов» для получения гидрологических прогнозов должны быть представлены материалы, содержащие характеристики, необходимые для прогностической модели (состояния подстилающей поверхности: глубины промерзания почвогрунтов, их увлажнения, изменения характера землепользования; погодных условий, использования водных ресурсов и др.) Сама модель, как правило, совпадает с оценочной, но в уравнениях меняются параметры.

В заключение отметим, что в работах отечественных и зарубежных ученых подчеркивается наличие существенного сходства между региональными проблемами и задачами в самых различных по специфике регионах, что позволяет говорить о типовых региональных задачах [6]. Использование в разработанной системе дистанционной информации и геоинформационной технологии компьютерного картографического синтеза служит оптимальным способом поиска решений в условиях огра-

ниченности систематических наблюдений за гидрологическими объектами.

ЛИТЕРАТУРА

1. Методы анализа и синтеза модульных информационно-управляющих систем / Н. А. Кузнецов, В. В. Кульба, С. С. Ковалевский, С. А. Косяченко. — М.: Наука, 2002. — 798 с.
2. Гладкий Ю. Н., Чистобаев А. И. Регионоведение. — М.: Гардарики, 2000. — 382 с.
3. Опыт применения геоинформационных технологий для прогнозирования стока половодья / В. А. Бельчиков, С. В. Борщ, В. Ю. Вознесенский и др. // Метеорология и гидрология. — 2001. — № 10.
4. Проблемы создания региональных геоинформационных комплексов и опыт решения прикладных задач на основе аэрокосмической информации / Под ред. В. В. Лебедева. — М.: Наука, 2002. — 239 с.
5. Харитонов В. А. Новый метод картографического представления информации о динамике экосистем / В кн.: Аэрокосмические методы исследований при мелиоративном и водохозяйственном строительстве. — М.: Союзводпроект, 1990. — С. 92—99.
6. Гранберг А. Г. Основы региональной экономики. Европейская комиссия. — М., 2000. — 494 с.

☎ (495) 202-11-49;

e-mail: ngic@yandex.ru



УДК 658.5;334.012.64

ПОДХОД К СТРАТЕГИЧЕСКОМУ УПРАВЛЕНИЮ МАЛЫМ ИННОВАЦИОННЫМ ПРЕДПРИЯТИЕМ

С. Е. Шкуратов

Обнинский филиал Государственного университета управления

Выделены основные подходы к управлению малым инновационным предприятием и предложена модель стратегического управления, сформированная с учетом особенностей малых инновационных предприятий. В основу выбора стратегии заложен анализ степени развития продукта, анализ степени доступности ресурсов, а также оценка степени развития бизнес-компонента предприятия. Отмечено, что комбинации данных оценок позволяют сформировать широкую совокупность альтернатив развития предприятия — от рекомендации по самостоятельному продолжению работ и выходу на рынок до рекомендации поисков возможностей вступления в стратегический альянс с крупным предприятием.

Переход к инновационному развитию экономики возможен только при устойчивом развитии сектора малых инновационных предприятий. В связи с этим вызывают беспокойство стабилизация и даже сокращение числа малых предприятий, а также отсутствие ожидае-

мых сдвигов в отраслевой структуре, в первую очередь, в направлении повышения удельного веса малых инновационных предприятий. Данная тенденция свидетельствует не только о недостаточной эффективности системы поддержки, но и в большой степени о недостаточном



методическом обеспечении стратегического управления, учитывающего особенности малых инновационных предприятий.

Первоначальный анализ инструментария, закладываемого в основу решения задач на различных этапах процесса стратегического управления, позволил выявить, что инструментарий стратегического анализа, в особенности анализа портфеля продуктов и выбора стратегических альтернатив дальнейшего развития предприятия, наиболее чувствителен к специфике малого инновационного предприятия.

Основным инструментом анализа и управления является многомерная «матрица», позволяющая классифицировать и кластеризовать управляющие объекты в соответствии со значениями ключевых параметров, традиционно называемых осями матрицы.

Сегодня для осуществления работ на данных этапах предлагается ряд моделей: BCG, GE/McKinsey, ADL/LC, Shell/DPM, Hofer/Schendel [1–3]. Далее показаны недостатки наиболее известных моделей.

В модели BCG приняты допущения, которые не бесспорны, например: чем выше темпы роста, тем больше возможности развития; чем больше доля рынка, тем сильнее компания; чем больше доля рынка, тем больше прибыль; а также предположение о том, что бизнес-области фирмы независимы. Затруднено ее применение в отраслях, где невысок уровень конкуренции либо незначительны объемы производства. Кроме того, модель требует соизмеримости стратегических перспектив по показателям темпов роста и доли рынка всех сравниваемых бизнес-единиц. Ось «скорость роста объема продаж» как главной характеристики привлекательности рынка (что тоже сомнительно), не позволит нанести на неё позиции малого инновационного предприятия: о каких продажах и темпах роста можно говорить на ранних этапах развития, если компания только создала опытный образец, либо находится вовсе на этапе научно-исследовательских работ? В модели не предусматривается рассмотрение большинства факторов неопределенности внешней среды, которые могут изменить траекторию прошлой динамики роста организации [4].

Модель Hofer/Schendel предполагает, что все виды бизнеса связаны между собой и что их жизненные циклы схожи. Принято допущение, что любой рынок имеет жизненный цикл, соответствующий «классической» кривой, хотя на самом деле форма цикла может значительно варьироваться. Основные стратегии, предлагаемые в модели, соответствуют сразу нескольким позициям на матрице. Стратегии достаточно схематичны.

Модель GE/McKinsey дает только самые общие направления для формулирования стратегий. Сосредотачивая внимание в основном на краткосрочном эффекте от инвестиций, сложно оценить, каковы будут долгосрочные перспективы от вложения средств. Выбор переменных для анализа очень условен, трудно оценить, какие из переменных более значимы. Модель не позволяет получить ясного ответа на вопрос о том, как следует перестроить структуру бизнес-портфеля корпорации. Конкурентный статус экстраполируется на стратегический период: он устанавливается на текущий период и без существенных изменений переносится на будущее;

на практике все может значительно меняться. Процесс стратегического выбора в данной модели сводится к тому, что организация предугадывает будущее, которое должно состояться как бы без ее участия [5].

Рассмотрение лишь некоторых недостатков позволяет сделать вывод, что применение большинства моделей в условиях малого инновационного предприятия ограничено. Задачи, на решение которых они направлены, зачастую не актуальны для малого инновационного предприятия либо применение данных моделей в условиях малого инновационного предприятия не позволяет решать задачи, на которые они рассчитаны. Таким образом, механический перенос методологии, которая в большинстве случаев разрабатывалась для крупных предприятий, на малые инновационные предприятия не приводит к получению ожидаемого эффекта.

В отличие от крупных компаний, имеющих широкий ассортимент продуктов, малое инновационное предприятие чаще всего имеет одну базовую технологию и единичные направления, идеи о развитии данной технологии. Считать «полипродуктностью» наличие в малом инновационном предприятии результатов НИР, ОКР и т. д. в понимании авторов рассмотренных моделей нельзя [6]. Кроме того, жизненный цикл малого инновационного предприятия тесным образом связан с жизненным циклом продукта, причем с его ранними этапами, с многоэтапным периодом разработки, который не выделяется в классической совокупности циклов. Исходя из этого, стратегический анализ и один из его компонентов — анализ портфеля продуктов — должны иметь разные акценты в условиях малого и крупного бизнеса. Если для крупного предприятия необходимо обеспечение сбалансированного портфеля продуктов, сбалансированного портфеля объектов инвестирования, то для малого инновационного предприятия важна оценка перспективности развития технологии, оценка различных альтернатив по ее развитию и оценка объема необходимых или достаточных вложений для каждого сценария развития.

Можно согласиться с тем, что в разрабатываемом инструментарии должны учитываться заложенные в основу существующего инструментария существенные параметры (жизненный цикл продукта, предприятия, отрасли, параметры конкурентоспособности предприятия, привлекательности внешней среды, отрасли и т. д.). В то же время, их измерение и трансляция в инструментарий должна учитывать особенности малых инновационных предприятий.

Возможным способом учета перечисленных характеристик может быть построение трехмерной матрицы выбора стратегии, в основу построения осей которых закладываются различные оценки:

- степени развития продуктового компонента предприятия, что фактически связано с оценкой жизненного цикла бизнес-идеи и (или) технической идеи, заложенной в ее основу;
- жесткости «бизнес-фундамента» компании, которая определяется степенью развития внутренних элементов компании, формирующих потенциал предприятия, его конкурентоспособность;
- доступности и достаточности ресурсов как ключевой характеристики привлекательности внешней среды.

В таблице представлены основные особенности малых инновационных предприятий, определившие характеристики и оси предложенной матрицы.

