

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

УПРАВЛЕНИЕ БОЛЬШИМИ СИСТЕМАМИ

Выпуск 34

Сентябрь 2011

**СБОРНИК
ТРУДОВ**

ISSN 1819-2467

Регистрационный номер Эл №ФС77-27285 от 22.02.2007

Москва – 2011

www.mtas.ru

ИНТЕРНЕТ-сайт теории управления организационными системами

ТЕОРИЯ УПРАВЛЕНИЯ ОРГАНИЗАЦИОННЫМИ СИСТЕМАМИ

ИНСТИТУТ ПРОБЛЕМ УПРАВЛЕНИЯ ИМ. В.А. ТРАПЕЗНИКОВА РАН



Целью сайта является предоставление специалистам по теории и практике управления организационными системами (ученым, преподавателям, аспирантам, студентам, а также реальным управленцам) доступа к ресурсам, отражающим современное состояние теории и возможности обмена идеями и результатами.

На сайте имеются разделы:

Теория – с обзором теории управления организационными системами, глоссарием, информацией для аспирантов;

Практика – с обзором результатов внедрения механизмов управления в реальных организациях;

Библиография – около 2500 публикаций по теории управления, снабжена классификатором и аннотациями;

Электронная библиотека – около 700 полнотекстовых монографий, статей и учебных пособий;

а также многое другое.

На сайте работает **форум**, на котором можно обсудить вопросы, относящиеся к математике, экономике, управлению организациями, узнать новости теории управления и ознакомиться с планируемыми конференциями и семинарами.

ubs.mtas.ru

Интернет-сайт электронного периодического
научного издания «Управление большими
системами: сборник трудов»

УПРАВЛЕНИЕ БОЛЬШИМИ СИСТЕМАМИ

ЭЛЕКТРОННЫЙ СБОРНИК НАУЧНЫХ ТРУДОВ
ИНСТИТУТА ПРОБЛЕМ УПРАВЛЕНИЯ ИМ. В.А. ТРАПЕЗНИКОВА РАН

С 1998 года Институт проблем управления РАН выпускает периодический сборник трудов ученых, занимающихся разработкой и исследованием математических моделей управления большими (социально-экономическими, организационными, организационно-техническими и др.) системами. Все статьи, публикуемые в сборнике, проходят рецензирование ведущими специалистами по теории управления.

С 2006 года сборник "Управление большими системами" вместе с ведущим журналом ИПУ РАН "Проблемы управления" – включены в Российский индекс научного цитирования (РИНЦ).

С июля 2007 года Сборник входит в список ВАК (перечень ведущих рецензируемых научных журналов и изданий, в которых должны быть опубликованы основные научные результаты диссертации на соискание ученой степени доктора и кандидата наук):

* по управлению, вычислительной технике и информатике (для докторов наук);

* по электронике, измерительной технике, радиотехнике и связи; по энергетике (для кандидатов наук).

**Уважаемые коллеги! Приглашаем Вас опубликовать
Вашу статью в очередном выпуске сборника
"Управление большими системами"!**

Периодичность сборника - 4 раза в год. Время выхода прошедшей рецензирование статьи - 3-4 месяца. Плата с авторов за публикацию рукописей не взимается.

cmm.ipu.ru

3-я Российская конференция с международным участием «Технические и программные средства систем управления, контроля и измерения» (УКИ-12)

Уважаемые дамы и господа, приглашаем Вас принять участие в этой конференции, она состоится с 16 по 19 апреля 2012 года в Институте проблем управления РАН, г. Москва.

Тематика конференции

- Теория, методы исследования и проектирования, опыт применения технических средств (от датчиков до исполнительных механизмов), основанных на различных физических и схематических принципах.
- Теория, алгоритмы и программное обеспечение систем УКИ.
- Методы компьютерного и физического моделирования технических и программных средств и в целом систем автоматизации.
- более, смотрите на сайте конференции.

Принимаются предложения по пленарным докладам, по организации специальных заседаний конференции и тематике дискуссий за круглым столом.

Даты конференции

- 28 ноября 2011** - завершение приема заявок на участие и кратких текстов докладов
- 20 января 2012** - результаты рассмотрения докладов; последний срок подачи пленарных докладов.
- 10 марта 2012** - окончание приема финальных версий принятых докладов и внесения оргвзноса.
- 16-19 апреля 2012** - проведение конференции.

В качестве заявки на участие достаточно прислать по электронной почте **cmm-conf@mail.ru** краткую версию текста не менее 4 страниц.

Электронный адрес конференции: **cmm-conf@mail.ru**

Официальный сайт конференции: **<http://cmm.ipu.ru>**

РОССИЙСКАЯ АКАДЕМИЯ НАУК
Институт проблем управления
им. В.А. Трапезникова

**УПРАВЛЕНИЕ
БОЛЬШИМИ
СИСТЕМАМИ**

СБОРНИК ТРУДОВ

Выпуск 34

Москва – 2011

КООРДИНАЦИОННЫЙ СОВЕТ

Академики РАН: Васильев С.Н., Емельянов С.В., Коровин С.К., Куржанский А.Б., Федосов Е.А., Черноусько Ф.Л.; члены-корреспонденты РАН: Желтов С.Ю., Каляев И.А., Пархоменко П.П., Попков Ю.С.; д-ра техн. наук: Бутковский А.Г., Дорофеев А.А., Кузнецов О.П., Кульба В.В., Кротов В.Ф., Лотоцкий В.А., Павлов Б.В., Поляк Б.Т., Рутковский В.Ю.

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

Главный редактор: член-корр. РАН Новиков Д.А. **Отв. секретарь:** к.т.н. Губко М.В. Д-ра техн. наук: проф. Алескерев Ф.Т. (ГУ ВШЭ), проф. Артамонов Е.И. (ИПУ РАН), д-р экон. наук, проф. Архипова М.Ю. (ИПИ РАН), д-ра техн. наук: проф. Афанасьев В.Н. (МИЭМ), проф. Бахтадзе Н.Н. (ИПУ РАН), проф. Бурков В.Н. (ИПУ РАН), проф. Вишневский В.М. (ИППИ РАН), д-р экон. наук, проф. Голиченко О.Г. (ЦЭМИ РАН), д-р физ.-мат. наук, проф. Добровидов А.В. (ИПУ РАН), д-ра техн. наук: проф. Заложнев А.Ю. (ИПУ РАН), проф. Ириков В.А. (МФТИ), проф. Калянов Г.Н. (ИПУ РАН), проф. Касаткин С.И. (ИПУ РАН), проф. Каравай М.Ф. (ИПУ РАН), канд. техн. наук Квинто Я.И. (ИПУ РАН), д-р экон. наук, проф. Клочков В.В. (ИПУ РАН), д-ра техн. наук: проф. Кононенко А.Ф. (ВЦ РАН), проф. Курдюков А.П. (ИПУ РАН), проф. Лебедев В.Г. (ИПУ РАН), к-т техн. наук, доцент Лебедев В.Н. (ИПУ РАН), д-р экон. наук, проф. Ловчиновский Э.В. (ИПУ РАН), д-р техн. наук, проф. Мандель А.С. (ИПУ РАН), д-р экон. наук, проф. Нижегородцев Р.М. (ИПУ РАН), д-ра техн. наук: проф. Новосельцев В.Н. (ИПУ РАН), проф. Орлов А.И. (МВТУ), д-р физ.-мат. наук, проф. Рапопорт Л.Б. (ИПУ РАН), д-р техн. наук, проф. Рыков А.С. (МИСИС), д-р экон. наук, проф. Секерин В.Д. (ИПУ РАН), д-ра техн. наук: проф. Сидельников Ю.В. (МАИ), проф. Совлуков А.С. (ИПУ РАН), д-р экон. наук, проф. Сухарев О.С. (Ин-т экономики РАН), д-ра техн. наук: проф. Уткин В.А. (ИПУ РАН), проф. Хоботов Е.Н. (МВТУ), д-ра физ.-мат. наук: доцент Чеботарев П.Ю. (ИПУ РАН), проф. Чхартишвили А.Г. (ИПУ РАН), проф. Щербаков П.С. (ИПУ РАН).

РЕГИОНАЛЬНЫЕ РЕДАКЦИОННЫЕ СОВЕТЫ

Волгоград – д-ра физ.-мат. наук: проф. Воронин А.А., проф. Лосев А.Г. (ВолГУ); **Воронеж** – д-р техн. наук, проф. Баркалов С.А., д-р физ.-мат. наук, проф. Головинский П.А. (ВГАСУ), д-р техн. наук, проф. Подвальный С.Л. (ВГТУ); **Ижевск** – д-р физ.-мат. наук, проф. Непейвода Н.Н., к-т физ.-мат. наук, проф. Родионов В.И. (УдмГУ); **Иркутск** – д-ра физ.-мат. наук: проф. Бычков И.В., проф. Лакеев А.В. (ИДСТУ СО РАН); **Казань** – д-р физ.-мат. наук, проф. Маликов А.И., д-р техн. наук, проф. Сиразетдинов Р.Т. (КГТУ-КАИ); **Липецк** – д-ра техн. наук: проф. Кузнецов Л.А., проф. Погодаев А.К. (ЛГТУ); **Самара** – д-ра экон. наук: проф. Богатырев В.Д., проф. Гераськин М.И., д-р техн. наук, проф. Засканов В.Г. (СГАУ); **Санкт-Петербург** – д-ра физ.-мат. наук: проф. Петросян Л.А. (СПбГУ), проф. Фрадков А.Л. (ИПМ РАН); **Старый Оскол** – д-р техн. наук, проф. Еременко Ю.И. (СТИ); **Тверь** – д-ра техн. наук: проф. Кузнецов В.Н., проф. Палюх Б.В. (ТГТУ).

Адрес редакции: 117997, г. Москва, ул. Профсоюзная, д. 65.

Адрес в Интернет: ubs.mtas.ru.

Номер гос. регистрации электронного научного издания (ЭНИ): 0420900023.

© ИПУ РАН, 2011

СОДЕРЖАНИЕ

Системный анализ

- Салтыков С. А.**
Снижение затрат на построение функции ценности при использовании решающих правил теории важности критериев по сравнению с традиционным подходом теории полезности 5
- Юдицкий С. А.**
Алгебраическое представление модели многоагентных сетей.... 30

Математическая теория управления

- Агаев Р. П.**
Дискретная процедура согласования характеристик с помощью минимального цикла, объединяющего базовые бикомпоненты..... 46
- Карташов В. Я., Новосельцева М. А.**
Обнаружение структурно-параметрических изменений в стохастических системах в реальном масштабе времени алгоритмами непрерывных дробей и структурного анализа 62

Информационные технологии в управлении

- Каравай М. Ф., Подлазов В. С.**
Распределенный полный коммутатор как «идеальная» системная сеть для многопроцессорных вычислительных систем 92
- Яшкин А. В.**
Применение онтологий для поддержки версионности серверных операций..... 117

Управление в социально-экономических системах

- Богатырев В. Д., Морозова С. А.**
Модель и методика решения задачи оптимизации графика финансирования инвестиционного проекта на графах работ 130
- Варюхина Е. В., Клочков В. В.**
Анализ эффективности экономических стимулов повышения безопасности авиационной техники..... 146

Мишин С. П. <i>Свойства оптимального делегирования управления в организации</i>	165
Спиро А. Г., Дорофеев Ю. А., Alperovich Ed. <i>Индексные паевые инвестиционные фонды (анализ доходности)</i>	200
Эйфельд А. А. <i>Об одном подходе к разработке социально ориентированных тарифов на электроэнергию для населения</i>	213
Управление в медико-биологических и экологических системах	
<hr/> Романюха А. А., Носова Е. А. <i>Модель распространения ВИЧ-инфекции в результате социальной дезадаптации</i>	227
Управление техническими системами и технологическими процессами	
<hr/> Васильев Д. А., Колоколов М. В. , Иващенко В. А. <i>Модели автоматизированного прогнозирования электрических нагрузок промышленных предприятий</i>	254
Денисов А. Р., Белянкин М. В. <i>Система календарного планирования процессов конструкторско-технологической подготовки мелкосерийного машиностроительного производства</i>	267
Куров Б. Н. <i>Сравнение эффективности алгоритмов управления с учётом точности данных и реализации решений</i>	279
Надёжность и диагностика средств и систем управления	
<hr/> Дубов А. В. <i>Определение информационной модели объекта диагностики, заданного сигнальными графами функциональных элементов....</i>	292

УДК 519.816

ББК 22.18

СНИЖЕНИЕ ЗАТРАТ НА ПОСТРОЕНИЕ ФУНКЦИИ ЦЕННОСТИ ПРИ ИСПОЛЬЗОВАНИИ РЕШАЮЩИХ ПРАВИЛ ТЕОРИИ ВАЖНОСТИ КРИТЕРИЕВ ПО СРАВНЕНИЮ С ТРАДИЦИОННЫМ ПОДХОДОМ ТЕОРИИ ПОЛЕЗНОСТИ

Салтыков С. А.¹

(Учреждение Российской академии наук
Институт проблем управления РАН, Москва)

По результатам вычислительного эксперимента использование решающих правил теории важности критериев позволяет снизить затраты времени на построение функции ценности примерно в 2–5 раз по сравнению с традиционным подходом теории полезности. Для отбора из двух вариантов на основании двух критериев доказана теорема, определяющая количественно, во сколько раз при использовании теории важности критериев снижается доля случаев, когда нужно строить функцию ценности.

Ключевые слова: многокритериальный анализ решений, теория важности критериев, функция ценности, снижение затрат, имитационное моделирование.

1. Введение

Решать многокритериальные задачи можно разными методами: методом теории полезности (с построением функции полезности или ценности) [3], теории важности критериев [4], методом взвешенной суммы.

¹ Сергей Анатольевич Салтыков, младший научный сотрудник (ssaltykov@mail.ru).

Разграничение терминов «полезность» и «ценность» требует пояснения. Согласно [12, с. 266], разделение концептов ценности и полезности было введено Кини и Райфа в 1976 году и с тех пор основательно укоренилось в литературе: было предложено использовать термин «функция полезности» для случаев принятия решения в условиях риска и «функция ценности» для порядкового сравнения (*ordinal comparisons*). Уже здесь видна некая условность введенного разделения, так как ситуации делятся не по одному основанию, а по разным на «рисковые» и «порядковые». В дальнейшем, согласно [12, с. 293], такое разделение привело к тому, что теперь фактически есть три различных метода, базирующихся на разных исходных допущениях (аксиоматике): метод порядковой (ординальной) аддитивной функции ценности, метод измеримой (*measurable*) аддитивной функции ценности и метод, использующий функцию полезности. Последний используется исключительно для принятия решений в условиях риска и не может быть применен в безрисковом случае [12, с. 267], это несмотря на то, что существует очень большое число классов многокритериальных задач, где нет предположений о риске [12, с. 267], что говорит об ограниченности применения *MAUT* (*multiattribute utility theory* – многокритериальная теория полезности).

Идея *MAVT* (*multiattribute value theory* – многокритериальная теория ценности), куда входит использование порядковой и измеримой функции ценности, состоит в следующем. Предполагается, что существует ряд критериев, которые ЛПР учитывает при отборе из множества вариантов принятия решений, и характер влияния этих критериев в их взаимосвязи на конечное решение ЛПР неясен и сложен. Тогда может существовать способ не определять характер этой взаимосвязи в явном виде: если эти критерии удовлетворяют определенным условиям, можно построить некую специфическую функцию в аддитивном виде, называемую функцией ценности, большим значениям которой соответствуют более предпочтительные варианты, а меньшим – менее предпочтительные. Специфичность этой функции заключается в том, что, в общем случае, эта ординальная функция ценности напрямую не связана с «силой» предпочтения ЛПР, именно поэтому значения ординальной функции ценности счи-

таются измеренными в шкале порядка (следовательно, в частности, их нельзя делить друг на друга). С измеримой аддитивной функцией ценности принципиально ситуация такая же с той лишь разницей, что условия её существования строже, но если она существует, то над ее значениями допустим более широкий диапазон действий, так как она является измеренной в шкале интервалов. Построение функции ценности является весьма сложным [12, с. 275], утомительным [12, с. 278] и громоздким [12, с. 293], поэтому если проверка условий существования функции ценности и ее построение является более легким, чем определение характера взаимосвязанного влияния критериев на конечный отбор, то *MAVT* имеет смысл использовать. Если же из специфики предметной области характер этой взаимосвязи виден достаточно ясно, то в использовании специфической функции в аддитивном виде, которая не соотносится в содержательном плане с «силой» предпочтения ЛПП, нет смысла. Как видим, и в основе *MAVT* лежат довольно сильные допущения, которые существенным образом ограничивают класс ситуаций, к которым целесообразно этот метод применять. В [12, с. 293] также отмечается, что в методе анализа иерархий (АНР), считается, что «сила» предпочтения ЛПП может быть оценена прямо или косвенно экспертом или аналитиком в шкале отношений. Этот подход, применяемый в методе анализа иерархий, не покрывается полностью ни *MAUT*, ни *MAVT*, так как не использует их вышеперечисленных допущений, и, кроме того, распространяется на важные классы прикладных задач, не покрываемых *MAUT* и *MAVT*. Парадоксально, но этот подход к оценке «силы» предпочтений ЛПП, применяемый в широко распространенном методе анализа иерархий, сам по себе никак не именуется.

Видимо, именно поэтому за пределами МКАР сформировалась другая линия исследований, в рамках которой функцией полезности называется не то же, что и в МКАР. Так, в институтском учебнике по теории управления организационными системами [2, с. 30] читаем: «Предположим, что предпочтения этого субъекта описываются *функцией полезности* $f(y): A \rightarrow R^1$ (или целевой функцией, функцией предпочтения – будем использовать эти термины как синонимы), которая отображает множество его *действий* (альтернатив) A на числовую ось R^1 ». В рамках

этого подхода «сила» предпочтений ЛПР определяется непосредственно (без введения специфических функций и предположений о риске) и является синонимом полезности, а возможно, и ценности. Идея этого подхода предельно проста: изучать влияние не просто самого значения величины (скажем, в килограммах или миллиметрах) на принятие решения ЛПР, а «полезности» значения этой величины (то есть, «полезности» этих килограммов и миллиметров).

Не имея прямой необходимости базироваться на исходных допущениях *MAUT* или *MAVT*, солидаризируемся в данной работе с последним описанным подходом и будем считать ценность, полезность и «силу» предпочтений ЛПР синонимами (в отличие от [12, с. 266], где ценность и полезность выступают частными случаями «силы» предпочтения). Еще раз подчеркнем, что использование такого подхода в МКАР не является новшеством (он «защит» в метод анализа иерархий), но по стечению обстоятельств до сих пор не имеет в МКАР адекватного устоявшегося названия (встречаются «вынужденные» именованья этой «силы» предпочтения, которая не есть функция ценности, например – «*true preferences*», «*true preference function*» [12, с. 276]).

Из всего синонимического ряда в связи с подходом, теорией в название данной работы вынесено словосочетание «теория полезности», так как оно, во-первых, соотносится с определением функции полезности, данным в [2, с. 30], а во вторых, более знакомо и интуитивно понятно специалистам за пределами МКАР, так как словосочетание «теория ценности» не является употребительным. С другой стороны, в названии работы фигурирует именно «функция ценности», ввиду того, что функция ценности, в том смысле, как она употребляется в настоящей работе, может быть интерпретирована как аддитивная измеримая функция ценности, и в то же время этот термин свободен от ассоциаций с риском.

Итак, в работе [10] были проанализированы подход к решению многокритериальной задачи с построением функции ценности и метод взвешенной суммы, и было показано, что доля случаев, когда использование метода взвешенной суммы может привести к ошибочному отбору из множества вариантов, весьма высока – может достигать 35–40%.

Поясним, что подразумевается под ошибочным отбором из множества вариантов. В метод взвешенной суммы «зашито» неявное предположение, что предпочтения ЛПП равномерно растут вдоль шкалы, что в действительности не всегда верно. Поэтому может сложиться ситуация, когда у некоторого варианта средневзвешенное больше других, однако он не является наилучшим, так как его интегральная ценность (определенная с помощью функции ценности, отражающей неравномерный рост предпочтений) окажется меньше интегральной ценности другого варианта. В этом случае отбор варианта с наибольшим средневзвешенным, но не с самой большой интегральной ценностью будет ошибочным отбором варианта.

Согласно проведенному вычислительному эксперименту [10], эта доля ошибочных отборов при использовании метода взвешенной суммы не опускается ниже полутора процентов. Если из специфики предметной области известно, что «цена» одной такой ошибки значительна, то, по-видимому, придется использовать методы, корректно учитывающие характер роста предпочтений ЛПП при движении вдоль шкалы. Например, методы теории полезности или методы теории важности критериев. Поскольку построение функции ценности – всегда весьма трудоемкая операция, актуальным становится снижение доли случаев, когда придется строить функцию ценности. В работе [10] показано, что использование решающих правил теории важности критериев позволяет определять наилучший вариант, не строя функцию ценности примерно в 39–66% случаев. В данной работе мы пойдем дальше и выясним, во сколько раз использование теории важности критериев позволяет снизить затраты времени на построение функции ценности по сравнению с традиционным подходом теории полезности. На наш взгляд, количественная оценка снижения затрат времени будет еще более способствовать технико-экономическому обоснованию применения решающих правил теории важности критериев. В данной работе анализируется класс ситуаций, когда варианты оцениваются по нескольким критериям, имеющим общую балльную шкалу.

Работа является продолжением и развитием исследований [1, 7, 8, 9, 10].

2. Описание используемой многокритериальной модели

Многокритериальная модель, рассматриваемая в статье, может быть представлена следующим образом:

$$(1) \langle S, K_1, \dots, K_m, R \rangle,$$

где S – это множество вариантов решений (стратегий, планов, альтернатив и т.д.), далее называемое множеством вариантов; K_1, \dots, K_m – критерии (целевые функции и т.д.); R – отношение нестрогого предпочтения.

Поясним представленную модель (1). Каждый вариант s из множества S всех (данных) вариантов характеризуется значениями $m \geq 2$ критериев K_i . Под критерием K_i мы понимаем функцию, определенную на множестве S и принимающую значения из множества X_i , называемого шкалой (а также множеством оценок, шкальных градаций, значений критериев). Без ограничения общности будем считать, что все оценки выражены в численном виде и большие значения предпочтительней меньших. Таким образом, каждый вариант s характеризуется значениями $K_i(s)$ всех критериев, формирующих вектор оценок этого варианта, или его векторную оценку $x(s) = (K_1(s), \dots, K_m(s))$. Следовательно, варианты сравниваются по предпочтительности посредством сопоставления их векторных оценок. Множество всех векторов оценок $X = X_1 \times \dots \times X_m$. Мы предполагаем далее, что критерии являются однородными, т. е. имеют одинаковую (общую) шкалу $X_0 = X_1 = \dots = X_m$ (так что $X = X_0^m$); более того, если критерий K_j заменить на $x(K_j)$, где x – некоторое допустимое преобразование, определяемое типом шкалы, то и все остальные критерии K_i следует заменить на $x(K_i)$. Примем также, что множество X_0 конечно: $X_0 = \{1, \dots, q\}$. Элементы этого множества будем называть шкальными градациями.

Предпочтения ЛПП моделируются отношением предпочтения R на X : xRy означает, что вектор оценок x не менее предпочтителен, чем y . Отношение R порождает отношения безразличия I и (строгого) предпочтения P : xIy имеет место, когда справедливо xRy и yRx ; xPy выполнено, когда верно xRy и неверно yRx .

В данной работе рассматриваются отношения нестрогого предпочтения Парето P^0 ; отношение предпочтения количественной важности критериев с порядковой шкалой R^\ominus [5] и отношение предпочтения со шкалой первой порядковой метрики при замедлении роста предпочтений вдоль шкалы $R^{\ominus\&D}$ [6].

Поясним представленные отношения предпочтения. Отношение R^\ominus определяет более предпочтительный вариант на основании информации о значении критериев для данных вариантов и важности критериев. Стоит подчеркнуть, что при этом не делается никаких предположений о характере роста предпочтений ЛППР при движении от меньших шкальных градаций к большим. В теории важности критериев разработаны решающие правила [5], позволяющие определить более предпочтительный вариант из двух на основании информации лишь о значении критериев для этих вариантов и важности критериев, или констатировать, что этой информации недостаточно для определения более предпочтительного из них. В последнем случае говорят, что варианты несравнимы по отношению предпочтения R^\ominus .

Отношение $R^{\ominus\&D}$ определяет более предпочтительный вариант на основании информации о значении критериев для данных вариантов, важности критериев и информации о «затухающем» характере роста предпочтений ЛППР при движении от меньших шкальных градаций к большим. Стоит подчеркнуть, что «сила» «затухания» роста предпочтений никак не ограничивается и не задается, известно лишь, что затухание роста существует. В теории важности критериев разработаны решающие правила [6], позволяющие определить более предпочтительный вариант из двух на основании информации о значении критериев для этих вариантов, важности критериев и информации, что предпочтения растут «с затуханием», или констатировать, что этой информации недостаточно для определения более предпочтительного из них. В последнем случае говорят, что варианты несравнимы по отношению предпочтения $R^{\ominus\&D}$.

Информацию о количественной важности критериев будем использовать в форме значений важности критериев – чисел β_i , $1 \leq i \leq m$, так как это более удобно для целей проводимого вычислительного эксперимента. Обобщенную (интегральную) функцию ценности будем задавать в аддитивном виде, сопос-

ставляя каждой шкальной градации k её ценность $v(k)$ (критерии предполагаются однородными, т.е. у них одна общая шкала).

Необходимо сделать некоторые пояснения относительно целесообразности исследования многокритериальных задач с однородными критериями. Когда аналитик сопоставляет по предпочтительности два варианта, он или делает это напрямую, определяя замещения, или опосредованно, сводя предпочтительности этих двух исходных вариантов к предпочтительности (или предпочтительностям) некоего промежуточного объекта (или объектов) и сопоставляет уже последние. Делается это тогда, когда сравнение исходных вариантов (альтернатив) оказывается проблематичным для ЛПР, отчасти из-за того, что они с познавательной точки зрения являются «слишком разными», слишком далекими друг от друга семантически для их прямого сопоставления через определение замещений. Тогда аналитик вынужден предложить ЛПР оценивать по предпочтительности некие «промежуточные» объекты, к которым могут быть сведены исходные варианты, тем самым несколько «сближая» изначально слишком (для использования замещений) семантически «далекие» варианты. Такими «промежуточными объектами» могут быть значения обобщенных критериев, а также варианты (альтернативы), у которых исходные шкалы критериев, по которым они изначально оценивались, сведены к одной общей шкале (т.е. варианты теперь сравниваются по однородным критериям).

Таким образом, получается, что для альтернатив, условно говоря, семантически близких, имеет смысл определять замещения для сравнения их по предпочтительности, а для семантически далеких одним из способов их некоторого «сближения» (для повышения адекватности сравнения по предпочтительности) является сведение критериев к общей шкале (к однородным критериям). В этом свете оказывается, что класс многокритериальных задач с однородными критериями является весьма обширным и представительным, так как выступает адекватным подходом для сравнения семантически неблизких альтернатив. Поэтому мы считаем, что исследование в данной работе многокритериальных задач именно с однородными критериями не является сильным сужением сферы применимости её выводов.

Для большей ясности необходимо привести пример: такой

подход к разграничению применения методов многокритериального отбора с определением замещений и методов решения многокритериальных задач с однородными критериями был использован в работе [11]. Там эти два подхода используются на разных этапах многокритериальной процедуры отбора прогнозных экспертных методов. Сначала производится отбор между различными экспертными методами, а затем на следующем этапе процедуры отбор между различными разновидностями экспертных методов. Выбирая между содержательно столь различными методом Дельфи и, скажем, методом морфологического анализа, ЛПР сложно определять замещения: непросто сказать, насколько прирост быстроты метода Дельфи по сравнению с морфологическим анализом может компенсировать его (Дельфи) меньшую вероятность (возможность) получения правильного ответа. Для ЛПР проще и адекватнее оказывается привести критерии быстроты и возможности правильного решения, так сказать, к общему знаменателю – свести критерии к однородным по определенным правилам и решать многокритериальную задачу с однородными критериями. Здесь дополнительная операция, выполняемая аналитиком, – сведение критериев к однородным – позволяет компенсировать изначальное содержательное (семантическое) «расстояние» между непохожими Дельфи и морфологическим анализом.

Напротив, когда на следующем этапе многокритериальной процедуры производится отбор, скажем, между разновидностями метода Дельфи – Дельфи I и Дельфи II, то здесь имеет смысл определять замещения и не сводить критерии к однородным, так как эти разновидности во многом похожи, близки друг другу. Аналитик/ЛПР сможет оценить, насколько выгоды от использования многоточечной экспертной оценки, «защитой» в одну из разновидностей Дельфи, могут компенсировать некоторую сложность ее применения.

Еще раз подчеркнем, что в свете вышеописанного мы считаем, что многокритериальные задачи с однородными критериями покрывают весьма обширный класс задач, и использование однородных критериев в вычислительном эксперименте и аналитических выкладках несильно сужают сферу применимости практических выводов данной статьи.

3. Описание вычислительного эксперимента с использованием отношений предпочтения R^0 , R^\ominus , $R^{\ominus \& D}$ и его результатов

Итак, проведем вычислительные эксперименты для получения ответов на вопросы, поставленных во введении. Сначала исследуем случай, когда используется отношение предпочтения Парето. В ходе проведения нового имитационного моделирования (вычислительного эксперимента) было решено 64 миллиона случайно сгенерированных многокритериальных задач. Вычислительный эксперимент строился следующим образом. Фиксировалась размерность многокритериальной задачи некоторой размерности (например, 7 вариантов и 5 критериев). Генератор случайных чисел выдавал балльные векторные оценки для каждого из вариантов. Баллы использовались от 2 до 5 по аналогии со школьными оценками, понятными и знакомыми большинству ЛПП¹.

Появление любой из балльных оценок по каждому критерию было равновероятным. После этого генератором случайных чисел определялась важность критериев. Она определялась не через коэффициенты важности, а через N -модель [4]. Если (для простоты интерпретации) представлять одну из многокритериальных задач, решаемых в эксперименте, задачей выбора лучшего студента по его оценкам по разным предметам, тогда элемент N -модели может иметь следующий смысл: это число равноважных разделов в данном предмете. Также предполагалось, что предмет может иметь любое число «разделов» от 1 до максимального с равной вероятностью. При решении многокритериальных задач одной размерности максимальное число разделов в предмете считалось постоянным. Таким образом, гене-

¹ В данном исследовании, как и в работе [10], используется «школьная» пятибалльная шкала («урезанная» до четырехбалльной). Более подробно возможность ее использования обсуждается в [10]. Для других стран результаты, полученные на ее основе, могут быть менее репрезентативными.

ратором случайных чисел определялись значения важности критериев, а не коэффициенты их важности.

Если удастся выделить единственный недоминируемый вариант, т. е. лучший, (точнее, один недоминируемый класс эквивалентности вариантов), это значит, что имеющейся информации для данной задачи достаточно и функцию ценности можно не строить.

Решался миллион таких задач одной размерности (случайно сгенерированных). И определялась доля случаев, когда для задачи данной размерности использование соответствующей информации достаточно для выделения одного лучшего (с точностью до эквивалентности) варианта.

Предполагается, что отношение Парето является «очевидным», точнее очевидным для ЛПР является, доминирует ли какой-либо вариант над другим по Парето или нет. В то время как для других, более «сильных» отношений предпочтения почти всегда является не очевидным, доминирует ли данный вариант по ним над другим вариантом. То есть если есть два варианта, один оценивается по двум критериям на «отлично» и «отлично», а второй на «удовлетворительно» и «плохо», очевидно без построения функции ценности, что первый вариант лучше. Поэтому для того чтобы определить долю случаев, когда приходится строить функцию ценности в традиционном подходе теории полезности, надо от ста процентов отнять долю случаев, когда отношения Парето достаточно для определения единственного (с точностью до эквивалентности) недоминируемого варианта. Вычислительный эксперимент для определения такой доли случаев был проведен в данном исследовании, его результаты представлены в таблице 1.

Для определения того, в какой доле случаев придется строить функцию ценности при использовании решающих правил теории важности критериев, нужно от ста процентов отнять долю случаев, когда отношения предпочтения R^{\ominus} или $R^{\ominus \& D}$ достаточно для определения единственного (с точностью до эквивалентности) недоминируемого варианта. Вычислительные эксперименты по определению вышеуказанных долей были проведены в работе [10], и для удобства читателя мы продублируем результаты из того эксперимента в таблицах 2 и 3 соответственно.

Таблица 1. Доля случаев, когда отношения Парето достаточно для определения одного (с точностью до эквивалентности) недоминируемого варианта

Число вариантов	Число критериев							
	2	3	4	5	6	7	8	9
2	0,718	0,473	0,301	0,190	0,119	0,075	0,047	0,029
3	0,589	0,298	0,143	0,068	0,032	0,015	0,007	0,003
4	0,533	0,229	0,092	0,036	0,014	0,006	0,002	0,001
5	0,511	0,197	0,070	0,024	0,008	0,003	0,001	0,000
6	0,503	0,182	0,060	0,019	0,006	0,002	0,001	0,000
7	0,506	0,176	0,055	0,016	0,005	0,002	0,000	0,000
8	0,515	0,174	0,052	0,015	0,004	0,001	0,000	0,000
9	0,529	0,176	0,051	0,015	0,004	0,001	0,000	0,000

Таблица 2. Доля случаев, когда количественной важности критериев достаточно для определения одного (с точностью до эквивалентности) недоминируемого варианта; 5 различных значений важности критериев

Число вариантов	Число критериев							
	2	3	4	5	6	7	8	9
2	0,850	0,738	0,669	0,627	0,602	0,585	0,574	0,566
3	0,778	0,624	0,536	0,484	0,454	0,435	0,422	0,415
4	0,754	0,574	0,472	0,413	0,380	0,359	0,345	0,336
5	0,753	0,554	0,439	0,374	0,336	0,313	0,298	0,287
6	0,763	0,549	0,422	0,350	0,307	0,281	0,265	0,255
7	0,778	0,550	0,412	0,333	0,288	0,259	0,243	0,231
8	0,796	0,554	0,406	0,323	0,273	0,243	0,224	0,212
9	0,813	0,560	0,404	0,315	0,263	0,230	0,210	0,197

Напомним, что варианты сравниваются по отношению предпочтения R^0 , когда информации о скорости роста предпочтений вдоль шкалы критериев нет и, следовательно, можно применить методы теории количественной важности критериев с порядковой шкалой.

И если предполагать наличие информации о том, что предпочтения ЛПП растут вдоль шкалы (при движении от меньших

градаций к большим) с затуханием (так называемая информация D), то можно применить методы теории количественной важности критериев со шкалой первой порядковой метрики. То есть варианты в этом случае будут сравниваться по отношению нестрогого предпочтения $R^{\Theta \& D}$.

Таблица 3. Доля случаев, когда количественной важности критериев и информации D достаточно для определения одного (с точностью до эквивалентности) недоминируемого варианта; 5 различных значений важности критериев

Число вариантов	Число критериев							
	2	3	4	5	6	7	8	9
2	0,938	0,888	0,863	0,832	0,813	0,816	0,806	0,807
3	0,903	0,831	0,783	0,743	0,731	0,720	0,710	0,713
4	0,898	0,801	0,739	0,698	0,660	0,649	0,637	0,633
5	0,890	0,788	0,705	0,659	0,635	0,609	0,594	0,585
6	0,897	0,773	0,696	0,637	0,603	0,581	0,551	0,558
7	0,907	0,778	0,691	0,628	0,589	0,564	0,535	0,530
8	0,910	0,787	0,694	0,618	0,575	0,539	0,517	0,508
9	0,920	0,781	0,683	0,605	0,550	0,533	0,501	0,481

Расчеты проводились при помощи специально разработанной компьютерной программы.