Использование данной матрицы позволит отнести инновационные предприятия к определенным классам,

Особенности малых инновационных предприятий и обусловленные ими оси матрицы

Особенность МИП	Характеристика (ось) матрицы
Постоянный дефицит ресурсов	Доступность ресурсов
Необходимость различной поддержки на разных этапах развития	Доступность ресурсов. Состояние продукта
Непредсказуемость рыночной инфраструктуры	Бизнес-фундамент
Малое количество информации о продажах и т. п. Возможно измерение только потенциала	То же
Специфический кадровый состав, особенности личности предпринимателя	Простота использования, простота показателей
Высокорисковая деятельность	Состояние продукта, обуславливающее пошаговое прохождение стратегического анализа
Монопродуктность	Доступность ресурсов. Состояние продукта
Специфические способы создания предприятия	Бизнес-фундамент. Простота использования, простота показателей
Неравномерность темпов развития предприятия	Бизнес-фундамент. Доступность ресурсов. Состояние продукта
Преобладание нематериальных активов в стоимости предприятия	Бизнес-фундамент
Небольшой возраст предприятия	Бизнес-фундамент
Эволюционное развитие продукта	Состояние продукта
Возможность продажи результатов инновационной деятельности	Доступность ресурсов. Состояние продукта
Уязвимость перед силовым давлением	Доступность ресурсов

для которых известны базовые стратегии управления и тем самым получить реальную оценку текущего положения малого инновационного предприятия, а также выработать альтернативы дальнейшего развития бизнес-идеи и самого предприятия.

Оригинальная совокупность осей дает направления в стратегическом анализе в целом, а матричная структура и совокупность позиций дают поле для выбора стратегии: модель предусматривает 45 возможных комбинаций — стратегических позиций:

- продолжение бизнеса самостоятельно: с собственным производством, с использованием субконтрактинга, аутсорсинга по основным производственным операциям и др.;
- продажа лицензии на разработку: полная (эксклюзивная) или частичная;
- совместное предприятие: с малым и (или) крупным предприятием;
- развитие еще одной идеи («снизиться» по оси «доступность и достаточность ресурсов» при сохранении имеющейся сильного бизнес-компонента и «взять на борт» еще одну идею для раскрутки);
- отложить в архив и др.

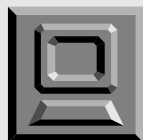
Использование предложенной матрицы возможно как начинающими свою деятельность компаниями, так и компаниями, которые уже осуществляют деятельность в течение определенного периода времени. Данная модель может быть с некоторыми допущениями использована с целью управления внутрикорпоративными инновационными группами.

ЛИТЕРАТУРА

1. Ансофф И. Новая корпоративная стратегия. — СПб.: Питерком, 1996. — 416 с.
2. Ефремов В. С. Стратегия бизнеса. Концепции и методы планирования: Учебное пособие. — М.: Финпресс, 1998. — 192 с.
3. Минцберг Г. Школа стратегии — Стратегическое сафари: экскурсии по дебрям стратегий менеджмента. — СПб.: Питер, 2000. — 330 с.
4. Белкина Е. Л. Стратегический подход на начальном этапе жизнедеятельности малого предприятия / Дисс. ... канд. экон. наук. Новосибирск: 2003. — 167 с.
5. Анненков М. Е. Формирование стратегии конкуренции в условиях переходной экономики // Менеджмент в России и за рубежом. — 1997. — № 2. — С. 12—25.
6. Управление инновациями: 17-модульная программа для менеджеров / В. Н. Гунин, В. П. Баранчев, В. А. Устинов, С. Ю. Ляпина. — М.: ИНФРА-М, 2000. — Т. 7.

☎ (48439)4-41-28;

sshkuratov@mail.ru



Для вас, подписчики журнала "Проблемы управления"!

С 2006 г. электронная версия журнала «Проблемы управления» (включая архивные выпуски) размещена на платформе Научной электронной библиотеки. Условия доступа к массиву выложены на сайте <www.elibrary.ru>. Сотрудникам ИПУ РАН предоставлена возможность открытого доступа.

Мы рады предоставить Вам комфортные условия для работы с нашим журналом, используя современные технологии поиска научной информации, обработки и сохранения полученных материалов в электронной форме.

Научная электронная библиотека eLIBRARY.RU



ОПРЕДЕЛЕНИЕ ОПТИМАЛЬНОГО ВРЕМЕНИ ФУНКЦИОНИРОВАНИЯ ИНФОРМАЦИОННОЙ СИСТЕМЫ

Н. В. Лисицын, В. К. Викторов

ООО «Наука, технология, информатика, контроль», г. Санкт-Петербург

Рассмотрена задача определения интервала времени, по истечении которого существующая информационная система в силу выработки своих ресурсов должна быть заменена на новую. Предложен алгоритм ее решения, основанный на принципе оптимальности Беллмана. Приведены примеры.

Сегодня любое промышленное предприятие не обходится без информационных систем (ИС), предназначенных для сбора и обработки технологической, экономической, транспортной, материально-технической и другого рода информации и обеспечения быстрого доступа к ней. Благодаря ИС предприятия получают дополнительный доход [1]. Обозначим этот доход $d(n, t)$, где n — номер момента времени начала функционирования ИС, t — «возраст» ИС. Отметим, что n, t могут принимать значения из натурального ряда чисел ($n, t = 1, 2, \dots, N$). Время может измеряться числом определенных промежутков времени — месяцев, лет и др. Очевидны следующие свойства функции $d(n, t)$. Она возрастает по n по мере появления со временем все более совершенных ИС и убывает по t в результате морального и физического их старения. Более совершенные ИС стоят дороже, поэтому функция стоимости ИС $c(n)$ монотонно возрастающая. Что касается конкретного вида функций $d(n, t)$ и $c(n)$, то они могут быть получены в результате статистической обработки экспериментальных данных и представлены в виде матриц (см. Пример).

Возникает задача определения времени, когда следует внедрить новую ИС взамен старой, чтобы получить максимальный доход от их применения.

Процесс принятия решения о замене ИС является многошаговым марковским процессом, для решения указанной задачи может быть применен принцип оптимальности Беллмана [2]. Обозначим $f(n, t)$ — максимальный суммарный доход от работы ИС возраста t за время $N - n + 1$, начиная с момента времени n и до фиксированного срока N окончания работы ИС. Составим следующие рекуррентные функциональные уравнения:

$$f(n, t) = \max \begin{cases} d(n, 0) - c(n) + f(n+1, 1) \\ d(n-t, t) + f(n+1, t+1) \end{cases}, \quad (1)$$

$$\begin{aligned} n, t &= 1, \dots, N, \\ f(N+1, *) &= 0. \end{aligned} \quad (2)$$

На каждом из N шагов может быть принято два решения: внедрить новую ИС — верхняя строка в уравнении (1), сохранить старую ИС — нижняя строка. На каждом шаге выбирается максимальное из двух строк. В обеих строках первые слагаемые — это доход от ИС на n -м шаге, а последующие слагаемые — это доход от работы ИС на последующих $N - n$ шагах: $n+1, \dots, N$.

Условие (2) означает, что за пределами времени N ИС перестает работать (можно сказать, что N — это срок службы ИС) и доход от нее равен нулю.

Алгоритм решения задачи на основе уравнений (1) может быть построен следующим образом. Вначале при $n = N$ рассчитывается величина

$$f(N, t) = \max \begin{cases} d(N, 0) - c(N) \\ d(N-t, t) \end{cases}, \quad (3)$$

$$t = 1, \dots, N.$$

так как $f(N+1, t) = 0$ и $f(N+1, t+1) = 0$.

Затем рассчитывается величина

$$f(N-1, t) = \max \begin{cases} d(N-1, 0) - c(N-1) + f(N, 1) \\ d(N-1-t, t) + f(N, t+1) \end{cases}, \quad (4)$$

$$t = 1, \dots, N-1$$

и так далее:

$$f(N-i, t) = \max \begin{cases} d(N-i, 0) - c(N-i) + f(N-i+1, 1) \\ d(N-i-t, t) + f(N-i+1, t+1) \end{cases}, \quad (5)$$

$$i = 2, \dots, N-1; t = 1, \dots, N-i.$$

При $i = N-1$ получается значение $f(1, 1)$ — максимальный доход от использования старой и новой ИС при оптимальной стратегии замены систем. В процессе работы алгоритма (3)—(5) определяется и время замены ИС.

Пример. Проиллюстрируем работу алгоритма. Пусть матрица $D = \{d(n, t)\}$ задана следующим образом ($N = 5$):

n	t					
	0	1	2	3	4	5
0	—	60	45	33	23	15
1	65	50	40	32	26	22
2	73	60	50	42	38	—
3	80	70	63	60	—	—
4	84	76	70	—	—	—
5	85	80	—	—	—	—

В матрице отражены упомянутые свойства функции $d(n, t)$: числа растут по строкам и убывают по столбцам.

Задана также таблица значений функции стоимости $c(n)$:

n	1	2	3	4	5
c	100	110	118	124	126

В результате применения алгоритма (3)—(5) к задаче (6), (7) получается решение для $\{f(n, t)\} = F$ в виде матрицы F :

n	t				
	1	2	3	4	5
1	200	—	—	—	—
2	148	140	—	—	—
③	152	98	95	—	—
4	133	92	58	38	—
5	76	63	42	26	15

В матрице (8) кружком отмечен момент времени 3 — момент замены ИС. Таким образом, оптимальная стратегия состоит в том, что первые два отрезка времени работает старая ИС, в начале третьего отрезка времени внедряется новая система и она работает 3-, 4- и 5-й отрезки времени. Доход от такой стратегии складывается следующим образом:

$$f(1, 1) = d(0, 1) + d(0, 2) + d(3, 0) - c(3) + d(3, 1) + d(3, 2) = 60 + 45 + 80 - 118 + 70 + 63 = 200.$$

Если бы работала старая ИС — первая строка матрицы (6), тогда суммарный доход был бы:

$$\sum_{i=1}^5 d(0, i) = 60 + 45 + 33 + 23 + 15 = 176.$$

Отсюда следует вывод о том, что на третьем отрезке времени ИС необходимо заменить на новую. ♦

С помощью компьютерной программы, реализующей алгоритм (3)—(5), можно решить, кроме основной задачи определения времени t_c замены ИС, и ряд других технико-экономических задач.

В качестве примера рассмотрим задачу определения политики замены ИС в зависимости от экономической конъюнктуры — от колебаний стоимости новых ИС.

Для удобства предположим, что доход от применения ИС задан не в виде матрицы (6), а в виде функции $d(n, t) = 70 + 5n - 10t$, а $c(n)$ задается в виде таблицы.

В последних двух столбцах таблицы приведены решения исходной задачи определения момента t_c и максимальный доход $f(1, 1)$, полученные по алгоритму (3)—(5).

Колебаниям подвержена цена ИС, разработанной на 2- и 3-м отрезках времени. Из сравнения первой и второй строк таблицы следует, что цена 105 является критической: достаточно ей вырасти на 1 до 106, как меняется решение о времени замены t_c со 2-го отрезка времени на 3-й.

Если же падает цена $c(3)$ с 120 до 115, то даже при $c(2) = 105$ замена на $t_c = 2$ — невыгодна, а максимальный доход достигается при $t_c = 3$.

Слагаемые суммарного дохода для трех вариантов из таблицы:

$$\begin{aligned} \text{1-й вариант: } & d(0, 1) + d(2, 0) - c(2) + d(2, 1) + \\ & + d(2, 2) + d(2, 3) = 215; \end{aligned}$$

Стоимость ИС $c(n)$

№ варианта	n						t_c	$f(1, 1)$
	1	2	3	4	5			
1	90	105	120	130	137	2	2	215
2	90	106	120	130	137	3	3	215
3	90	105	115	130	137	3	3	220

$$\begin{aligned} \text{2-й вариант: } & d(0, 1) + d(0, 2) + d(3, 0) - c(3) + \\ & + d(3, 1) + d(3, 2) = 215; \end{aligned}$$

$$\begin{aligned} \text{3-й вариант: } & d(0, 1) + d(0, 2) + d(3, 0) - c(3) + \\ & + d(3, 1) + d(3, 2) = 220. \end{aligned}$$

Подобного рода результаты могут быть получены предлагаемым методом и при других рыночных вариациях в стоимостях ИС $c(n)$ и в доходах от их применения $d(n, t)$.

В работе [3] предложен метод оценки экономической эффективности применения ИС. Его идея состоит в том, что реальный выигрыш от применения ИС заключается в более точном прогнозировании прибыли предприятия по переработки нефти. Идеальная ИС дает безошибочный прогноз. Можно ввести коэффициент ошибки прогноза K . Для идеальной системы $K = 1$. Чем система хуже, тем меньше K . Прогноз прибыли предлагается вычислять следующим образом: $d = 360KpP$, где d — прибыль в год, млн долл./г.; p — средняя прибыль от переработки одного барреля, долл./баррель; P — производительность предприятия по сырой нефти, млн баррелей/сут.; $K \in [0, 1]$.

Предположим, что коэффициент ошибки прогноза уменьшается при увеличении n и t следующим образом: $K(t, n) = K_0 + K_1t - K_2n$.

Годовая прибыль, таким образом, будет измеряться как

$$d(n, t) = 360K(n, t)pP = K(n, t)R.$$

Для конкретных значений $K_0 = 0,85$, $K_1 = 0,01$, $K_2 = 0,03$, $p = 2$ долл./баррель, $P = 0,1$ млн баррелей/сут., $R = 72$ долл./год и для $c(n) = (20 + n)$ млн долл. была поставлена задача: определить, насколько должна снизиться стоимость $c(4)$ ИС в результате технической революции на четвертом интервале времени, чтобы было выгодно произвести замену информационной системы на 4-м интервале, а не на 3-м, как следует из решения задачи для $c(4) = 24$.

С помощью упомянутой компьютерной программы удалось получить значение $c(4) = 19,1$, удовлетворяющее условиям задачи. Решение без замены: $t_c = 0$, $d = 273,6$ млн долл./год; решение при $c(4) = 24$: $t_c = 3$, $d = 277,52$ млн долл./год, решение при $c(4) = 19,1$: $t_c = 4$, $d = 277,54$ млн долл./год. Выигрыши составляют 3,94 млн долл./год и 0,02 млн долл./год, соответственно.

Предложенный подход к исследованию эффективности информационных систем отличается от традиционных, основанных на анализе их функций, и может быть полезен для оценки работоспособности крупных автоматизированных комплексов.

ЛИТЕРАТУРА

1. Садчиков И. А., Сомов В. Е. Киришинефтеоргсинтез — от ПО к... ПО. — СПб.: Химия, — 1997. — 271 с.
2. Беллман Р., Дрейфус С. Прикладные задачи динамического программирования. — М.: Наука, — 1965. — 475 с.
3. Уайт Д. К. Определение истинных экономических преимуществ современной информационной системы на крупном НПЗ // Нефтегазовые технологии. — 2005. — № 5. — С. 73—76.

☎ (812) 550-40-79

e-mail: lisitsyn@ntic.24





АДАПТИВНЫЕ МЕХАНИЗМЫ ОЦЕНКИ И КЛАССИФИКАЦИИ ДАЛЬНОВИДНЫХ АКТИВНЫХ СИСТЕМ

Н. Ф. Сирина⁽¹⁾, В. В. Цыганов⁽²⁾

⁽¹⁾ Уральский государственный университет путей сообщения, г. Екатеринбург;

⁽²⁾ Институт проблем управления им. В. А. Трапезникова РАН, г. Москва

Поставлена задача оптимального синтеза комплексного адаптивного механизма, основанного на количественной и качественной оценках результатов функционирования дальновидной активной системы. Доказано, что для прогрессивности комплексного адаптивного механизма оценки и классификации достаточно прогрессивности адаптивного механизма оценки.

Для эффективного управления развитием организации необходимо регламентировать процессы ее адаптации к изменениям. Механизм адаптации организации (или, кратко, адаптивный механизм) формирует ее управляющий орган (Центр). На практике широко применяются адаптивные механизмы количественной и качественной оценок результатов работы элементов организации [1, 2]. В теоретических исследованиях обычно рассматривается двухуровневая модель организации, на верхнем уровне которой находится Центр, а на нижнем — дальновидный элемент.

Принятые сокращения:

АМК — адаптивный механизм классификации;
АМО — адаптивный механизм оценки;
АМОК — адаптивный механизм оценки и классификации;
ДЭ — дальновидный элемент.

Традиционно, особое внимание уделяется определению условий прогрессивности адаптивного механизма, обеспечивающего раскрытие потенциала организации. Например, в работе [1] для получения количественной оценки результатов работы и стимулирования применяются адаптивные механизмы оценки, для которых найдены условия прогрессивности. В работах [1, 2] для получения качественной оценки (класса) применяются адаптивные механизмы обучения классификации, для которых также определены условия прогрессивности. На практике управляющие органы одновременно используют как количественные, так и качественные

оценки результатов работы элементов системы. В данной работе исследуются условия прогрессивности комплексного адаптивного механизма функционирования двухуровневой организации, в котором одновременно проводятся как количественная, так и качественная оценки.

Обозначим состояние ДЭ в периоде t как y_t . Это состояние ограничено: $y_t \leq p_t$, $t = \overline{1, T}$, где p_t — потенциал ДЭ в периоде t , зависящий от случайных факторов; $t = \overline{1, T}$, T — число рассматриваемых периодов. Значение p_t в периоде t известно ДЭ, но не известно Центру ($p_t \in P$, $t = \overline{1, T}$). Для эффективного использования потенциала, осуществляется количественная и качественная оценки работы ДЭ.