Необходимо отдельно остановиться на вышеприведенных в тексте предположениях относительно равномерного распределения некоторых величин, участвующих в вычислительном эксперименте и при доказательстве теорем. Мы считаем, что отличие реального распределения этих величин от равномерного хоть и влияет на результаты, полученные в эксперименте, и на формулы в теоремах, но не настолько, чтобы изменить принципиальный вывод данной статьи о том, что снижение затрат на построение функции ценности, получаемое при использовании решающих правил теории важности критериев, является весьма существенным. Доказанные теоремы в их текущем виде, хоть и даны для частного подкласса случаев, но все равно хорошо подходят для ориентировочного определения величины этого снижения затрат.

И, наконец, несколько слов относительно возможности использования статистического подхода в настоящем исследовании. Мы понимаем, что каждая практическая задача многокритериального отбора по-своему уникальна, и нередко бывают случаи, когда аналитик и ЛПР собираются разово для решения данной многокритериальной задачи. Но также считаем, что довольно часто бывает ситуация, когда одной группе экспертов, аналитиков, ЛПР приходится решать типичные для некоторой предметной области задачи многокритериального отбора на системной основе. Таким задачами являются отбор перспективных для выполнения проектов, отбор контрагентов, подрядчиков и многие другие. И для каждой задачи может быть выбран в общем случае свой многокритериальный метод, когда-то более подходящий и удачный, когда-то менее. И в среднем деятельность этой рабочей группы по отбору (проектов, подрядчиков и т.д.) может быть более или менее эффективной. Поэтому знание о возможности снижения затрат может повысить эту среднюю эффективность на системной основе. В этой связи считаем, что использование статистического подхода в данном исследовании оправданным, хотя и признаем существование отдельных уникальных многокритериальных задач.

4. Количественная оценка снижения затрат на построение функции ценности при использовании решающих правил теории важности критериев

Определим, во сколько раз снижаются затраты на построение функции ценности при использовании решающих правил теории важности критериев по сравнению с традиционным подходом теории полезности. Другими словами, во сколько раз снижается доля случаев, когда не удастся определить единственный недоминируемый вариант по отношению P^{\ominus} и $P^{\ominus \& D}$ по сравнению долей случаев, когда не удастся определить единственный недоминируемый вариант по отношению Парето. Предполагается, что операция определения доминированности по отношению Парето является «очевидной» для ЛПР в отличие от отношений P^{\ominus} и $P^{\ominus \& D}$.

Например, для случая отбора из трех вариантов на основании трех критериев вероятность определить наилучший вариант по отношению Парето – 29,8%, а по отношению P^{\ominus} – 62,4%. То есть при следовании традиционному подходу теории полезности функцию ценности придется строить в 100 – 29,8 = 70,2% случаев, а при использовании решающего правила теории важности критериев, не налагающего никаких ограничений на характер возрастания предпочтений ЛПР вдоль шкалы, 100 – 62,4 = 37,6% случаев. Получается, что доля случаев, когда нужно тратить время на построение функции ценности уменьшается в $70,2/37,6 \approx 1,87$ раза! Аналогично, можно посчитать, что если для данной балльной шкалы критериев верен закон «убывающей предельной полезности» (полезности/ценности для ЛПР), то эти затраты времени по сравнению с традиционным подходом сократятся в $(100 - 29,8)/(100 - 83,1) \approx 4,15$ раза! Результаты, полученные таким образом, представлены в таблицах 4 и 5.

Таблица 4. Отношение долей случаев, когда функцию ценности придется строить в подходе теории полезности и при использовании количественной важности критериев; 5 различных значений важности критериев

Число вариантов	Число критериев							
	2	3	4	5	6	7	8	9
2	1,880	2,011	2,112	2,172	2,214	2,229	2,237	2,237
3	1,851	1,867	1,847	1,806	1,773	1,743	1,718	1,704
4	1,898	1,810	1,720	1,642	1,590	1,551	1,524	1,505
5	1,980	1,800	1,658	1,559	1,494	1,451	1,423	1,403
6	2,097	1,814	1,626	1,509	1,434	1,388	1,359	1,342
7	2,225	1,831	1,607	1,475	1,397	1,347	1,321	1,300
8	2,377	1,852	1,596	1,455	1,370	1,320	1,289	1,269
9	2,519	1,873	1,592	1,438	1,351	1,297	1,266	1,245

Таблица 5. Отношение долей случаев, когда функцию ценности придется строить в подходе теории полезности и при использовании количественной важности критериев с информацией D ; 5 различных значений важности критериев.

Число вариантов	Число критериев							
	2	3	4	5	6	7	8	9
2	4,548	4,705	5,102	4,821	4,711	5,027	4,912	5,031
3	4,237	4,154	3,949	3,626	3,599	3,518	3,424	3,474
4	4,578	3,874	3,479	3,192	2,900	2,832	2,749	2,722
5	4,445	3,788	3,153	2,862	2,718	2,550	2,461	2,410
6	4,825	3,604	3,092	2,702	2,504	2,382	2,225	2,262
7	5,312	3,712	3,058	2,645	2,421	2,289	2,151	2,128
8	5,389	3,878	3,098	2,579	2,344	2,167	2,070	2,033
9	5,888	3,763	2,994	2,494	2,213	2,139	2,004	1,927

5. Аналитический вывод

Для класса ситуаций отбора из двух вариантов на основе двух критериев произведем следующие аналитические выводы, представленные двумя нижеследующими теоремами и следствиями из них. Для других классов ситуаций получить применением одной формулы числа, приведенные в таблицах 4 и 5, пока не удастся.

Теорема 1. Пусть l – максимальное число различных значений важности критериев; n – число вариантов; m – число критериев, а $Q(n, m, l)$ – доля случаев, когда функцию ценности можно не строить (без наложения ограничений на скорость роста предпочтения для ЛПР). Тогда

$$Q(2, 2, l) = \frac{107l + 9}{128l}.$$

Доказательство. Доказательство основано на переборе сопоставлений всех векторных оценок с четырьмя градациями

друг с другом по отношению предпочтения P^\ominus . Это сопоставление основывается на применении решающего правила, приведенного в [5]. Рассмотрим его применение на примере сравнения по предпочтительности двух векторных оценок $s_1 = (5; 3)$ и $s_2 = (4; 4)$. Предположим, что значения важности критериев есть $x > 0$ и $y > 0$. Их сравнение по отношению предпочтения P^\ominus сводится к сравнению по отношению Парето оценок $e_1 = (x; x; x + y)$ и $e_2 = (0; x + y; x + y)$. Эти оценки получены следующим образом: первая компонента векторной оценки показывает, есть ли в исходной векторной оценке компонента, равная 5. Если она есть, то она учитывается столько раз, каково значение важности критерия. Вторая компонента оценки e показывает, есть ли в векторной оценке s компонента, большая или равная 4. Учитывается столько раз, каково значение важности критериев, где она присутствует. Аналогично определяется третья компонента векторной оценки e . Легко видеть, что оценки e_1 и e_2 не сравнимы по Парето, из этого следует, что оценки s_1 и s_2 не сравнимы по отношению P^\ominus . Результаты, полученные таким образом для случая, когда $x > y$, занесены в таблицу, изображенную на рисунке в Приложении 2. А для ситуации, когда важности двух критериев равны между собой – в Приложении 1. Темно-серым цветом выделены ячейки, соответствующие случаям, когда варианты несравнимы по данному отношению предпочтения.

Таким образом, если важности критериев равны между собой, то доля случаев, когда функцию ценности можно не строить, будет

$$1 - \frac{12 \times 2}{16^2} = 1 - \frac{3}{32} = \frac{29}{32}.$$

А если важности критериев не равны, то

$$1 - \frac{21 \times 2}{16^2} = 1 - \frac{21}{128} = \frac{107}{128}.$$

Предполагаем, что важность критерия представляется через значения важности критериев и принимает значения от 1 до l , тогда вероятность того, что значение важности второго критерия совпадет с первым (т.е. важности критериев будут равны) будет $1/l$, соответственно, что важности критериев будут не равны – $(l - 1)/l$.

Таким образом:

$$Q(2,2,l) = \frac{1}{l} \times \frac{29}{32} + \frac{l-1}{l} \times \frac{107}{128} = \frac{4 \times 29 + (l-1) \times 107}{128l} = \\ = \frac{116 + 107l - 107}{128l} = \frac{107l + 9}{128l}.$$

Что и требовалось доказать.

Следствие 1 из теоремы 1. Пусть l – максимальное число различных значений важности критериев; n – число вариантов; m – число критериев; $Q(n, m, l)$ – доля случаев, когда функцию ценности можно не строить. Тогда $Q(2, 2, l)$ заключено в следующих пределах:

$$\frac{107}{128} \leq Z(2, 2, l) < \frac{29}{32}.$$

Следствие 2 из теоремы 1. Сравнение результатов теоремы 1 и экспериментальных расчетов.

$$Q(2, 2, 5) = \frac{17}{20} = 0,85,$$

что в точности совпадает с опубликованными экспериментальными данными [10].

$$Q(2, 2, 9) = \frac{27}{32} = 0,84375.$$

В работе [9] было опубликовано значение, округленное до трех знаков после запятой, 0,844. Отметим также, что теорема 1 позволяет не только получить данные, совпадающие с результатами вычислительного эксперимента для 5-ти и 9-ти различных значений важности критериев, но и получить результаты для любого числа различных значений важности критериев. В этом смысле эта теорема дает более общие результаты, нежели вычислительный эксперимент.

Следствие 3 из теоремы 1. Сравнение результатов теоремы 1 и ранее сделанных предположений: в статье [10] по результатам экспериментов было сделано предположение, что число различных значений важности критериев не влияет на содержательные выводы статьи, поэтому, где не указано обратное, далее использовалось 5 различных значений важности критериев. Теперь это показано аналитически. Действительно, при

любом числе различных значений важности критериев доля случаев для двух вариантов и двух критериев, когда функцию ценности можно не строить колеблется от примерно 0,836 до 0,906, что не меняет содержательных выводов.

Теорема 2. Для данного класса ситуаций снижение затрат $Z(2, 2, l)$ на построение функции ценности при использовании решающих правил Подиновского (без наложения ограничений на скорость роста предпочтения ЛПР) по сравнению с традиционным подходом теории полезности составляет

$$Z(2, 2, l) = \frac{12l}{7l-3}.$$

Доказательство.

Согласно перебору сравнения вариантов по отношению Парето, приведенному в таблице в Приложении 3 вероятность того, что при данных условиях взятия отношения предпочтения Парето будет достаточно для определения единственного недоминируемого варианта (с точностью до эквивалентности) будет равна:

$$1 - \frac{36 \times 2}{256} = 1 - \frac{9}{32} = \frac{23}{32}.$$

В этой доле случаев строить функцию ценности не придется.

А отношение доли случаев, когда функцию ценности придется строить в традиционном подходе теории полезности и при использовании решающего правила теории важности критериев (без наложения ограничений на скорость роста предпочтения ЛПР) будет

$$\begin{aligned} Z(2, 2, l) &= \frac{1 - \frac{23}{32}}{1 - \frac{107l+9}{128l}} = \frac{9}{32} \div \frac{128l-107l-9}{128l} = \\ &= \frac{9}{32} \times \frac{128l}{21l-9} = \frac{3}{1} \times \frac{4l}{7l-3} = \frac{12l}{7l-3}. \end{aligned}$$

Что и требовалось доказать.

Следствие 1 из теоремы 2. Для различных значений l (максимального значения различных значений важности критериев) $Z(l)$ заключено в пределах:

$$1,714 \approx \frac{12}{7} < Z(l) \leq 3.$$

Следствие 2 из теоремы 2.

$$Z(5) = \frac{60}{32} = \frac{15}{8} = 1,875,$$

что совпадает с экспериментально полученными данными в этой работе с точностью до ошибок округления.

6. Выводы

При принятии несильных допущений, приведенных ранее в тексте для многокритериальных задач с однородными критериями с числом вариантов от 5 до 9 и числом критериев от 5 до 9:

1. Применение решающих правил теории важности критериев позволяет снизить затраты времени на построение функции ценности в 2–5 раз.

2. Одно лишь предположение о том, что закон убывающей «предельной полезности» верен для имеющейся в данной задаче балльной шкалы, позволяет существенно снизить затраты на построение функции ценности – преимущество перед теорией полезности растет примерно с 1,2–2 раз до 2–5 раз. Это обуславливает актуальность поиска операциональных способов констатирования наличия затухания скорости роста предпочтений ЛПП вдоль балльной шкалы.

Проведенное исследование позволяет более обоснованно применять теорию важности критериев как эффективный инструмент анализа многокритериальных задач принятия решений.

Литература

1. БАРЫШНИКОВ Ю. М. *О среднем числе вариантов, недоминируемых по сравнению В.В. Подиновского* // Автоматика и телемеханика. – 1990. – №6. – С. 161–167.

2. БУРКОВ В.Н., КОРГИН Н.А., НОВИКОВ Д.А. *Введение в теорию управления организационными системами: Учебник* / Под ред. Д.А. Новикова. – М.: Книжный дом «ЛИБРОКОМ», 2009. – 264 с.
3. КИНИ Р.Л., РАЙФА Х. *Принятие решений при многих критериях: предпочтения и замещения* / Пер. с англ. – М.: Радио и связь, 1981.
4. ПОДИНОВСКИЙ В.В. *Введение в теорию важности критериев в многокритериальных задачах принятия решений.* – М.: Физматлит, 2007.
5. ПОДИНОВСКИЙ В.В. *Количественная важность критериев* // Автоматика и телемеханика. – 2000. – №5. – С. 110–123.
6. ПОДИНОВСКИЙ В.В. *Количественная важность критериев с дискретной шкалой первой порядковой метрики* // Автоматика и телемеханика. – 2004. – №8. – С. 196–203.
7. ПОДИНОВСКИЙ В.В. *Об оценке эффективности решающих правил в многокритериальных задачах* // Изв. АН СССР. Техн. кибернетика. – 1987. – №1. – С. 3–9.
8. ПОДИНОВСКИЙ В.В. *Оценка эффективности решающих правил в дискретных многокритериальных задачах* // Методы оптимизации в экономико-математическом моделировании. – М.: Наука, 1991. – С. 308–324.
9. САЛТЫКОВ С.А. *Экспериментальное сопоставление различных многокритериальных подходов* // Материалы XXXVI Международной конференции «Информационные технологии в науке, социологии, экономике и бизнесе», Ялта, Гурзуф, май 2009 г. Приложение к журналу «Открытое образование». – С. 315–317.
10. САЛТЫКОВ С.А. *Экспериментальное сопоставление методов взвешенной суммы, теории полезности и теории важности критериев для решения многокритериальных задач с балльными критериями* // Управление большими системами. – 2010. – Вып. 29. – С. 16–41.
11. СИДЕЛЬНИКОВ Ю.В., САЛТЫКОВ С.А. *Процедура отбора наиболее приемлемых разновидностей экспертных методов* // Управление большими системами. – 2010. – Вып. 30. – С. 35–66.

12. DYER J.S. *MAUT – Multiattribute Utility Theory* // Multiple Criteria Decision Analysis: State of the Art Surveys. Springer Science + Business Media, Inc., 2005. – P. 265 – 295.

REDUCING COSTS OF CONSTRUCTION OF VALUE FUNCTION USING DECISION RULES OF CRITERIA IMPORTANCE THEORY AS COMPARED WITH TRADITIONAL APPROACH OF UTILITY THEORY

Sergey Saltykov, Institute of Control Sciences of RAS, Moscow, junior researcher, (ssaltykov@mail.ru).

Abstract: According to the results of a computational experiment using decision rules of criteria importance theory can reduce 2–5 times the time spent on constructing a value function compared with a traditional approach of utility theory. For the case of two-variant and two-criterion choice the theorem is proved to quantify how many times criteria importance theory allows to reduce the proportion of cases when the value must be constructed.

Keywords: criteria importance theory, value function, reducing costs.

Статья представлена к публикации членом редакционной коллегии А. И. Орловым

Приложение 1

	5; 5	5; 4	5; 3	5; 2	4; 5	4; 4	4; 3	4; 2	3; 5	3; 4	3; 3	3; 2	2; 5	2; 4	2; 3	2; 2
5; 5	k;k;k;k k;k;k;k	k;k;k;k k;k;k;k	k;k;k;k k;k;k;k	k;k;k;k k;k;k;k	k;k;k;k k;k;k;k	k;k;k;k k;k;k;k	k;k;k;k k;k;k;k	k;k;k;k k;k;k;k	k;k;k;k k;k;k;k	k;k;k;k k;k;k;k	k;k;k;k k;k;k;k	k;k;k;k k;k;k;k	k;k;k;k k;k;k;k	k;k;k;k k;k;k;k	k;k;k;k k;k;k;k	k;k;k;k k;k;k;k
5; 4																
5; 3																
5; 2																
4; 5																
4; 4																
4; 3																
4; 2																
3; 5																
3; 4																
3; 3																
3; 2																
2; 5																
2; 4																
2; 3																
2; 2																

	5,5;4	5,3	5,2	4,5	4,4	4,3	4,2	3,5	3,4	3,3	3,2	2,5	2,4	2,3	2,2
5,5	хххххххх	хххххххх	ххх	хххххххх	хххххххх	хххххххх	ххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх
5,4		хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх
5,3		хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх
5,2				хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх
4,5					хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх
4,4					хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх
4,3						хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх
4,2								хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх
3,5								хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх
3,4									хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх
3,3										хххххххх	хххххххх	хххххххх	хххххххх	хххххххх	хххххххх
3,2											хххххххх	хххххххх	хххххххх	хххххххх	хххххххх
2,5												хххххххх	хххххххх	хххххххх	хххххххх
2,4													хххххххх	хххххххх	хххххххх
2,3														хххххххх	хххххххх
2,2															хххххххх

Приложение 3

	5,5	5,4	5,3	5,2	4,5	4,4	4,3	4,2	3,5	3,4	3,3	3,2	2,5	2,4	2,3	2,2
5,5	Black															
5,4	Grey	Black														
5,3		Black	Black													
5,2		Black	Black	Black												
4,5			Grey													
4,4			Grey		Black											
4,3					Grey	Black										
4,2						Black	Black	Black								
3,5								Grey								
3,4									Grey							
3,3										Grey						
3,2											Grey	Black	Black	Black		
2,5													Grey			
2,4														Grey		
2,3															Grey	
2,2																Grey

УДК 021.8 + 025.1
ББК 78.34

АЛГЕБРАИЧЕСКОЕ ПРЕДСТАВЛЕНИЕ МОДЕЛИ МНОГОАГЕНТНЫХ СЕТЕЙ

Юдицкий С. А.¹

(Учреждение Российской академии наук
Институт проблем управления РАН, Москва)

Предлагается алгебраический аппарат описания ориентированных графов в виде строки символов, дополненной нумерованными вертикальными стрелками (язык слогов), предназначенный для отображения статической составляющей модели многоагентных сетей. Динамическая составляющая модели описывается алгебраическими индикаторными выражениями на основе рекуррентных уравнений (продукционных правил и правил алгебры логики). Введены операции над слогами. Рассматривается индивидуальная и коллективная деятельность агентов в многоагентной сети, в том числе взаимодействия между агентами.

Ключевые слова: многоагентная сеть, индивидуальная и коллективная деятельность агентов, триадная модель агента, слог, операции над слогами, индикаторное выражение, индикаторная логическая формула.

1. Введение

Статья является развитием работ [9, 10] и посвящена применению языка слогов (Статических Линейных Описаний Графов) и индикаторных выражений при моделировании деятельности многоагентных сетей (в дальнейшем именуемых просто сетями), представляющих собой множество взаимосвязанных агентов [2] (субъектов или объектов), в котором:

¹ Семен Абрамович Юдицкий, доктор технических наук, профессор, главный научный сотрудник ИПУ РАН, yuseab@yandex.ru.

- каждый агент сохраняет свою индивидуальность, а именно имеет собственные (индивидуальные) цели, выполняет направленные на достижение этих целей индивидуальные действия, характеризуется индивидуальными показателями;
- связанность агентов заключается в том, что их деятельность может координироваться во времени, и в определенные моменты они могут передавать друг другу ресурсы;
- результатом индивидуальной деятельности агентов является достижение определенных коллективных целей и определенная динамика коллективных показателей.

Задача синтеза сети в общем виде формулируется следующим образом. Известен набор агентов, представленных, например, «модернизированными» триадными структурами [10], отображающими взаимовлияние подструктур индивидуальных целей, действий, показателей (термин «триадный» определяет единство и взаимозависимость трех базовых концептов сети – целей, действий, показателей), и заданы коллективные цели и показатели. Требуется определить, можно ли путем организации связей между агентами создать сеть, в которой наряду с индивидуальными целями достигались бы и желаемые коллективные цели при допустимых значениях индивидуальных и коллективных показателей. Если желаемых целей достичь нельзя, то какие коллективные цели реально достижимы? Аналитическое решение задачи синтеза сети с получением соответствующих необходимых и достаточных условий представляется чрезвычайно трудным и может стать предметом специальной работы. В данной статье автор ограничивается построением модели, позволяющей подбирать параметры сети и проверять эффективность такого подбора с помощью имитационного моделирования.

Триадная модель агента строится в два этапа. Вначале создается статическая составляющая модели в виде двудольных ориентированных графов с вершинами-позициями, соответствующими параметрам агента, и вершинами-переходами, выражающими отношения на множестве параметров (в двудольном

графе дуги могут проводиться только от позиций к переходам и обратно).

Затем формируется динамическая составляющая модели путем «нагружения» переходов графа рекуррентными уравнениями вида $S(\tau + 1) = F(S(\tau), V(\tau))$, где $\tau = 0, 1, \dots, N$ – дискретное время, $S(\tau)$, $V(\tau)$ – состояние агента и внешнее воздействие на него в момент τ соответственно, F – функциональное преобразование (отображение). Символы S , V представляются индикаторными выражениями [10] на основе правил алгебры логики [1, 6], преобразование F отображается формулой ЕСЛИ – ТО (продукцией [4]).

Таким образом, статическая составляющая представляется в графовой форме, а динамическая – в алгебраической, что делает модель агента и сети неоднородной. Кроме того, сложность (а тем самым и наглядность) графового описания экспоненциально возрастает при увеличении его размерности. Следовательно, для моделирования статической составляющей имеет смысл применить алгебраическое описание, линейно зависящее от размерности агентов и сети. В работе для этого введен формальный язык слогов, отображающих ориентированный граф в виде строки, составленной из символов вершин (применительно к двудольному графу – из чередующихся символов позиций и переходов) и нумерованных вертикальных стрелок. Описание структуры показателей базируется на арифметических и алгебраических нотациях (сложение, вычитание, умножение, деление, возведение в степень, логарифмирование и т.д.).

Раздел 2 статьи посвящен моделированию статистики многоагентных сетей. Дается пояснение к технике слогов, показано как переходить от слогов к графам и обратно. Введены операции над слогами, позволяющие выполнять их преобразования. В разделе 3 рассматривается моделирование динамики индивидуальной и коллективной деятельности агентов сети на основе индикаторных выражений [10], в том числе путем установления межагентных временных и ресурсных связей.

2. Статическое моделирование сетей на языке слогов

Алгебраическое описание проиллюстрируем на примере сети, состоящей из двух агентов (табл. 1). В строках 1, ..., 4 даны соответственно двудольные слоги для коллективной цели сети, индивидуальных целей агентов, индивидуальных действий агентов и взаимовлияния индивидуальных показателей. В строках 5, 6 приведены арифметические формулы для интегральных показателей агентов и сети в целом.

Таблица 1. Табличная форма статической модели многоагентной сети.

1. Слог для интегральной цели сети	$\downarrow^1 x_1 \downarrow^2 e_1 \uparrow^1 \uparrow^3 x! \downarrow^3 x_2 \uparrow^2$	
2. Слоги для целей агентов	$\downarrow^1 c_1 q_1 \uparrow^1 \downarrow^2 \downarrow^6 c_4 \downarrow^4 q_3 \uparrow^2 \uparrow^5$ $x_1! \downarrow^3 c_2 q_2 \uparrow^3 \uparrow^6! \downarrow^5 c_3 \uparrow^4$	$\downarrow^1 c_5 q_4 \uparrow^1 \downarrow^2 x_2! \downarrow^3 c_6 q_5 \uparrow^2 \uparrow^3$
3. Слоги для действий агентов	$\downarrow^1 d_0 t_1 \uparrow^2 d_1 t_2 d_3 \downarrow^3 \uparrow^4 t_4 \uparrow^1 \downarrow^2$ $d_2 t_3 d_4 \uparrow^3 \downarrow^4 t_5 d_5!$	$\downarrow^1 d_6 t_6 d_7 \uparrow^2 t_7 d_8 t_9 \uparrow^1 \downarrow^2 t_8 d_9!$
4. Слоги взаимовлияния показателей агентов	$\downarrow^1 \downarrow^2 p_1 r_1 p_2 r_2 \downarrow^3 p_3 r_3 \downarrow^4 p_4 r_4 \uparrow$ $\uparrow^3 p_5 r_5 \uparrow^2 \uparrow^4$	$\downarrow^1 p_6 r_6 p_7 r_7 \uparrow^1$
5. Структура показателей агентов	$V_1 = p_1 + p_2 + p_3 -$ $- \alpha(p_4 + p_5)$	$V_2 = p_6^2 + p_7^2$
6. Интегральный показатель сети	$V = v_1 + v_2$	

Слог взаимно однозначно отображает ориентированный граф – вершины, дуги, функцию инциденций. Примем, что вершины обозначаются символами из алфавита $A = \{a_i, i = 1, \dots, n\}$. Если из вершины a_i исходят дуги, ведущие в вершины

$a_{j_1}, a_{j_2}, \dots, a_{j_k}$, то первая дуга отображается последовательностью $a_i a_{j_1}$ символов в строке слога, остальные – упорядоченными парами пронумерованных вертикальных стрелок. Стрелки подразделяются на выходные, направленные вверх, и на входные, направленные вниз. Выходная стрелка проставляется справа от символа a_i , входные стрелки – слева от символов a_{j_2}, \dots, a_{j_k} . Пара стрелок, соответствующая одной дуге, нумеруется одинаковыми верхними индексами. Применительно к дуге $a_i a_{j_1}$ стрелки, находящиеся между a_i и a_{j_1} , во внимание не принимаются. Если таких стрелок несколько, то вначале проставляются подряд все выходные стрелки, а затем подряд все входные стрелки. Справа от терминальной вершины, из которой не исходит ни одной дуги, ставится знак «!». Если в слоге расположенные последовательно символы a_i, a_j не должны образовывать дугу на графе, то между ними также ставится знак «!».

Структуру слогов проиллюстрируем на примере формулы целей для агента 1, помещенной в строке 2 в среднем столбце табл. 1. Позиции (цели) c_i чередуются с переходами q_j . Через x_1 обозначена интегральная цель агента 1, являющаяся терминальной позицией слога. Следование символа q_1 непосредственно за c_1 отображает дугу $c_1 q_1$, пара стрелок с номером 1 – дугу $q_1 c_1$. В позицию c_4 ведут входные стрелки 2 и 6, стрелки с аналогичным номером выходят соответственно из переходов q_3, q_2 . Следовательно, слог отображает дуги $q_3 c_4, q_2 c_4$. В слоге вслед за переходом q_1 следует позиция c_4 (с точностью до разделяющих их стрелок), это говорит о наличии дуги $q_1 c_4$ и т.д. Знак «!» в интервале между q_2 и c_3 запрещает дугу $q_2 c_3$. Условием корректности слога является то, что номер каждой пары «выходная – входная стрелка» должен отличаться от номеров других пар стрелок.

Ввод слога в компьютер удобно выполнять на основе «технологии», позволяющей сократить число перенастроек режима клавиатуры. На первом шаге формируется строка символов (для двудольного слога – позиций и переходов) в виде букв с нижними индексами. На втором шаге расставляются стрелки, на третьем вводятся номера стрелок.

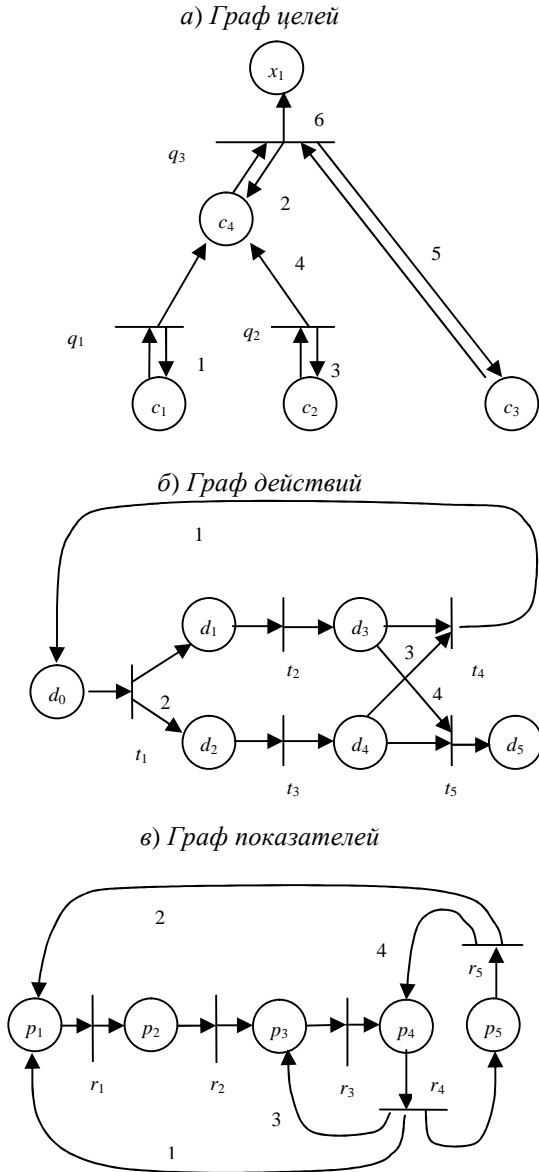


Рис. 1. Статическая составляющая триадной модели агента 1 в форме графов

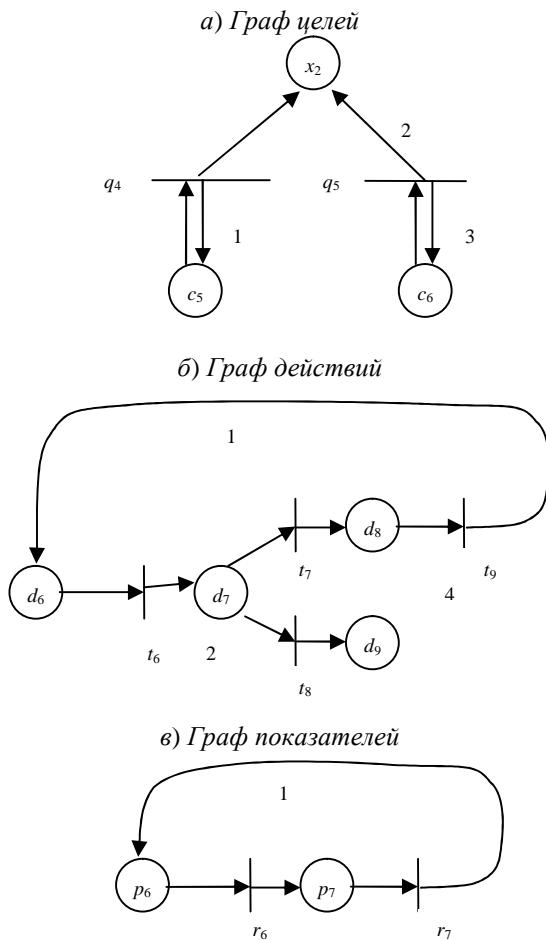


Рис. 2. Статическая составляющая триадной модели агента 2 в форме графов

Рассмотрим представление слогов, приведенных в табл. 1, в форме ориентированных графов. На рис. 1 и рис. 2 даны компоненты триадной модели – графы целей, действий, показателей соответственно для агентов 1 и 2, на рис. 3 показан граф для интегральной цели сети. Дуги этих графов, которые сопоставлены парам стрелок в слого, помечены номерами стрелок. Осталь-

ные дуги (отображают последовательности «позиция – переход» и «переход – позиция» в слоге) в графе не помечаются.

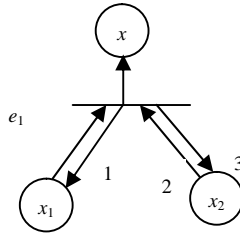


Рис. 3. Граф для интегральной цели многоагентной сети

Над слогами могут выполняться *алгебраические операции*, подобные описанным в [9]. Введем и проиллюстрируем на примерах набор таких операций.

1. *Конкатенация слогов* $S_1 S_2$, где для первого слога задан конечный символ (выделен буквой k в верхнем индексе), а для второго – начальный символ (выделен буквой n). Слоги не содержат совпадающих символов. Операция заключается в приписывании справа к первому слогу второго слога, и введении пары нумерованных стрелок, направленных от конечного символа слога S_1 к начальному символу слога S_2 . Слоги-операнды заключаются в круглые скобки. Пример выполнения конкатенации проиллюстрирован алгебраическим выражением (1), графовая интерпретация результата операции дана на рис 4а. Дуги графа, помеченные цифрами, соответствуют парам стрелок в слоге, непомеченные дуги – последовательному расположению символов. В правую часть соотношения (1) введена дополнительная пара стрелок с номером 5.

$$(1) (\downarrow^2 a_1 \uparrow^1 a_2^k \downarrow^1 a_3 \uparrow^2) (\downarrow^4 a_4^n \uparrow^3 a_5 \downarrow^3 a_6 \uparrow^4) \rightarrow \\ \rightarrow \downarrow^2 a_1 \uparrow^1 a_2 \uparrow^5 \downarrow^1 a_3 \uparrow^2 \downarrow^4 \downarrow^5 a_4 \uparrow^3 a_5 \downarrow^3 a_6 \uparrow^4$$

2. *Выделение подслога с заданным начальным и конечным символом*, в частном случае цикла длины $n > 1$. Слог в правой части (1) с началом и концом, соответственно a_1 и a_4 , содержит, например, подслоги $\{a_1 a_2 a_3 a_4\}$, $\{a_1 a_3 a_4\}$, $\{a_1 a_4\}$.