Адаптивный механизм оценки формирует количественную оценку ДЭ путем определения степени выполнения адаптивных нормативов, регламентирующих его работу. Именно, в АМО $\Sigma_E = (X, E)$ на основе текущего норматива x_t и состояния ДЭ y_t формируется адаптивный норматив оценки x_{t+1} на следующий период $(t+1)x_{t+1} = X(x_t, y_t)$, где $X(x_t, y_t)$ — процедура адаптивного нормирования $t = \overline{1, T}$, $x_1 = x^1$. Состояние y_t сопоставляется с нормативом x_t и определяется количественная оценка ДЭ $e_t = E(x_t, y_t)$, где E — процедура оценивания. Если дальновидный элемент заинтересован в увеличении текущей и будущей оценок, то его целевая

функция зависит от вектора состояний $\bar{y} = (y_1, \dots, y_T)$ и определяется по формуле:

$$V_E(\bar{y}) = \sum_{t=1}^T k^{t-1} E(x_t, y_t), \quad (1)$$

где k — коэффициент дисконтирования, используемый для приведения будущих оценок к текущему периоду.

Предположим, что классификация заключается в отнесении ДЭ к одному из двух классов. Тогда АМК формирует качественную оценку (класс) ДЭ, путем сопоставления полученной в АМО оценки e_t с адаптивной нормой n_t . Формально АМК $\Sigma_R = (N, R)$, где N — процедура нормирования, R — процедура классификации. С помощью процедуры нормирования $n_{t+1} = N(n_t, e_t)$ формируется адаптивная норма классификации n_t . Процедура классификации $r_t = R(n_t, e_t)$ определяет класс r_t ДЭ

$$r_t = R(n_t, e_t) = \begin{cases} 1 & \text{при } e_t \geq n_t, \\ 0 & \text{при } e_t < n_t. \end{cases} \quad (2)$$

Рассмотрим процедуру нормирования $n_{t+1} = N(n_t, e_t)$. Обозначим через Δ_1 и Δ_2 некоторое разбиение множества состояний ДЭ Δ , $\bigcup_{k=1}^2 \Delta_k = \Delta$. Задача Центра заключается в классификации состояний ДЭ путем отнесения его к множеству Δ_1 или Δ_2 . При $y_t < x_t$ состояние ДЭ относится к классу «удовлетворительное» (в этом случае

ДЭ можно поощрить), в противном случае — к классу «неудовлетворительное» (и тогда ДЭ можно наказать). Каждое из этих решений связано для Центра с определенным риском. В первом случае его потери F_1 возрастают при увеличении состояния ДЭ y_t (например, из-за занижения поощрения). Во втором случае потери F_2 возрастают с уменьшением состояния y_t (из-за завышения поощрения). Как показано в работе [2], настраиваемая норма классификации минимизирует средний риск, если процедура нормирования имеет вид:

$$n_{t+1} = N(n_t, e_t) = \begin{cases} n_t + \gamma & \text{при } y_t < n_t, \\ n_t - \gamma u & \text{при } y_t \geq n_t, \end{cases} \quad (3)$$

где γ — шаг адаптации, $\gamma > 0$, u — параметр процедуры нормирования, $u > 0$. Если ДЭ заинтересован в увеличении текущего и будущих классов, то его целевая функция определяется по формуле:

$$V_R(\bar{y}) = \sum_{t=1}^T k^{t-1} R(n_t, y_t). \quad (4)$$

Адаптивный механизм оценки и классификации позволяет последовательно определять количественную и качественную оценки ДЭ (см. рисунок). В АМОК состояние y_t сопоставляется с нормативом x_t и определяется оценка ДЭ e_t . Далее на основе оценки e_t корректируется норма классификации n_{t+1} , используемая для определения ранга r_{t+1} . Адаптивность АМОК обеспечивается непрерывной настройкой нормативов оценивания и норм классификации.

Если ДЭ заинтересован в увеличении текущих и будущих оценок и классов, то его целевая функция определяется по формуле:

$$W(\bar{y}) = V_E(\bar{y}) + V_R(\bar{y}) = \sum_{t=1}^T k^{t-1} [E(x_t, y_t) + R(n_t, y_t)]. \quad (5)$$

Задача синтеза АМОК заключается в выборе совокупности процедур, обеспечивающих состояния ДЭ, предпочтительные для Центра. В общем случае задача оптимального синтеза АМОК $\Sigma = \{\Sigma_E, \Sigma_R\}$ в условиях неопределенности относительно потенциала ДЭ имеет вид

$$\min_{p_t \in P, t=1, T} \min_{\bar{y} \in G(\Sigma, p)} \Psi(\bar{x}, \bar{n}, \bar{y}) \xrightarrow{\Sigma} \max, \quad (6)$$

где $G(\Sigma, p) = \arg \max_{\bar{y}} W(\bar{y})$ — множество оптимальных состояний ДЭ y_t^* ; $\bar{x} = (x_1, \dots, x_T)$ — совокупность адаптивных нормативов оценки; $\bar{n} = (n_1, \dots, n_T)$ — совокупность адаптивных норм на весь срок дальновидности T . Предположим, что целевая функция Центра монотонно возрастает с увеличением выхода ДЭ: $\Psi(\dots, y_{1T}, \dots) \leq \Psi(\dots, y_{1T+1}, \dots)$

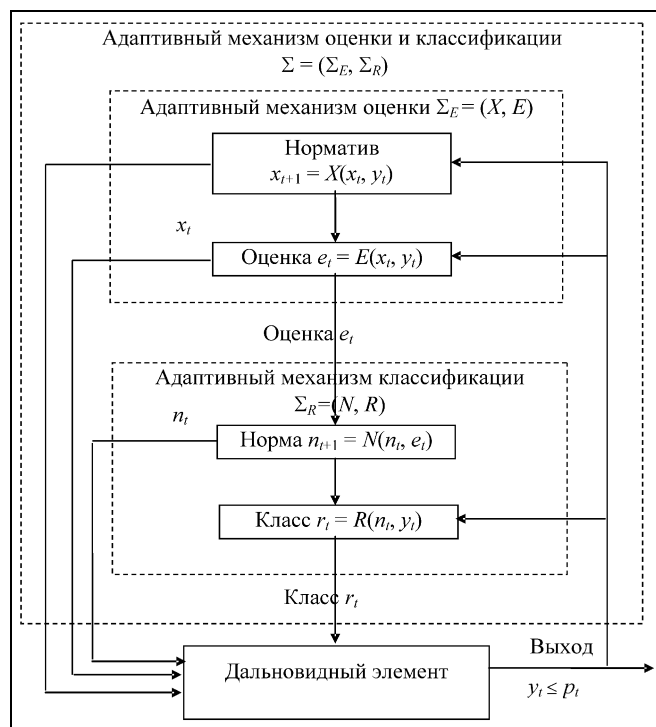


Схема адаптивного механизма оценки и классификации



$y_{2t}, \dots, y_{1t} \leq y_{2t}, t = \overline{1, T}$. Тогда ее максимум достигается, если состояние ДЭ y_t^* соответствует потенциалу p_t :

$$y_t^* = p_t, \quad t = \overline{1, T}. \quad (7)$$

Как принято в работах [1, 2] будем предполагать, что ДЭ благожелателен по отношению к Центру: если $p_t \in G(\Sigma, p)$, то $y_t^* = p_t$.

Механизм, обеспечивающий выполнение условия (7), т. е. раскрытие потенциала p_t , называют прогрессивным [1, 2]. Поэтому задача оптимального синтеза АМОК $\Sigma = \{\Sigma_E, \Sigma_R\}$ (6), при условии (7), сводится к задаче синтеза прогрессивного механизма.

Теорема. Для прогрессивности АМОК $\Sigma = \{\Sigma_E, \Sigma_R\}$ достаточно прогрессивности АМО Σ_E .

Доказательство. По условиям теоремы АМО Σ_E прогрессивен. Поэтому с ростом показателя y_t , текущая оценка e_t , получаемая из АМО Σ_E , возрастает. Далее, в АМОК используется АМК Σ_R с процедурой классификации (2). Поэтому, с ростом y_t , текущий класс ДЭ (r_t) также возрастает. Исследуем теперь зависимость будущих классов ДЭ ($r_{t+1}, \dots, r_{t+T}$) от y_t . В АМК используется процедура нормирования (3), согласно которой с ростом оценки e_t будущие нормы n_t убывают при $\tau = \overline{1, T}$. Следовательно, с ростом показателя y_t будущие нормы ($n_{t+1}, \dots, n_{t+T}$) убывают. Но при процедуре классификации (2) класс r_t монотонно возрастает с уменьшением нормы n_t при заданном y_t , $\tau = \overline{1, T}$. Таким образом с ростом показателя y_t будущие классы ДЭ ($r_{t+1}, \dots, r_{t+T}$) возрастают. Согласно выражениям (1), (4) и (5) целевая функция ДЭ — монотонно возрастающая функция текущих и будущих оценок и классов. Но как было показано ранее, $e_t, r_t, \dots, e_{t+T}, r_{t+T}$ — монотонно возрастающие функции показателя y_t . Следовательно, с ростом показателя y_t , возрастает и целевая функция ДЭ. В силу неравенства $y_t \leq p_t$ ее максимум достигается при состоянии y_t^* , равном потенциалу p_t . Поскольку ДЭ благожелателен по отношению к Центру, то $y_t^* = p_t, t = \overline{1, T}$. Следовательно, АМОК $\Sigma = \{\Sigma_E, \Sigma_R\}$ является прогрессивным. ♦

Таким образом, решение задачи синтеза оптимального АМОК $\Sigma = \{\Sigma_E, \Sigma_R\}$ (6) при условии (7) сводится к решению задачи синтеза прогрессивного АМО Σ_E . В свою очередь, прогрессивность АМО Σ_E обеспечивается известными методами, например, выбором подходящей процедуры оценки [1]. Рассмотрим, например, АМОК

$\Sigma^1 = \{\Sigma_E^1, \Sigma_R\}$ с линейным АМО Σ_E^1 , в котором процедура нормирования имеет вид: $x_{t+1} = \alpha y_t + \beta x_t, t = 1, 2, \dots, x_1 = x^1$, а процедура оценки — $e(x_t, y_t) = by_t - dx_t$, где b, d, α и β — неотрицательные величины. Как показано в работе [1], для прогрессивности линейного АМО Σ_E^1 достаточно выполнения неравенства:

$$b \geq \alpha dk[1 - (k\beta)^T]/(1 - k\beta). \quad (8)$$

Следствие. Линейный АМОК $\Sigma^1 = \{\Sigma_E^1, \Sigma_R\}$ прогрессивен, если выполняется неравенство (8).

Доказательство. В соответствии с теоремой для прогрессивности АМОК $\Sigma^1 = \{\Sigma_E^1, \Sigma_R\}$ достаточно обеспечить прогрессивность АМО Σ_E^1 . Но для прогрессивности линейного АМО Σ_E^1 достаточно выполнения неравенства (8), что и требовалось доказать. ♦

Заметим, что условия прогрессивности АМОК не зависят от параметров процедуры нормирования (в том числе шага адаптации).

В заключение отметим, что доказанная теорема и следствие существенно упрощают синтез АМОК. Действительно, для прогрессивности АМОК, как комплексного механизма, включающего в себя два локальных механизма — АМО и АМК — достаточно обеспечить лишь прогрессивность АМО, используя наработанные в работах [1, 2] методы. Этот результат заранее не очевиден. Более того, ранее считалось, что для обеспечения прогрессивности комплекса (композиции) адаптивных механизмов следует обеспечить прогрессивность *всех* его составляющих. Это предположение даже было сформулировано в виде принципа декомпозиции: «Комплексный механизм будет прогрессивным, если прогрессивен каждый его локальный механизм» (с. 119 работы [1]). При выполнении условий указанной теоремы и следствия можно отказаться от соблюдения этого принципа и, тем самым, упростить проектирование и реализацию комплексных прогрессивных адаптивных механизмов.

ЛИТЕРАТУРА

1. Цыганов В. В. Адаптивные механизмы в отраслевом управлении. — М.: Наука, 1991. — 165 с.
2. Цыганов В. В., Бородин В. А., Шишкин Г. Б. Интеллектуальное предприятие: механизмы овладения капиталом и властью (теория и практика управления эволюцией организации). — М.: Университетская книга, 2004. — 768 с.

☎ (343) 245-14-91, 334-91-91;

e-mail: nsirina@nis.usurt.ru, bbc@ipu.ru



Уважаемые читатели!

С экстратекстом журнала "Проблемы управления" вы можете ознакомиться в Интернете, посетив сайт <http://www.extratext.by.ru/>.

Экстратекст — это новый инструмент информационной поддержки инноваций.

Экстратекстом научной статьи мы называем информационный объект, элементами которого являются: Библиографическое описание, Аннотация, Введение, Заключение (выводы) и Список литературы.

Экстратекст дает сжатое представление о перечне и сути рассматриваемых вопросов, полученных результатах, позиции и эрудиции автора.

Редакция

Указатель статей, опубликованных в журнале в 2006 г.

- Абрамов О. В.** Методы и алгоритмы параметрического синтеза стохастических систем. № 4.
- Абрамова Н. А., Коврига С. В.** О рисках, связанных с ошибками экспертов и аналитиков. № 6.
- Анатолий Николаевич Шубин**. № 1.
- Аржаков М. В.** Механизмы выполнения договоров. № 5.
- Аржакова Н. В.** Механизмы партнерства при конкуренции. № 5.
- Артемьева И. Л., Гаврилова Т. Л., Грибова В. В.** и др. Мультидисциплинарная система управления информационными ресурсами различных уровней общности. № 4.
- Аршинский Л. В.** Основания векторизации истинности в логико-математических моделях обработки данных. № 5.
- Асратян Р. Э.** Интернет-служба для поддержки межсерверных взаимодействий в распределенных информационных системах. № 5.
- Бакусов Л. М., Ильясов Б. Г., Рамазанов М. Д., Сафин Ш. М.** Биологические вычисления: общие принципы. № 1.
- Билик Р. В., Вертлиб В. А., Гуденко А. А.** Методология автоматизированных систем массового обслуживания — база современной сетевой экономики. № 2.
- Билик Р. В., Мясоедова З. П., Неснова Н. Н., Шпанов С. Н.** Электронные платежные системы и речевые технологии. № 2.
- Бунич А. Л.** Системы управления с идентификатором. Ч. II. № 2.
- Буянов Б. Б., Лубков Н. В., Поляк Г. Л.** Система поддержки принятия управленческих решений с применением имитационного моделирования. № 6.
- Ведешенков В. А.** Способ выделения подсистем достаточного размера для параллельного диагностирования больших цифровых систем с регулярной структурой. № 5.
- Вступление к подборке.** № 2.
- Вторая российско-турецкая конференция «Математические модели социально-экономического развития».** № 1.
- Выборнов Р. А.** Стимулирование по результатам коллективной деятельности в организационных системах с манипулированием информацией. № 5.
- Выхованец В. С.** Разнесенный грамматический разбор. № 1.
- Гаврилова Т. Л., Клещев А. С.** Внутренняя модель математической практики для систем автоматизированного конструирования доказательств теорем. Ч. 1. Общее описание модели. № 4.
- Гаврилова Т. Л., Клещев А. С.** Внутренняя модель математической практики для систем автоматизированного конструирования доказательств теорем. Ч. 2. Модель математического диалекта. № 5.
- Гаврилова Т. Л., Клещев А. С.** Внутренняя модель математической практики для систем автоматизированного конструирования доказательств теорем. Ч. 3. Модель доказательств. № 6.
- Гафорова Е. Б.** Вопросы создания единой системы менеджмента предприятия на основе качества. № 3.
- Генкин А. Л., Куделин А. Р.** Проблемы энергосберегающего управления листопркатным комплексом. Ч. 1. № 6.
- Головинский П. А.** Когерентный нейрон и распознавание образов. № 5.
- Гребенюк Г. Г.** Особенности разработки автоматизированных систем компьютерной поддержки принятия решений в задачах управления отраслями жизнеобеспечения города. № 5.
- Грибова В. В.** Проблемно-независимый генератор текстов, управляемый онтологией. № 4.
- Грибова В. В., Клещев А. С.** Управление проектированием и реализацией пользовательского интерфейса на основе онтологий. № 2.
- Девятисильный А. С., Кислов Д. Е.** Исследование разрешимости задач определения спутниковых орбит по измерениям. № 4.
- Девятисильный А. С., Числов К. А.** К задаче выставки инерциальной навигационной системы на неподвижном основании в условиях гравитационной неопределенности. № 4.
- Десова А. А., Легович Ю. С., Разин О. С.** Принципы формирования диагностически значимых признаков ритмической структуры пульсового сигнала. № 1.
- Дорожко В. М., Лебедева А. Н.** Экспертные представления об основных ситуационных моделях коллективного движения судов. № 4.
- Егоров А. Ф., Савицкая Т. В., Михайлова П. Г.** Модели и методы решения задач оперативного управления безопасностью непрерывных химико-технологических систем. Ч. 2. Продукционные модели представления знаний в системах поддержки принятия решений. № 3.
- Елисеев В. В., Игнатушенко В. В.** Проблема надежного выполнения сложных наборов задач в управляющих параллельных вычислительных системах. № 6.
- Жирабок А. Н.** Аналог ганкелевой матрицы для нелинейной динамической системы. № 2.
- Жирабок А. Н., Летенко А. А.** Логико-динамический подход к диагностированию билинейных систем. № 4.
- Жожикашвили В. А., Петухова Н. В.** Современные технологии управления в диспетчерской службе такси
- Жожикашвили В. А., Петухова Н. В., Фархадов М. П.** Компьютерные системы массового обслуживания и речевые технологии. № 2.
- Жучков А. В., Твердохлебов Н. В.** Эволюция архитектуры информационно-вычислительных ресурсов большого научно-технического проекта. № 3.
- Зенькова Л. П.** Инвестиционная теория циклов и трансформационная экономика Беларуси. № 1.
- Зыков С. В.** Управление жизненным циклом корпоративных информационных ресурсов. № 3.
- Ивашенко А. А.** Многокритериальные системы стимулирования. № 6.
- Ивашенко А. А.** Модели управления карьерой сотрудника в организации. № 5.
- Ивери Варламович Прангишвили**. № 2.
- Index of papers published in 2006.** N 6.
- Институту автоматизации и процессов управления** Дальневосточного отделения РАН — 35 лет. № 4.
- Иткин Г. П.** Состояние проблемы создания искусственного сердца в России. № 3.
- Карибский А. В., Мишутин Д. Ю., Шишорин Ю. Р.** Методы моделирования процесса бюджетирования в интегрированных компаниях. № 2.
- Касимов А. М., Мамедли Э. М., Чернявский Л. Т.** и др. Обеспечение стойкости систем управления летательных аппаратов к внешним возмущающим факторам. № 1.
- Касимов А. М., Мамедли Э. М., Чернявский Л. Т.** и др. Обеспечение точности и непрерывности управления летательным аппаратом в условиях действия дестабилизирующих факторов. № 2.
- Киншт Н. В., Петрунько Н. Н.** Логические формализмы обработки данных электромагнитной обстановки высоковольтной подстанции. № 4.