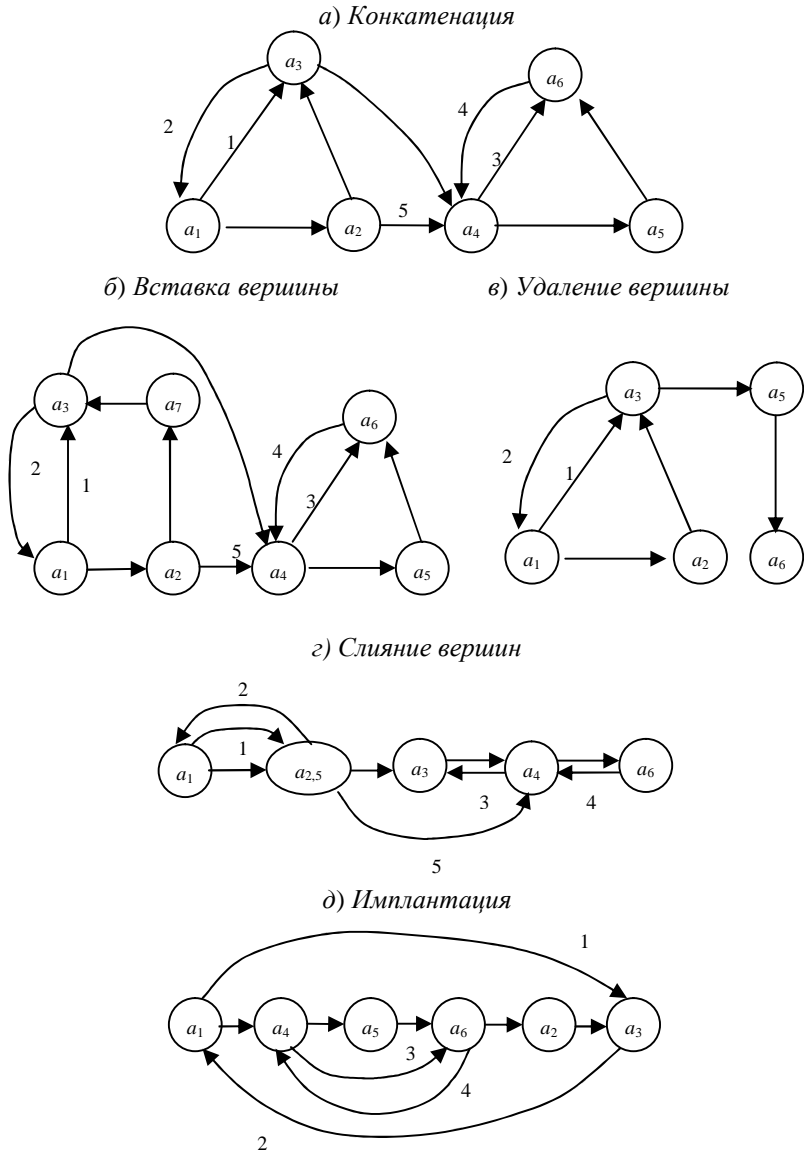


Рис. 4. Графовая интерпретация результатов операций над слогами

3. *Вставка вершины a_k в дугу $a_i a_j$* . Пример: в слове в правой части (1) выделяем подслог $a_2 a_3$, не содержащий знака «!», и в него между первым и вторым символами вставляем a_7 : $\downarrow^2 a_1 \uparrow^1 a_2 \uparrow^5 a_7 \downarrow^1 a_3 \uparrow^2 \downarrow^4 \downarrow^5 a_4 \uparrow^3 a_5 \downarrow^3 a_6 \uparrow^4$ (рис. 4б).

4. *Удаление вершины вместе с ведущими в нее и исходящими из нее парами стрелок*. Пример: в слове в правой части (1) удаляем символ a_4 и три пары стрелок с номерами 3, 4, 5, т.к. непосредственно слева от a_4 проставлены входные стрелки 4, 5, а справа – выходная стрелка 3: $\downarrow^2 a_1 \uparrow^1 a_2 \downarrow^1 a_3 \uparrow^2 a_5 a_6$ (рис. 4в).

5. *Слияние пары вершин $a_i, a_j \rightarrow a_{i,j}$* . Символ a_i заменяем символом $a_{i,j}$, а символ a_j удаляем. Непосредственно слева от $a_{i,j}$ дополнительно вставляем входные стрелки для a_j , а справа – выходные стрелки для a_j . В рассматриваемом примере слияние вершин a_2 и a_5 порождает слог: $\downarrow^2 a_1 \uparrow^1 a_{2,5} \downarrow^1 \downarrow^3 \uparrow^5 a_3 \uparrow^2 \downarrow^4 \downarrow^5 a_4 \uparrow^3 a_6 \uparrow^4$, которому соответствует граф на рис. 4г.

6. *Имплантация слогов $S_1 \times S_2$* . В слове S_1 косой чертой выделяется «место разреза», в которое вставляется слог S_2 :

$$(2) (\downarrow^2 a_1 \uparrow^1 / a_2 \downarrow^1 a_3 \uparrow^2) \times (\downarrow^4 a_4 \uparrow^3 a_5 \downarrow^3 a_6 \uparrow^4) \rightarrow \\ \rightarrow \downarrow^2 a_1 \uparrow^1 \downarrow^4 a_4 \uparrow^3 a_5 \downarrow^3 a_6 \uparrow^4 a_2 \downarrow^1 a_3 \uparrow^2 \text{ (рис. 4д).}$$

Из описанных выше операций конкатенация и имплантация являются бинарными, остальные – унарными. Приведенные операции – это лишь примеры возможных преобразований над слогами, список которых может быть продолжен.

В заключение раздела 2 выскажем некоторые соображения по развитию языка слогов:

- в статье операции определены на слогах, соответствующих классическим ориентированным графам. Следует модифицировать операции применительно к двудольным слогам (двудольным графам), являющимся основой статической составляющей триадной модели сетевых структур [10];
- оправдано построение операторов, выражаемых суперпозицией операций над слогами, и проведение преобразований с целью упрощения структуры этих операторов.

Указанные вопросы в статье не рассматривались.

3. Динамическое моделирование сетей на языке индикаторных выражений

Динамика сетей определяется динамикой внутреннего поведения агентов и динамикой внешнего поведения - взаимодействий между агентами, а также агентами и внешней средой, на заданном интервале (горизонте) дискретной временной шкалы $\tau = 0, 1, \dots, N$. При этом изменение параметров системы (достижение целей, смена действий, изменение величины показателей) происходит в моменты срабатывания вершин-переходов двудольного графа (слога), а значения параметров в промежутке между срабатыванием переходов фиксируется в вершинах-позициях. Переходы помечаются продуктами – выражениями вида «ЕСЛИ (выполнение условия), ТО (реализация оператора)», где условие и оператор описываются индикаторными логическими формулами – ИЛФ [6, 7]. Индикатором назван двухместный предикат $(x\#y)$, где x, y – числовые переменные, одна из которых может быть константой, интерпретируемые как значение дискретного времени, факт достижения цели либо выполнения действия, балльное значение показателя [3] и т.д., $\#$ – знак сравнения, принадлежащий множеству $\{=, \neq, >, \geq, <, \leq\}$. Продукции такого типа, сопоставленные переходам двудольного графа, далее будем называть индикаторными выражениями и описывать формулами причинно-следственной связи ИЛФ1 \rightarrow ИЛФ2.

Индикатор равен 1, если соотношение $\#$ выполняется, и равен 0, если нет. ИЛФ формируется на базе индикаторов путем применения к ним логических операций конъюнкции, дизъюнкции, отрицания. Переход двудольного графа срабатывает в момент τ , если в соответствующем ему индикаторном выражении ИЛФ1 = 1. В результате в момент $\tau + 1$ будет выполняться соотношение ИЛФ2 = 1, инициирующее реализацию оператора перехода.

Дадим примеры индикаторных выражений для переходов двудольных графов, описывающих компоненты триадной структуры агентов (рис. 1 и рис. 2). Условие ИЛФ1 и оператор ИЛФ2 будем разделять горизонтальной стрелкой.

Выражение для перехода $t4$ графа действий и перехода $r5$ графа показателей агента 1 имеет вид:

$$(3) \quad F(t4): (d3 = 1) \ \& \ (d4 = 1) \ \& \ (v = 0) \ \rightarrow \\ \rightarrow (d3 = 0) \ \& \ (d4 = 0) \ \& \ (d0 = 1),$$

где v – булева переменная, соответствующая внешнему воздействию,

$$(4) \quad \Pi(r5): (p5(\tau) > 7) \ \rightarrow \ (p1(\tau + 1) = \\ = p1(\tau) - 1) \ \& \ (p4(\tau + 1) = p4(\tau) - 1),$$

где значение показателей представлено функцией дискретного времени τ . Оператор индикаторного выражения (4) при наступлении момента $\tau + 1$ уменьшает на единицу значение показателей $p1, p4$ в момент τ .

Моделирование динамики внутреннего поведения агента на заданном временном горизонте производится на основе индикаторных выражений, отображающих триадную структуру агента, следующим образом:

- последовательно проверяем выполнение условий ИЛФ1 для индикаторных выражений всех переходов графов целей, действий и показателей агента;
- для тех выражений, условия которых выполнены, вычисляем оператор ИЛФ2 и находим значение выходных переменных.

Таким образом, применяется метод имитационного моделирования. Возможны два исхода имитационного эксперимента: либо доходим до конечного момента горизонта моделирования, не встречая нарушений в поведении агента, либо фиксируем нарушения, например:

- «зависание» агента, если выявлен момент, в который не выполняется условие ни для одного перехода;
- противоречивость, если некоторой переменной в один и тот же момент присваиваются разные значения;
- выход переменной за границы допустимого для нее интервала значений.

При обнаружении вышеуказанных, а возможно и иных нарушений в деятельности агентов и сети, производится корректировка триадной модели проекта (управление моделью) на

уровне слогов, графов, индикаторных выражений. Далее цикл «имитационное моделирование – управление» повторяется и т.д.

Перейдем к моделированию динамики внешнего поведения сети, определяемой взаимодействиями между агентами. Как уже говорилось во введении, два агента считаются связанными, если их индивидуальная деятельность координируется во времени и в определенные моменты они могут передавать друг другу ресурсы. Взаимодействие агентов в составе сети будем интерпретировать как проявление их связанности. Два перехода, относящиеся к разным агентам, назовем синхронизированными, если они срабатывают одновременно, и только одновременно (по терминологии [5] находятся в связке). Синхронизированными могут быть переходы графов действий и графов показателей агентов. Для синхронизации переходов соответственно с условиями ИЛФ_{*i*}, ИЛФ_{*j*} надо эти условия в индикаторных выражениях заменить конъюнкцией ИЛФ_{*i*} & ИЛФ_{*j*}. Если в момент срабатывания синхронизированных переходов, принадлежащих разным агентам, от одного из них надо передать другому ресурс *R*, то оператор передающего агента в этот момент вычитает из соответствующей переменной величину *R*, а оператор принимающего агента прибавляет *R*.

Проиллюстрируем взаимодействие агентов на примере. Пусть ресурс $R = 2$ передается от агента 1 при срабатывании перехода 5 (рис. 1), описываемого индикаторным выражением (4), а принимается агентом 2 при срабатывании синхронизированного с *r*5 перехода *r*6 (рис. 2), описываемого выражением:

$$(5) \quad П(r6): (p6(\tau) < 3) \rightarrow (p7(\tau + 1) = p7(\tau) + 1).$$

Выражения (4) и (5) заменяем соответственно на (4*), (5*):

$$(4^*) \quad П(r5): (p5(\tau) > 7) \& (p6(\tau) < 3) \rightarrow$$

$$\rightarrow (p1(\tau + 1) = p1(\tau) - 2) \& (p4(\tau + 1) = p4(\tau) - 2),$$

$$(5^*) \quad П(r6): (p5(\tau) > 7) \& (p6(\tau) < 3) \rightarrow (p7(\tau + 1) = p7(\tau) + 3).$$

Из выражений (4*) и (5*) следует, что при взаимодействии агентов в момент $\tau + 1$ из позиций *p*1 и *p*4 агента 1 изымается по одной единице ресурса, а их сумма (две единицы) вносится в позицию *p*7 агента 2.

4. Заключение

В статье изложен один из возможных подходов к построению «прозрачной» (наглядной) модели многоагентных сетей, предваряющей организацию управления такими системами. Основные идеи подхода в тезисной форме выражаются следующим образом.

1. Разделены статическая и динамическая составляющие модели, представленные обе алгебраическими конструкциями – слогами и индикаторными выражениями. Статическая составляющая является основой динамической составляющей: для каждого перехода задает его входные и выходные позиции.

2. Для описания статической составляющей введен язык слогов (Статических Линейных Описаний Графов), который характеризуется линейной зависимостью численной (балльной) оценки «прозрачности» сети от ее размерности (при графовом представлении такая зависимость экспоненциальна). Следует добавить, что традиционно «барьер сложности» преодолевается путем иерархического описания с ограничением на размерность каждого уровня [10]. Но для реальных сложных сетей и число уровней растет экспоненциально, так что применение языка слогов представляется прагматичным. Кроме того, использование формализма слогов упрощает общение человека с компьютером, в частности ручной ввод информации.

3. Для описания динамической составляющей модели применен язык индикаторных выражений, представляющий собой интеграцию продукционных формализмов [4] и формализмов алгебры логики [1]. Язык обеспечивает:

- применение единого алгебраического способа описания как для статической, так и для динамической составляющей модели (принцип однородности [10]);
- описание продукций в терминах переменных, соответствующих целям, действиям и показателям системы (принцип триадности [8]);
- формирование условия и оператора продукций на базе индикаторных логических формул – ИЛФ;

- описание механизма взаимодействия агентов в многоагентной системе на основе синхронизации переходов в графах и передачи ресурсов от агента к агенту.

4. Перспективы развития сформулированного в работе сло-го-индикаторного подхода связаны, в частности, с разработкой:

- алгебры слогов;
- алгебры индикаторных выражений, включая подалгебру индикаторных логических формул;
- классификатора многоагентных сетей с адаптацией подхода для различных классов таких сетей.

Литература

1. ГИЛЬБЕРТ Д., АККЕРМАН В. *Основы теоретической логики*. М.: URSS, 2010.
2. ГОРОДЕЦКИЙ В.И., ГРУШИНСКИЙ М.С., ХАБАЛОВ А.В. *Многоагентные системы (обзор)* // *Новости искусственного интеллекта*. 1998, №2, С. 64 – 116.
3. КУЗНЕЦОВ О.П., КУЛИНИЧ А.А., МАРКОВСКИЙ А.В. *Анализ влияний при управлении слабоструктурированными ситуациями на основе когнитивных карт*. М.: КомКнига, 2006, С. 313 – 344.
4. ПОСПЕЛОВ Д.А. *Ситуационное управление: Теория и практика*. М.: Наука, 1986.
5. ТАЛЬ А.А., ЮДИЦКИЙ С.А. *Иерархия и параллелизм в сетях Петри I,II* // *Автоматика и телемеханика*. 1982, №7, С. 113 – 122, №9, С. 83– 88.
6. ЮДИЦКИЙ С.А., РАДЧЕНКО Е.Г. *Алгебра потокособытий и сети Петри – язык потокового моделирования многоагентных иерархических систем*. // *Приборы и системы. Управление, контроль, диагностика*. 2004, №9, С. 61 – 66.
7. ЮДИЦКИЙ С.А., ВЛАДИСЛАВЛЕВ П.Н. *Основы предпроектного анализа организационных систем*. М.: Финансы и статистика, 2005.
8. ЮДИЦКИЙ С.А., ВЛАДИСЛАВЛЕВ П.Н., ТОЧ Д.С. *Триадный подход к моделированию систем сетевидного типа*

управления // Управление большими системами, Выпуск 28, 2010, С. 24 – 39.

9. ЮДИЦКИЙ С.А. *Графодинамическое имитационное моделирование развития сетевых структур / Управление большими системами*. Выпуск 33. М.: ИПУ РАН, 2011. С.21 – 34.
10. ЮДИЦКИЙ С.А., ГОРЮНОВА В.В. *Однородная многоуровневая триадная модель развития больших систем на индикаторных сетях*. // подано в журнал «Управление большими системами», 2011.

ALGEBRAIC REPRESENTATION OF MULTIAGENT NETWORK MODEL

Semen Yuditskiy, Institute of Control Sciences of RAS, Moscow, Doctor of Science, professor (Moscow, Profsoyuznaya st., 65, tel. (495) 339-59-10, yuseab@yandex.ru).

Abstract: An algebraic appliance is suggested which describe oriented graphs as a string of symbols appended with numbered vertical arrows (syllable language). This appliance is intended to represent a static component of a model of multiagent networks. A dynamic component of the model is described with algebraic indicator expressions based on recurrent equations (producing rules and Boolean algebra rules). Syllable operations are introduced. Considered individual and collective functioning of agents in multiagent networks, as well as interaction between agents.

Keywords: multi-agent net, individual and collective functioning of agents, triad model of agent, syllable, syllable operations, indicator expression, indicator logical formula.

Статья представлена к публикации членом редакционной коллегии Н.Н. Бахтадзе

УДК 519.177+519.217.2+517.977.1

ББК 22.18

ДИСКРЕТНАЯ ПРОЦЕДУРА СОГЛАСОВАНИЯ ХАРАКТЕРИСТИК С ПОМОЩЬЮ МИНИМАЛЬНОГО ЦИКЛА, ОБЪЕДИНЯЮЩЕГО БАЗОВЫЕ БИКОМПОНЕНТЫ¹

Агаев Р. П.²

*(Учреждение Российской академии наук Институт проблем
управления им. В.А. Трапезникова РАН, Москва)*

Статья посвящена задаче дискретного согласования характеристик в многоагентных системах, в которых оргграф влияний Γ состоит только из несвязанных сильных компонент. Показано, что каждый блок предела правильной матрицы влияний для Γ пропорционален соответствующему блоку предела матрицы влияний для оргграфа Γ^h , полученного из Γ объединением сильных компонент с помощью минимального цикла. Установлено, что итоговая матрица процедуры ортогональной проекции, примененной к оргграфу влияний Γ , совпадает с пределом матрицы влияний для оргграфа Γ^h при определенных весах дуг объединяющего цикла.

Ключевые слова: многоагентные системы, децентрализованное управление, граф коммуникаций, консенсус, лапласовская матрица, матрица Кирхгофа, модель Де Гроота, управление.

¹ Работа выполнена при поддержке РФФИ, грант № 09-07-00371а и Программы Президиума РАН «Математическая теория управления».

² Рафиг Паашевич Агаев, к.т.н., с.н.с. (agaraf@rambler.ru, Москва, ул. Профсоюзная, д. 65, тел. (495) 334-88-69).

1. Введение

Согласно модели Де Гроота [8] если $s(0) = (s_1^0, \dots, s_n^0)^T$ – вектор начальных мнений членов группы, а $s(k) = (s_1^k, \dots, s_n^k)^T$ – вектор мнений после k -го шага согласования, то $s(k) = Ps(k-1)$, $k = 1, 2, \dots$, где P – стохастическая матрица влияний, элемент p_{ij} которой задает степень влияния мнения j -го агента на мнение i -го. В матричной форме модель Де Гроота имеет следующее представление:

$$(1) \quad s(k) = P^k s(0), \quad k = 1, 2, \dots$$

Известно, что согласие достижимо при любых начальных мнениях в том и только том случае [8], если существует предельная матрица $P^\infty = \lim_{k \rightarrow \infty} P^k$ и все ее строки равны, что равносильно регулярности матрицы P .

Если матрица P нерегулярна, то согласие может быть достигнуто при векторах начальных мнений, принадлежащих определенному подпространству. В [3] получена характеристика этого подпространства и предложен метод ортогональной проекции, обобщающий процедуру Де Гроота. Показано, что небазовые агенты в методе ортогональной проекции, как и в процедуре Де Гроота, не влияют на конечный результат.

Статья имеет следующую структуру. После введения приведены основные определения и обозначения. В разделе 3 доказано, что нормированная матрица исходящих лесов является однородной относительно минимального цикла, объединяющего все базовые бикомпоненты. В разделе 4 установлено, что итоговая матрица процедуры ортогональной проекции совпадает с пределом матрицы влияний после объединения всех базовых бикомпонент минимальным объединяющим циклом с определенными весами дуг.

2. Основные термины и обозначения

Стохастической матрице P , входящей в модель Де Гроота, поставим в соответствие *орграф влияний* Γ с множеством вершин $V(\Gamma) = \{1, \dots, n\}$, в котором при $p_{ij} > 0$ (т.е. если j -й агент влияет на i -го) от вершины j к вершине i проводится дуга (j, i) с весом $w_{ji} = p_{ij}$.

Матрица Кирхгофа $L = L(\Gamma) = (\ell_{ij})$ орграфа Γ определяется следующим образом: при $j \neq i$ полагают $\ell_{ij} = -w_{ji}$, если в Γ имеется дуга (j, i) , и $\ell_{ij} = 0$ в противном случае; $\ell_{ii} = \sum_{k \neq i} w_{ki}$, $i, j = 1, \dots, n$. Нередко вместо матрицы Кирхгофа строится *лапласовская матрица*. Она определяется соотношениями $\ell_{ij} = -w_{ij}$, если $j \neq i$, и $\ell_{ii} = -\sum_{k \neq i} \ell_{ik}$, $i, j = 1, \dots, n$.

Через I будем обозначать единичную матрицу.

В силу приведенных определений для орграфа Γ , отвечающего матрице P , имеем

$$(2) \quad L(\Gamma) = I - P.$$

С другой стороны, для любой матрицы Кирхгофа $L(\Gamma) = (\ell_{ij})$ взвешенного орграфа Γ (веса дуг – произвольные положительные числа) следующим образом определим стохастическую матрицу влияний:

$$(3) \quad P = I - \epsilon L,$$

где $\epsilon < (\max \ell_{ii})^{-1}$.

Любой максимальный по включению сильный подграф орграфа называют его *сильной компонентой* или *бикомпонентой*. *Базовая бикомпонента* – такая бикомпонента, в которую не входят дуги извне. Через ν будем обозначать число базовых бикомпонент.

Будем говорить, что матрица A имеет предел, если A^m стремится к некоторой матрице при $m \rightarrow \infty$.

Если стохастическая матрица P орграфа влияний имеет предел P^∞ , то

$$(4) \quad P^\infty = \bar{J},$$

где $\bar{J} = (j_{kr})$ – нормированная матрица максимальных исходящих лесов соответствующего взвешенного орграфа Γ (следствие матричной теоремы о деревьях для цепей Маркова [6], см. также теорему 7 из [1]).

Элементы матрицы $\bar{J} = (j_{kr})$ определяются следующим образом:

$$(5) \quad j_{kr} = \frac{q_{kr}}{\sigma},$$

где q_{kr} – вес множества максимальных исходящих из вершины r лесов, в которых вершина k достижима из r , σ – вес множества всех максимальных исходящих лесов в орграфе Γ .

Поскольку предел матрицы влияний равен нормированной матрице исходящих лесов орграфа влияний, этот предел можно определить рекурсивно как многочлен от L (см. [2, раздел 4]) с помощью метода Леверье-Фаддеева или же как (теорема 6 в [1])

$$(6) \quad \bar{J} = \lim_{\tau \rightarrow \infty} (I + \tau L)^{-1}.$$

Для матриц влияний, Кирхгофа и предела степеней матрицы влияний i -ю бикомпоненту обозначим соответственно через P_i , L_i и P_i^∞ . В данном случае матрицы L и P^∞ всей системы имеют вид

$$(7) \quad L = \begin{pmatrix} L_1 & & & \\ & L_2 & & \\ & & \ddots & \\ & & & L_\nu \end{pmatrix}, \quad P^\infty = \begin{pmatrix} P_1^\infty & & & \\ & P_2^\infty & & \\ & & \ddots & \\ & & & P_\nu^\infty \end{pmatrix},$$

где блоки соответствуют базовым бикомпонентам, а не входящие в них элементы равны нулю.

Матрицы P_i^∞ соответствуют сильно связным орграфам и представляются в виде

$$(8) \quad P_i^\infty = \mathbf{1}(\pi^i)^T, \quad i = 1, \dots, \nu,$$

где $\mathbf{1} = (1, \dots, 1)^T$, $(\pi^i)^T$ – любая строка P_i^∞ .

Через t^i обозначим сумму весов³ всех остовных исходящих деревьев i -й бикомпоненты орграфа Γ , а через t_k^i – сумму весов тех из них, которые имеют корень в k -й вершине i -й бикомпоненты. Отметим, что согласно матричной теореме о деревьях (см., например, теорему 16.9' в [7], где результат формулируется для матрицы L^T и невзвешенных орграфов) t_k^i равно алгебраическому дополнению любого элемента k -й строки матрицы L_i .

3. Однородность нормированных матриц исходящих лесов относительно минимального объединяющего цикла

В [3] для дискретных моделей был предложен метод согласования, сводящийся к 1) преобразованию вектора начальных мнений в вектор, принадлежащий определенной области, с помощью ортогональной проекции и 2) дальнейшей коррекции мнений посредством преобразования с использованием стохастической матрицы. В [4] было доказано, что любая сходящаяся процедура согласования может быть приближена процедурой Де Гроота, орграф влияний которой является гамильтоновым циклом. Но при этом воспроизводится лишь конечный результат, т.е. итоговый вектор влиятельности агентов, матрица же связей между агентами может сильно отличаться от аппроксимируемой.

Предположим, что матрица влияний P имеет предел. Пусть при этом орграф влияний $\Gamma(V, E)$ состоит только из базовых бикомпонент. Подграфы, соответствующие бикомпонентам, обозначим через $\Gamma_1(V_1, E_1), \dots, \Gamma_\nu(V_\nu, E_\nu)$; каждый из них – сильный. Поскольку общий орграф влияний не содержит остовного исходящего дерева, согласие в системе достигается не для любого

³ Вес дерева определяется как произведение весов всех его дуг.

вектора начальных мнений.

В каждой k -й бикомпоненте зафиксируем произвольную вершину v_k , $k = 1, \dots, \nu$. Эти вершины соединим минимальным циклом $H = (e_1, \dots, e_\nu)$, где $e_i = (v_i, v_{i+1})$, $i = 1, \dots, \nu$ и $v_{\nu+1} = v_1$. Такой цикл соединяет бикомпоненты в одну. Полученный орграф обозначим через $\Gamma^h(V, E_h)$ (рис.1). В силу связности $\Gamma^h(V, E_h)$ последовательность степеней его матрицы влияний имеет предел – матрицу с одинаковыми строками. Эту регулярную положительную матрицу обозначим через $\bar{J}^{(h)}$ и представим в следующем блочном виде

$$\bar{J}^{(h)} = \begin{pmatrix} \bar{J}_1^{(h)} & * & * & * \\ * & \bar{J}_2^{(h)} & * & * \\ * & * & * & * \\ * & * & * & \bar{J}_s^{(h)} \end{pmatrix},$$

где все блоки, включая блоки, обозначенные $*$ – состоят из положительных чисел.

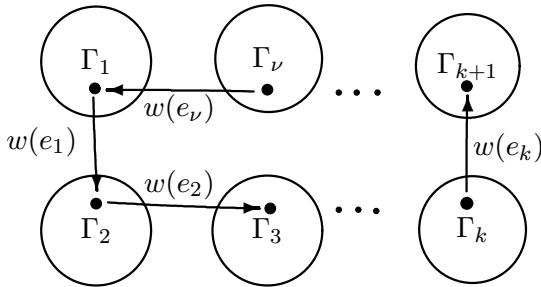


Рис. 1.

Предложение 1. Для матрицы $\bar{J}^{(h)} = (j_{kr}^{(h)})$, определенной выше, каждая функция $j_{kr}^{(h)}(w(e_1), \dots, w(e_\nu))$ является

ся однородной 0-й степени⁴, т.е. $j_{kr}^{(h)}(xw(e_1), \dots, xw(e_\nu)) = j_{kr}^{(h)}(w(e_1), \dots, w(e_\nu))$ для любого $x \in \mathbb{R}_+$.

Доказательство предложения 1. Рассмотрим матрицу Кирхгофа $L^{(h)}$ орграфа влияний Γ^h .

Пусть q_r – суммарный вес деревьев, исходящих из вершины r в Γ^h . Пусть r принадлежит базовой бикомпоненте k . Тогда

$$q_r = t_r^k \prod_{s=0}^{\nu-2} (w(e_{k+s})t_{v_{k+s+1}}^s),$$

где t_r^k , как и ранее, – сумма весов остовных исходящих деревьев k -й бикомпоненты орграфа Γ , в которых r является корнем⁵.

Поскольку для всех $r \in \{1, \dots, n\}$ и любого $x \in \mathbb{R}_+$

$$t_r^k \prod_{s=0}^{\nu-2} (xw(e_{k+s})t_{v_{k+s+1}}^s) = x^{\nu-1} t_r^k \prod_{s=0}^{\nu-2} (w(e_{k+s})t_{v_{k+s+1}}^s)$$

и вес множества всех исходящих деревьев t орграфа Γ^h равен $\sum_{r=1}^n q_r$, согласно (5) имеем

$$j_{kr}^{(h)}(xw(e_1), \dots, xw(e_\nu)) = q_r t^{-1} = j_{kr}^{(h)}(w(e_1), \dots, w(e_\nu)).$$

□

Предложение 2. Каждый блок $\bar{J}_k^{(h)}$ матрицы $\bar{J}^{(h)}$ пропорционален соответствующему блоку P_k^∞ матрицы P^∞ .

⁴ Функция $f(x_1, \dots, x_n)$ от n аргументов называется однородной функцией m -й степени, если при умножении всех ее аргументов на множитель μ функция приобретает этот же множитель m -й степени, т.е. если тождественно выполняется равенство $f(\mu x_1, \dots, \mu x_n) = \mu^m f(x_1, \dots, x_n)$.

⁵ Если индекс $k + s$ больше ν , то вместо него используется $k + s \bmod \nu$. Аналогично, если $k + s + 1$ больше ν , то используется $k + s + 1 \bmod \nu$.

Доказательство предложения 2. Не уменьшая общности, докажем, что первые блоки $P_1^\infty = (p_{kr}^{(1)})$ и $\bar{J}_1^{(h)} = (j_{kr}^{(1)})$ пропорциональны. Поскольку каждый блок состоит из одинаковых строк, для элементов первой строки матрицы P_1^∞ согласно (4) и (5) имеет место:

$$\frac{p_{1r}^{(1)}}{p_{1k}^{(1)}} = \frac{t_r^1}{t_k^1}.$$

Пропорциональность двух матриц P_1^∞ и $\bar{J}_1^{(h)}$ следует из соотношения элементов первой строки матрицы $\bar{J}_1^{(h)}$:

$$\frac{j_{1r}^{(1)}}{j_{1k}^{(1)}} = \frac{t_r^1 \prod_{s=2}^{\nu} w(e_{s-1}) t_{v_s}^s}{t_k^1 \prod_{s=2}^{\nu} w(e_{s-1}) t_{v_s}^s} = \frac{t_r^1}{t_k^1}.$$

□

Отметим, что при объединении базовых бикомпонент была использована одна вершина из каждой бикомпоненты. Будет ли справедливым предложение 2, если вместо минимального цикла задействовать более одной вершины хотя бы в одной бикомпоненте? Построенные примеры показывают, что при увеличении длины цикла предложение 2, вообще говоря, перестает быть верным.

4. Метод ортогональной проекции как частный случай процедуры минимального объединяющего цикла

Следующее предложение позволяет добавлением минимального числа дуг построить орграф влияний агентов, реализующий предельную матрицу влияний, совпадающую с матрицей процедуры ортогональной проекции для исходного орграфа влияний.

Предложение 3. Пусть $\Gamma(V, E)$ и $\Gamma^h(V, E_h)$ — орграфы, определенные выше. Тогда для некоторого минимального цикла

$H = (e_1, \dots, e_\nu)$, соединяющего все базовые бикомпоненты, при определенных весах дуг $w(e_1), \dots, w(e_\nu)$, предел матрицы влияний $P^{(h)}$, соответствующей орграфу $\Gamma^h(V, E_h)$ (см. формулу (3)), совпадает с итоговой матрицей процедуры ортогональной проекции при орграфе влияний $\Gamma(V, E)$.

Доказательство предложения 3. Рассмотрим весовой вектор $\alpha = (\alpha_1, \dots, \alpha_n)$ процедуры ортогональной проекции [3]

$$(9) \quad \bar{P} = P^\infty S = 1\alpha^T.$$

Предположим, что матрица $\bar{J}^{(h)}$ определена с помощью матрицы Кирхгофа $L^{(h)}$). Поскольку ранг матрицы $\bar{J}^{(h)}$ равен единице (предложение 11 из [1]), рассмотрим ее первую строку $j_{11}^{(h)}, \dots, j_{1n}^{(h)}$. Для совпадения матриц $\bar{J}^{(h)}$ и \bar{P} согласно предложению 2 достаточно выполнение следующего равенства:

$$(10) \quad \frac{j_{1v_r}^{(h)}}{j_{1v_k}^{(h)}} = \frac{\alpha_{1v_r}}{\alpha_{1v_k}}, \quad k = 1, \dots, \nu.$$

Пусть t – вес множества всех исходящих деревьев орграфа Γ^h . Тогда согласно (5)

$$j_{1v_r}^{(h)} = \frac{\prod_{s=1}^{\nu} w(e_s) t_{v_s}^s}{w(e_{r-1}) t} \quad (w(e_0) \equiv w(e_\nu))$$

и

$$(11) \quad \frac{j_{1v_r}^{(h)}}{j_{1v_k}^{(h)}} = \frac{w(e_{k-1})}{w(e_{r-1})}.$$

Положим $w(e_\nu) = 1$ и определим веса других дуг e_k по формуле

$$(12) \quad w(e_k) = \frac{\alpha_{v_1}}{\alpha_{v_{k+1}}}, \quad k = 1, \dots, \nu - 1.$$

Стохастическая матрица влияний $P^{(h)}$ определяется по формуле (3):

$$P^{(h)} = I - \epsilon L^{(h)},$$

где $\epsilon < (\max \ell_{ii}^{(h)})^{-1}$.

Поскольку предел матрицы $P^{(h)}$ совпадает с $\bar{J}^{(h)}$, в силу (11) и (12) этот предел совпадает и с итоговой матрицей процедуры ортогональной проекции при орграфе влияний $\Gamma(V, E)$.

□

Замечание 1. В предложении 3 веса добавленных дуг были вычислены с помощью компонент весового вектора $\alpha = (\alpha_1, \dots, \alpha_n)$ процедуры ортогональной проекции. Согласно формуле (П.6) из [3] для компонент вектора α имеет место

$$(13) \quad \frac{\alpha_k}{\alpha_r} = \frac{t_k^i}{t_r^j} \cdot \frac{W_j}{W_i},$$

где W_i – определитель матрицы T_i , полученной из L_i заменой первого столбца вектором π^i , i и j номера компонент, которым соответственно принадлежат вершины k и r .

Из (12) и (13) получим:

$$(14) \quad w(e_k) = \frac{t_{v_1}^1}{W_1} \cdot \frac{W_{k+1}}{t_{v_{k+1}}^{k+1}}, \quad k = 1, \dots, \nu - 1.$$

Предложение 4. Пусть диагональные элементы p_{v_r, v_r} , $r=1, \dots, \nu$, матрицы влияний P положительны и все базовые бикомпоненты объединены минимальным циклом $H=(e_1, \dots, e_\nu)$. Тогда при некоторых значениях весов дуг $w^o(e_1), \dots, w^o(e_\nu)$, предел матрицы влияний $P^{(h)}$, отличающейся от P только теми элементами, которые соответствуют добавленным дугам, и диагональными элементами, совпадет с итоговой матрицей метода ортогональной проекции для P .

Доказательство предложения 4. Пусть веса дуг вычислены как в предложении 3. Согласно предложению 1 умножение весов всех дуг, входящих в минимальный цикл, на одно и то же число не влияет на нормированную матрицу исходящих лесов. Поэтому переопределим значения $w^o(e_1), \dots, w^o(e_\nu)$ следующим образом:

$$(15) \quad w^o(e_i) = \theta \left(\max_{1 \leq k \leq \nu} w(e_k) \right)^{-1} \cdot w(e_i),$$

где $0 < \theta < \min_{1 \leq k \leq \nu} p_{v_r v_r}$.