- Клепарский В. Г., Клепарская Ек. В.** Адекватность управления саморазвивающимися системами социально-экономической природы (на примере предприятий черной металлургии). № 5.
- Колемаев В. А.** Исследование поведения модели Самуэльсона—Хикса. № 1.
- Колоколов А. С.** Обработка сигнала в частотной области при распознавании речи. № 3.
- Колотников А. В., Мясоедова З. П.** Применение языка Simscript для описания систем массового обслуживания. № 2. Contents and abstracts. № 1—6.
- Краснов А. Е., Умеренков Д. Е.** Среднесрочная дискретная динамическая модель хозяйственной деятельности предприятия, производящего товары ограниченного срока хранения, в условиях нестабильности спроса. № 3.
- Крупорницкая И. А.** Совершенствование управления экспортной стратегией компании «Газпром». № 2.
- Кудинов Ю. И., Кудинов И. Ю.** Построение адаптивного нечеткого комбинированного регулятора в многосвязных системах. № 5.
- Кульчин Ю. Н., Денисов И. В., Панов А. В., Рыбальченко Н. А.** Применение персептронов для нелинейной реконструктивной томографии. № 4.
- Кульчин Ю. Н., Ким А. Ю.** Распознавание динамических образов распределенной информационно-измерительной системой сегментарного типа. № 5.
- Лебедев В. В., Гансвинд И. Н., Горохова И. Н.** и др. Модульная распределенная система решения типовых региональных задач (на примере класса гидрологических задач). № 6.
- Лейбов Р. Л.** Линейная модель с неопределенными собственными значениями. № 3.
- Лейбов Р. Л.** Линейная модель с неопределенными элементами матриц. № 5.
- Лисицын Н. В., Викторов В. К.** Определение оптимального времени функционирования информационной системы. № 6.
- Мандель А. С.** Экспертно-статистические методы обработки информации в интегрированных системах управления производством и технологическими процессами. № 6.
- Матросов В. М., Баранов В. В.** Проблема превентивной безопасности. Модель и методы принятия решений. № 5.
- Микрин В. Е.** Репланирование сельскохозяйственного производства озимых культур в условиях чрезвычайных ситуаций. № 3.
- Миллер А. М., Готгельф Г. П., Левин М. А., Ловский В. Ю.** Система резервирования «Сирена-2.3». № 2.
- Мишин С. П.** Оптимальность древовидной иерархии управления симметричной производственной линией. № 6.
- Мокров Д. В., Бахтадзе Н. Н.** Программные анализаторы в системах управления транспортно-экспедиционной компанией: экспериментальное исследование. № 1.
- Научно-практический семинар** «Современные методы управления сложными социальными и экономическими системами». № 5.
- Нижегородцев Р. М.** Проблемы управления инфляцией: современные подходы. № 6.
- Павельев С. В.** Задачи оптимального оперативного резервирования информационных массивов и программных модулей в корпоративных вычислительных сетях, построенных с использованием каналов Интернета. № 3.
- Панкова Л. А., Пронина В. А.** Способы создания универсального инструментария для компьютерного моделирования. № 6.
- Печерский М. П., Лившиц Б. Ю.** Общегородская АСУ дорожным движением в Москве (система «Старт») и ее дальнейшее развитие. № 2.
- Подлазов В. С.** Нарастаемые многокольцевые некоммутируемые сети связи для многопроцессорных вычислительных систем. № 2.
- Подлазов В. С., Соколов В. В.** Однокаскадные коммутаторы большой размерности для многопроцессорных и многомашинных вычислительных систем. № 6.
- Подлазов В. С., Стецюра Г. Г.** Регулярные плоские сети для суперкомпьютеров. № 1.
- V Международная конференция** «Когнитивный анализ и управление развитием ситуаций» CASC'2005. № 3.
- Радько С. Г.** Оценка трудового потенциала специалистов управленческого звена. № 1.
- Рахматуллин И. Ф.** Моделирование управления информационными ресурсами как часть промышленной политики. № 3.
- Рудько И. М.** Определение дистанции до цели, находящейся в зоне Френеля антенны. № 6.
- Рывкин С. Е., Изосимов Д. Б., Байда С. В.** Регуляризация переключений компонент многомерного разрывного управления в реальном скользящем режиме. № 1.
- Саакян В. Г., Левин М. А., Ловский В. Ю., Миллер А. М.** Автоматизированная система продажи автобусных перевозок «Автовокзал-2». № 2.
- Санникова И. Н.** Информационная поддержка обучающейся организации. № 2.
- Сачук Т. В.** Разработка деловых стратегий развития территории. № 1.
- Сирина Н. Ф., Цыганов В. В.** Адаптивные механизмы оценки и классификации дальновидных активных систем. № 6.
- Соколов С. В., Погорелов В. А.** Решение задачи идентификации структуры стохастического процесса. № 2.
- Сомов Е. И.** Аналитический синтез программного гиросилового управления свободнолетающим космическим роботом в режиме транспортировки полезного груза. № 6.
- Столбов В. Ю., Федосеев С. А.** Модель интеллектуальной системы управления производством. № 5.
- Стукач О. В.** Сигнально-параметрическая инвариантность систем управления. № 2.
- Торгашов А. Ю.** Итерационный синтез робастного многомерного ПИД-регулятора для управления реакционно-ректификационной колонной. № 4.
- Третья международная конференция** по проблемам управления. Аналитический обзор. № 5.
- Указатель статей**, опубликованных в 2006 г. № 6.
- Фархадов М. П., Жожикашвили А. В.** Математическая модель для расчета среднего числа переспросов при компьютерном распознавании речи. № 2.
- Филаретов В. Ф.** Разработка и исследование методов синтеза высокоточных систем управления сложными динамическими объектами в условиях параметрической неопределенности. № 4.
- Шелков А. Б., Пелихов В. П., Гладков Ю. М.** Методы анализа и оценки аудиторского риска при проведении выездных документальных проверок. № 5.
- Шелков М. А.** Интегрированные системы регистрации информации для отражения удаленных атак. № 3.
- Шкуратов С. Е.** Подход к стратегическому управлению малым инновационным предприятием. № 6.
- Шлаин В., Глейбман А., Ваганов Д.** Базовые алгоритмы системы автоматического обнаружения и классификации дефектов для управления производством полупроводниковых микросхем. № 1.
- Шполянская И. Ю.** Модель управления малым предприятием как открытой системой: методологический аспект. № 3.
- Щетинин В. Г., Смолякова М. К., Бражников А. И.** Нейросетевой синтез стратегий развития промышленного производства в регионе. № 3.
- Эпштейн В. Л.** Как увеличить индекс цитирования научной публикации. № 6.
- Эпштейн В. Л.** Тезисы В. А. Трапезникова в контексте теории постиндустриализма. № 1.
- Якубовская Н. Н., Викторов В. К., Лисицын Н. В.** Задача оперативного управления транспортной производственной системой. № 1.