Для орграфа влияний $\Gamma^h(V, E_h)$ построим матрицу $P^{(h)}$. Пусть L' – матрица Кирхгофа для $\Gamma'(V, H)$. Заметим, что $P^{(h)} = P - L'$ и является стохастической.

Поскольку, нормированная матрица исходящих деревьев для $\Gamma^h(V, E_h)$ совпадает с пределом матрицы $P^{(h)}$, очевидно, что последняя совпадает также с итоговой матрицей ортогональной проекции для P . □

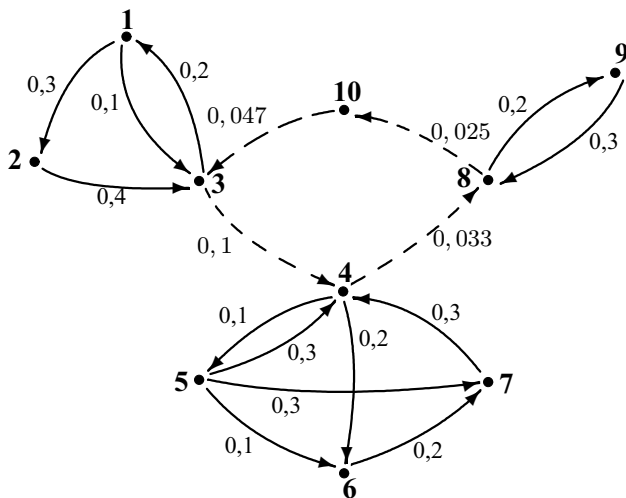


Рис. 2.

$$w^o(e_1) \approx 0,1, \quad w^o(e_2) \approx 0,033, \quad w^o(e_3) \approx 0,025, \quad w^o(e_4) \approx 0,047.$$

Пример. Применим предложения 3 и 4 к орграфу влияний, приведенному на рис. 2, где для простоты не показаны петли. Вначале предположим, что нет дуг, соединяющих вершины 3, 4, 8, 10. Тогда подграфы $\Gamma_1, \Gamma_2, \Gamma_3, \Gamma_4$ соответственно на мно-

жествах вершин $\{1, 2, 3\}$, $\{4, 5, 6, 7\}$, $\{8, 9\}$, $\{10\}$, являются базовыми бикомпонентами.

Определим стохастическую матрицу влияний для соответствующего орграфа влияний.

$$P = \begin{pmatrix} 0,8 & 0 & 0,2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0,3 & 0,7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0,1 & 0,4 & 0,5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,4 & 0,3 & 0 & 0,3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,1 & 0,9 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,2 & 0,1 & 0,7 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0,3 & 0,2 & 0,5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,7 & 0,3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,2 & 0,8 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Определим матрицу P^∞ по одному из вышеуказанных способов (например, по формуле (6)):

$$P^\infty = \bar{J} = \begin{pmatrix} 0,517 & 0,276 & 0,207 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0,517 & 0,276 & 0,207 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0,517 & 0,276 & 0,207 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,139 & 0,722 & 0,056 & 0,083 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,139 & 0,722 & 0,056 & 0,083 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,139 & 0,722 & 0,056 & 0,083 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,139 & 0,722 & 0,056 & 0,083 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,4 & 0,6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,4 & 0,6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Теперь предположим, что базовые бикомпоненты соединены циклом $3 \rightarrow 4 \rightarrow 8 \rightarrow 10 \rightarrow 3$. На рис. 2 эти связи указаны пунктирными дугами. Полученный орграф обозначим через Γ^h . Определим веса добавленных дуг таким образом, что нормиро-

ванная матрица исходящих лесов $\bar{J}^h = (j_{kr}^h)$ была равна \bar{P}^∞ – итоговой матрице процедуры ортогональной проекции.

По формуле (14) вычислим $w(e_k)$. Для этого положим $w(e_4) = 1$ и определим матрицы:

$$T_1 = \begin{pmatrix} 0,5172 & 0 & -0,2 \\ 0,2759 & 0,3 & 0 \\ 0,2069 & -0,4 & 0,5 \end{pmatrix}; T_2 = \begin{pmatrix} 0,1389 & -0,3 & 0 & -0,3 \\ 0,7222 & 0,1 & 0 & 0 \\ 0,0556 & -0,1 & 0,3 & 0 \\ 0,0833 & -0,3 & -0,2 & 0,5 \end{pmatrix};$$

$$T_3 = \begin{pmatrix} 0,4 & -0,3 \\ 0,6 & 0,2 \end{pmatrix}.$$

Вычислим $W_1 = \det(T_1) = 0,1121$, $W_2 = \det(T_2) = 0,0595$, $W_3 = \det(T_3) = 0,26$, $W_4 = 1$.

Применив матричную теорему о деревьях для матриц L_i определяем следующие значения: $t_3^1 = 0,06$; $t_4^2 = 0,015$; $t_8^3 = 0,2$.

Итак, $w(e_1) = 2,1237$, $w(e_2) = 0,696$, $w(e_3) = 0,5354$
 $w(e_4) = 1$.

По формуле (15) определим $w^o(e_1), \dots, w^o(e_\nu)$. Положив $\theta=0,1$, находим $\max_{k \in \{1, \dots, \nu\}} w(e_k) = 2,1237$. Далее находим:

$$w^o(e_1) = 0,1 \cdot 2,1237^{-1} \cdot 2,1237 \approx 0,1;$$

$$w^o(e_2) = 0,1 \cdot 2,1237^{-1} \cdot 0,696 \approx 0,033;$$

$$w^o(e_3) = 0,1 \cdot 2,1237^{-1} \cdot 0,5354 \approx 0,025;$$

$$w^o(e_4) = 0,1 \cdot 2,1237^{-1} \cdot 1 \approx 0,047.$$

Эти значения приведены на рис. 2.

Для орграфа Γ^h , приведенного на рис. 2, построим матрицу Кирхгофа L' и определим матрицу $P^{(h)} = P - L'$.

Предел матрицы влияний $P^{(h)}$ имеет следующее представление.

$$(P^{(h)})^\infty = \mathbf{1} \cdot (0,1827; 0,0974; 0,0731; 0,0344; 0,1789; 0,0138; 0,0206; 0,1050; 0,1575; 0,1365).$$

Матрица $(P^{(h)})^\infty$ совпадает с \bar{P}^∞ .

5. Заключение

В работе доказано, что итоговая матрица процедуры ортогональной проекции, применяемой к орграфу влияний Γ , совпадает с пределом матрицы влияний для орграфа Γ^h , который получен из Γ объединением всех сильных компонент с помощью минимального цикла. При этом веса всех дуг, входящих в минимальный цикл, определяются однозначно с точностью до множителя, а сам цикл содержит по одной вершине из каждой бикомпоненты. Если все сильные компоненты соединить минимальным циклом из дуг с произвольными весами, то соответствующая матрица влияний также будет иметь предел. Этот предел в общем случае не совпадет с матрицей ортогональной проекции, но его строки являются, как и в методе проекции, выпуклыми комбинациями линейно независимых строк матрицы P^∞ .

В работе не изучалось влияние порядка бикомпонент в объединяющем их минимальном цикле. Скорее всего, при объединении бикомпонент не имеет значения, из какой бикомпоненты приходит дуга в данную бикомпоненту, важен лишь вес дуги и вершина, в которую дуга входит. Предложенный метод может быть интерпретирован как процедура управления со стороны «центра», и было бы интересно сравнить результаты, приведенные в разделе 3 «Динамические модели информационного управления. Анализ» работы [5], с результатами настоящей работы. Указанное сравнение выходит за рамки данной статьи и требует отдельного исследования.

Литература

1. АГАЕВ Р.П., ЧЕБОТАРЕВ П.Ю. *Матрица максимальных исходящих лесов орграфа и ее применения* // Автоматика и телемеханика. – 2000. – № 9. – С. 15–43.

2. АГАЕВ Р.П., ЧЕБОТАРЕВ П.Ю. *Остовные леса орграфа и их применение* // Автоматика и телемеханика. – 2001. – № 3. – С. 108–133.
3. АГАЕВ Р.П., ЧЕБОТАРЕВ П.Ю. *Метод проекции в задаче о консенсусе и регуляризованный предел степеней стохастической матрицы* // Автоматика и телемеханика. – 2011. – № 12. – (в печати).
4. АГАЕВ Р.П., ЧЕБОТАРЕВ П.Ю. *Представление дискретной процедуры согласования характеристик с помощью циклического орграфа* // Автоматика и телемеханика. – 2012. – (в печати).
5. БАРАБАНОВ И.Н., КОРГИН Н.А., НОВИКОВ Д.А., ЧХАРТИШВИЛИ А.Г. *Динамические модели информационного управления в социальных сетях* // Автоматика и телемеханика. – 2010 № 11. – С. 172–182.
6. ВЕНТЦЕЛЬ А.Д., ФРЕЙДЛИН М.И. *О малых случайных возмущениях динамических систем* // Успехи мат. наук. – 1970. – Т. 25. – С. 3–55.
7. ХАРАРИ Ф. *Теория графов*. – М.: Мир, 1973. – 300 с.
8. DeGROOT M.H. *Reaching a consensus* // J. Amer. Statist. Assoc. – 1974. – Vol. 69, No. 345. – P. 118–121.

DISCRETE CONSENSUS CONVERGENCE PROCEDURE VIA MINIMAL CYCLE COMBINING STRONG COMPONENTS

Rafiq Agaev, Institute of Control Sciences of RAS, Moscow,
Candidate of Science, senior researcher (agaraf@rambler.ru,
Moscow, Profsoyuznaya str., 65, (495)334-88-69).

Abstract: This paper is devoted to consensus problems in discrete multi-agent systems whose communication digraphs consist of disjoint strong components. It is shown that any block in the power limit of a decomposable and aperiodic influence matrix P of a digraph Γ is proportional to the corresponding block in the power limit of the influence matrix of the digraph Γ^h obtained from Γ by combining the strong components by means of a minimal cycle. It is proved that for some arc weights in this minimal cycle, the power limit of the influence matrix of Γ^h coincides with the resulting matrix of the orthogonal projection procedure applied to Γ .

Keywords: multi-agent systems, decentralized control, communication digraph, consensus, Laplacian matrix, Kirchhoff matrix, DeGroot model, control.

*Статья представлена к публикации
членом редакционной коллегии Д. А. Новиковым*

УДК 681.51
ББК 32.811.1

ОБНАРУЖЕНИЕ СТРУКТУРНО-ПАРАМЕТРИЧЕСКИХ ИЗМЕНЕНИЙ В СТОХАСТИЧЕСКИХ СИСТЕМАХ В РЕАЛЬНОМ МАСШТАБЕ ВРЕМЕНИ АЛГОРИТМАМИ НЕПРЕРЫВНЫХ ДРОБЕЙ И СТРУКТУРНОГО АНАЛИЗА

Карташов В. Я.¹, Новосельцева М. А.²
(*ГОУ ВПО «Кемеровский государственный
университет», Кемерово*)

В статье разработан алгоритм обнаружения структурно-параметрических изменений в стохастических процессах и системах в режиме реального времени. Алгоритм основан на совместном использовании структурного анализа и теории непрерывных дробей.

Ключевые слова: структурная функция, непрерывная дробь, идентифицирующая матрица.

1. Введение

Обработка и анализ временных рядов представляет собой интенсивно развивающееся направление в математических, технических, экономических, политических, социальных и других науках. Традиционным при анализе временных рядов является предположение о том, что статистические свойства наблюдаемого ряда или свойства порождающей его системы

¹ Владимир Яковлевич Карташов, заведующий кафедрой автоматизации исследований и технической кибернетики, доктор технических наук, профессор (kartash@kemsu.ru).

² Марина Александровна Новосельцева, доцент кафедры автоматизации исследований и технической кибернетики, кандидат технических наук, доцент (aapov@pochta.ru).

сохраняют определенное постоянство во времени. Однако многие практические задачи текущего контроля производства, экономики, медицинской и технической диагностики, гидроакустики, геофизики и др. связаны с оценкой риска наступления той или другой непредвиденной ситуации, вызванной некоторым устойчивым изменением свойств наблюдаемого временного ряда, происходящим в неизвестный момент времени.

Интерес к проблематике данных задач стал возрастать с середины 60-х годов, что вызывалось потребностями практических приложений и появлением ЭВМ, давшей возможность автоматизировать данный процесс. При этом основные усилия исследователей направлялись на то, чтобы разработать методы, использующие как можно меньше априорной информации о наблюдениях и свойствах временного ряда. Популярность данной проблематики несомненна.

2. Постановка задачи

Определим круг практических задач и их постановки, в которых возникает потребность обнаруживать изменение свойств случайных процессов [7, 10, 12, 26].

Первый класс рассматривает ретроспективные (апостериорные) задачи. Пусть представлена реализация случайного процесса. Корректность методов анализа основных свойств случайных процессов, а также интерпретация результатов анализа в значительной степени влияют на правильность получения модели, оценку ее параметров и прогнозирование будущих значений временного ряда. Поэтому статистическая обработка реализации случайного процесса должна быть основана на предположении, что его свойства в процессе сбора данных не изменялись. Предварительным этапом любой статистической обработки должен быть этап проверки подобной неизменчивости. Если характеристики не изменялись, то следует заниматься статистической обработкой в зависимости от целей исследователя. В противном случае возникает задача обнаружения моментов изменения статистических характеристик и разбиения

исходной выборки на несколько интервалов стационарности и однородности. Такая задача решалась авторами в [19, 21, 27].

Второй класс непосредственно связан с ситуацией обработки измерений случайного процесса в реальном масштабе времени. Допустим, что в некоторый заранее неизвестный момент времени происходит изменение какой-либо статистической характеристики процесса. Как обнаружить произошедшее изменение скорейшим образом после того, как оно возникло? В данном случае запаздывание в обнаружении момента изменения свойств временных рядов может привести к наступлению рискованной ситуации в системе [17, 28]. Эта задача часто формулируется как задача обнаружения момента разладки процесса [9, 12, 26, 29]. Именно ее решению будет посвящено дальнейшее изложение.

3. Обзор методов обнаружения изменения свойств стохастических процессов и систем

Существующие методы решения данной задачи позволяют достаточно эффективно улавливать момент изменения характеристик случайного процесса. Однако развитая теория имеет своим основанием сложный математический аппарат, не самый простой с точки зрения практического применения, поскольку его можно использовать, имея дополнительную априорную информацию о данных измерений [4, 5, 7, 9, 22]. В качестве априорных знаний или предположений при обнаружении изменений свойств стохастических сигналов и систем обычно используют: знание распределения момента появления изменений [33]; знание законов или классов распределений данных измерений в нормальном состоянии и после появления изменений [22, 29, 33]; знание вида модели процесса до и после изменений с указанием меняющихся параметров процесса [29]; задание величины измененного параметра либо пороговых значений процесса [7, 22, 29]; стационарность данных измерений до и после разладки [4, 29] и т.д.

В связи с отсутствием какой-либо априорной информации при обнаружении момента изменения свойств временных рядов

желательно применение непараметрических критериев, не использующих априорную информацию о распределениях. Кроме того, ограничения области применимости критериев одним типом разлаженного состояния могут привести к запаздыванию в принятии решения по управлению стохастическими объектами и системами [18].

Алгоритм кумулятивных сумм (АКС) [26, 29] является наиболее популярным, теоретически обоснованным и исследованным методом обнаружения изменений в последовательности наблюдений. Данный метод обобщен на различные типы данных измерений и имеет несколько модификаций: для независимых отсчетов, для зависимых отсчетов, для линейных объектов, для процессов авторегрессии проинтегрированного скользящего среднего и т.д. Однако методика АКС имеет ряд недостатков, заключающихся в следующем.

1. Использование разного рода априорной информации, часто недоступной на практике для малоизученных явлений и при работе с новыми уникальными объектами.

2. Выбор шага дискретизации случайного процесса не ясен, подчеркивается его важность, но рекомендаций и правил его выбора не дается.

3. Использование малоэффективных корреляционных методик визуального анализа стационарности методом Бокса–Дженкинса, недостатки которых подробно описаны в [21, 27].

4. Идентификация временных рядов на основе перебора пробных моделей методикой Бокса–Дженкинса. Подробный анализ этой методики приведен в [16, 21, 27]. Авторы сами отмечают [26], что на качество работы алгоритма влияет «...неточность восстановления модели авторегрессии», заключающаяся в неверном задании порядка или определении параметров модели.

5. Метод основан на вычислении логарифма правдоподобия авторегрессионной модели, на корреляционной теории случайных процессов, которые применимы лишь в стационарных условиях.

Необходимость наличия большого количества априорной информации приводит к ограничению использования АКС.

Кроме того, в некоторых условиях нет возможности проводить длительную настройку алгоритма по какому-то конкретному параметру даже с наименьшей информацией о нем и с уточняющими процедурами. Следовательно, актуальной становится задача применения таких методов обработки данных измерений, которые не включали бы в себя вышеописанные недостатки. Представляется, что в таких условиях наиболее эффективными окажутся непараметрические методы, не использующие значительное большинство из перечисленных выше априорных сведений.

Одним из таких критериев является инверсионный алгоритм разделения данных на интервалы стационарности [16, 24, 25]. Алгоритм является одним из наиболее эффективных непараметрических критериев обнаружения изменений характеристик случайных процессов в условиях априорной неопределенности. В [16, 27] данный алгоритм был применен для ретроспективного анализа стационарности случайных процессов при идентификации стохастических объектов. Метод может использоваться в реальном масштабе времени и позволяет обнаружить изменения среднего, дисперсии, а также появление нестационарности. Недостатком метода, как отмечает сам автор [24, 25], является то, что данный метод не позволяет выявлять момент смены корреляционной функции процесса, т.е. появления структурной неоднородности случайного процесса [16, 27].

4. Использование структурного анализа и теории непрерывных дробей для обнаружения структурно-параметрических изменений стохастических сигналов и систем

Для разработки критерия обнаружения изменений структурно-параметрических свойств процессов и стохастических систем воспользуемся структурным анализом – разделом теории случайных процессов, связанным с изучением последних на основе структурных функций [16, 30]. Структурный анализ случайных процессов в большинстве случаев приводит к более устойчивым характеристикам по сравнению с корреляционны-

ми. Эффективность структурного анализа заключается в том, что параметры структурной функции обладают свойствами инвариантности относительно некоторых форм нестационарности, проявляющихся, например, при смещенности по математическому ожиданию, а также в случае квазистационарного характера случайного процесса.

Структурная функция [16, 30], предложенная А.Н. Колмогоровым, определяется как математическое ожидание квадрата разности сечений случайного процесса, соответствующим значениям аргумента t и $t + \tau$:

$$(1) \quad C_x(t, t + \tau) = M \{ x(t) - x(t + \tau) \}^2,$$

где $x(t)$ – некоторый случайный процесс. Очевидно, что функция всегда неотрицательна, четна и удовлетворяет условию $C_x(t, t) = 0$.

Рассмотрим частный случай, когда $x(t)$ представляет собой стационарный случайный процесс с нулевым средним. Тогда структурная функция будет иметь вид

$$(2) \quad \begin{aligned} C_x(t, t + \tau) &= M \{ x(t) - x(t + \tau) \}^2 = \\ &= M(x^2(t)) - 2M(x(t)x(t + \tau)) + \\ &\quad + M(x^2(t + \tau)) = 2R_{xx}(0) - 2R_{xx}(\tau) = C_x(\tau). \end{aligned}$$

Как видно из (2), структурная функция стационарного случайного процесса не зависит от текущего момента времени. На основании (2) можно утверждать, что структурная функция стационарного случайного процесса с течением времени стремится к установившемуся значению:

$$(3) \quad C_x(\tau) \xrightarrow{\tau \rightarrow \infty} 2R_{xx}(0).$$

Практическое построение структурной функции более надежно по сравнению с корреляционной, поскольку на нее не влияют ошибки определения среднего значения процесса $x(t)$ [30]. Оно осуществляется по формуле

$$(4) \quad C_x(k) = \frac{1}{N - k} \sum_{i=1}^{N-k} (x(i) - x(i + k))^2,$$

где N – число измерений процесса $x(t)$.

Проведенные ранее исследования [16, 19, 21, 27] позволили использовать структурную функцию для анализа стационарности случайного процесса, т.е. для решения ретроспективной задачи. Поскольку данная методика ляжет в основу алгоритма обработки реализации случайного процесса в реальном масштабе времени, опишем ее более подробно.

Для анализа стационарности использовались следующие закономерности [16, 19, 27]:

- структурная функция стационарного процесса ограничена и с течением времени выходит на установившееся значение;
- структурная функция нестационарного процесса с течением времени неограниченно возрастает.

С целью обеспечения эффективной алгоритмической и программной реализации метода анализа стационарности, а также исключения визуального анализа кривых структурной функции был разработан алгоритм получения модели структурной функции сигнала на основании ее значений [16, 19, 21, 27]. Кроме того, свойство стационарности случайного процесса отождествлялось со свойством устойчивости полученной модели структурной функции.

Алгоритм основан на использовании нетрадиционного математического аппарата – теории непрерывных дробей [15]. В частности наибольшее развитие в этой теории получил класс правильных C -дробей следующего вида:

$$(5) \quad \left[b_0 ; \frac{a_n x}{1} \right]_1^\infty = b_0 + \frac{a_1 x}{1 + \frac{a_2 x}{1 + \frac{a_3 x}{1 + \dots}}},$$

где a_1, \dots, a_n, b_0 – постоянные величины.

Наиболее приемлемым и простым способом перехода к непрерывной дроби является модифицированный алгоритм В. Висковатова [15, 13].

При нахождении структурной функции используется виртуальный подход [16], заключающийся в том, что процесс определения структурной функции $C_x(k)$ можно представить как реакцию динамического виртуального объекта на единичное

входное воздействие. В этом случае $C_x(k)$ для фиксированных k порождает передаточную функцию виртуального объекта, знаменатель которой представляет собой его характеристическое уравнение. Исследуя это уравнение на устойчивость любым известным методом, делается вывод о стационарности или нестационарности случайного процесса.

Критерий проверки стационарности случайного процесса $x(t)$ с помощью структурной функции включает в себя следующие этапы.

1. На основании значений случайного процесса $x(t)$ вычисляются значения структурной функции по формуле (4).

2. На основании значений структурной функции процесса определяется идентифицирующая матрица:

(-1) строка	1	1	1	1	1	1
0 строка	$C_x(0)$	$C_x(\Delta t)$	$C_x(2\Delta t)$...	$C_x(n\Delta t)$...
1 строка	$a_1(0)$	$a_1(\Delta t)$	$a_1(2\Delta t)$...	$a_1(n\Delta t)$...
2 строка	$a_2(0)$	$a_2(\Delta t)$	$a_2(2\Delta t)$...	$a_2(n\Delta t)$...
...
m строка	$a_m(0)$	$a_m(\Delta t)$	$a_m(2\Delta t)$...	$a_m(n\Delta t)$...
...

в которой (-1)-строка содержит значения единичной функции $1(t)$, а (0)-строка – значения структурной функции входного сигнала $C_x(k\Delta t)$ в моменты времени $\{n\Delta t\}_0^N$; Δt – шаг дискретизации, а элементы $a_m(n\Delta t)$ последовательно определяются с помощью соотношения:

$$(7) \quad a_m(n\Delta t) = \frac{a_{m-2}((n+1)\Delta t)}{a_{m-2}(0)} - \frac{a_{m-1}((n+1)\Delta t)}{a_{m-1}(0)},$$

где $a_{-1}(n) = 1(n\Delta t)$; $a_0(n) = C_x(n\Delta t)$; $m = 1, 2, 3, \dots, n = 0, 1, 2, \dots$. Вычисление элементов $a_m(n\Delta t)$ продолжается до появления нулевой строки.

3. Элементы первого столбца матрицы (6) порождают частные числители правильной C -дроби [13, 15], что позволяет получить модель структурной функции сигнала в форме дискретной передаточной функции (ДПФ) виртуального объекта:

$$(8) \quad G_{C_x}(z) = \left[\frac{a_0(0)/a_{-1}(0)}{1}; \frac{a_i(0)z^{-1}}{1} \right]_1^\infty = \frac{C_x(1)z^{-1}}{1 + \frac{a_1(0)z^{-1}}{1 + \frac{a_2(0)z^{-1}}{1 + \frac{a_2(0)z^{-1}}{\dots}}}}.$$

Если в некоторой i -й строке ($i = 0, 1, 2, \dots$) матрицы (6) конечное число k_i первых элементов равны нулю, а последующие элементы отличны от нуля, то необходимо осуществить сдвиг влево на k_i элементов до появления в нулевом столбце ненулевого элемента и далее продолжить определение других элементов матрицы (6) по правилу (7). Для i -й строки при восстановлении модели структурной функции элемент a_i умножается на z^{-k_i} .

4. Производится проверка устойчивости полученного виртуального объекта по математической модели ДПФ (8) с помощью любого из известных критериев устойчивости [31]. В случае устойчивости объекта следует утверждать, что данный сигнал стационарен. В противном случае сигнал является нестационарным.

Пример 1. Пусть $x(t)$ – стационарный случайный процесс с нулевым средним и корреляционной функцией $R_{xx}(t) = e^{-0,3t}$. Структурная функция процесса имеет вид $C_x(\tau) = 2 - 2e^{-0,3\tau}$. На основании значений структурной функции, взятых с шагом дискретизации $\Delta t = 1$, построим идентифицирующую матрицу (6):

1	1	1	1	1	1
0,51836	0,90238	1,18686	1,39761	1,55374	1,66940
-0,74082	-1,28963	-1,69620	-1,99739	-2,22052	
0	0	0	0		

С учетом того, что в нулевой строке был осуществлен сдвиг влево на один элемент, получаем модель структурной функции в форме ДПФ виртуального объекта

$$G_{C_x}(z) = \frac{0,51836z^{-1}}{1 - 0,74082z^{-1}}.$$

ДПФ объекта имеет один полюс $z^n = 0,74082$, который лежит внутри единичной окружности с центром в начале координат плоскости z . Следовательно, объект устойчив, а это означает, что случайный процесс $x(t)$ является стационарным.

Чтобы убедиться в том, что структурная функция может служить индикатором структурной неоднородности, рассмотрим пример.

Пример 2. Имеется стохастический процесс $y^1(t)$, взятый из [3]:

$$(9) \quad y^1(t) = 10 + x(t) + x(t-1), \quad t = 1, \dots, 50,$$

где $x(t)$ – белый шум с нулевым средним и единичной дисперсией.

В момент времени $t = 51$ происходит изменение параметров модели случайного процесса, и она принимает вид [3]:

$$(10) \quad y^2(t) = 10 + x(t) - x(t-1), \quad t = 51, \dots, 100.$$

Графики процессов (9) и (10) представлены на рис. 1. Нетрудно показать, что математические ожидания и дисперсии процессов $y^1(t)$ и $y^2(t)$ одинаковы:

$$M(y^1(t)) = M(y^2(t)) = 10,$$

$$D(y^1(t)) = D(y^2(t)) = 2,$$

поэтому применение инверсионного критерия [16, 24, 25] не является эффективным.

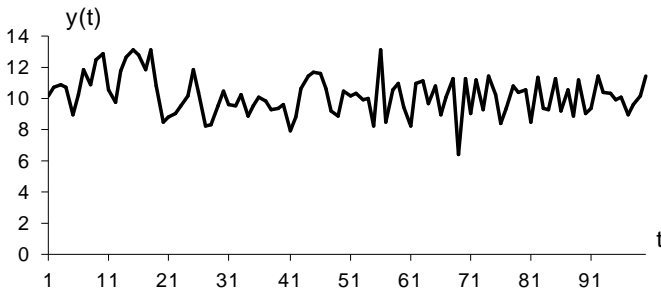


Рис. 1. Реализации случайных процессов (9) при $t = 1, \dots, 50$ и (10) при $t = 51, \dots, 100$

Корреляционные функции процессов (9) и (10) соответственно равны

$$(11) R_{y_1}(t) = \begin{cases} 2, & t = 0, \\ 1, & t = 1, \\ 0, & t \geq 2; \end{cases}$$

$$(12) R_{y_2}(t) = \begin{cases} 2, & t = 0, \\ -1, & t = 1, \\ 0, & t \geq 2. \end{cases}$$

Тогда структурные функции процессов (9) (рис. 2а) и (10) (рис. 2б) соответственно равны:

$$(13) C_{y_1}(t) = \begin{cases} 0, & t = 0, \\ 2, & t = 1, \\ 4, & t \geq 2; \end{cases}$$

$$(14) C_{y_2}(t) = \begin{cases} 0, & t = 0, \\ 6, & t = 1, \\ 4, & t \geq 2. \end{cases}$$

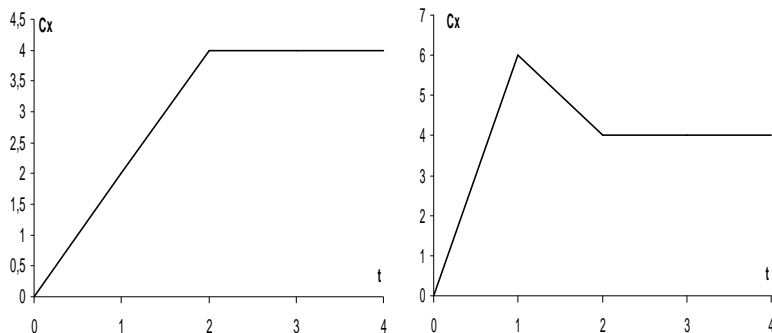


Рис. 2. а) Структурная функция (13),
б) структурная функция (14)

Определим структурную функцию стационарного случайного процесса на всем интервале $t = 1, \dots, 100$ (рис. 3.) по формуле (4). Это доказывает, что объединение двух стационарных реализаций с разными структурными функциями приводит к

стационарной реализации с отличной от исходных структурной функцией. Аналогичное явление будет постепенно происходить при добавлении значений случайного процесса в режиме реального времени.

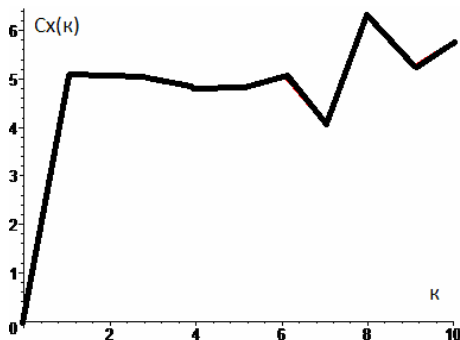


Рис. 3. Структурная функция объединенного процесса $t = 1, \dots, 100$

Методы экспериментального исследования характеристик объектов, явлений, процессов, как правило, предусматривают проведение определенного числа опытов с фиксацией результатов [11]. Лишь после окончания последнего опыта, используя тот или иной алгоритм, находится нужная характеристика. Это относится и к структурной функции, определяемой на практике по формуле (4). В системах реального времени с использованием мониторинга наблюдений временных рядов необходимо отслеживать изменения их свойств на основе итерационных характеристик. Если какая-либо характеристика изменяется, то итерационные алгоритмы «следят» за этими изменениями [11, 23]. Так, например, итерационная формула для вычисления начального s -го момента [11] после N -го измерения имеет вид:

$$(15) \quad m_N^s(x) = m_{N-1}^s(x) + \frac{1}{N} (x_N^s - m_{N-1}^s(x)).$$

Получим рекуррентную формулу для вычисления структурной функции случайного процесса. Подставляя в данное

равенство выражение, задающее структурную функцию (1), получим:

$$(16) C_x^N(k) = C_x^{N-1}(k) + \frac{1}{N-k} \left[(x(N) - x(N-k))^2 - C_x^{N-1}(k) \right],$$

где $C_x^{N-1}(k)$ – оценка структурной функции, полученная на основе $(N-1)$ измерения, $k = 0, 1, 2, \dots$

Для проверки полученного равенства используем основную формулу (4) практического определения структурной функции. Раскроем скобки в полученном выражении (16):

$$\begin{aligned} C_x^N(k) &= \frac{1}{N-k-1} \sum_{i=1}^{N-1-k} (x(i) - x(i+k))^2 + \\ & \frac{1}{N-k} \left[(x(N) - x(N-k))^2 - \frac{1}{N-k-1} \sum_{i=1}^{N-1-k} (x(i) - x(i+k))^2 \right] = \\ &= \frac{1}{N-k} \sum_{i=1}^{N-1-k} (x(i) - x(i+k))^2 + \frac{1}{N-k} (x(N) - x(N-k))^2 = \\ &= \frac{1}{N-k} \sum_{i=1}^{N-k} (x(i) - x(i+k))^2. \end{aligned}$$

Таким образом, итерационная формула структурной функции тождественна вычислению структурной функции по N измерениям.

Использование итерационной формулы расчета структурной функции дает возможность получать сведения об изменении характеристик процесса в реальном масштабе времени. Все преимущества структурного анализа продолжают иметь место, кроме того, структурная функция может служить индикатором изменения как математического ожидания, дисперсии, так и корреляционной функции случайного процесса, а также наличия в нем нестационарной компоненты.

Разовьем идею ретроспективного анализа стационарности случайного процесса [16, 19, 21, 27] на случай его применения в реальном масштабе времени.

Пусть имеется реализация случайного процесса в режиме нормального функционирования системы. Шаг дискретизации Δt выбран исходя из условия структурно-параметрической идентифицируемости [16, 27]. Необходима предварительная

проверка данной реализации на стационарность и однородность. Кроме того, следует получить идентифицирующую матрицу и модель виртуального объекта для дальнейшего контроля и обнаружения изменений свойств данных измерений. Далее следует осуществлять контроль времени начала возможных изменений внутренних свойств виртуального объекта на основе значений выходной структурной функции. Для этого осуществляется перерасчет структурной функции случайного процесса по итерационной формуле (16). Поскольку наличие нулевой строки в идентифицирующей матрице связано с определением порядка функции и показывает механизм аппроксимации непрерывными дробями ДПФ объекта [13], то различные изменения, которые могут появляться в нулевой строке, будут свидетельствовать об изменениях, происходящих в объекте. Теория непрерывных дробей относится к разделу алгоритмической математики, поэтому теоретически доказать этот факт не представляется возможным. В [13] данное утверждение доказано для апериодического звена 1-го порядка. Для объектов 2-го порядка справедливость утверждения доказана на ЭВМ. Из-за чрезвычайной громоздкости доказательства в общем случае данный факт можно показать лишь на тестовых примерах.

Таким образом, для решения задачи обнаружения структурно-параметрических изменений в реальном масштабе времени будем использовать итерационный алгоритм, в котором при добавлении каждого i -го измерения процесса производится перерасчет выходной структурной функции, т.е. уточняется ее значение. Если свойства объекта, формирующего на выходе структурную функцию, изменяются, то будет меняться и его модель, следовательно, нулевая строка идентифицирующей матрицы перестает быть таковой [13].

При решении практических задач важным моментом является формирование критериев определения нулевых строк матрицы (6), т.е. определение модели виртуального объекта. На основе модельных исследований были установлены качественные признаки выявления нулевых строк. Таковыми являются [14]:

- близость к нулю элементов строки матрицы;
- изменение элементов строки по величине и знаку;

– резкое увеличение по модулю элементов следующей строки.

Возможна также дополнительная проверка гипотезы о равенстве нулю среднего значения элементов нулевой строки идентифицирующей матрицы с помощью любого непараметрического критерия [2, 6].

Если при добавлении новых данных измерений в идентифицирующей матрице нулевая строка сохраняет свое местоположение, полученная модель виртуального объекта остается неизменной, следовательно, с некоторой заданной вероятностью можно утверждать, что структурно-параметрические изменения отсутствуют. Если же при добавлении новых измерений нулевая строка в матрице отсутствует либо некоторые (или все) ее элементы резко возрастают (по модулю), полученная ранее модель не является адекватной виртуальному объекту. Таким образом, с некоторой заданной вероятностью p_d можно утверждать, что в объекте произошли параметрические и (или) структурные изменения, связанные с изменениями характеристик временного ряда.

5. Алгоритм обнаружения структурно-параметрических изменений стохастических процессов и систем

Проведенные исследования позволили сформулировать алгоритм обнаружения изменения характеристик случайного процесса $x(t)$ с помощью структурной функции, который включает в себя следующие этапы.

Подготовительный этап

А) Имеется $(N - 1)$ измерение случайного процесса $x(i\Delta t)$ в режиме нормального функционирования системы, где Δt - шаг дискретизации, $i = 0, \dots, N - 1$. На данном этапе также необходимо осуществить проверку измерений $x(i\Delta t)$ на стационарность с помощью указанных выше критериев.

Б) На основе измерений случайного процесса $x(i\Delta t)$ осуществляется вычисление значений структурной функции для количества отчетов $(N - 1)$ по формуле (4).

В) На основании значений структурной функции сигнала рассчитывается идентифицирующая матрица (6). Вычисления продолжаются до появления нулевой строки в матрице (обозначим ее за m -ю). Затем определяется модель виртуального объекта вида (8).

Г) Осуществляется проверка значений нулевой строки на независимость с заданным уровнем значимости α . Если гипотеза о независимости не отвергается, то строится α -процентный доверительный интервал $(e(k) - \alpha; e(k) + \alpha)$ для значений нулевой строки. Если гипотеза отвергается, то данная строка не является нулевой, и получить модель виртуального объекта не представляется возможным. В этом случае вопрос определения модели виртуального объекта и структурной функции на его выходе остается открытым.

Этап контроля возможных изменений случайного процесса

А) При добавлении N -го наблюдения оценка структурной функции пересчитывается по рекуррентной формуле (16).

Б) Осуществляется перерасчет идентифицирующей матрицы (6).

В) Производится оценка покрытия элементов m -й строки идентифицирующей матрицы α -процентным интервалом. Если хотя бы одно из значений вышло за границы интервала, с вероятностью $p_d = 1 - \alpha$ можно утверждать, что модель виртуального объекта изменилась. Последнее означает смену какой-либо характеристики случайного процесса (математического ожидания, дисперсии, корреляционной функции, появления нестационарности). В случае покрытия всех элементов нулевой строки α -процентным интервалом можно утверждать, что характеристики случайного процесса не изменялись.

6. Примеры использования алгоритма для обнаружения структурно-параметрических изменений стохастических процессов и систем

Пример 3. Имеется 250 измерений стационарного случайного процесса, заданного конечно-разностным уравнением:

$$(17) \quad x(k) = 0,9x(k-1) + n(k),$$

где $n(k)$ – белый шум с нулевым математическим ожиданием и дисперсией, равной 0,25. Получим значения структурной функции случайного процесса по формуле (4). На основе этих значений рассчитаем идентифицирующую матрицу (6):

№ столбца	1	2	3	4	5	6
(-1) строка	1	1	1	1	1	1
0 строка	0,2821	0,5387	0,7588	0,9771	1,1426	1,2870
1 строка	-0,9099	-1,6899	-2,4639	-3,0507	-3,5626	-4,1004
2 строка	0,0527	-0,0181	0,1112	0,1354	0,0561	-0,1124
№ столбца	7	8	9	10	11	12
(-1) строка	1	1	1	1	1	1
0 строка	1,4387	1,6200	1,7822	1,4387	1,6200	1,7822
1 строка	-4,7432	-5,3179	-5,8010	-6,086	-6,3733	
2 строка	-0,1013	-0,0575	0,1124	0,0816		

Вторая строка матрицы является нулевой. С учетом того, что в нулевой строке был осуществлен сдвиг влево на один элемент, получаем модель структурной функции в форме ДПФ виртуального объекта

$$G_{C_x}(z) = \frac{0,2821z^{-1}}{1 - 0,9099z^{-1}}.$$

Произведем проверку гипотезы о независимости значений $e(k)$, стоящих в нулевой строке и представляющих собой погрешности вычислений. Для этого используем критерий серий [2, 16]. Составим вариационный ряд $\tilde{e}(k)$ из $e(k)$, $k = 1, \dots, 10$, и вычислим выборочное значение медианы по формуле:

$$med = \frac{1}{2}(\tilde{e}(k/2) + \tilde{e}(k/2 + 1)) = 0,0544.$$

Сравнивая каждое $e(k)$ с медианой, ставим знак «+», если $e(k) \geq med$ и знак «-», если $e(k) < med$. Образуется последовательность вида: -- + + --- + +. *Серией* называется последовательность однотипных наблюдений, после которой следует наблюдения противоположного типа или же вообще нет никаких наблюдений. В вышеприведенном примере число серий $S = 4$.

Принятие гипотезы H_0 о независимости значений производится, если верно условие $S(10/2; 1 - a/2) < S < S(10/2; a/2)$, где критические значения определяются нормальным распределе-

нием S и приводятся в [2], α – уровень значимости. Для нашего случая при уровне значимости $\alpha = 0,05$ условие верно: $2 < 4 < 9$. Последнее свидетельствует об отсутствии корреляции остатков в нулевой строке.

Зададим пятипроцентный интервал $(e(k) - \alpha; e(k) + \alpha)$, в который должны попадать значения, стоящие в нулевой строке, в случае отсутствия структурно-параметрических изменений (таблица 1). Если при добавлении нового измерения хотя бы одно из значений, стоящих во 2-й строке выйдет за границы этого интервала, то с вероятностью $p_d = 0,95$ можно утверждать, что модель объекта изменилась.

Таблица 1. Интервал нулевой строки к примеру 3

№ столбца	1	2	3	4	5
Верхняя граница	0,1027	0,0319	0,1612	0,1854	0,1061
Нижняя граница	0,0027	-0,0681	0,0612	0,0854	0,0061
№ столбца	6	7	8	9	10
Верхняя граница	-0,0624	-0,0513	-0,0075	0,1624	0,1316
Нижняя граница	-0,1624	-0,1513	-0,1075	0,0624	0,0316

Добавим 251-е значение случайного процесса $x(k)$ без изменения его характеристик. Пересчитав структурную функцию по формуле (16), построим идентифицирующую матрицу:

№ столбца	1	2	3	4	5	6
(-1) строка	1	1	1	1	1	1
0 строка	0,2810	0,5371	0,7561	0,9732	1,1380	1,2856
1 строка	-0,9116	-1,6911	-2,4639	-3,0504	-3,5754	-4,1428
2 строка	0,0566	-0,0117	0,1178	0,1284	0,0311	-0,0826
№ столбца	7	8	9	10	11	12
(-1) строка	1	1	1	1	1	1
0 строка	1,4449	1,6193	1,7800	1,9105	1,9914	2,0713
1 строка	-4,7636	-5,3357	-5,8000	-6,0879	-6,3725	
2 строка	-0,0893	-0,0265	0,1219	0,0977		

Значения, стоящие во 2-й строке, попадают в интервал (таблица 1), следовательно, нулевая строка сохраняется, и характеристики процесса не изменились.

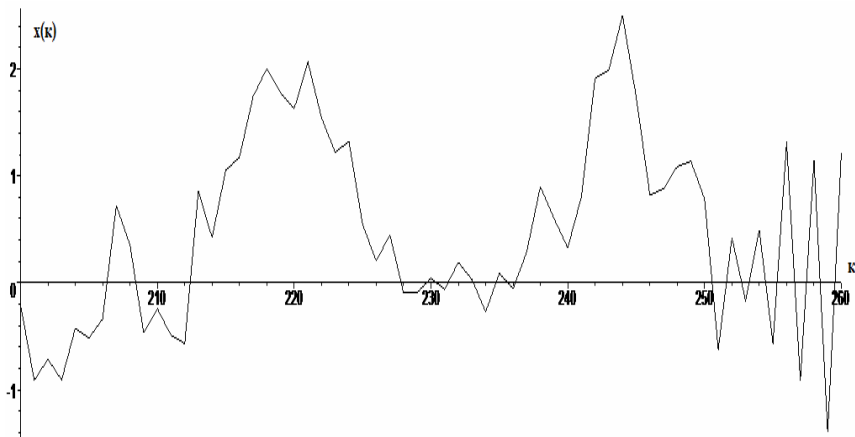


Рис. 4. График процессов к примеру 3

На 252-м шаге изменяется корреляционная функция случайного процесса. Модель процесса принимает вид:

$$(18) \quad x(k) = -0,9x(k-1) + n(k).$$

Графики процессов (17) и (18) приведены на рис. 4. Пересчитаем структурную функцию по рекуррентной формуле (16). Матрица на основе полученных значений примет вид:

№ столбца	1	2	3	4	5	6
(-1) строка	1	1	1	1	1	1
0 строка	0,2889	0,5490	0,7675	0,9823	1,1464	1,3046
1 строка	-0,9006	-1,6570	-	-2,9689	-3,5162	-4,0987
			2,4006			
2 строка	0,0608	-0,0084	0,1042	0,0647	-0,0346	-0,1002

№ столбца	7	8	9	10	11	12
(-1) строка	1	1	1	1	1	1
0 строка	1,4728	1,6414	1,8013	1,9190	1,9942	2,0774
1 строка	-4,6823	-5,3179	-5,2360	-5,9039	-6,1918	
2 строка	-0,1013	-0,1313	-0,0298	0,0880		

Значения в 5, 8, 9 столбцах 2-й строки выходят за границы интервала. Следовательно, изменение модели обнаружено без запаздывания. Добавим дополнительно 253-е наблюдение и посмотрим, какие изменения произойдут в матрице. Она принимает вид:

№ столбца	1	2	3	4	5	6
(-1) строка	1	1	1	1	1	1
0 строка	0,3024	0,5435	0,7659	0,9805	1,1408	1,2863
1 строка	-0,7974	-1,5329	-2,2456	-2,7729	-3,2541	-3,8297
2 строка	-0,1249	-0,2794	-0,2347	-0,3080	-0,5486	-0,7766
№ столбца	7	8	9	10	11	12
(-1) строка	1	1	1	1	1	1
0 строка	1,4604	1,6542	1,8079	1,9303	1,9899	2,0600
1 строка	-4,4706	-4,9789	-5,3836	-5,5807	-5,8124	
2 строка	-0,7732	-0,7724	-0,6148	-0,7084		

Вторая строка матрицы претерпевает значительные изменения и перестает быть нулевой. Ни одно из ее значений не попадает в интервал. Таким образом, изменение корреляционной функции обнаружено без запаздывания.

Пример 4. Имеется 50 значений временного ряда $x(k)$, математическое ожидание которого равно нулю. В 51-й момент времени происходит изменение среднего значения ряда, которое становится равным 2 (рис. 5).

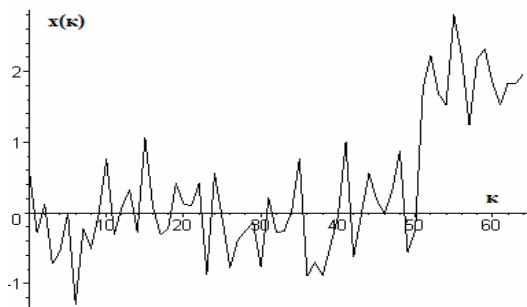


Рис. 5. Значения временного ряда $x(k)$ к примеру 4

На основе имеющихся 50 измерений рассчитаем структурную функцию (4) и идентифицирующую матрицу (6):

1	2	3	4	5	6	7
1	1	1	1	1	1	1
0,5389	0,5114	0,4726	0,5891	0,5458	0,5619	0,5973
0,0511	0,1230	-0,0931	-0,0127	-0,0426	-0,1084	

Первая строка нулевая, ее значения представляют собой независимые ошибки вычислений. Пятипроцентный интервал представлен в таблице 2.

Таблица 2. Интервал нулевой строки к примеру 4

№ столбца	1	2	2	4	5	6
Верхняя граница	0,1011	0,1730	-0,0431	0,0373	0,0074	-0,0584
Нижняя граница	0,0011	0,0730	-0,1431	-0,0627	-0,0926	-0,1584

После добавления 51 измерения структурная функция пересчитывается по формуле (16) и заполняется матрица:

1	2	3	4	5	6	7
1	1	1	1	1	1	1
0,6352	0,5168	0,5062	0,6432	0,5875	0,5813	0,6560
0,1863	0,2030	-0,0126	0,0752	0,0849	-0,0327	

Элементы 1–5 столбцов выходят за границы интервала. Изменение характеристики случайного процесса обнаружено без запаздывания.

Для сравнения работоспособности критериев данная последовательность была также обработана с помощью инверсионного алгоритма [16, 24, 25], который на 56-м измерении обнаружил возрастающий тренд (запаздывание 6 значений).

Пример 5. 350 измерений случайного процесса $x(k)$
 (19) $x(k) = 0,5x(k-1) + 2n(k-1)$

производились в нормальном режиме, где $n(k)$ – белый шум с нулевым математическим ожиданием и дисперсией, равной 0,09. На основании значений структурной функции была рассчитана идентифицирующая матрица:

№ столбца	1	2	3	4
(-1) строка	1	1	1	1
0 строка	0,5076	0,7593	0,8840	0,9731
1 строка	-0,4960	-0,7415	-0,9171	-0,9313
2 строка	0,0011	-0,1074	0,0395	0,0012
№ столбца	5	6	7	8
(-1) строка	1	1	1	1
0 строка	0,9803	0,9935	0,9546	0,9891
1 строка	-0,9574	-0,8807	-0,9486	
2 строка	0,1820	-0,0317		

Вторая строка матрицы нулевая, ее элементы представляют собой независимые погрешности вычислений. Пятипроцентный интервал для нулевой строки представлен в таблице 3.

Таблица 3. Интервал нулевой строки к примеру 5

№ столбца	1	2	3	4	5	6
Верхняя граница	0,0511	-0,0574	0,0895	0,0512	0,2320	0,0183
Нижняя граница	-0,0489	-0,1574	-0,0105	-0,0488	0,1320	-0,0817

На 351-ом измерении появляется нестационарность в форме линейного тренда и значения стохастического процесса становятся равными:

$$(20) \quad x(k) = 0,5x(k-1) + 2n(k-1) + 0,04(k-290).$$

Графики процессов (20) и (19) представлены на рис. 6. Матрица после 351-го наблюдения имеет вид:

№ столбца	1	2	3	4
(-1) строка	1	1	1	1
0 строка	0,5063	0,7573	0,8826	0,9711
1 строка	-0,4957	-0,7432	-0,9180	-0,9309
2 строка	-0,0035	-0,1088	0,0401	0,0001
№ столбца	5	6	7	8
(-1) строка	1	1	1	1
0 строка	0,9777	0,9909	0,9566	0,9909
1 строка	-0,9572	-0,8893	-0,9570	
2 строка	0,1632	-0,0413		

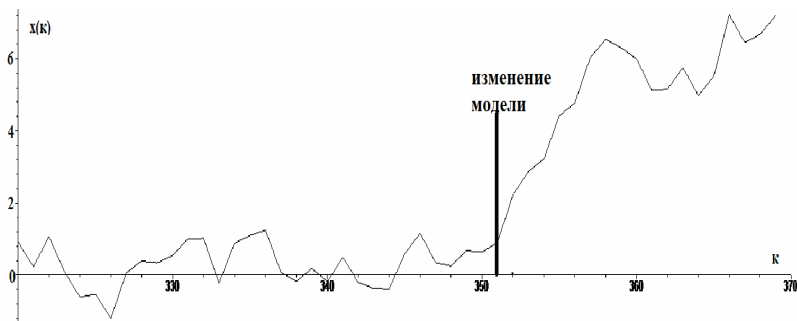


Рис. 6. График процесса до и после появления тренда к примеру 5

Все значения нулевой строки попали в интервал. Пересчитав структурную функцию и идентифицирующую матрицу с поступлением 352-го наблюдения, получим:

№ столбца	1	2	3	4
(-1) строка	1	1	1	1
0 строка	0,5099	0,7624	0,8869	0,9795
1 строка	-0,4951	-0,7394	-0,9210	-0,9317
2 строка	0,0016	-0,1208	0,0391	0,0246
№ столбца	5	6	7	8
(-1) строка	1	1	1	1
0 строка	0,9850	0,9913	0,9616	1,0078
1 строка	-0,9442	-0,8858	-0,9764	
2 строка	0,1550	-0,0864		

Значение, стоящее в 6-ом столбце матрицы, не попало в интервал. Добавим 353-е наблюдение:

№ столбца	1	2	3	4
(-1) строка	1	1	1	1
0 строка	0,5096	0,7713	0,8984	0,9904
1 строка	-0,5133	-0,7633	-0,9435	-0,9657
2 строка	0,0265	-0,0745	0,0623	0,2264
№ столбца	5	6	7	8
(-1) строка	1	1	1	1
0 строка	1,0018	1,0068	0,9672	1,0200
1 строка	-0,9755	-0,8979	-1,0014	
2 строка	-0,0527	-0,2060		

Значения в 4, 5, 6 столбцах не попали в интервал. Изменение модели обнаружено после 352-го наблюдения (запаздывание – 1 наблюдение).

Пример 6. Имеется 250 измерений стационарного случайного процесса, имеющего модель

$$(21) \quad x(k) = 0,85x(k-1) + n(k-1) + 0,4n(k),$$

где $n(k)$ – белый шум с нулевым математическим ожиданием и дисперсией, равной 0,04. На основе имеющихся наблюдений была получена идентифицирующая матрица для структурной функции:

1	2	3	4	5	6
1	1	1	1	1	1
0,0512	0,1287	0,1943	0,2514	0,2957	0,3308
-1,5163	-2,7969	-3,9141	-4,7804	-5,4665	
0,6717	1,2156	1,7614	2,1753		
0,0349	-0,0409	-0,0857			

Третья строка матрицы нулевая. Пятипроцентный интервал для нулевой строки представлен в таблице 4.

Таблица 4. Интервал нулевой строки к примеру 6

№	1	2	3
Верхняя граница	0,0849	0,0091	-0,0357
Нижняя граница	-0,0151	-0,0909	-0,1357

На 251-м наблюдении происходит параметрическое изменение модели случайного процесса. Модель процесса принимает вид:

$$(22) \quad x(k) = 0,95x(k-1) + 3n(k-1) + 0,8n(k).$$

График процессов (21) и (22) представлен на рис. 7.

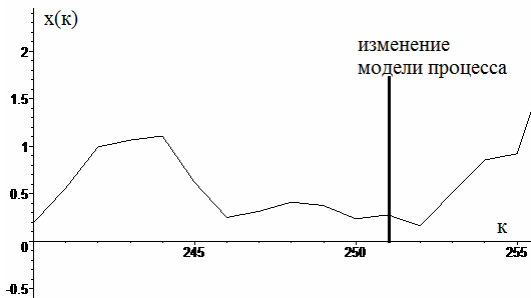


Рис. 7. График процессов (21) и (22) к примеру 6

После пересчета структурной функции на 251 шаге идентифицирующая матрица имеет вид:

1	2	3	4	5	6
1	1	1	1	1	1
0,0510	0,1283	0,1935	0,2504	0,2945	0,3300
-1,5167	-2,7978	-3,9133	-4,7794	-5,4745	
0,6720	1,2176	1,7621	2,1621		
0,0328	-0,0419	-0,0777			

Все значения 3-ей строки попадают в интервал. Пересчитаем матрицу после добавления 252 наблюдения:

1	2	3	4	5	6
1	1	1	1	1	1
0,0508	0,1278	0,1930	0,2496	0,2934	0,3286
-1,5145	-2,7973	-3,9130	-4,7750	-5,4679	
0,6675	1,2136	1,7602	2,1645		
0,0287	-0,0534	-0,0901			

Нулевая строка по-прежнему третья. Добавим следующее 253 наблюдение:

1	2	3	4	5	6
1	1	1	1	1	1
0,0511	0,1275	0,1925	0,2487	0,2923	0,3275
-1,4948	-2,7670	-3,8673	-4,7202	-5,4087	
0,6437	1,1798	1,7095	2,1018		
0,0183	-0,0685	-0,0107			

Все значения нулевой строки попали в интервал. Пересчитав структурную функцию и идентифицирующую матрицу на 254 шаге, получим:

1	2	3	4	5	6
1	1	1	1	1	1
0,0509	0,1276	0,1940	0,2494	0,2930	0,3274
-1,5067	-2,8103	-3,8979	-4,7548	-5,4296	
0,6417	1,2235	1,7423	2,1514		
-0,0416	-0,1284	-0,1972			

Все значения 3-ей строки выходят за границы интервала. Третья строка перестает быть нулевой. Параметрическое изменение модели процесса зафиксировано на 254-м наблюдении. Запаздывание составило 3 измерения.

7. Выводы

Предложенный алгоритм обнаружения структурно-параметрических изменений в моделях динамических объектов, процессов и систем, основанный на использовании модели структурной функции случайного процесса в форме непрерывной дроби, позволяет улавливать изменения либо в момент их появления, либо с небольшим запаздыванием (1–3 значения), что подтверждено рядом модельных примеров. Метод является непараметрическим, не требует знания априорной информации и ввиду его формализованности может применяться для решения различных практических задач контроля над функционированием объектов и систем в режиме реального времени. В отличие от других существующих методов, предложенный метод не требует настройки на конкретный тип изменений, стационарности данных измерений, знания законов распределений данных. Кроме того, на сложность и структуру метода не оказывает влияния неизвестность момента (или его распределения) появления изменения. Метод позволяет эффективно отслеживать изменения всех характеристик случайного процесса: среднего, дисперсии, корреляционной функции и появления нестационарности. Развитая теория структурно-параметрической идентифи-

кации [16, 19–21, 27] на подготовительном этапе позволяет осуществлять эффективную проверку данных на стационарность и однородность, получать модели объектов и систем без их перебора. Помимо этого при проектировании системы дискретного контроля за функционированием непрерывного динамического объекта необходимо, чтобы ее неотъемлемой функцией была возможность изменения шага дискретизации. Принцип вариации шага дискретизации [16, 19–21, 27] является достаточным условием восстановления свойств и характеристик непрерывного объекта и процесса. Неизменность непрерывной модели объекта в случае выполнении принципа вариации шага дискретизации при обработке временных рядов может служить критерием достоверных оценок их свойств.

Литература

1. АНДЕРСОН Т. *Статистический анализ временных рядов*. – М.: Мир, 1976.
2. БЕНДАТ ДЖ., ПИРСОЛ А. *Прикладной анализ случайных процессов*. – М.: Мир, 1989.
3. БОКС ДЖ., ДЖЕНКИНС Г. *Анализ временных рядов. Прогноз и управление*. – М.: Мир, 1974. – Вып. 1., 1974. – Вып. 2.
4. БОРОДКИН Л.И., МОТГЛЬ В.В. *Алгоритм обнаружения моментов изменения параметров уравнения случайного процесса* // Автоматика и телемеханика. – 1976. – №6. – С. 23–31.
5. БРАВЕРМАН Э.М., МУЧНИК И.Б. *Структурные методы обработки эмпирических данных*. – М.: Наука, 1983.
6. БРИЛЛИНДЖЕР Д.Р. *Временные ряды: обработка данных и теория*. – М.: Мир, 1980.
7. БРОДСКИЙ Б.Е., ДАРХОВСКИЙ Б.С. *Непараметрический метод обнаружения моментов переключения двух случайных последовательностей* // Автоматика и телемеханика. – 1989. – №10. – С. 66–74.
8. ВАЛЬД А. *Последовательный анализ*. – М.: Физматгиз, 1960.
9. ВОРОБЕЙЧИКОВ С.Э., КОНЕВ В.В. *Об обнаружении разладок в динамических системах* // Автоматика и телемеханика. – 1990. – №3. – С. 56–58.

10. ДАРХОВСКИЙ Б.С. *Непараметрический метод для апостериорного обнаружения момента «разладки» последовательности независимых случайных величин* // Теория вероятностей и ее применения. – 1976. – Том 21. – С. 180–184.
11. ДУДНИКОВ Е.Г., БАЛАКИРЕВ В.С., КРИВСУНОВ В.С. *Построение математических моделей химико-технологических объектов*. – Ленинград: Химия, 1970.
12. ЖИГЛЯВСКИЙ А.А., КРАСНОВСКИЙ А.Е. *Обнаружение разладки случайных процессов в задачах радиотехники*. – Ленинград: Изд-во Ленинградского университета, 1988.
13. КАРТАШОВ В.Я. *Анализ и исследование аппроксимационных свойств непрерывных дробей при решении задачи структурно-параметрической идентификации динамических объектов* // Препринт №22. – Барнаул: Изд-во Алтайского госуниверситета, 1996.
14. КАРТАШОВА Л.В., КАРТАШОВ В.Я. *Построение причинно-следственных моделей социально-экономических процессов: монография*. – Томск: Издательство Томского государственного педагогического университета, 2008.
15. КАРТАШОВ В.Я. *Непрерывные дроби (определения и свойства)*. – Кемерово: Изд-во Кемеровского госуниверситета, 1999.
16. КАРТАШОВ В.Я., НОВОСЕЛЬЦЕВА М.А. *Идентификация стохастических объектов: учебное пособие*. – Томск: Изд-во Томского государственного педагогического университета, 2008.
17. КАРТАШОВ В.Я., НОВОСЕЛЬЦЕВА М.А. *Обнаружение изменений внутренних свойств динамических объектов* // Материалы XII Международной конференции по мягким вычислениям и измерениям, Санкт-Петербург, 2009. – С. 118–121.
18. КАРТАШОВ В.Я., НОВОСЕЛЬЦЕВА М.А. *Определение риска по структурным особенностям временных рядов* // Материалы X Международной конференции по мягким вычислениям и измерениям, Санкт-Петербург, 2007. – С. 73–76.
19. КАРТАШОВ В.Я., НОВОСЕЛЬЦЕВА М.А. *Способ идентификации линейного объекта* // Патент РФ №2189622. – 2002. –

- Бюл. №26.
20. КАРТАШОВ В.Я., НОВОСЕЛЬЦЕВА М.А. *Способ идентификации линейного объекта* // Патент РФ №2146063. – 2000. – Бюл. №6.
 21. КАРТАШОВ В.Я., НОВОСЕЛЬЦЕВА М.А. *Структурно-параметрическая идентификация стохастических объектов с использованием непрерывных дробей* // Управление большими системами. – 2008. – Вып. 21. – С. 27–48.
 22. КОНЕВ В.В. *Последовательные оценки параметров стохастических динамических систем.* – Томск: изд-во Томского ун-та, 1985.
 23. МАКС Ж. *Метода и техника обработки сигналов при физических измерениях.* – М.: Мир, 1983. – Том 1.; Том 2.
 24. *МикроЭВМ в информационно-измерительных системах.* – М.: Машиностроение, 1987.
 25. МИРОНОВ И.И., ОСИПОВ С.Н. *Многоконтурные системы обработки информации и активного управления.* – М.: Энергоатомиздат, 1997.
 26. НИКИФОРОВ И.В. *Последовательное обнаружение изменения свойств временных рядов.* – М.: Наука, 1983.
 27. НОВОСЕЛЬЦЕВА М.А. *Идентификация моделей совместных случайных процессов для систем контроля горной техники:* дис. канд. тех. наук. – Кемерово, 2001.
 28. НОВОСЕЛЬЦЕВА М.А. *Оценка риска в стохастических системах* // Материалы VII Всероссийской научно-практической конференции AS'2009 «Системы автоматизации в образовании, науке и производстве». – Новокузнецк: изд-во СибГИУ, 2009. – С.441–446.
 29. *Обнаружение изменения свойств сигналов и динамических систем* / Под ред. М. Бассвиль, А. Банвениста. – М.: Мир, 1989.
 30. РОМАНЕНКО А.Ф. СЕРГЕЕВ Г.А. *Вопросы прикладного анализа случайных процессов.* – М.: Советское радио, 1968.
 31. СОЛОДОВНИКОВ В.В., ПЛОТНИКОВ В.Н., ЯКОВЛЕВ А.В. *Основы теории и элементы систем автоматического регулирования:* учебное пособие для вузов. – М.: Машиностроение, 1985.

32. ШИРЯЕВ А.Н. *О марковских достаточных статистиках в неаддитивных байесовских задачах последовательного анализа* // Теория вероятностей и ее применения. – 1964. – №IX, 4. – С. 670–686.
33. ШИРЯЕВ А.Н. *Статистический последовательный анализ*. – М.: Наука, 1976.
34. PAGE E.S. *Continuous Inspecting Schemes* // Biometrika. – 1954. – Vol. 41. – P. 100–114.
35. WILLISKY A.S. *A Survey of Design Methods for Failure Detection in Dynamic Systems* // Automatica. – 1976. – Vol. 12. – P. 601–611.

DETECTION OF STRUCTURALLY-PARAMETRICAL CHANGES IN STOCHASTIC SYSTEMS IN REAL TIME BY ALGORITHMS OF CONTINUED FRACTIONS AND STRUCTURE ANALYSIS

Vladimir Kartashov, head of the Department of Automation of Researches and Technical Cybernetics of the Kemerovo State University, Doctor of Engineering, professor (kartash@kemsu.ru).

Marina Novoseltseva, associate professor of Department of Automation of Researches and Technical Cybernetics of the Kemerovo State University, Candidate of Engineering, (aanov@pochta.ru).

Abstract: An algorithm is developed for real-time detection of structural and parametric changes in stochastic processes and systems. The algorithm is based on the combined use of structural analysis and theory of continued fractions.

Keywords: structural function, the continued fraction, identifying matrix.

Статья представлена к публикации членом редакционной коллегии Я. И. Квинто

УДК 004.724.2 + 004.272.43
ББК 3.9.7.3.02

РАСПРЕДЕЛЕННЫЙ ПОЛНЫЙ КОММУТАТОР КАК «ИДЕАЛЬНАЯ» СИСТЕМНАЯ СЕТЬ ДЛЯ МНОГОПРОЦЕССОРНЫХ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ

Каравай М. Ф.¹, Подлазов В. С.²

*(Учреждение Российской академии наук
Институт проблем управления РАН, Москва)*

Рассматриваются методы построения распределенного полного коммутатора любого размера, составленного из полных коммутаторов и разветвителей каналов малого фиксированного размера. Распределенный коммутатор сохраняет свойства неблокируемости и самомаршрутизируемости полного коммутатора и образует функционально идеальную системную сеть.

Ключевые слова: многопроцессорные вычислительные системы, идеальные системные сети, распределенный полный коммутатор, канальная коммутация, червячная маршрутизация, неблокируемые сети, самомаршрутизируемые сети.

1. Введение

За сетями многопроцессорных вычислительных систем (МВС) в настоящее время утвердился термин системные сети [24]. В литературе по суперкомпьютерным технологиям часто встречается график зависимости времени счета гипотетической задачи от числа задействованных процессоров (рис. 1) [1]. Рост времени выполнения всей задачи при увеличении числа процессоров объясняется простоями процессоров, возникающих в зна-

¹ Михаил Федорович Каравай, доктор технических наук, доцент (mkaravay@ipu.ru, Москва, ул. Профсоюзная, д. 65, тел. (495) 334-90-00).

² Виктор Сергеевич Подлазов, доктор технических наук, доцент (podlazov@ipu.ru, Москва, ул. Профсоюзная, д. 65, тел. (495) 334-78-31).

чительной степени из-за задержек доставки пакетов данных по системной сети.

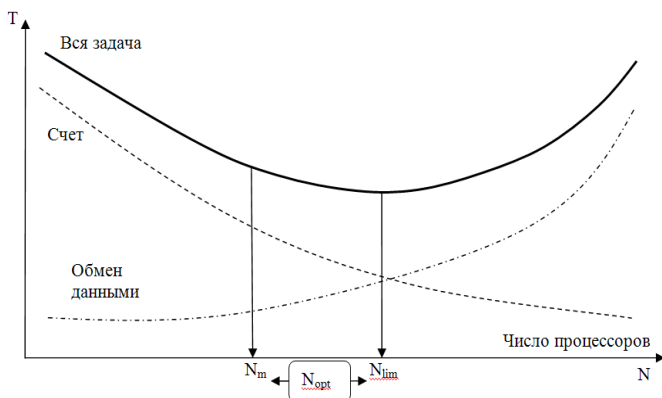


Рис. 1. График времени решения гипотетической задачи на вычислительном кластере

Задержки доставки сильно зависят от числа промежуточных буферизаций пакетов данных в системной сети на пути между источником и приемником. Современная системная сеть (СС) – это часто 2-х, 3-х, 4-хмерные торы [6, 14] или сеть Клоза [25]. Уже на подходе шестимерные торы³. Все они не свободны от конфликтов при параллельной передаче пакетов многими абонентами, которые разрешаются посредством буферизации пакетов.

Время доставки отдельного пакета по пустой СС составляет величину $T_0 = \alpha + b/v$, где α (сек) – это латентность сети, v (бит/сек) – скорость передачи и b (бит) – размер пакета. При наличии очередей пакетов в промежуточных буферах время доставки составляет величину $T = T_0 + Q \cdot b/v$, где Q – суммарный размер очередей на отдельном пути от источника к приемнику в СС. Из теории массового обслуживания известно, что при высокой загрузке сети величина $Q \cdot b/v$ может быть много больше T_0 . Поэтому в литературе [21] появилось предложение считать идеальной ту СС, в которой обеспечиваются прямые каналные со-

³ Ласточка в облаках // Суперкомпьютеры. 2010. № 2. С. 17 – 19.

единения (без промежуточной буферизации) для любой пары абонентов сети при параллельной передаче пакетов от всех абонентов, т.е. в которой $Q = 0$. Правильнее такую СС называть функционально идеальной, т.к. здесь не учитывается вопрос о ее сложности. Однако для краткости в дальнейшем используется термин «идеальная СС».

Общепринятой моделью параллельной передачи данных по СС является произвольная перестановка пакетов данных между всеми абонентами, поскольку она наилучшим образом характеризует логические возможности сети. Какая СС на N абонентов является функционально идеальной? Очевидно, это СС со структурой полного графа, которая может быть как распределенной, так и сосредоточенной. В первом случае предполагается наличие у каждого абонента $N - 1$ дуплексных портов и использование в СС $N(N - 1)$ дуплексных каналов. Во втором случае предполагается использование СБИС полного коммутатора $N \times N$ с N дуплексными портами. В обоих случаях невозможно построить СС с большим числом абонентов либо вследствие большого числа портов у каждого абонента и каналов между ними, либо из-за невозможности создания СБИС с необходимым числом портов.

Поставим задачу создания функционально идеальной СС, построенной в элементной базе, состоящей из коммутаторов и разветвителей каналов малого размера. Такая элементная база имеется, например, для интерфейса PCI-Express [2] и технологии Space Wire [11, 12].

Пусть имеется исходная идеальная СС (рис. 2), объединяющая K абонентов – ИС(K).

Ставится задача расширить ИС(K) до идеальной сети РС(R), объединяющей $R > K$ абонентов. Более конкретно, мы хотим строить расширенную сеть РС(R), соблюдая следующие условия:

1. РС(R) должна сохранять маршрутные свойства ИС(K).
2. РС(R) должна строиться из тех же схемных компонент, что и ИС(K).
3. РС(R) должна быть расширяема до любого сколь угодно большого R при любом K при сохранении свойств пп. 1-2.