Index of papers published in 2006

- 3rd International conference on control problems. N 5.
- 5th International conference «Cognitive analysis and situation control» (CASC'2005). N 3.
- Abramov O. V.** Methods and algorithms for parametric synthesis of stochastic systems. N 4.
- Abramova N. A., Kovriga S. V.** On the risks related with experts' and analysts' mistakes. N 6.
- Anatoly N. Shoubin**. N 1.
- Arshinsky L. V.** Grounds for truth vectorization in the logic-mathematical data processing models. N 5.
- Artemieva I. L., Gavrilova T. L., Gribova V. V., et al.** An interdisciplinary system for information resources management at various generality levels. N 4.
- Arzhakov M. V.** Contract implementation mechanisms. N 5.
- Arzhakova N. V.** Partnerships mechanisms under competition. N 5.
- Asratian R. E.** Internet service for server-server communications support in distributed information systems. N 5.
- Automated queuing and service systems: a thematic collection.** N 2.
- Bakusov L. M., Ilyasov B. G., Ramazanov M. D., Safin Sh. M.** Biological calculations: general principles. N 1.
- Bilik R. V., Myasoyedova Z. P., Nesnova N. N., Shpanov S. N.** Electronic payment systems and speech recognition technologies. N 2.
- Bilik R. V., Vertlib V. A., Gudenko A. A.** The methodology of automated service systems — the basis for modern network economy. N 2.
- Bunich A. L.** Control systems with identifier. Part II. N 2.
- Buyanov B. B., Lubkov N. V., Polyak G. L.** A management decision-making support system with simulation. N 6.
- Contents and abstracts.** N 1–6.
- Desova A. A., Legovich Yu. S., Razin O. S.** Principles of forming diagnostically significant values of the features of pulse signal's rhythmical structure. N 1.
- Dev'yatisil'ny A. S. and Chislov K. A.** On the problem of inertial navigator alignment on immovable mount under gravitational uncertainty. N 4.
- Dev'yatisil'ny A. S. and Kislov D. E.** Solvability investigation for satellite orbit determination problem based on measurements. N 4.
- Dorozhko V. M., Lebedeva A. N.** Expert judgments about basic situational models of collective movements of ships. N 4.
- Epstein V. L.** How to increase the citation index of a scientific publication. N 6.
- Epstein V. L. V. A.** Trapeznikov's theses in the context of postindustrialism theory. N 1.
- Farkhadov M. P., Zhozhikashvili A. V.** Mathematical model for calculating the average number of retries in a speech recognition dialogue. N 2.
- Filaretov V. F.** Development and investigation of synthesis methods for high-precision control systems of compound dynamic objects under parametric uncertainty. N 4.
- Gafforova E. B.** The problems of unified quality-based enterprise management system development. N 3.
- Gavrilova T. L., Kleshev A. S.** An internal model of mathematical practice for interactive systems of theorem proof construction. Part 1. General model description. N 4.
- Gavrilova T. L., Kleshev A. S.** An internal model of mathematical practice for interactive systems of theorem proof construction. Part 2. A mathematical dialect model. N 5.
- Gavrilova T. L., Kleshev A. S.** An internal model of mathematical practice for interactive systems of theorem proof construction. Part 3. A proof model. N 6.
- Ghenkin A. L., Kudelin A. R.** The problems of energy-efficient control of a sheet-rolling complex. Part 1. N 6.
- Golovinsky P. A.** Coherent neuron and pattern recognition. N 5.
- Grebenyuk G. G.** On the development of computer-aided decision-making support systems for city operation control tasks. N 5.
- Gribova V. V.** A Domain-independent ontology-driven text generator. N 4.
- Gribova V. V., Kleshchev A. S.** The management of user interface design and implementation on the basis of ontologies. N 2.
- Shpolyanskaya I. Yu.** A model of small enterprise management as an open system: the methodological aspect. N 3.
- Institute for automation and control processes of Far-eastern branch of Russian academy of sciences is 35.** N 5.
- Itkin G. P.** Artificial heart development problem: the current status. N 3.
- Ivashchenko A. A.** Models of employee's career management in an organization. N 5.
- Ivashchenko A. A.** Multiobjective motivation systems. N 6.
- Iveri V. Prangishvili**. N 2.
- Index of papers published in 2006.** N 6.
- Karibsky A. V., Mishutin D. Yu., Shishorin Yu. R.** Budgeting modeling methods for integrated companies. N 2.
- Kasimov A. M., Mamedli E. M., Chernyavsky L. T., et al.** Ensuring aircraft control systems resistance to external perturbations. N 1.
- Kasimov A. M., Mamedli E. M., Chernyavsky L. T., et al.** Ensuring aircraft control accuracy and continuity under external perturbations. N 2.
- Kinsht N. V., Petrun'ko N. N.** Logical formalisms of data processing for electromagnetic environment of HV substation. N 4.
- Kleparsky V. G., Kleparskaya Ek. V.** The adequacy of the management of self-developing socio-economic systems (with a case study of ferrous metallurgical works). N 5.
- Kolemaev V. A.** Investigation of Samuelson—Hicks model behavior. N 1.
- Kolokolov A. S.** Frequency domain signal processing in speech recognition. N 3.
- Kolotnikov A. V., Myasoyedova Z. P.** Simscript simulation language application for queuing systems description. N 2.
- Krasnov A. E., Umerenkov D. E.** Medium-term discrete dynamic model of an enterprise manufacturing products with limited shelf life under stochastic demand. N 3.
- Krupornitskaya I. A.** Gazprom JSC management improvement. N 2.
- Kudinov Yu. I., Kudinov I. Yu.** Design concepts of multivariable adaptive on-off control systems. N 5.
- Kul'chin Yu. N., Kim A. Yu.** Dynamic patterns recognition by an segmental type measuring system. N 5.
- Kulchin Yu. N., Denisov I. V., Panov A. V., Rybal'chenko N. A.** Perceptron application in nonlinear reconstruction tomography. N 4.
- Lebedev V. V., Gansvind I. N., Gorokhova I. N., et al.** A distributed modular system for solving standard regional problems (with the example of hydrologic problems). N 6.
- Leibov R. L.** A linear model with uncertain matrices. N 5.
- Leibov R. L.** Linear model with uncertain eigenvalues. N 3.
- Lisitsyn N. V., Victorov V. K.** Evaluation of information system's optimal lifetime. N 6.
- Mandel' A. S.** Expert-statistical methods for information processing in integrated production management and process control systems. N 6.
- Matrosov V. M., Baranov V. V.** Preventive safety problem. The model and decision-making methods. N 5.
- Mikrin V. E.** Winter crops production replanning in emergency situations. N 3.
- Miller A. M., Gotgelf G. P., Levin M. A., Lovsky V. Yu.** Sirena-2.3 computer-based reservation system. N 2.
- Mishin S. P.** Optimality of tree-type control hierarchy for a symmetric production line. N 6.
- Mokrov D. V., Bakhtadze N. N.** Soft sensors in forwarding company's management systems: an experimental study. N 1.
- Nizhegorodtsev R. M.** Inflation control: contemporary approaches. N 6.
- Pankova L. A., Pronina V. A.** The ways to create a versatile computer modeling toolkit. N 6.



- Paveliev S. V.** Optimal on-line backup of information arrays and software modules in corporate networks based on Internet channels. N 3.
- Pechersky M. P., Livshits B. Yu.** The Moscow traffic management system (the *Start* system) and its further evolution. N 2.
- Podlazov V. S.** Scalable multiring fixed plants for multi-processor computation systems. N 2.
- Podlazov V. S., Sokolov V. V.** High dimensionality single-stage switches for multiprocessor and multimachine systems. N 6.
- Podlazov V. S., Stetsyura G. G.** Regular flat networks for supercomputers. N 1.
- Rad'ko S. G.** Evaluation of labor potential of management specialists. N 1.
- Rakhmatullin I. E.** Information resource management modeling as a part of industrial policy. N 3.
- Rud'ko I. M.** Evaluation of the distance from a target located in antenna's Fresnel region. N 6.
- Rykin S. E., Izosimov D. B., Bayda S. V.** Regularization of components switching for multidimensional discontinuous control in a real sliding mode. N 1.
- Saakian V. G., Levin M. A., Lovsky V. Yu., Miller A. M.** Auto-vokzal-2 bus conveyance sales system. N 2.
- Sachuk T. V.** The development of business strategies for territory growth. N 1.
- Sannikova I. N.** Information support for a learning organization. N 2.
- Schetinin V. G., Smolyakova M. K., Brazhnikov A. I.** The neural-network based synthesis of developing strategies for regional industrial production. N 3.
- Scientific and practical seminar «Modern control methods for complex social and economic systems».** N 5.
- Second Russian-Turkish Conference on Mathematical Models for Socio-Economic Progress.** N 1.
- Shelkov A. B., Pelikhov V. P., Gladkov Yu. M.** The techniques of travelling auditor's risks analysis and evaluation. N 5.
- Shelkov M. A., Gladkov M. Yu.** Integrated information registration systems for remote attacks repulsion. N 3.
- Shkuratov S. E.** Approach to strategic management of a small innovative enterprise. N 6.
- Shlain V., Gleibman A., Vagapov D.** Basic algorithms for an automatic faults detection and classification system in microchip manufacturing control. N 1.
- Sirina N. F., Tsyganov V. V.** Adaptive mechanisms of the prescient systems estimation and classification. N 6.
- Sokolov S. V., Pogorelov V. A.** Identification of stochastic process structure. N 2.
- Somov E. I.** Analytical synthesis of gyro-force program control of a free-flying space robot in net load transportation mode. N 6.
- Stolbov V. Yu., Fedoseev S. A.** A model of an intelligent production control system. N 5.
- Stukach O. V.** Signal-parametric invariance of control systems. N 2.
- Torgashov A. Yu.** Sequential synthesis of a robust multi-loop PID controller for a reactive distillation column. N 4.
- Vedeshenkov V. A.** A method for parcelling out the subsystems of sufficient dimension for parallel diagnosis of large digital systems with regular structure. N 5.
- Vybornov R. A.** Collective outcome incentive mechanisms in organizational systems with information manipulation. N 5.
- Vykhoanets V. S.** Separated parsing. N 1.
- Yakubovskaya N. N., Victorov V. K., Lisitsyn N. V.** Transportation system operation problem. N 1.
- Yegorov A. F., Savitskaya T. V., Mikhailova P. G.** Models and methods for on-line safety management of chemical processes. Part 2. Production models of knowledge representation in decision-making support systems. N 3.
- Yeliseev V. V., Ignatushchenko V. V.** On the problem of reliable execution of complex task sets in parallel control computer systems. N 6.
- Zen'kova L. P.** Investment cycles theory and transformational economy of Byelorussia. N 1.
- Zhirabok A. N.** An analog of Hankel matrix for a nonlinear dynamic system. N 2.
- Zhirabok A. N., Letenko A. A.** Logic-dynamical approach to bilinear systems diagnosis. N 5.
- Zhozhikashvili V. A., Petukhova N. V., Farkhadov M. P.** Computerized queuing systems and speech technologies. N 2.
- Zhozhikashvili V. A., Petukhova N. V., Zatsepin A. N., Azarov V. V.** Time control technologies in taxi dispatching center. N 2.
- Zhuchkov A. V., Tverdokhlebov N. V.** Cyber-resources architecture evolution in a large-scale scientific project. N 3.
- Zykov S. V.** Lifecycle management of corporate information resources. N 3.

CONTENTS & ABSTRACTS

THE WAYS TO CREATE A VERSATILE COMPUTER MODELING TOOLKIT. 2

Pankova L. A., Pronina V. A.

The paper reviews modern computer modeling technologies and classifies computer models. The approaches to creating a versatile computer modeling toolkit based on state-of-the-art technologies are outlined.

ON THE PROBLEM OF RELIABLE EXECUTION OF COMPLEX TASK SETS IN PARALLEL CONTROL COMPUTER SYSTEMS. 6

Yeliseev V. V., Ignatushchenko V. V.

To provide the reliable execution of complex sets of interdependent tasks with random realization times in the parallel control computer systems, a new computer technology relying on static forecasting of task execution times is proposed. The reliable execution of the required task sets is regarded as their execution with the desired probability in a time not greater than the prescribed «directive» time. The technology offers a basically new approach to intelligent dynamic control of parallel computing processes.

HIGH DIMENSIONALITY SINGLE-STAGE SWITCHES FOR MULTIPROCESSOR AND MULTIMACHINE SYSTEMS. 19

Podlazov V. S., Sokolov V. V.

A method for regular building of high-dimensionality single-stage switches (flat switched networks) for modern supercomputers is considered.

INFLATION CONTROL: CONTEMPORARY APPROACHES 25

Nizhegorodtsev R. M.

The paper outlines the key problems of modern inflation processes control. The main ways to restrain the inflation are discussed. The failure of monetary methods of overcoming the inflation is shown, and a myth about the overheating of current Russian economy is destroyed. The recommendations on how to reduce the inflation in Russian economy based on technological modernization are made.

MULTIOBJECTIVE MOTIVATION SYSTEMS 31

Ivashchenko A. A.

The paper describes game-theory models of multi-criterion motivation systems where the activities of each controlled subject is de-

scribed by several indicators whose values determine the reward paid to the subject by the authority.

OPTIMALITY OF TREE-TYPE CONTROL HIERARCHY FOR A SYMMETRIC PRODUCTION LINE 36

Mishin S. P.

A model of a multilayer hierarchy controlling a «production line», i. e., a business process with workers interacting sequentially with the same intensity, is considered. Manager's costs are described by a power function dependent on the intensity of the sub-process controlled by the manager. The paper proves that the optimal tree-type hierarchy is the one for which the search is reduced to a constrained optimization problem. The optimal tree for a convex cost function is derived analytically.

A MANAGEMENT DECISION-MAKING SUPPORT SYSTEM WITH SIMULATION 43

Buyanov B. B., Lubkov N. V., Polyak G. L.

Functional and algorithmic structures of a decision-making support system comprising a mathematical model of management activity scenarios and a multi-criterion model of alternatives comparison are developed. A software suite implementing the developed procedures is described. The software provides the user interface enabling interactive formation of management scenarios and a preference system for the best decision-making.

THE PROBLEMS OF ENERGY-EFFICIENT CONTROL OF A SHEET-ROLLING COMPLEX. Part I. 50

Ghenkin A. L., Kudelin A. R.

Theoretical background of energy-efficient control of a sheet-rolling «furnaces—flattening-mill» system is outlined. The system's distinctive feature is the redistribution of energy between different parts of the processing train.

EXPERT-STATISTICAL METHODS FOR INFORMATION PROCESSING IN INTEGRATED PRODUCTION MANAGEMENT AND PROCESS CONTROL SYSTEMS . . . 55

Mandel' A. S.

The paper examines the possibility of expert-statistical methods application for information processing in production management and process control tasks within the integrated control system available at the enterprise.

ON THE RISKS RELATED WITH EXPERTS' AND ANALYSTS' MISTAKES 60

Abramova N. A., Kovriga S. V.

A new approach to the analysis of the situations traditionally treated as experts' analysts' and decision-makers' mistakes is presented. The approach is based on the analysis of experts' practical activities and on the theoretical models accounting for both practice and psychological knowledge. Risk factors caused by inadequate interpretation of «mistakes» are analyzed by the known examples of preferences' transitivity violation.

AN INTERNAL MODEL OF MATHEMATICAL PRACTICE FOR INTERACTIVE SYSTEMS OF THEOREM PROOF CONSTRUCTION. Part 3. A proof Model 68

Gavrilova T. L., Kleschev A. S.

The paper completes the series of 3 articles describing the model of mathematical practice for an automated theorem proving system. Reasoning rules used in proof construction are formulated. A model of a complete proof is determined. The examples of mathematical statements in MMD language — a formal model of a mathematical dialect — and a model of correct complete proof for one of them are presented.

ANALYTICAL SYNTHESIS OF GYRO-FORCE PROGRAM CONTROL OF A FREE-FLYING SPACE ROBOT IN NET LOAD TRANSPORTATION MODE. 72

Somov E. I.

A method for analytical synthesis of spatial angular program motion of a free-flying robotized space module in net load transportation mode is presented. The paper shows that when the module performs a turn maneuver with a non-rigid load under general-form boundary conditions, weak excitation of elastic vibrations of the transported structure arise owing to analytically calculated gyro-force control.

EVALUATION OF THE DISTANCE FROM A TARGET LOCATED IN ANTENNA'S FRESNEL REGION 79

Rud'ko I. M.

The paper discusses a method for determining the distance from a target under space-time signal processing in the mode of passive location of targets positioned in the antenna's Fresnel region in case of the signal emitted by the target containing a narrowband component.

HOW TO INCREASE THE CITATION INDEX OF A SCIENTIFIC PUBLICATION. 83

Epstein V. L.

Comparative data on the citation frequency of printed and electronic publications are presented. The paper suggests to undertake Internet marketing of «extra-texts» from scientific papers. The proposed approach can presumably accelerate the dissemination of information about new scientific achievements, put into scientific and economic use the intelligent capital stored in our libraries as forestalled or underrated and undeservedly forgotten publications, create public scientific and informational environment for innovative economy.

EVALUATION OF INFORMATION SYSTEM'S OPTIMAL LIFETIME 85

Lisitsyn N. V., Victorov V. K.

The problem of information system's optimal lifetime evaluation is discussed. The algorithm for problem solving is based on the Bellman's optimality principle. Some examples are included.

A DISTRIBUTED MODULAR SYSTEM FOR SOLVING STANDARD REGIONAL PROBLEMS (with the example of hydrologic problems) 88

Lebedev V. V., Gansvind I. N., Gorokhova I. N., et al.

Based on the analysis of the general structure of a regional socio-economic system and its environmental subsystem, a functional structure of an automated system for solving standard regional hydrologic problems is offered.

APPROACH TO STRATEGIC MANAGEMENT OF A SMALL INNOVATIVE ENTERPRISE 91

Shkuratov S. E.

Basic approaches to the management of a small innovative enterprise are identified, and a strategic management model created in view of distinctive features of such enterprises is proposed. The strategy selection is based on the analysis of product design, resources availability, and business development status. The paper notes that the combination of these estimates enables the formation of a wide set of enterprise development alternatives from self-dependent work and market entry up to looking for a strategic alliance with a large-scale enterprise.

ADAPTIVE MECHANISMS OF THE PRESIDENT SYSTEMS ESTIMATION AND CLASSIFICATION 93

Sirina N. F., Tsyganov V. V.

The problem of the optimal synthesis of the complex adaptive mechanism based on quantitative and qualitative estimation of the results of the president system is stated. It is proved that the progressiveness of the estimation adaptive mechanism is sufficient for the progressiveness of the complex adaptive mechanism.