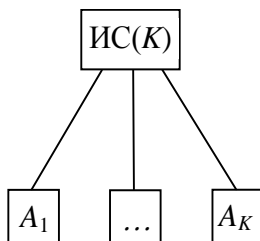


Рис. 2. Исходная сеть на K абонентов

Основными маршрутными свойствами идеальной СС со структурой полного графа являются ее неблокируемость и самомаршрутизируемость. Неблокируемость означает возможность бесконфликтно осуществлять произвольную перестановку пакетов данных между абонентами при параллельной передаче пакетов от всех абонентов, а самомаршрутизируемость – возможность прокладки маршрута при перестановке пакетов каждым абонентом самостоятельно независимо от других абонентов.

Пусть теперь идеальная исходная СС – это полный коммутатор $m \times m$ с m дуплексными портами. Обычный метод расширения полного коммутатора на большее число абонентов – это построение многокаскадной сети Клоза [23] или гиперкуба [18]. Они привлекательны тем, что имеют меньшую сложность, чем полный коммутатор того же размера, и обладают свойством перестраиваемости. Перестраиваемая сеть имеет отдельное бесконфликтное расписание для любой перестановки данных между входными и выходными портами. Однако его построение требует много больше времени, чем реализация самой перестановки. Поэтому на практике для произвольной перестановки обычно используется самомаршрутизация, например червячная маршрутизация [19, 20]. Она допускает возникновение конфликтов, снижающих пропускную способность сети и увеличивающих задержки передачи пакетов. В итоге оказывается, что сеть Клоза и гиперкуб не сохраняют идеальности полного коммутатора.

Известно, что расширение полного коммутатора в виде двумерного обобщенного гиперкуба или двумерного полного мультикольца [8, 25] сохраняет свойства неблокируемости и самомаршрутизируемости, т.е. обеспечивает сохранение идеальности расширенной сети. Однако при этом выполняются только

пп. 1-2. Отметим, что [25] – это единственная иностранная публикация, известная авторам, в которой имеется частичное пересечение с рассматриваемым методом.

Кроме того, авторами был разработан новый метод построения $PC(R)$, удовлетворяющий условиям 1-2 [3, 4, 7-10], основанный на математической теории неполных уравновешенных блок-схем, исследуемых в комбинаторике.

В дальнейшем эти методы рассматриваются отдельно и дополняются каскадным их применением, которое позволяет удовлетворить условию 3.

Методы, рассматриваемые в разделах 2.1 и 2.3, являются полностью оригинальными, а метод из раздела 2.2 частично пересекается с методом построения самой большой коммутаторной СБИС в [25].

2. Распределенные полные коммутаторы

2.1. РАСПРЕДЕЛЕННЫЙ ПОЛНЫЙ КОММУТАТОР НА БАЗЕ МУЛЬТИКОЛЬЦА

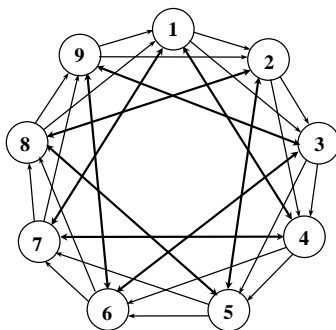


Рис. 3. Полное двумерное мультикольцо с $N = 9$ узлами

Двумерное полное мультикольцо определяется как кольцевой орграф с $N = m^2$ узлами, в котором из каждого узла выходят набор дуг с длинами $(1, 2, \dots, m - 1, m, 2m, \dots, (m - 1)m)$. Длиной дуги мы называем разницу номеров по $\text{mod } N$ инцидентных ей узлов. В таком мультикольце все узлы имеют одинаковую степень $2(m - 1)$.

На рис. 3 приводится пример мультикольца с девятью узлами. Оно имеет дуги с набором длин $\{1, 2, 3, 6\}$. Дуги с длинами 3 и 6 обозначены двунаправленными стрелками. Каждый его узел содержит абонента с $m = 3$ портами и коммутатор $m \times m$ (рис. 4).

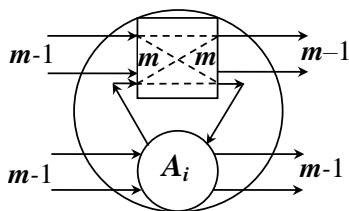


Рис. 4. Коммутатор $m \times m$ и абонент A_i (процессор) в составе i -го узла

Маршрутизация в двумерном мультикольце осуществляется как червячная маршрутизация, т.е. путем прокладки прямого канала между абонентом-источником и абонентом-приемником через промежуточный коммутатор. Эта прокладка осуществляется путем отправки пилотного пакета, содержащего адрес абонента-приемника. Она осуществляется в два этапа – сначала по дугам малых длин (от 1 до $m - 1$), а затем – по дугам больших длин (от m до $(m - 1)m$). На первом этапе используются каналы от абонентов к коммутаторам, а на втором – от коммутаторов к абонентам (или наоборот). Нельзя только смешивать в одном этапе передачи по каналам малых и больших длин, т.к. это может привести к возникновению конфликта и, как следствие, к возникновению тупиковой ситуации.

Схема подсоединений дуг двумерного мультикольца может быть перерисована в виде двудольного орграфа (рис. 5). Одну его долю составляют абоненты, а другую – коммутаторы. Степень всех вершин в каждой доле одинакова и равна m . Значение m выбирается минимальным, при котором любые два абонента связаны одним путем длины два через один и только коммутатор в другой доле. В этом случае число вершин в каждой доле N задается равенством $N = m^2$. Такой орграф авторы называют минимальным квазиполным орграфом [5]. На рис. 4 приведен пример этого орграфа для $m = 3$ ($N = 9$).

Схема соединений между коммутаторами и абонентами при $N = 9$ задается в табл. 1. В любой СС со структурой минимального квазиполного орграфа любые два абонента связаны одним путем длины два через один и только один коммутатор.

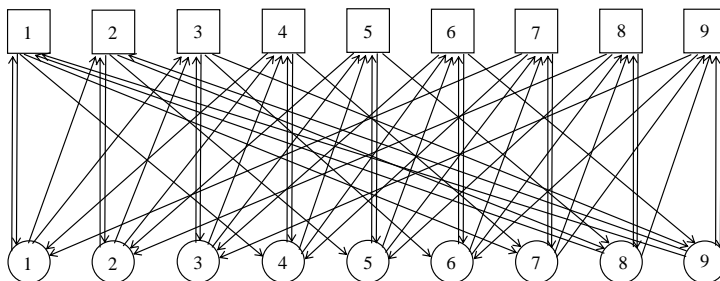


Рис. 5. Минимальный квазиполный орграф для двумерного мультикольца при $t = 3$

Можно сформулировать следующую теорему, которая приводится без формального доказательства, но фактически доказывается всем текстом данного раздела.

Теорема 1. *Системная сеть, построенная по схеме минимального квазиполного орграфа, является неблокируемой и самомаршрутизируемой посредством червячной маршрутизации на произвольной перестановке пакетов данных за счет прокладки отдельных каналов между любыми двумя абонентами.*

Минимальный квазиполный граф описывает схему расширения идеальной СС ИС(m) до идеальной СС РС(m^2), в которой ИС(m) представляет собой полный коммутатор $m \times m$. В табл. 2 приводится схема межсоединений абонентов и коммутаторов при произвольном m . Здесь на пересечении i -ой строки ($1 \leq i \leq N$) и j -го столбца ($1 \leq j \leq m$) в левой части таблицы содержится номер $(i - j) \bmod (N + 1)$, а в правой – номер $(i + (j - 1)m) \bmod (N + 1)$. При этом приемные и передающие порты в каждом абоненте или в каждом коммутаторе могут принадлежать разным дугам. Поэтому дуплексные порты здесь использовать невозможно.

Таблица 1. Схема соединений в квазиполном графе мультикольца при $t = 3$.

Коммутаторы	Симплексные каналы от абонентов			Симплексные каналы к абонентам		
	1	1	9	8	1	4
2	2	1	9	2	5	8
3	3	2	1	3	6	9
4	4	3	2	4	7	1
5	5	4	3	5	8	2
6	6	5	4	6	9	3
7	7	6	5	7	1	4
8	8	7	6	8	2	5
9	9	8	7	9	3	6

Таблица 2. Схема соединений в квазиполном графе мультикольца при произвольном t .

$t \times t$	Входы коммутаторов $t \times t$					Выходы коммутаторов $t \times t$				
1	1	N	$N-1$...	$N-(m-2)$	1	$1+m$	$1+2m$...	$1+(m-1)m$
2	2	1	$N-2$...	$N-(m-3)$	2	$2+m$	$2+2m$...	$2+(m-1)m$
...
t	t	$m-1$	$m-2$...	1	t	$2m$	$3m$...	N
$t+1$	$t+1$	t	$m-1$...	2	$t+1$	$2m+1$	$3m+1$...	1
...
$N-1$	$N-1$	$N-2$	$N-3$...	$N-m$	$N-1$	$m-1$	$2m-1$...	$N-m-1$
N	N	$N-1$	$N-2$...	$N-(m-1)$	N	t	$2m$...	$N-m$

Отличительным свойством идеальной СС РС(m^2) является то, что она является неблокируемой и самомаршрутизируемой на любой однородной t -перестановке, при которой (по определению) в портах каждого абонента имеются пакеты, адресованные только разным абонентам или поступившие от разных аabo-

нентов. Это свойство позволяет каждому абоненту параллельно передавать и принимать до m разных пакетов. Это же свойство позволяет реализовать групповую операцию «все – всем» за два сеанса [26, 22] с суммарной длительностью, равной времени передачи $m + 1$ пакета. Во время первого сеанса каждый источник передает свой пакет с каждого порта m разным приемникам, а каждый приемник получает m пакетов на каждый порт от разных источников. Это занимает время передачи одного пакета. Во время второго сеанса каждый источник передает каждый пакет, принятый в первом сеансе, так же как в первом сеансе. Это занимает время передачи m пакетов. В результате каждый абонент получит **все** пакеты других абонентов.

Выделим у каждого абонента схему формирования m портов (рис. 6), которую составляет разветвитель/объединитель m симплексных каналов. Будем различать два вида таких схем. Первая (рис. 6а) позволяет всем портам работать параллельно и независимо. Она обозначается POK_m^* и изображается с заливкой. Это схема типа многопортовой сетевой карты в PCI-Express. Вторая (рис. 6б) позволяет работать только одному порту. Это схема демультиплексора-мультиплексора. Такие схемы есть в технологии Space Wire.

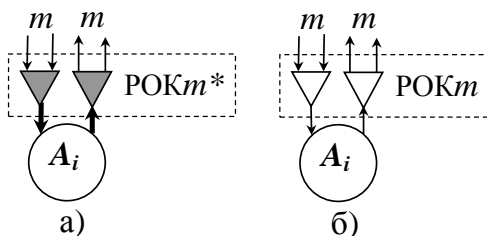


Рис. 6. Разветвитель/объединитель m симплексных каналов POK_m^* (а) и POK_m (б)

Теперь схему $CC\ PC(m^2)$ (рис. 5) можно преобразовать в схему CC с двумя коммутирующими каскадами – коммутаторов $m \times m$ и POK_m^* или POK_m . Пример такой схемы при $m = 3$ приведен на рис. 7. Часть такой схемы CC , располагающаяся выше интерфейса абонент- POK_m^* (POK_m) и заключенная в пунктир-

ный прямоугольник, представляет собой полный распределенный коммутатор $N \times N$, где $N = m^2$. Он обладает свойствами неблокируемости и самомаршрутизируемости. При этом неблокируемость достигается на любой однородной m -перестановке при использовании схемы РОК m^* и на любой обычной 1-перестановке при использовании схемы РОК m .

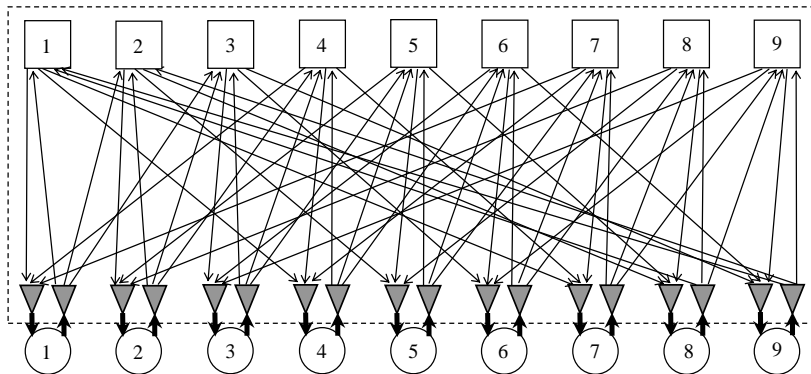


Рис. 7. Распределенный полный коммутатор при $N = 9$

Рассмотрим две такие характеристики распределенного полного коммутатора $PC(N) = PC(m^2)$ как схемную и портовую (канальную) сложность. Базовый способ создания СС в виде полного коммутатора – это соединение абонентов прямыми каналами. При этом схемная сложность S составляет $S = N(N - 1)$. Схемная сложность однокаскадного распределенного коммутатора s составляет $s = 2mN + Nm^2 = 2mN + N^2$. Их отношение: $s/S \approx 1 + 2/N^{1/2}$. Поэтому распределенный коммутатор немного сложнее. Совсем другая картина складывается по числу портов или каналов. Для СС в виде полного коммутатора число портов составляет величину $W = N(N - 1)$, а для распределенного коммутатора – $w = mN$. Их отношение: $w/W \approx 1/N^{1/2}$. Поэтому распределенный коммутатор имеет много меньше каналов и занимает много меньшую площадь при реализации в СБИС или в ПЛИС.

2.2. РАСПРЕДЕЛЕННЫЙ ПОЛНЫЙ КОММУТАТОР НА БАЗЕ ОБОБЩЕННОГО ГИПЕРКУБА

Мультикольцо является не единственной сетевой структурой, которая обеспечивает неблокируемость и самомаршрутизируемость СС. Другой такой структурой является двумерный m -ичный (обобщенный) гиперкуб. В нем каждая строка или столбец из m узлов образует полный граф. Он имеет $N = m^2$ узлов. Пример двумерного обобщенного гиперкуба с $N = 9$ узлами приведен на рис. 8. Хотя он выглядит как граф, но его можно представить как орграф, если учесть, что каждый узел содержит кроме абонента еще и коммутатор $m \times m$ (рис. 9) [5, 25]. Здесь хорошо видно, что каналы могут быть дуплексными, а порты – только симплексными.

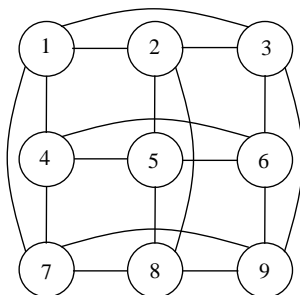


Рис. 8. Двумерный троичный гиперкуб как граф

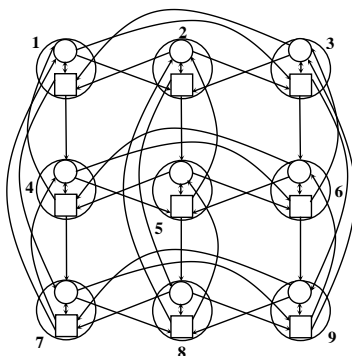


Рис. 9. Двумерный троичный гиперкуб как орграф

Его также можно представить в виде минимального квази-полного орграфа, одну долю которого составляют абоненты, а другую – коммутаторы. Степень всех вершин в каждой доле одинакова и равна m . Число вершин в каждой доле N задается равенством $N = m^2$. В этом орграфе любые два абонента связаны одним путем длины 2 (через один и только один коммутатор). Схема соединений в нем между коммутаторами и абонентами задается в табл. 3. Она отличается от аналогичной схемы для мультикольца.

Таблица 3. Схема соединений в квазиполном орграфе обобщенного гиперкуба при $m = 3$.

Коммутаторы	Симплексные каналы от абонентов			Симплексные каналы к абонентам		
	1	2	3	1	4	7
1	1	2	3	1	4	7
2	1	2	3	2	5	8
3	1	2	3	3	6	9
4	4	5	6	1	4	7
5	4	5	6	2	5	8
6	4	5	6	3	6	9
7	7	8	9	1	4	7
8	7	8	9	2	5	8
9	7	8	9	3	6	9

При произвольном m на пересечении i -й строки ($1 \leq i \leq N$) и j -о столбца ($1 \leq j \leq m$) в левой части таблицы содержится номер $\lfloor i/m \rfloor m + j$, а в правой – номер $(i) \bmod (m + 1) + (j - 1)m$.

Любой минимальный квазиполный граф описывает схему расширения идеальной СС ИС(m) до идеальной СС РС(m^2), в которой ИС(m) представляет собой полный коммутатор $m \times m$. За счет использования разветвителей/объединителей каналов РОК m^* или РОК m СС РС(m^2) можно представить как полный распределенный коммутатор $N \times N$, где $N = m^2$. В данном случае такой распределенный коммутатор представлен на рис. 10. Такой коммутатор будем называть однокаскадным (по числу каскадов схем РОК m^* или РОК m) и обозначать как РК $_1(N)$.

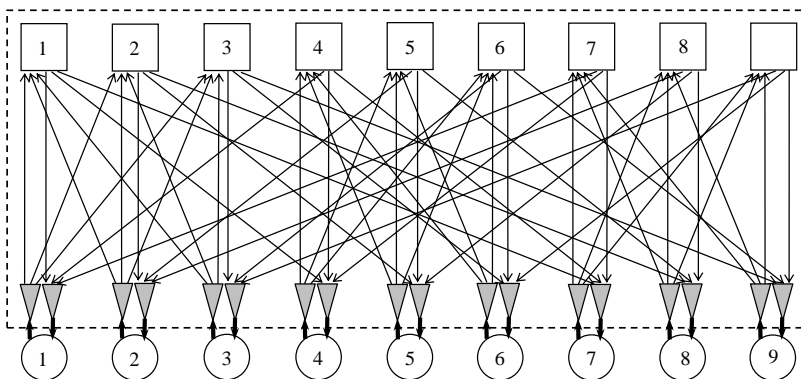


Рис. 10. Еще один распределенный полный коммутатор при $N = 9$

Здесь возникает вопрос: можно ли его расширять дальше за счет увеличения числа каскадов в постоянном схемном базисе коммутаторов $m \times m$ и схем РОК m^* или РОК m . Ответ положительный. Делается это следующим образом. Пусть в качестве исходного коммутатора выступает k -каскадный распределенный полный коммутатор $R_k \times R_k$ на R_k абонентов, где $R_k = m^{k+1}$ и $R_1 = N$. Берем N таких коммутаторов. Разобьем каждый из них на m^k равных частей – каждая из m портов. К первой части всех коммутаторов $R_k \times R_k$ подсоединяется абоненты с номерами от 1 до R_k как при построении коммутатора $R_1 \times R_1$. К i -ой части ($1 < i \leq m^k$) всех коммутаторов $R_k \times R_k$, подсоединим абонентов с номерами от $1 + iR_1$ до $(i + 1)R_1$. При этом абонент с номером $j + iR_1$ ($1 \leq j \leq R_1$) подсоединяется к тем и только тем коммутаторам, что и абонент с номером $j + R_k$. Так создается распределенный коммутатор $R_{k+1} \times R_{k+1}$ на $R_{k+1} = (R_k/m)R_1 = m^{k+2}$ абонентов.

В табл. 4 приводится пример схемы межсоединений для двухкаскадного распределенного коммутатора 27×27 при $m = 3$. Он расширен из полного коммутатора на рис. 10.

Теорема 2. *Расширенная сеть $PC_k(m^{k+1})$ в виде построенного выше k -каскадного распределенного коммутатора $PK_k(m^{k+1})$ является неблокируемой и самомаршрутизируемой СС. В ней любые два порта любых абонента связаны только одним путем длины $2k$, проходящим последовательно только*

через один коммутатор $t \times t$ и разные РОКт. Таким образом $PC_k(m^{k+1})$ является идеальной СС.

Теорема 2 доказывается по индукции. Основанием индукции является теорема 1. При переходе от k -каскадного коммутатора к $(k + 1)$ -каскадному коммутатору неблокируемость и самомаршрутизируемость обеспечивается в каждой группе по теореме 1 и по построению, а между группами – их свойствами по индуктивному предположению.

Таблица 4. Схема межсоединений в двухкаскадном коммутаторе при $t = 3$.

3×3	Входы коммутаторов 3×3										Выходы коммутаторов 3×3								
1	1	2	3	10	11	12	19	20	21	1	4	7	10	13	16	19	22	25	
2	1	2	3	10	11	12	19	20	21	2	5	8	11	14	17	20	23	26	
3	1	2	3	10	11	12	19	20	21	3	6	9	12	15	18	21	24	27	
4	4	5	6	13	14	15	22	23	21	1	4	7	10	13	16	19	22	25	
5	4	5	6	13	14	15	22	24	21	2	5	8	11	14	17	20	23	26	
6	4	5	6	13	14	15	22	24	21	3	6	9	12	15	18	21	24	27	
7	7	8	9	16	17	18	25	26	27	1	4	7	10	13	16	19	22	25	
8	7	8	9	16	17	18	25	26	27	2	5	8	11	14	17	20	23	26	
9	7	8	9	16	17	18	25	26	27	3	6	9	12	15	18	21	24	27	

В k -каскадном полном коммутаторе червячная маршрутизация осуществляется по адресу, состоящему из цуга $(k + 1)$ -го локальных адресов. Первый адрес используется для выбора порта в первом каскаде, i -й ($1 < i \leq k$) – для выбора выходного порта в i -м каскаде, а последний – для выбора выходного порта в хребте. На обратном пути от хребта к абоненту адреса не используются. Однако здесь в общем случае приходится осуществлять множественный доступ к одному выходному порту [9].

Рассмотрим еще один пример построения двухкаскадного распределенного полного коммутатора при $t = 2$, который приводится на рис. 11-12. Из них видно, что в распределенном k -

каскадном полном коммутаторе имеется один хребтовый каскад, который состоит только из коммутаторов $m \times m$, и k каскадов, каждый из которых состоит только из схем РОК m^* или РОК m .

В заключение данного раздела еще раз подчеркнем, что отличительным свойством идеальной СС РС $_k(m^{k+1})$ в виде распределенного полного коммутатора РК $_k(m^{k+1})$ является то, что она является неблокируемой и самомаршрутизируемой на любой однородной m -перестановке, при которой во всех портах каждого абонента имеются пакеты, адресованные разным абонентам или поступившие от разных абонентов.

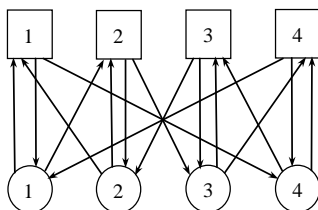


Рис. 11. Минимальный квазиполный орграф для $m = 2$

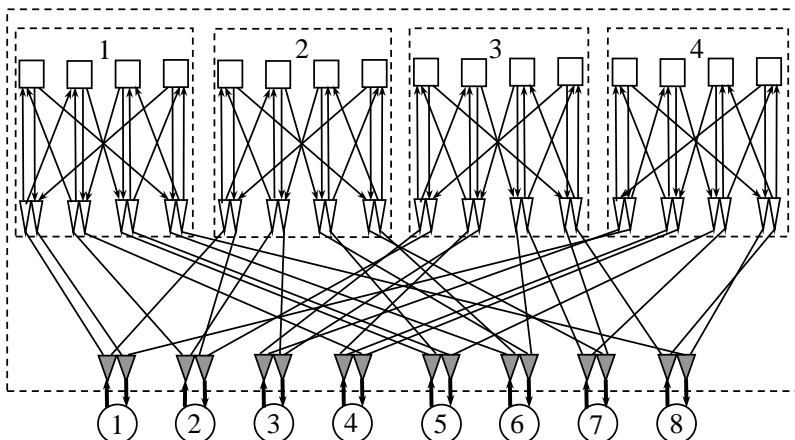


Рис. 12. Двухкаскадный распределенный полный коммутатор 8×8

2.3. РАСПРЕДЕЛЕННЫЙ ПОЛНЫЙ КОММУТАТОР НА БАЗЕ СИММЕТРИЧНЫХ БЛОК-СХЕМ

Рассмотрим еще одну модель расширения идеальной СС ИС(m) (рис. 2). Ее задает однородный двудольный **граф**, одну

долю которого составляют коммутаторы $m \times m$, а другую – m -портовые абоненты. Значение m выбирается минимальным, при котором любые два узла в одной доле связаны σ путями длины два через разные узлы в другой доле. В одной доле имеется N коммутаторов, а в другой – N абонентов. Каждый такой путь проходит через один коммутатор, и разные пути проходят через разные коммутаторы. Двудольный однородный граф с описанными свойствами мы называем минимальным квазиполным графом [3]. Пример такого графа приведен на рис. 13 для $m = 4$, $N = 7$ и $\sigma = 2$. На рис. 13 толстыми линиями выделены пути между абонентами, выделенными одинаковой заливкой. Нетрудно видеть, что их два для каждой пары абонентов.

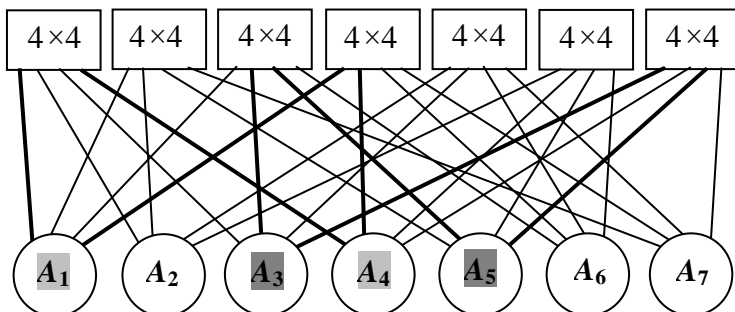


Рис. 13. Минимальный квазиполный граф с $m = 4$, $N = 7$ и $\sigma = 2$

Здесь возникает вопрос о существовании минимальных квазиполных графов и об их параметрах. Оказывается, что он уже давно решен в комбинаторике [3, 13]. Такие графы описываются на языке неполных уравновешенных блок-схем, в частности, симметричных блок-схем [3, 4, 7-10].

Симметричная блок-схема $B(N, m, \sigma)$ состоит из элементов, составляющих одну долю графа, и блоков, составляющих другую долю графа. Число элементов и блоков одинаково и равно N . Параметр m задает число блоков, в которые входит каждый элемент, и число элементов, входящих в каждый блок. Вхождение некоторого элемента в некоторый блок задает ребро на двудольном графе между соответствующими вершинами разных долей. Параметр $\sigma < m$ задает число блоков, в которые входит

каждая пара элементов. Указанные параметры связаны соотношением $N = m(m - 1) / \sigma + 1$.

Следует отметить, что попарно сбалансированное расположение элементов и блоков, задаваемое симметричными блок-схемами, уже нередко используется в вычислительной технике для описания различного рода взаимодействий процессоров и блоков памяти [17, 22, 26].

Любая блок-схема описывается таблицей, в которой строчки задают блоки, а ячейки – вхождения элементов. Блоки и элементы задаются своими номерами. Теперь проинтерпретируем блок как коммутатор $m \times m$ с дуплексными портами, элемент – как абонент с m дуплексными портами, а вхождение элемента в блок – как подсоединение абонента к коммутатору дуплексным каналом через один из своих портов. Тогда σ интерпретируется как число коммутаторов, через которые любые два абонента соединены разными каналами. Вся блок-схема интерпретируется как минимальный квазиполный граф, одна доля которого состоит из абонентов, а другая – из коммутаторов. Он описывает идеальную системную сеть с σ -кратным резервированием каналов – $PC(N, m, \sigma)$. Задающая блок-схему таблица описывает схему межсоединений абонентов и коммутаторов. В таблице 5 приводится пример $B(7, 4, 2)$ и $PC(7, 4, 2)$.

Таблица 5. Схема межсоединений в $PC(7, 4, 2)$.

Блоки 4×4	$B(7, 4, 2)$ $PC(7, 4, 2)$			
	0	0	1	2
1	0	1	4	6
2	0	2	4	5
3	0	3	5	6
4	1	2	5	6
5	1	3	4	5
6	2	3	4	6

Таблица 6. Параметры N и m при $\sigma = 1$.

$B(N, m, 1)$ и ПРК($N, m, 1$)											
m	2	3	4	5	6	7	8	9	10	11	12
N	3	7	13	21	31	43	57	73	91	111	133

Таблица 7. Параметры N и m при $\sigma = 2$ и $\sigma = 3$.

$B(N, m, 2)$ и ПРК($N, m, 2$)										
m	2	3	4	5	6	7	8	9	10	11
N	2	4	7	11	16	22	29	37	46	56
$B(N, m, 3)$ и ПРК($N, m, 3$)										
m	3	4	5	6	7	8	9	10	11	12
N	3	5	–	11	15	–	25	31	–	45

Для блок-схем существует проблема их построения [3, 13]. В таблицах 6 и 7 приводятся параметры блок-схем $B(N, m, \sigma)$ при малых m и σ . Светлой заливкой выделены блок-схемы, которые не существуют по теории [13], а темной заливкой – блок-схемы, которые еще не построены.

Введение в $PC(N, m, \sigma)$ разветвителей/объединителей каналов РОК m^* или РОК m превращает ее в однокаскадный распределенный полный коммутатор $PK_1(N, m, \sigma)$. На рис. 14 приводится схема $PK_1(7, 4, 2)$, состоящая из коммутаторов 4×4 и разветвителей/объединителей дуплексных каналов (РОК4*).

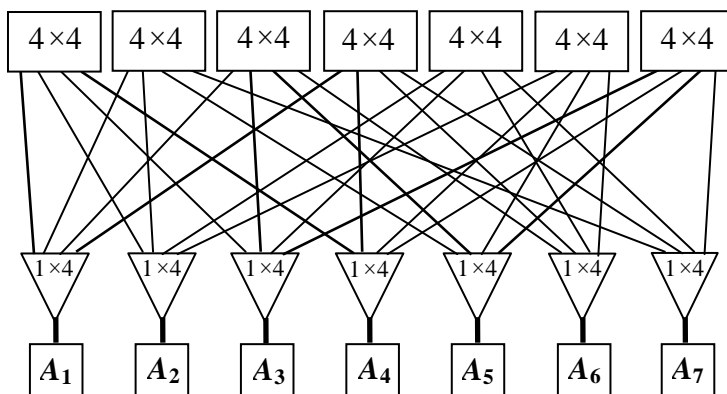


Рис. 14. Схема $PK_1(7, 4, 2)$ из коммутаторов 4×4 и РОК4*.

$PK_1(N, m, \sigma)$ может быть расширен в k -каскадный распределенный коммутатор $PK_k(R_k, m, \sigma)$ с $R_k > R_1 = N$, в котором любые два абонента связаны путями длины $2k$ не меньше чем σ разными путями через разные коммутаторы $m \times m$. Делается это тем

же методом, что и в предыдущем разделе [4, 7]. Единственное отличие состоит в том, что последняя группа содержит меньше m столбцов. Поэтому в ней размещаются только те абоненты, которые помещаются в имеющихся столбцах и «вручную» добавляется еще один абонент согласно определению блок-схемы.

В таблице 8 приводится пример построения двухкаскадного распределенного полного коммутатора при $m = 4$ и $\sigma = 2$. Здесь «вручную» удалось подключить одиннадцатого абонента. Видно, что некоторые абоненты (например 2 и 8) связаны $\sigma = 2$ путями через однокаскадные коммутаторы $R_1 \times R_1$ (7×7), а некоторые абоненты (3 и 10) – $m = 4$ путями. Можно доказать [10, 7], что m путями связаны те и только те абоненты, номера которых равны по $\text{mod } N$. Это заставляет обозначать k -каскадный распределенный коммутатор $R_k \times R_k$ как $\text{PK}_k(R_k, m, \sigma/m)$. Поскольку $\sigma < m$, то число линий ввода/вывода, по которым соединены любые два абонента, не стало меньше σ .

Таблица 8. Табличное описание двухкаскадного $\text{PK}_2(11, 4, 2 \setminus 4)$.

$\text{PK}_2(11, 4, 2 \setminus 4)$							
7×7	1-ый $\text{PK}_1(7, 4, 2)$				2-ый $\text{PK}_1(7, 4, 2)$		
1	1	2	3	4	8	9	10
2	1	2	5	7	8	9	11
3	1	3	5	6	8	10	11
4	1	4	6	7	8		
5	2	3	6	7	9	10	
6	2	4	5	6	10		11
7	3	4	5	7	10		11

В таблице 9 приводится пример построения трехкаскадного распределенного полного коммутатора при $m = 4$ и $\sigma = 2$. Здесь «вручную» удалось подключить восемнадцатого абонента. Любые два абонента связаны не менее чем двумя путями длины 6.

Описанная картина сохраняется при добавлении каждого следующего каскада. Поэтому в k -каскадном коммутаторе любые два абонента связаны только путями длины $2k$, проходящими через разные коммутаторы. Число таких путей, проходящих через разные коммутаторы $m \times m$, задается набором $\sigma^r m^{t-r}$

$(1 \leq r \leq k)$, в котором нет членов меньше σ . Можно показать [7, 4], что число абонентов R_k задается выражением $R_k \approx m \lfloor (N/m)^k \rfloor \approx m \lfloor (m-1)/\sigma^k \rfloor$, которое имеет точность в несколько процентов при $k > 2$.

Таблица 9. Табличное описание трехкаскадного РК₃(18, 4, 2\4).

РК ₃ (18, 4, 2\4)											
11×11	1-ый ПРК(7, 4, 2)				2-ый ПРК(7, 4, 2)				3-ый ПРК(7, 4, 2)		
1	1	2	3	4	8	9	10	11	15	16	17
2	1	2	5	7	8	9	12	14	15	16	18
3	1	3	5	6	8	10	12	13	15	17	18
4	1	4	6	7	8	11	13	14	15		
5	2	3	6	7	9	10	13	14	16	17	
6	2	4	5	6	9	11	12	13	16		18
7	3	4	5	7	10	11	12	14	17		18

Полные распределенные коммутаторы на основе минимального квазиполного **графа** и на основе минимального квазиполного **орграфа** отличаются числом абонентов, используемыми каналами и числом разных путей между любыми двумя абонентами.

Число абонентов в первых в $\sim [m/(m-1)]^k$ раз меньше, чем во вторых. Они используют дуплексные порты вместо симплексных. И, самое главное, они могут иметь σ разных канальных путей между любыми двумя абонентами. Правда, при этом число абонентов уменьшается в $\sim \sigma$ раз.

Последнее свойство можно использовать как для обеспечения отказоустойчивости идеальной СС, так и для повышения ее коммутационных возможностей. Такая идеальная СС является неблокируемой и самомаршрутизируемой при произвольной (неоднородной) σ -перестановке пакетов данных между абонентами, т.е. позволяет любому абоненту вести обмен пакетами данных одновременно по σ каналам с любыми другими абонентами.

3. Заключение

В работе рассмотрен метод построения функционально идеальной СС на любое число абонентов в виде распределенного многокаскадного полного коммутатора. Он обладает свойством неблокируемости и самомаршрутизируемости на произвольной перестановке пакетов данных. Он имеет квадратичную схемную сложность, но содержит значительно меньшее число проводников, чем цельный полный коммутатор.

Построенные идеальные СС ориентированы на две основных области применения – в многоядерных СБИС и в отказоустойчивых МВС реального времени. В первом случае квадратичная схемная сложность не играет особой роли, т.к. сложность коммутатора много меньше сложности ядра, но существенно важно значительное сокращение числа проводников. Во втором случае особую роль выполняет полная однородность СС, которая позволяет иметь любое число резервных процессоров, и возможность резервирования каналов.

Для создания подлинно идеальной СС необходимо сокращение схемной сложности до сложности сетей Клоза или многомерных обобщенных гиперкубов при сохранении свойства неблокируемости и самомаршрутизируемости. Классические сети Клоза не обеспечивают выполнения одновременно обоих этих свойств. Надежду на успех поддерживает наличие подобной сети в виде d -мультиплицированной сети Бенеша (двоичной сети Клоза) [18]. Она, однако, малопригодна для практических применений из-за малого размера коммутаторов ($2d \times 2d$) и, как следствие, сравнительно большой глубины сети ($\sim 2\log_2 N$ каскадов), тогда как современные тенденции [25] требуют построения системных сетей с малой глубиной за счет использования связанных СБИС максимально больших коммутаторов.

Отражением этой же тенденции является также сетевая структура самого нового суперкомпьютера Blue Water (IBM) [15], в которой связь между процессорными узлами выполняется не более чем за 3 скачка между связными узлами с промежуточной буферизацией пакетов. Эта сетевая структура строится на основе двухуровневой иерархии полных графов (!) для системы с десятками тысяч процессоров. Замена полных графов на

квазиполные (ор)графы, как в идеальной СС, может дать значительный эффект в части увеличения числа процессоров и/или снижения числа используемых каналов. Это – новая и неожиданная для авторов область применимости идеальных СС.

Литература

1. ГОРБУНОВ В.С. *Архитектура хорошо масштабируемого вычислительного кластера* // Труды международной научно-технической конференции «Суперкомпьютерные технологии: разработка, программирование, применение» (СКТ-2010) Дивноморское. Сентябрь 2010. Т.1. С. 48 – 54.
2. ГОРБУНОВ В.С., ЛАЦИС А.О., ИВАНОВ А.Н. *О построении суперкомпьютеров на основе интерфейса PCI-Express*. // Труды международной научно-технической конференции «Суперкомпьютерные технологии: разработка, программирование, применение» (СКТ-2010) Дивноморское. Сентябрь 2010. Т.1. С. 55 – 57.
3. КАРАВАЙ М.Ф., ПАРХОМЕНКО П.П., ПОДЛАЗОВ В.С. *Комбинаторные методы построения двудольных однородных минимальных квазиполных графов (симметричных блок-схем)* // Автоматика и телемеханика. 2009. №. 2. С. 153 – 170.
4. КАРАВАЙ М.Ф., ПОДЛАЗОВ В.С. *Метод инвариантного расширения системных сетей многопроцессорных вычислительных систем* // Автоматика и телемеханика. 2010. №. 12. С. 166 – 176.
5. КАРАВАЙ М.Ф., ПАРХОМЕНКО П.П., ПОДЛАЗОВ В.С. *Универсальная сетевая структура для отказоустойчивых многопроцессорных систем реального времени* // Труды конференции «Технические и программные средства систем управления, контроля и измерения» (УКИ'10). М. 2010. С. 583 – 597. URL: <http://cmm.ipu.ru/proc/index.html> (дата доступа – 26.09.2011)..
6. КОРЖ А.А., МАКАГОН Д.В., БОРОДИН А.А. и др. *Отечественная коммуникационная сеть 3D-тор с поддержкой глобально адресуемой памяти для суперкомпьютеров транснафтафлопсного уровня производительности* // Междуна-

- родная научная конференция «Параллельные вычислительные технологии 2010» г.Уфа, март – апр. 2010. С. 227 – 237.
7. НИКОЛАЕВ А.Б., ПОДЛАЗОВ В.С. *Отказоустойчивое расширение системных сетей многопроцессорных вычислительных систем* // Автоматика и телемеханика.. 2008. № 1. С. 162 – 170.
 8. ПОДЛАЗОВ В.С., СОКОЛОВ В.В. *Однокаскадные коммутаторы большой размерности для многопроцессорных и многомашинных вычислительных систем* // Проблемы управления. 2006. № 6. С. 19 – 24.
 9. ПОДЛАЗОВ В.С., СОКОЛОВ В.В. *Схемотехника однокаскадных коммутаторов большой размерности* // Датчики и системы. 2006. № 9.С. 12 – 17.
 10. ПОДЛАЗОВ В.С., СОКОЛОВ В.В. *Метод однородного расширения системных сетей многопроцессорных вычислительных систем* // Проблемы управления. 2007. № 2. С. 22 – 27.
 11. СОЛОХИНА Т.В., ПЕТРИЧКОВИЧ Я.Я., ШЕЙНИН Ю.Е. *Технология Space Wire и бортовых распределенных комплексов* // ЭЛЕКТРОНИКА: Наука, Технология, Бизнес. 2007. №. 1. С. 38 – 49.
 12. ШЕЙНИН Ю.Е., СОЛОХИНА Т.В., ПЕТРИЧКОВИЧ Я.Я. *Технология Space Wire и бортовых распределенных комплексов* // ЭЛЕКТРОНИКА: Наука, Технология, Бизнес. 2006. №. 6. С. 64 – 75.
 13. ХОЛЛИ М. *Комбинаторика*. // М.: Мир. 1970. 421 С.
 14. ALVERSON R., ROWETH D. AND KAPLAN L., CRAY INC. // *The Gemini System Interconnect* // 18th IEEE Symposium on High Performance Interconnects. 2009. P. 83 – 87.
 15. ARIMILLI B. ARIMILLI R., CHUNG V., et al, *The PERCS High-Performance Interconnect* // 18th IEEE Symposium on High Performance Interconnects. 2009. P. 75 – 82.
 16. ARORA S., LEIGHTON F.T., MAGGS B.M. *On-line algorithm for path selection in nonblocking network* // SIAM Journal of Computing 1996. 25(3). P. 600 – 652.
 17. BERCOVICH E., BERCOVICH S. *A combinatorial architecture for instruction-level parallelism* // Microprocessors and Microsystems. 1998. V. 32. P. 23 – 31.

18. GU Q.P., TAMAKI H. *Routing a permutation in hypercube by two sets of edge-disjoint paths* // Journal of parallel and distributed computing. 1997. V. 44. No. 2. P. 147 – 152.
19. NI L.M., MCKINLEY P.K. *A survey of wormhole routing techniques in direct networks* // IEEE Computer 1993. V.26. No. 2. P. 62 – 73.
20. *Guide to myrinet-2000 switches and switch networks* // URL: <http://www.myti.com/myrinet/m3switch/guide/> (дата доступа – 26.09.2011)
21. KUMAR A., PEH L-S., KUNDU P., JHA N.K. *Toward ideal on-chip communication using express virtual channels* // IEEE Micro. 2008. Jan/Feb. P. 80 – 90.
22. OKBIN L., SANGHO L., SEONGYEOL K., ILYONG CH. *An Efficient Load Balancing Algorithm Employing a Symmetric Balanced Incomplete Block Design* // Lecture Notes on Computer Science 3046. 2004. P. 647 – 654.
23. PIPENGER N. *On rearrangeable and non-blocking switching networks* // Journal Of Computer and Systems Science. 1978. V. 17. P. 307 – 311.
24. RZYMIAHOWICZ L. *Designing efficient network interfaces for system area networks* // URL: http://bibserv7.bib.uni-mannheim.de/madoc/volltexte/2002/54/pdf/54_1.pdf (дата доступа – 26.09.2011).
25. SCOTT S., ABTS D., KIM J., AND DALLY W. *The black widow high-radix Clos network* // Proc. 33rd International Symposium on Computer Architecture. (ISCA'2006). 2006. Рукопись доступна на сайте URL: <http://cva.stanford.edu/people/jjk12/isca06.pdf> (дата доступа – 26.09.2011).
26. YOUNGJOO CH., CHANGKYUN CH., ILYONG CH. *An Efficient Conference Key Distribution System Based on Symmetric Balanced Incomplete Block Design* // Lecture Notes on Computer Science 2657. 2003. P. 147 – 154.

DISTRIBUTED FULL SWITCH AS IDEAL SYSTEM AREA NETWORK FOR MULTIPROCESSOR COMPUTERS

Mikhail Karavay, Institute of Control Sciences of RAS, Moscow, Doctor of Science, assistant professor (mkaravay@ipu.ru, Moscow, Profsoyuznaya st., 65, (495)334-90-00).

Viktor Podlazov, Institute of Control Sciences of RAS, Moscow, Doctor of Science, assistant professor (podlazov@ipu.ru, Moscow, Profsoyuznaya st., 65, (495)334-78-31).

Abstract: We consider a way to build distributed full switches of arbitrary size consisting of fixed-size switches and channel splitters. The distributed switch preserves nonblocking and self-routing properties of a complete switch and forms an ideal system area network.

Keywords: massive parallel multiprocessor computer, ideal system area networks, distributed full switch, channel switching, wormhole routing techniques, non-blocking networks, self-routing networks.

Статья представлена к публикации членом редакционной коллегии В.Н. Лебедевым

*3-я Российская конференция
с международным участием
«Технические и программные средства
систем управления, контроля и измерения»
(УКИ-12)*

ИПУ РАН, Москва, 16-19 апреля 2012г.

<http://cmm.ipu.ru>

УДК 004.822
ББК 32.81

ПРИМЕНЕНИЕ ОНТОЛОГИЙ ДЛЯ ПОДДЕРЖКИ ВЕРСИОННОСТИ СЕРВЕРНЫХ ОПЕРАЦИЙ

Яшкин А. В.¹

(Владимирский государственный университет, Владимир)

В статье описана задача поддержания версииности операций в больших информационных системах при изменении программного обеспечения. Рассмотрен способ решения этой задачи с помощью создания онтологий семантического веб, его достоинства и недостатки.

Ключевые слова: информационная система, онтология, клиент-серверная архитектура, семантический веб.

Введение

Во многих больших информационных системах крупных территориально-распределенных организаций приходится сталкиваться с частым обновлением версий программного обеспечения. Если же при этом разные отделения функционируют не автономно, а активно взаимодействуя друг с другом, то приходится учитывать различия в установленных на них версиях. В сложных системах это вызывает проблемы не только при обновлениях, но и при последующем сопровождении системы.

1. Постановка задачи поддержки версииности ПО

В клиент-серверной архитектуре взаимодействие клиентского процесса с серверным относительно простое: клиентская

¹ Александр Владимирович Яшкин, магистр техники и технологии по направлению «Информатика и вычислительная техника», аспирант кафедры вычислительной техники Владимирского государственного университета (моб. тел.: +79190275047, yashkinaalexandr@mail.ru).

программа занимается созданием пользовательского интерфейса и формирует запросы к серверу, а сервер обрабатывает эти запросы, осуществляет выборку и/или модификацию данных на основе механизма транзакций и возвращает результат обратно клиенту. Это может быть как двух-, так и трехуровневая модель «клиент–сервер» (рис. 1). В двухуровневом клиент-серверном приложении, как правило, все функции по формированию пользовательского интерфейса реализуются на клиенте, все функции по управлению данными – на сервере, а вот бизнес-правила можно реализовать как на сервере, используя механизмы программирования сервера (хранимые процедуры, триггеры, представления и т.п.), так и на клиенте. В трехуровневом приложении появляется третий, промежуточный уровень, реализующий бизнес-правила, которые являются наиболее часто изменяемыми компонентами приложения [1].

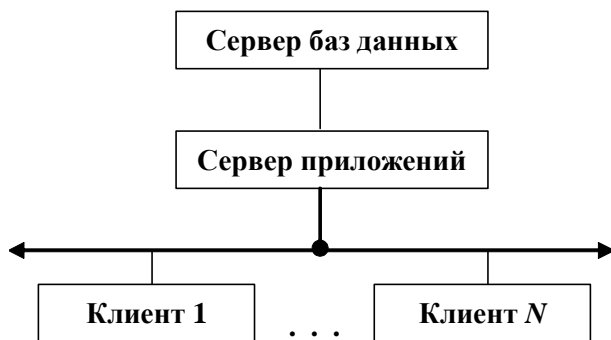


Рис. 1. Трехуровневая модель «клиент-сервер»

Однако такая простая схема годится не для всех приложений. В сложных информационных системах встречаются ситуации, когда сам сервер имеет сложную сетевую структуру, т.е. состоит из нескольких серверов, взаимодействующих друг с другом. Конечный пользователь при этом может и не подозревать об этой архитектуре, так как она скрыта от него так называемой единой «точкой входа» в систему («*single sign-on*»), и работать с ней, как с обычной клиент-серверной технологией

[8]. Примерами подобной системы могут служить грид-приложения на SOA (сервис-ориентированная архитектура) [6] или же клиент-серверная система какой-нибудь крупной территориально распределенной компании (например, банка), в которой в процессе развития понадобилась связь между серверами разных отделений фирмы (рис. 2).

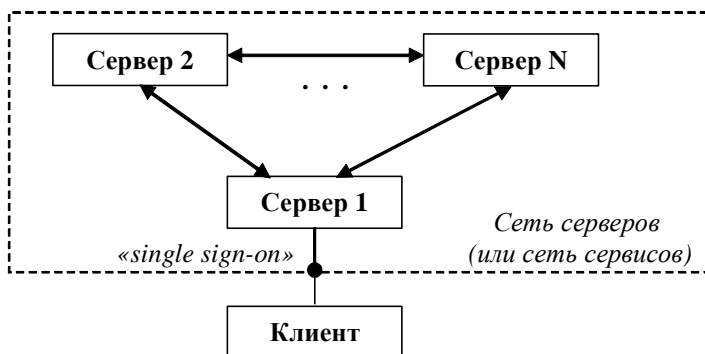


Рис. 2. Полносвязная система n-серверов с единой «точкой входа» в систему

В подобных системах возникают некоторые специфические задачи, например, задача поддержания версии программного обеспечения (ПО). Данная задача возникает как следствие обновления программного обеспечения серверов и/или клиента. Как правило, внедрение обновлений в крупной организации выполняют сначала на одном/двух отделениях, и лишь после нескольких недель успешной работы распространяют *upgrade* на остальные отделения компании. Поэтому в этом случае необходимо сохранять те версии операций, которые понадобятся обновленному серверу для взаимодействия с серверами, на которых установлено старое программное обеспечение. Кроме того, нужно отслеживать корректность прохождения запроса клиента через всю цепочку серверов. (В обычной клиент-серверной архитектуре (рис. 1) подобная задача не возникала, так как там имеется лишь одна версия сервера и одна версия клиента, которые обновляются одновременно).

2. Способы решения задачи поддержки версионности ПО

Самый простой способ решения данной проблемы заключается в разбивке всей цепочки прохождения запроса по серверам на отдельные простые части (рис. 3). Такое разделение возможно, так как вызов серверной операции с другого сервера ничем принципиально не отличается от вызова с клиента. Рассматривая таким образом каждый вызов, как будто имеем дело с обычным клиент-серверным приложением и можем решить задачу версионности, вызывая минимальную из максимальных версий операции 1 серверов 1 и 2 (рис. 3).

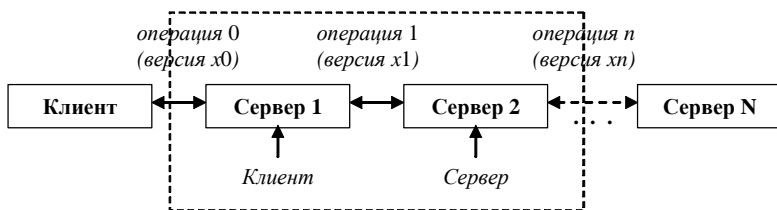


Рис. 3. Цепочка прохождения запроса по серверам

Достоинством этого решения является его простота. Для реализации понадобится лишь дополнительная серверная операция, возвращающая максимальную версию остальных «рабочих» серверных операций.

Недостатками решения являются:

- Ограничения при вызове новых версий операций, накладываемые предыдущими «шагами» цепочки операций. Это происходит из-за того, что не анализируется вся цепочка операций и логика их взаимодействия.
- Сложность сопровождения и мониторинга информационной системы из-за отсутствия семантической информации о связях между операциями.

Длина цепочки на практике, конечно, редко составляет более 3–5 операций. Сложность мониторинга и сопровождения за-

ключается в наличии взаимных блокировок между операциями, с учетом того, что подобных цепочек может быть несколько. Тогда системным администраторам трудно определить тот самый сервер и ту самую операцию, которая привела к блокировке, и приходится снимать все сервера, которые долгое время остаются заняты, в то время как на некоторых из них просто идут длительные операции, а разрешение проблемы блокировок, как правило, состоит в прерывании операций всего лишь на одном – максимум двух серверах. Таким образом, наличие семантической информации о цепочках операций позволит сократить количество прерванных операций пользователей при мониторинге информационной системы.

- Невозможность автоматического отслеживания и удаления устаревших версий операций.

Другой способ решения задачи версионности, заключается в сохранении семантической информации обо всех версиях операций всех серверов и связях между ними и использование этой информации при определении вызываемых версий операций. Для практической реализации этого решения хорошо подходит технология семантического веба.

3. Разработка модуля анализа версионности

Создание подобной подсистемы можно разбить на два этапа:

- разработка онтологии;
- разработка модуля определения версии.

3.1. СОЗДАНИЕ ОНТОЛОГИИ

В науке об «искусственном интеллекте» онтология – это «спецификация концептуализации предметной области», или упрощенно, документ или файл, формально задающий отношения между терминами. Это своего рода словарь понятий предметной области и совокупность явным образом выраженных предположений относительно смысла этих понятий [4].

Создание онтологии разобьем на несколько этапов.

Во-первых, нужно получить полный список терминов, не беспокоясь о пересечении понятий, которые они представляют,

об отношениях между терминами, о возможных свойствах понятий [7].

Во-вторых, из составленного списка выбираем термины, которые описывают объекты, существующие независимо, а не термины, которые описывают эти объекты. В онтологии эти термины будут классами и станут точками привязки в иерархии классов [7]. Для рассматриваемой задачи иерархия классов очень проста и состоит всего из четырех классов: «сервер», «серверная операция», «версия операции», «территориальное отделение».

В-третьих, опишем свойства классов, их тип и ограничения (фацеты) [5]. На орграфе (рис. 4) свойства представлены ребрами графа.

Последний шаг – это создание отдельных экземпляров классов в иерархии [7]. В отличие от предыдущих этапов создание экземпляров для онтологии версий можно автоматизировать при помощи написания соответствующего парсера – программы, которая бы по «исходникам» строила бы часть онтологии.

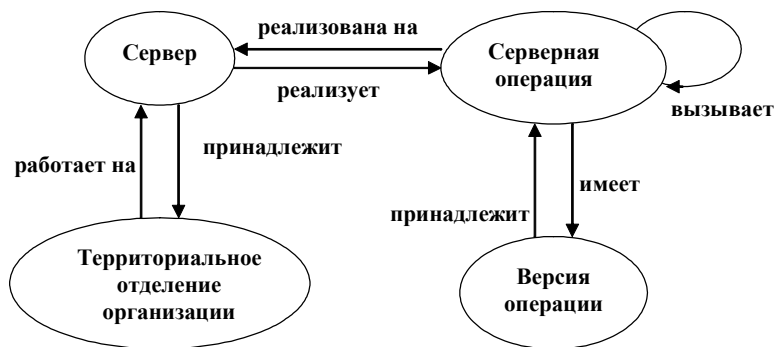


Рис. 4. Онтология: классы и свойства

3.2. РАЗРАБОТКА МОДУЛЯ ОПРЕДЕЛЕНИЯ ВЕРСИИ

Для определения версий операций с использованием онтологии применим простой рекурсивный алгоритм (рис. 5). Особенность реализации алгоритма при использовании описания версии с помощью семантического веба заключается в

том, что вся рекурсивная часть будет реализована в запросе к онтологии, а не в коде программы.



Рис. 5. Алгоритм определения версий операций

Таким образом, создание программной части разобьется на два этапа:

- разработка менеджера онтологии;
- разработка модуля, реализующего бизнес-логику (в данном случае модуль получения версии операции).

Функции менеджера онтологий можно разделить на три группы:

- функции, связанные с *TBox*-запросами к онтологии (т.е. получение иерархии классов, свойств класса, ограничений свойств);
- функции выполнения *SPARQL*-запросов к онтологии;
- функции добавления и изменения информации в онтологии (необходимы только при изменении версий на серверах).

Менеджер онтологии является, таким образом, классом доступа к данным в нашем приложении, поэтому его разработка и отладка и предшествует разработке бизнес-логики приложения.

Рассмотренные модули по работе с онтологиями интегрируются с существующей системой посредством модуля, в который вынесена часть функциональности самой системы, по получению версий операций, который для каждой конкретной информационной системы является уникальным.

Работу модуля получения версии операции тогда можно свести к этапам:

- в первую очередь нужно, используя менеджер онтологий, загрузить онтологию предметной области и определить структуру этой предметной области;
- на втором шаге вводим через пользовательский интерфейс или каким-либо иным образом получаем данные об интересующей нас операции;
- формируем на основе полученных данных запрос к онтологии и выполняем его с помощью того же менеджера онтологий;
- анализируем полученный результат. В результате чего можем сделать вывод о доступности на всех необходимых для выполнения операции серверах последних обновлений, соответствующих клиентской версии или об их недостаточности.

3.3. ПРИМЕР

На рис. 6 представлен пример выполнения операции по нарезке файлов с данными о зачислениях клиентов по отделениям банка в зависимости от установленных настроек на центральном сервере.

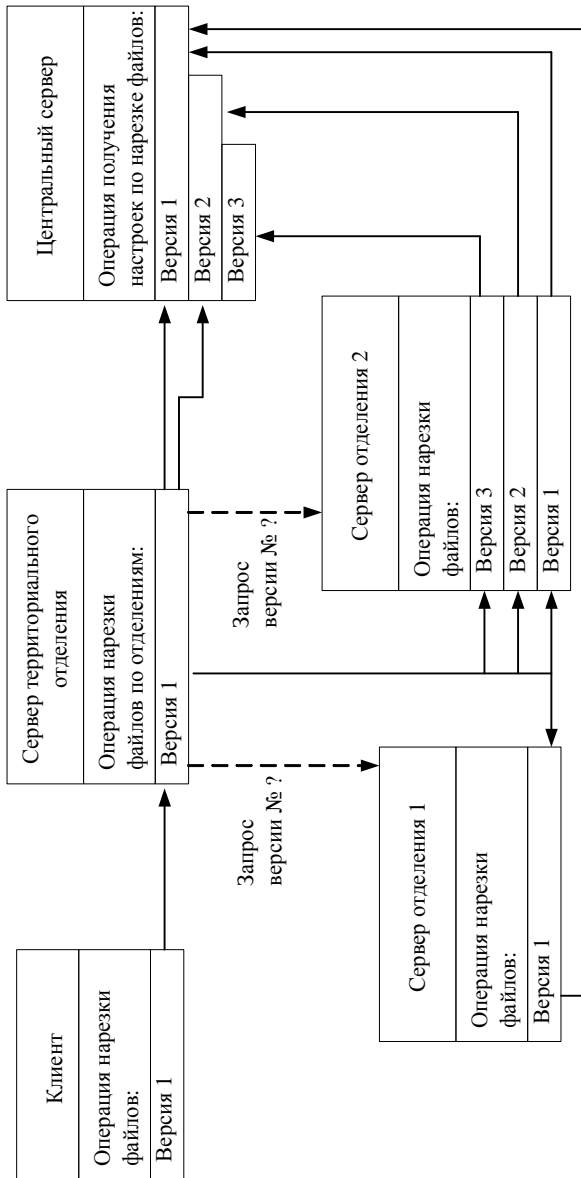


Рис. 6. Пример выполнения операции

Порядок вызова операций следующий:

1) клиент запрашивает выполнение операции на территориальном отделении (ТБ); 2) ТБ обращается к центральному серверу за настройками для выполнения нарезки; 3) ТБ вызывает операцию нарезки на отделении1 и отделении2.

Нюансом данного примера является то, что каждая из конечных версий, выполняемая на отделениях, требует соответствующих настроек из центральной базы. В результате, несмотря на наличие связи (возможность вызова) между операцией версии1 на ТБ и операцией версии3 на отделении2, в этом отделении должна быть вызвана более ранняя версия2.

Каждая версия всех представленных операций представляется в онтологии соответствующим индивидом (*ABOX*-объекты). Стрелками на рис. 6 обозначены установленные связи между этими индивидами онтологии. Выборки будут осуществляться примерно следующими запросами:

```
PREFIX servers: <http://example.org/servers>
PREFIX operations: <http://example.org/operations>
PREFIX links: <http://example.org/links>
SELECT ?serverN ?operationN ?version
WHERE {
  ?server :servers:title "Сервер ТБ"
  ?operation :operations:number 1134
  ?operationN :links:suboperations ?operation
  ?operationN :servers:onServer ?serverN
  ?version :operations:isVersion ?operationN
}
ORDER BY ?version
```

Естественно, запросы несколько варьируются в зависимости от цели, но основные элементы остаются неизменны: *?server* – сервер, с которого начинается операция (в данном случае сервер ТБ); *?operation* – первая операция в цепочке (в нашем примере операция с номером 1134); *?operationN* – все операции связанные с ней (здесь свойство *:links:suboperations* – транзитивное); *?version* – версия каждой конкретной операции.

Запрос, приведенный выше, представлен в самом общем виде и должен быть скорректирован в зависимости от конкретной задачи управления.

Так для мониторинга в большинстве случаев (например, при возникновении блокировок) нужны лишь цепочки операций на серверах, т.е. последовательность пар: *?serverN*, *?operationN*, а версии нужны лишь когда от них зависит вызов следующей в цепочке операции.

Для разбора нештатных ситуаций необходимо осуществить сверку уже трёх параметров: *?serverN*, *?operationN*, *?version*, полученных в результате *SPARQL*-запроса с теми значениями, которые получаются в результате реального прохождения операции. Различия в этих цепочках укажут на тот сервер(а), где неправильно установлено или/и сконфигурировано или/и функционирует программное обеспечение.

Заключение

Итак, что дает предложенный подход к построению системы и оправданы ли затраты на построение онтологии. Конечно, на первых этапах функционирования системы может показаться, что построение онтологии было избыточным, но при больших и частых изменениях функциональности системы наличие онтологии и программного определения версий цепочки операций будет значительно облегчать переход к новым обновлениям.

Недостатком подхода является сложность реализации, которая для небольших и редко изменяемых систем может быть избыточной.

Достоинства:

- сокращение времени на разбор нештатных ситуаций, возникающих из-за рассогласованности версий на разных серверах;
- облегчение сопровождения системы за счет возможности мониторинга версий запросов;
- возможность автоматического отслеживания и удаления устаревших версий операций.

Таким образом, можно определить область применения данного подхода к поддержанию версииности как большие территориально-распределенные системы с частым обновлением функциональности.

Литература

1. АЛЬПЕРОВИЧ М. *Еще раз об архитектуре «клиент–сервер»* [Электронный ресурс]. – Режим доступа: http://www.ci.ru/inform2_97/astr1.htm (дата обращения: 25.05.10)
2. ГЕЛОВАНИ В.А., БАШЛЫКОВ А.А., БРИТКОВ В.Б., ВЯЗИЛОВ Е.Д. *Интеллектуальные системы поддержки принятия решений в нестандартных ситуациях*. – М.: Эдиториал УРСС, 2001. – 356 с.
3. КУЧЕРЕНКО Е.И., ПАВЛОВ Д.А. *О проблемах выявления неполноты и избыточности в онтологических пространствах объектов исследования*. – 2008. . [Электронный ресурс]. – Режим доступа: <http://shcherbak.net/protivorechivost/>.
4. ЛАНДЭ Д.В. *Семантический веб: от идеи к технологии*. – Телеком, 2005. – URL: <http://www.visti.net/~dwl/art/sw/index1.html> (дата обращения: 25.05.10).
5. BAADER F., CALVANESE D., MCGUINNESS D. L., NARDI D., PATEL-SCHNEIDER P. F. *The Description Logic Handbook: Theory, Implementation, Applications*. – Cambridge University Press, Cambridge, UK, 2003.
6. CHANNABASAVAIAN K., TUGGLE E., HOLLEY K. *Migrating to a service-oriented architecture // IBM*. [Электронный ресурс]. – Режим доступа: <http://www.ibm.com/developerworks/webservices/library/ws-migratesoa/>.
7. NOY N.F., MCGUINNESS D.L. *Ontology Development 101: A Guide to Creating Your First // Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, March 2001*.
8. SOTOMAYOR B. *The Globus Toolkit 4 Programmer's Tutorial // University of Chicago, 2005*. [Электронный ресурс]. – Режим доступа: <http://gdp.globus.org/gt4-tutorial/>.
9. *SPARQL Protocol for RDF: W3C Candidate Recommendation, 6 April 2006 // Kendall Grant Clark, eds.* – URL: <http://www.w3.org/TR/rdf-sparql-protocol/>.

10. *SPARQL Query Language for RDF, W3C Working Draft, 4 October 2006* // Eric Prud'hommeaux, Andy Seaborne, eds. – URL: <http://www.w3.org/TR/rdf-sparql-query/>.
11. *SPARQL Query Results XML Format, W3C Candidate Recommendation, 6 April 2006* // Eric Prud'hommeaux, Andy Seaborne, eds. – URL: <http://www.w3.org/TR/rdf-sparql-XMLres/>.

USING ONTOLOGY TO SUPPORT VERSIONING OF SERVER OPERATIONS

Alexander Yashkin, master of engineering and technology in area “Computer science and computer engineering”, post-graduate student of sub-faculty Computer engineering by Vladimir State University (mob. tel. +79190275047, yashkinaalexandr@mail.ru).

Abstract: In the article the problem is considered of large-scale information systems operations versioning support in the process of software upgrade. The solution is proposed that bases on creating ontology of semantic web. Merits and demerits of the approach are discussed.

Keywords: information system, ontology, client-server architecture, semantic web.

Статья представлена к публикации членом редакционной коллегии М.Ф. Караваем

УДК 338.24.01
ББК 65.9(2).23

МОДЕЛЬ И МЕТОДИКА РЕШЕНИЯ ЗАДАЧИ ОПТИМИЗАЦИИ ГРАФИКА ФИНАНСИРОВАНИЯ ИНВЕСТИЦИОННОГО ПРОЕКТА НА ГРАФАХ РАБОТ

Богатырев В. Д.¹, Морозова С. А.²

(Самарский государственный аэрокосмический университет имени академика С.П. Королева (Национальный исследовательский университет), Самара)

Представлена экономико-математическая модель оптимизации графика финансирования инвестиционного проекта в условиях дефицита денежных средств. Приведена методика, которая позволяет оптимизировать график финансирования инвестиционного проекта с помощью изменений в плановом варианте графика реализации, включая перенос финансирования на более поздние сроки, привлечение заемных средств и учет «штрафных» санкций. Сформулированы рекомендации для принятия управленческих решений в области финансирования инвестиционных проектов.

Ключевые слова: инвестиционный проект, финансирование, сальдо, дефицит, профицит, чистый дисконтированный поток.

Введение

В научной литературе, посвященной управлению проектами, выделяют несколько обширных разделов, в числе основ-

¹ Владимир Дмитриевич Богатырев, доктор экономических наук, профессор, проректор по образовательной и международной деятельности (samelev@rambler.ru).

² Светлана Анатольевна Морозова, аспирант (morozova_s_a@mai.ru).

ных – календарно-сетевое планирование и управление (оптимизация расписания проекта на основе методов теории графов), а также методы управления проектами (разработка и анализ математических моделей проектов) [3, 4, 7-9].

Математические модели применяются для постановки и решения задач по оптимизации проекта по стоимости, времени и ресурсам. Огромный вклад в математические методы управления проектами внесли следующие ученые: Бурков В.Н., Новиков Д.А., Цветков А.В., Матвеев А.А., Шапиро В.Д., Разу М.Л., Мазур И.И., Якутин Ю.В., Воропаев В.И.

В настоящей статье предлагается для разработки математической модели и методики оптимизации графика финансирования инвестиционного проекта использовать методы теории графов и календарно- сетевого планирования и управления. При этом предполагается, что совокупность этапов работ инвестиционного проекта не формирует последовательную схему, оптимизация графика финансирования которой была рассмотрена ранее в работах [1, 2, 5, 6], а включает в себя ветвления с параллельным выполнением этапов и работ.

С точки зрения теории графов проект – совокупность операций и зависимостей между ними [3, 4, 8, 9]. Совокупность моделей и методов, использующих язык и результаты теории графов и ориентированных на решение задач управления проектами, получила название календарно- сетевого планирования и управления (КСПУ) [4,8,9]. В рамках КСПУ решаются задачи определения последовательности выполнения и распределения ресурсов между работами, оптимальных с точки зрения тех или иных критериев (времени выполнения, затрат, риска). Для описания и оптимизации проектов наиболее подходящими оказались сетевые модели, представляющие собой разновидность ориентированных графов [4, 8, 9].

В сетевой модели роль вершины графа могут играть события, определяющие начало и окончание отдельных работ, а дуги в этом случае будут соответствовать работам [4]. Такую сетевую модель принято называть сетевой моделью с работами на дугах (Activities on Arrows, AoA) [8, 9]. В тоже время, возможно, что в сетевой модели роль вершин графа играют работы, а

дуги отображают соответствие между окончанием одной работы и началом другой. Такую сетевую модель принято называть сетевой моделью с работами в узлах (Activities on Nodes, AoN) [8, 9].

Сетевые модели AoN появились позже графов AoA, вследствие чего они менее известны и сравнительно реже описываются в литературе. Тем не менее, они имеют свои преимущества, в частности, их легче строить и корректировать. При корректировке конфигурация сетевой модели AoN не меняется, однако для моделей AoA такие изменения исключить не удается. В моделях AoA при корректировке необходимо менять конфигурацию (топологию).

В настоящее время составление и корректировка сетевых моделей автоматизированы, и для пользователя, которому важно знать лишь последовательность работ и их резервы времени, не имеет особого значения, каким способом она сделана, т.е. какого она типа. В современных специализированных пакетах компьютерных программ планирования и оперативного управления в основном используются сетевые модели AoN [8, 9].

Корректировка сетевых моделей производится как на этапе их составления, так и использования. Она состоит в оптимизации работ по времени и по ресурсам. Корректировка по ресурсам производится путем построения линейных календарных графиков по ранним началам, соответствующих тому или иному варианту сетевого графика, и корректировки этого варианта.

Сетевым графиком AoN называется полное графическое отображение структуры сетевой модели на плоскости (рис. 1).

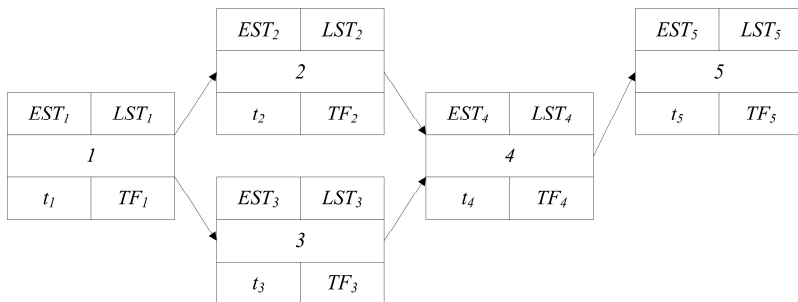


Рис. 1. Сетевой график модели AoN

В данной модели узлы сети, соответствующие работам, принято изображать прямоугольниками, в центре которых записывается наименование работы i . В левый верхний сектор прямоугольника, маркирующего вершину работы, записывается раннее время начала выполнения работы (EST_i). В правый верхний сектор прямоугольника записывается позднее время начала выполнения работы (LST_i). В левый нижний сектор прямоугольника записывается продолжительность работы (t_i). В правый нижний сектор прямоугольника записывается полный резерв выполнения работы (TF_i) [8,9].

Далее рассматривается построение экономико-математической модели оптимизации графика финансирования инвестиционного проекта на основе сетевых моделей типа AoN.

Разработка экономико-математической модели и методики

Любой инвестиционный проект от его зарождения до завершения проходит ряд этапов. Жизненный цикл инвестиционного проекта принято разделять на фазы (стадии). Фазы жизненного цикла могут различаться в зависимости от сферы деятельности и принятой системы организации работ.

Существует несколько вариантов выделения фаз жизненного цикла проекта. Наиболее распространенным является вариант, разделяющий жизненный цикл инвестиционного проекта на наиболее крупные фазы.

При данном варианте можно выделить прединвестиционную, инвестиционную и эксплуатационные фазы. Однако для простоты изложения экономико-математической модели инвестиционный проект будет рассматриваться едино, не подразделяясь на фазы.

В качестве целевой функции модели предлагается использовать чистый дисконтированный поток (NPV), рассчитываемый как разность дисконтированных притоков и оттоков:

$$(1) NPV = \sum_{t=1}^T \frac{(R^t - C^t)}{(1+r)^t} \cdot (1-h) \longrightarrow \max.$$

Функция NPV зависит от нескольких переменных: во-первых, от R^t – денежных средств, полученных в момент времени t ($t = 1, \dots, T$); во-вторых, от C^t – денежных средств, необходимых для оплаты работ в момент времени t ($t = 1, \dots, T$); в-третьих, от η – налога на прибыль; в-четвертых, от ρ – ставки дисконтирования.

Авансы от покупателей и заказчиков, а также авансы поставщикам и подрядчикам, в зависимости от учетной политики, принятой в организации, могут не являться доходами и расходами, соответственно. Для упрощения данной модели предполагается, что полученные денежные средства совпадают с доходами, а расходы учитываются по завершению работ.

В общем виде структуру расходов инвестиционного проекта можно представить в виде таблицы, ячейки которой отражают оплату работы i в момент времени t , для этого можно использовать $X = \|x_i^t\|$ – матрицу финансирования сетевой модели, элемент которой x_i^t характеризует оплату работы i в момент времени t , где $\forall i, \forall t \ 0 \leq x_i^t \leq 1, \sum_{t=1}^T x_i^t = 1$ (таблица 1).

Таблица 1. Структура расходов инвестиционного проекта

Работы i	Время t					
	1	...	t	...	T	
1	x_1^1	...	x_1^t	...	x_1^T	$\sum_{t=1}^T x_1^t = 1$
...
i	x_i^1	...	x_i^t	...	x_i^T	$\sum_{t=1}^T x_i^t = 1$
...
I	x_I^1	...	x_I^t	...	x_I^T	$\sum_{t=1}^T x_I^t = 1$

Тогда финансирование работ проекта в абсолютных/денежных единицах представляется в виде матрицы C (таблица 2).

Таблица 2. Финансирование работ инвестиционного проекта

Работы i	Время t					Итого
	1	...	t	...	T	
1	$C_1 \cdot x_1^1$...	$C_1 \cdot x_1^t$...	$C_1 \cdot x_1^T$	C_1
...
i	$C_i \cdot x_i^1$...	$C_i \cdot x_i^t$...	$C_i \cdot x_i^T$	C_i
...
I	$C_I \cdot x_I^1$...	$C_I \cdot x_I^t$...	$C_I \cdot x_I^T$	C_I
Итого	$\sum_{i=1}^I C_i \cdot x_i^1$...	$\sum_{i=1}^I C_i \cdot x_i^t$...	$\sum_{i=1}^I C_i \cdot x_i^T$	$\sum_{i=1}^I C_i$

Выполнение совокупности работ инвестиционного проекта зависит от финансирования. Введем функцию зависимости времени окончания работ от финансирования проекта $f_i(X_i)$, представляющую собой функцию вектора X_i , где $X_i = (x_i^1, \dots, x_i^t, \dots, x_i^T)$ – вектор финансирования i -ой работы, результатом которой является время окончания работ инвестиционного проекта. Однако помимо фактического времени выполнения работ существует плановое время \tilde{f}_i , при котором финансирование проекта условно безгранично. При дефиците источников финансирования инвестиционного проекта могут возникать задержки выполнения его работ, т.е. $f_i(X_i) \neq \tilde{f}_i$.

При наличии дефицита предлагается откладывать оплату работ и поставок на будущие периоды, т.е. «замораживать» этапы и их группы, не затрагивая взаимосвязь их выполнения, в результате чего для каждой работы определяется величина ее задержки $l_i = f_i(X_i) - \tilde{f}_i$, которая находится как разность между плановым и фактическим окончанием работ.

Математически задержки можно описать через вектор продолжительностей, где номер элемента вектора соответствует номеру работы инвестиционного проекта:

$$(2) l = (l_1, l_2, \dots, l_i, \dots, l_I) \in l_J,$$

где I – число работ инвестиционного проекта,

l_J – множество векторов.

Задержки работ инвестиционного проекта приносят не только отсрочку окончания инвестиционного проекта и получения валовой прибыли, но и дополнительные, помимо плановых выплат, расходы за невыполнение обязательств в срок – неустойку – которые моделируются функцией штрафа, зависящей от времени просрочки.

Однако так как не все задержки окончания выполнения работ приводят к задержке всего инвестиционного проекта, не каждая задержка будет вызывать неустойку.

Это можно учесть в модели, используя понятие резерва времени. Каждой работе инвестиционного проекта соответствует полный резерв времени ее выполнения (TF_i), о чем было сказано во введении. Поэтому если сдвиг работы не превышает резерв времени ее выполнения ($l_i \leq TF_i$), то отсутствует задержка реализации всего инвестиционного проекта, т.е. на предприятие, реализующее инвестиционный проект, не накладываются штрафные санкции. Если сдвиг работы превышает резерв времени ее выполнения ($TF_i < l_i$), то возникает задержка окончания инвестиционного проекта, следовательно, предприятие несет дополнительные расходы, которые моделируются функцией «штрафа». Данные условия математически можно описать через индикатор неустойки y_i следующим образом:

$$(3) \quad y_i = \begin{cases} 1, & \text{если } \forall i=1, \dots, I \quad TF_i < l_i; \\ 0, & \text{если } \forall i=1, \dots, I \quad l_i \leq TF_i < 0. \end{cases}$$

При своевременном финансировании работ инвестиционного проекта матрица X будет квадратной, при несвоевременном – прямоугольной. В исходную квадратную матрицу при несвоевременном финансировании к продолжительности критического пути T будут добавляться столбцы в количестве L :

$$(4) \quad L = \sum_{i=1}^I (l_i - TF_i) \cdot y_i,$$

где L – запаздывание инвестиционного проекта.

Тогда в общем виде функцию штрафа можно представить следующим образом:

$$(5) \ S(L) = \sum_{t=1}^L \frac{Q^t}{(1+r)^{T+t}},$$

где Q^t – сумма неустойки за задержку окончания инвестиционного проекта в t -ый период времени,
 ρ – ставка дисконтирования.

Устранение кассовых разрывов (ситуации, когда существует временный недостаток денежных средств, необходимых для финансирования расходов) возможно не только с помощью «замораживания» этапов инвестиционного проекта, но и с помощью привлечения заемных средств, что предлагается учесть в формируемой модели через W^t – платежи по кредиту в момент времени t . Если величина $W^t > 0$, то кредит привлекается, если $W^t < 0$, то кредит погашается.

Математически платежи по кредитам можно описать вектором W , где номер элемента вектора соответствует периоду инвестиционного проекта, в котором привлекается или погашается кредит:

$$(6) \ W = (W^1, \dots, W^t, \dots, W^T) \in W^q,$$

где W^q – множество векторов.

Привлекаемый кредит должен покрыть не только сложившийся дефицит, но и обеспечить необходимый минимальный остаток денежных средств предприятия s_{\min} в каждый из периодов.

Сумма непогашенного кредита (текущей задолженности перед банком) в каждый из периодов не может быть отрицательной величиной, а также не может превышать величины максимального объема кредита W^{\max} , то есть в зависимости от периода времени $t = 1, \dots, (T+L)$ должно выполняться условие:

$$(7) \ \forall t = 1, \dots, (T+L) \quad W^{\max} \geq \sum_{t=1}^t W^t \geq 0.$$

Если кредиты и займы предприятия планируется привлекать под $\beta\%$ годовых, то расходы, связанные с выплатой процентов за пользование кредитом в t -ом периоде, будут равны:

$$(8) \ b \sum_{t=1}^t W^t.$$

Сумма всех положительных и отрицательных финансовых потоков по кредиту должна равняться нулю. Это означает, что в момент времени T кредит должен быть выплачен полностью:

$$(9) \sum_{t=1}^T W^t = 0.$$

Таким образом, с учетом вышесказанного NPV в (1) будет выглядеть следующим образом:

$$(10) NPV = \left[\sum_{t=1}^{T+L} \left(\left[\frac{R^t - b \cdot \sum_{t=1}^t W^t}{(1+r)^t} - \sum_{i=1}^I \frac{C_i}{(1+r)^{f_i(X_i)}} \cdot (1-h) \right] \right) \right] - s(L).$$

Следует отметить, что задержки инвестиционного проекта приводят к возникновению «штрафов», которые выплачиваются из чистой прибыли, поэтому при выплате «штрафов» не учитывается налог на прибыль.

В разрабатываемой задаче оптимизации графика финансирования инвестиционного проекта предлагается учитывать ограничение на финансовые потоки – на конец каждого из периодов сальдо финансовых потоков, рассчитываемое нарастающим итогом, должно быть положительным, то есть в зависимости от периода времени $t = 1, \dots, (T + L)$ должно выполняться условие:

$$(11) \forall t = 1, \dots, (T + L) \left[R^t - \sum_{i=1}^I (C_i \cdot x_i^t) + W^t - b \cdot \sum_{j=1}^t W^j \right] \geq 0.$$

Данное ограничение обозначает, что в каждый период времени сумма притоков денежных средств превышает сумму оттоков, причем остаток денежных средств предыдущего периода добавляется к последующему.

В обобщенном виде экономико-математическую модель можно представить в следующем виде:

$$(12) \left\{ \begin{array}{l} NPV = \left[\sum_{t=1}^{T+L} \left[\left(\frac{R^t - b \cdot \sum_{t=1}^t W^t}{(1+r)^t} - \sum_{i=1}^I \frac{C_i}{(1+r)^{f_i(X_i)}} \right) \cdot (1-h) \right] \right] - \\ - S(L) \xrightarrow{x \in X, W \in W^q} \max, \\ \forall t = 1, \dots, (T+L) \quad \sum_{t=1}^t \left[R^t - \sum_{i=1}^I (C_i \cdot x_i^t) + W^t - b \cdot \sum_{j=1}^t W^j \right] \geq 0, \\ \forall t = 1, \dots, (T+L) \quad W^{\max} \geq \sum_{t=1}^t W^t \geq 0, \\ \forall i = 1 \dots I \quad \text{если } l_i \leq TF_i < 0, \text{ тогда } y_i = 0, \\ \forall i = 1 \dots I \quad \text{если } TF_i < l_i, \text{ тогда } y_i = 1, \\ S(L) = \sum_{t=1}^L \frac{Q^t}{(1+r)^{T+t}}, \\ L = \sum_{i=1}^I (l_i - TF_i) \cdot y_i, \\ \sum_{t=1}^T W^t = 0. \end{array} \right.$$

Исходной информацией для экономико-математической модели является массив *Works*, представляющий собой таблицу, содержащую следующие элементы сетевой модели типа AoN: *Num* – наименование работы *i*, *EST_i* – раннее время начала выполнения работы *i*, *LST_i* – позднее время начала выполнения работы *i*, *t_i* – продолжительность работы *i*, *TF_i* – полный резерв выполнения работы *i*.

Исходными переменными экономико-математической модели являются: *R^t* – притоки денежных средств инвестиционного проекта, *C^t* – оттоки денежных средств инвестиционного проекта, *η* – налог на прибыль и *ρ* – ставка дисконтирования. *X_i* = (*x_i¹*, ..., *x_i^t*, ..., *x_i^T*) – вектор искомых величин.

На основе вышеприведенной модели разработана методика оптимизации графика финансирования инвестиционного проекта, которая представлена на рис. 2. Методика построена на основе эвристического алгоритма и содержит пятнадцать этапов.

При реализации методики в виде программного обеспечения на первом этапе формируется массив данных *Works*, содержащий строки с исходными данными графа работ инвестиционного проекта.

На втором этапе сортируется массив *Works* по убыванию.

На третьем этапе рассчитывается размер дефицита

$$D = \sum_{t=1}^T \left(R^t - \sum_{i=1}^I C_i \cdot x_i^t \right) \text{ в период времени } t (t = 1, \dots, T).$$

На четвертом этапе оценивается рассчитанное значение дефицита, полученное на предыдущем этапе. При отсутствии дефицита ($D \geq 0$) или наличии профицита в t -ый период времени методика переходит на пятый этап, на котором период времени t сравнивается со значением продолжительности критического пути ($t < T$). Если условие выполняется, то методика переходит на шестой этап, где рассматривается следующий период времени $t = t + 1$, после чего методика переходит к третьему этапу. При окончании перебора всех временных периодов проекта методика переходит на седьмой этап. В случае наличия дефицита ($D < 0$) методика переходит на десятый этап, на котором рассматривается вариант переноса сроков финансирования работ проекта.

На седьмом этапе рассчитываются значения запаздывания инвестиционного проекта $L = \sum_{i=1}^I (l_i - TF_i) \cdot y_i$, функции «штрафа»

$$S(L) = \sum_{t=1}^L \frac{Q^t}{(1+r)^{T+t}}, \text{ а также значение чистого дисконтированного потока } \Pi_1 = NPV_1(R^t, C_i, I, \eta, \rho, \sigma).$$

На восьмом этапе оценивается запаздывание инвестиционного проекта. Если запаздывание существует ($L > 0$), то методика переходит на девятый этап, на котором происходит подбор необходимого кредита. Подбор осуществляется методом пере-

бора всех временных периодов инвестиционного проекта. В случае существования дефицита в t -ый период времени размер необходимого кредита рассчитывается по формуле

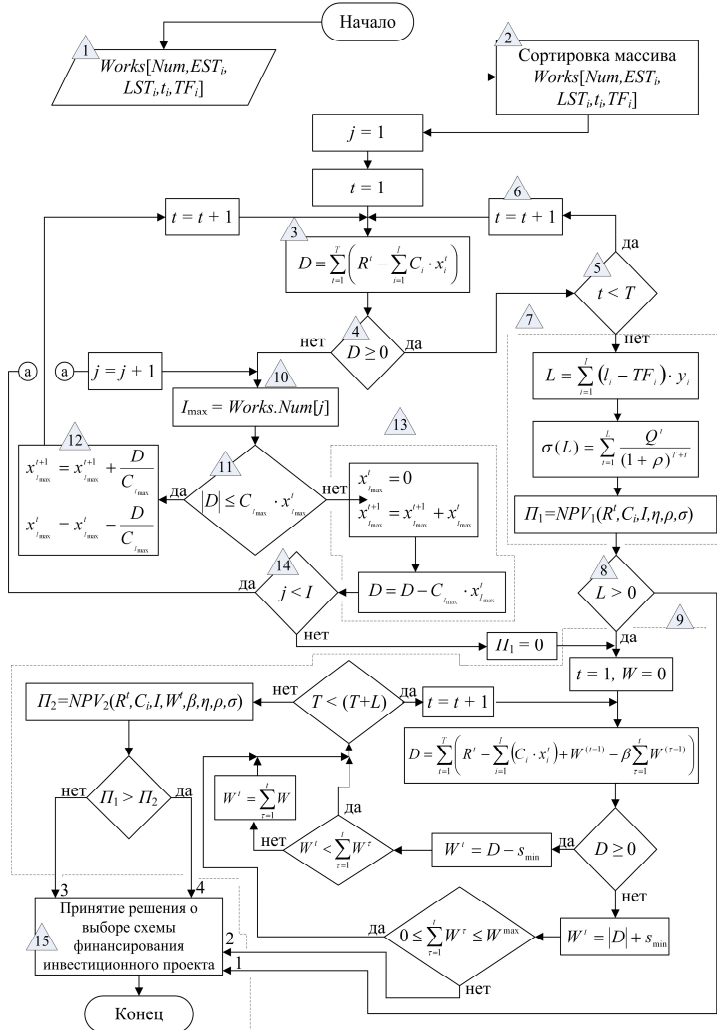


Рис. 2. Методика оптимизации графика финансирования на графах работ

$W^t = |D| + s_{\min}$. После чего происходит оценка величины необходимого кредита с величиной максимального объема кредита:

$$W^{\max} \geq \sum_{t=1}^t W^t \geq 0. \text{ Если кредит невозможен, методика переход}$$

ит на заключительный этап. По окончании перебора всех временных периодов происходит расчет чистого дисконтированного потока $\Pi_2 = NPV_2(R^t, C_i, W^t, \beta, I, \eta, \rho, \sigma)$ с учетом кредитования. Далее значения рассчитанных чистых дисконтированных потоков сравниваются ($\Pi_1 > \Pi_2$), и методика переходит на заключительный этап.

На десятом этапе показателю I_{\max} присваивается значение номера работы Num , стоящей в первой строке массива ($j = 1$) $Works$, т.е. с максимальным резервом времени.

На одиннадцатом этапе рассчитанный дефицит сравнивается с объемом финансирования работы I_{\max} $|D| \leq C_{I_{\max}} \cdot x_{I_{\max}}^t$.

Если на данном этапе финансирование работы I_{\max} превышает дефицит, то методика следует на двенадцатый этап, на котором объем финансирования работы I_{\max} в момент времени t частично уменьшается на объем дефицита за счет увеличения объема финансирования в момент времени $t + 1$, то есть

$$x_{I_{\max}}^{t+1} = x_{I_{\max}}^t + \frac{D}{C_{I_{\max}}} \text{ и } x_{I_{\max}}^t = x_{I_{\max}}^t - \frac{D}{C_{I_{\max}}}. \text{ Далее методика воз}$$

вращается на третий этап, на котором рассматривается следующий временной период.

Если же финансирование работы I_{\max} меньше дефицита, то методика переходит на тринадцатый этап, на котором объем финансирования работы I_{\max} в момент времени t полностью переносится на момент времени $t + 1$: $x_{I_{\max}}^t = 0$ и

$$x_{I_{\max}}^{t+1} = x_{I_{\max}}^{t+1} + x_{I_{\max}}^t. \text{ На данном этапе также рассчитывается новое}$$

значение дефицита после переноса расходов на будущий период, которое не удалось сократить путем «замораживания» оплаты работы с номером I_{\max} : $D = D - C_{I_{\max}} \cdot x_{I_{\max}}^t$. На четырнадцатом этапе происходит проверка, которая заключается в том, что пока

не достигнута последняя строка массива *Works*, то элемент j увеличивается на единицу ($j = j + 1$), т.е. показателю I_{\max} присваивается значение номера работы *Num*, стоящей в следующей строке массива *Works*, т.е. с меньшим резервом времени. В случае достижения последней строки массива *Works* значение чистого дисконтированного потока обнуляется $\Pi_1 = 0$. Это означает, что за счет переноса финансирования работы на будущий период времени невозможно избавиться от дефицита. После чего методика переходит к расчету объемов кредитования, т.е. на девятый этап.

На заключительном, пятнадцатом, этапе пользователю программного обеспечения необходимо принять решение о выборе схемы финансирования инвестиционного проекта.

Методика предполагает 4 сценария финансирования инвестиционного проекта, которые обозначены на рис. 2:

1 – При дефиците необходимо задерживать работы с максимальным полным резервом времени. Штрафов нет, кредиты не привлекаются.

2 – При дефиците необходимо задерживать работы с максимальным полным резервом времени. Возникают штрафы, так как кредит не реализуем.

3 – При дефиците необходимо привлекать кредиты, так как расходы по кредиту будут меньше штрафных санкций при задержках.

4 – При дефиците необходимо задерживать работы с максимальным полным резервом времени, так как штрафные санкции при задержках будут меньше, чем расходы по кредиту.

После выбора пользователя схемы финансирования инвестиционного проекта методика заканчивается.

Разработанная методика позволяет найти оптимальное решение модели оптимизации графика финансирования инвестиционного проекта на графах работ самым простым методом – путем перебора всех возможных значений. Методика трудоемка, так как использует самый затратный метод – метод перебора, однако обеспечивает сходимость к оптимальному решению за конечное число шагов.

Выводы

Разработанная экономико-математическая модель позволяет в условиях дефицита финансовых средств оптимизировать инвестиционные проекты, изменяя плановый вариант графика реализации, включающий ветвления с параллельным выполнением этапов работ. Оптимизация реализуется за счет изменений в финансировании этапов проекта, включая перенос финансирования на более поздние сроки, привлечение заемных средств и учет «штрафных» санкций. Разработанная методика включает одиннадцать этапов и позволяет оптимизировать график финансирования инвестиционных проектов на основе разработанной экономико-математической модели.

Литература

1. БОГАТЫРЕВ В.Д., ГРИЦЕНКО С.А. *Оптимизация инвестиционной программы строительной компании* // Труды международной научно-практической конференции «Управление большими системами – 2009», Т.2. М.: ИПУ РАН, 2009. С. 21 – 24.
2. БОГАТЫРЕВ В.Д., МОРОЗОВА С.А. *Экономико-математическая модель оптимизации графика финансирования с учетом сдвигов этапов инвестиционного проекта* // Вестник Самарского государственного университета путей сообщения. 2010. №3 (9). С. 76 – 84.
3. БУРКОВ В.Н. *Как управлять проектами: Научно-практическое издание* / В.Н. Бурков, Д.А. Новиков. М.: СИНТЕГ–ГЕО, 1997.
4. БУРКОВ В.Н. *Теория графов в управлении организационными системами* / В.Н. Бурков, А.Ю. Заложнев, Д.А. Новиков. М.: СИНТЕГ, 2001.
5. МОРОЗОВА С.А. *Модель и методика оптимизации графика финансирования инвестиционных проектов* // Вестник Самарского государственного аэрокосмического университета имени академика С.П. Королева (национального исследовательского университета). 2010. №3 (23). С. 82 – 93.

6. МОРОЗОВА С.А., ЕСИПОВА О.В. *Оптимизационная модель интеграции материальных и финансовых потоков инвестиционных проектов // Экономические науки. 2010. №8 (69). – С. 228 – 233.*
7. НОВИКОВ Д.А. *Управление проектами: организационные механизмы.* М.: ПМСОФТ, 2007.
8. РАЗУ М.Л. *Управление проектом. Основы проектного управления: учебник / М.Л. Разу, Т.М. Бронников, Б.М. Разу, С.А. Титов, Ю.В. Якутин.* М.: КНОРУС, 2006.
9. ТОВБ А.С. *Управление проектами: стандарты, методы, опыт / А.С. Товб, Г.А. Ципес.* М: ЗАО «Олимп-Бизнес», 2003.

MODEL AND TECHNIQUE OF INVESTMENT PROJECT FINANCING SCHEDULE OPTIMIZATION USING WORK CHARTS

Vladimir Bogatyrev, Samara State Aerospace University (National research university) (SSAU), Samara, Doctor of Science, professor, vice-rector of SSAU (samelev@rambler.ru).

Svetlana Morozova, Samara State Aerospace University (National research university) (SSAU), Samara, post-graduate student of Economics of SSAU (morozova_s_a@mail.ru).

Abstract: We propose an economic-mathematical model of investment project financing schedule optimization under cash deficiency. We develop a technique to optimize a investment project financing schedule by means of payment delays, loans drawing, and careful penalties calculation. At the end we draw policy recommendations for investment projects financing.

Keywords: investment project, financing, cash flow, deficit, surplus, net present value.

Статья представлена к публикации членом редакционной коллегии В.Н. Бурковым

УДК 62-78
ББК 35.9 + 65.050

АНАЛИЗ ЭФФЕКТИВНОСТИ ЭКОНОМИЧЕСКИХ СТИМУЛОВ ПОВЫШЕНИЯ БЕЗОПАСНОСТИ АВИАЦИОННОЙ ТЕХНИКИ

Варюхина Е. В.¹

(Московский физико-технический институт, Москва)

Клочков В. В.²

*(Учреждение Российской академии наук
Институт проблем управления РАН, Москва)*

Проведен экономико-математический анализ эффективности естественных рыночных механизмов, стимулирующих повышение безопасности авиационной техники. Показано, что авиационная промышленность, как отрасль, более уязвимая к спаду пассажирооборота после авиационных происшествий, заинтересована в повышении безопасности полетов сильнее, чем авиакомпания.

Ключевые слова: безопасность полетов, авиaperезовки, авиатехника, экономическая заинтересованность, стимулы, экономико-математическое моделирование.

1. Введение

На примере гражданской авиации рассматривается следующий вопрос: насколько эффективно естественные рыночные силы стимулируют повышение безопасности техники? После авиационных происшествий (катастроф), помимо непосредственных потерь (таких как потеря имущества, выплаты пострадавшим или их родственникам, судебные издержки и т.п.), авиакомпании также несут потери, связанные с сокращением спроса

¹ Екатерина Витальевна Варюхина, аспирант (eyebrow@yandex.ru).

² Владислав Валерьевич Клочков, доктор экономических наук, кандидат технических наук, ведущий научный сотрудник (vlad_klochkov@mail.ru).

на авиаперевозки. Падает доверие потенциальных пассажиров к воздушному транспорту, что заставляет их отказываться от поездок либо выбирать иные виды транспорта. Причем если риск прямых потерь является страхуемым, то риск сокращения спроса – нет. Поэтому, как полагали экономисты либерального направления, угроза сокращения спроса и доходов, наряду с угрозой прямых потерь, заставляет авиакомпании обеспечивать высокий уровень безопасности полетов. Таким образом, на первый взгляд, в данной сфере действуют естественные, рыночные обратные связи между уровнем безопасности полетов и доходами авиакомпаний, которые автоматически будут заинтересованы в повышении уровня безопасности. Однако в пионерской работе [6] было показано (эконометрическими методами), что потери самих авиакомпаний вследствие сокращения спроса на авиаперевозки после авиационных происшествий, как ни парадоксально, достаточно слабы. Кроме того, сокращения доходов носят краткосрочный характер и на долгосрочных интервалах нивелируются. Рассматривались изменения капитализации авиакомпаний, т.е. фактически изменения курсов их акций, которые, в свою очередь, представляют собой рыночную оценку их дисконтированных будущих доходов, и неудивительно, что краткосрочные спады этих доходов сглаживаются. Указанное исследование было проведено в начале так называемого *дерегулирования авиаперевозок* в США. По его результатам (а также по результатам более поздних исследований, опирающихся на названную основополагающую работу в методологическом отношении, см. [7, 9]) было признано, что рыночные силы недостаточно эффективно обеспечивают заинтересованность авиаперевозчиков в повышении безопасности полетов, и в этом отношении необходим жесткий государственный контроль.

В данной работе, в отличие от описанных, рассматриваются экономические механизмы, стимулирующие именно повышение безопасности авиационной техники. На первый взгляд, это не очень актуально в сложившейся ситуации, поскольку, по данным многих исследований (см., например, [3]) подавляющая доля авиационных происшествий связана с человеческим фактором в эксплуатации. То есть даже если сократить до нуля вероятность катастрофических отказов техники, радикального по-

вышения безопасности полетов ожидать не приходится, и, на первый взгляд, возможности ее повышения со стороны авиастроителей практически исчерпаны. Безопасность полетов – комплексная категория, она определяется всеми элементами авиатранспортной системы: воздушными судами, инфраструктурой технического обслуживания и ремонта (ТОиР) авиатехники, экипажами и системой их подготовки, системой управления воздушным движением (УВД), аэродромной инфраструктурой и т.д. Далеко не все эти элементы подконтрольны авиационной промышленности. Однако нередко существуют возможности сократить вероятность тех или иных негативных проявлений человеческого фактора в эксплуатации еще на этапе разработки изделий. Например, многие операторские ошибки вызваны эргономическим несовершенством авиатехники, и хотя с формальной точки зрения разработчик и производитель не виновны в них (оператор нарушил руководство по летной эксплуатации), нередко он способен исключить или существенно снизить их вероятность. Также известны конструктивные и технологические решения, затрудняющие даже преднамеренное нарушение дисциплины в процессе летной эксплуатации авиатехники, ее технического обслуживания и ремонта.

Проверяется следующая гипотеза: авиационная промышленность заинтересована в повышении безопасности полетов гораздо больше, чем эксплуатирующие ее изделия авиакомпании. Причем она может быть заинтересована в повышении безопасности не только в своей зоне ответственности (т.е. в минимизации конструктивно-производственных недостатков), но и за ее пределами, т.е. в период эксплуатации, сводя к минимуму возможные ошибки или преднамеренные небезопасные действия эксплуатантов, неконтролируемое воздействие на воздушные суда природных факторов и т.д. Для проверки этой гипотезы необходимо провести математическое моделирование соответствующих экономических механизмов. Если она подтвердится, необходимо провести анализ экономической заинтересованности авиакомпаний в приобретении новой, более безопасной авиатехники.

2. Анализ экономической заинтересованности авиастроительных компаний в повышении безопасности полетов

Прежде всего необходимо оценить влияние уровня безопасности полетов на конъюнктуру рынка авиаперевозок. Поскольку такие оценки необходимо получать не только при исторически достигнутых показателях безопасности, но и при ожидаемых в перспективе, здесь неприменим эконометрический подход, использованный в работах [6, 7]. Предлагается упрощенная структурная экономико-математическая модель зависимости спроса на авиаперевозки от уровня риска авиационного происшествия. Как и в других оптимизационных моделях спроса на перевозки, здесь рассматривается выбор пассажиров, дифференцированных по доходным группам, между воздушным и наземным транспортом в дальнем сообщении. Но, в отличие от традиционных моделей подобного рода (см., например, [2]), учитывается фактор риска попадания в авиационное происшествие. Как правило, структурные оптимизационные модели спроса на перевозки весьма сложны и требуют значительного объема исходной информации. Если учитывать, что в течение года пассажир может совершать несколько дальних поездок на различные расстояния, необходимо знать зависимость частоты поездок от дальности для каждой доходной группы. В ряде работ учитывается также стохастический характер поведения потребителей, в том числе путем использования рандомизированной функции полезности (см., например, [4]). Именно такой подход использован и в работе [9], однако такие модели весьма сложны для аналитического исследования в общем виде. В работе [1] предлагается более простой подход к прогнозированию спроса на перевозки: считается, что на определенную долю дохода сверх прожиточного минимума пассажир совершает единственную поездку в течение года, выбирая для ее совершения тот вид транспорта, который обеспечивает ему меньшее значение полной стоимости пассажира-километра (далее сокращенно *пкм*). В ее составе учитываются как прямые денежные затраты (тариф), так и упущенная выгода (стоимость часа рабочего времени пассажира, зависящая от дохода индивида). Здесь предлагается модифи-

кация упрощенной модели, предложенной в работе [1]: в суммарную стоимость пассажира-километра добавлено слагаемое, отражающее ожидаемые потери в случае происшествия. Простота принятой модели поведения пассажира делает получаемую зависимость спроса на перевозки от риска происшествия наглядной и пригодной для аналитического исследования, в отличие от весьма сложной модели [9]. Ниже приведена спецификация описанной модели.

Пусть пассажир оценивает потери при происшествии в R денежных единиц:

$$(1) \quad R = j \cdot M ,$$

где M – его среднемесячный доход; j – коэффициент связи между оценкой потерь и месячным доходом.

Тогда, если субъективная оценка вероятности происшествия в расчете на один пассажиро-километр равна ψ , дополнительные ожидаемые потери из-за низкой безопасности полетов составят

$$(2) \quad \psi \cdot R = j \cdot \psi \cdot M .$$

Эта добавка учитывается в полной стоимости пкм:

$$(3) \quad C_{\text{пкм}} = p + \frac{z}{v} + \psi \cdot R ,$$

где z – стоимость часа времени пассажира; v – средняя скорость транспорта; p – стоимость пассажиро-километра по тарифу.

Всё население разбивается на $n_{\text{дох}}$ групп по доходам. Бюджет, выделяемый на поездки представителем i -й группы, составляет

$$(4) \quad B_i = 12(M_i - M^{\min})a ,$$

где коэффициент a отражает долю разности между годовым доходом и прожиточным минимумом, затрачиваемую на поездки; M^{\min} – прожиточный минимум; M_i – месячный доход представителя i -й доходной группы, $i = 1, \dots, n_{\text{дох}}$.

Следовательно, максимальная дальность поездки пассажира из i -й доходной группы на k -м виде транспорта будет составлять

$$(5) \quad L_i^k = \frac{12(M_i - M^{\min}) \cdot \alpha}{p^k + \left(\frac{b}{v^k} + j \cdot \psi^k\right) \cdot M_i},$$

где k – вид транспорта (железнодорожный или авиационный); $b = z/M_i$.

Предполагаем, что представитель i -й доходной группы предпочтет авиатранспорт железнодорожному, если $L_i^{\text{авиа}} \geq L_i^{\text{жд}}$, т.е., как и в работе [1], ставится задача максимизации подвижности пассажира. Суммируя дальности поездок для тех доходных групп, чьи представители выберут авиатранспорт, и умножая на численность населения в i -й доходной группе n_i , получим пассажирооборот W воздушного транспорта, измеряемый в пассажиро-километрах и определяющий, в том числе, потребность в продукции авиастроения:

$$(6) \quad W = \sum_{i=1}^{n_{\text{дох}}} \Delta_i n_i L_i^{\text{авиа}},$$

$$\text{где } \Delta_i = \begin{cases} 1, & L_i^{\text{авиа}} \geq L_i^{\text{жд}}, \\ 0, & L_i^{\text{авиа}} < L_i^{\text{жд}}. \end{cases}$$

Как видно из выражений (5) и (6), пассажирооборот сокращается с ростом субъективной оценки риска попадания в авиационное происшествие, причем как по причине сокращения подвижности отдельного пассажира, так и по причине отказа представителей отдельных доходных групп (в порядке возрастания дохода) от услуг воздушного транспорта.

Для оценки потерь авиастроительных предприятий вследствие повышения оценки вероятности происшествия, необходима модель совокупного спроса на авиатехнику в зависимости от пассажирооборота гражданской авиации. В данной работе такая модель строится при следующих допущениях:

- авиастроение рассматривается как единое целое, конкуренция между отдельными предприятиями (в реальности чрезвычайная жесткая) не учитывается;
- динамика спроса на авиаперевозки и уровня мощностей авиационной промышленности задается экзогенным образом;
- все воздушные суда – однотипные;

- условия эксплуатации парка однородны;
- распределение воздушных судов в парке по остатку ресурса – равномерное.

Следует подчеркнуть, что говорить о спросе на авиатехнику в строгом смысле этого термина, т.е. о платежеспособной потребности, было бы правомерно лишь при условии анализа платежеспособности авиакомпаний после тех или иных изменений пассажиропотока. В данной работе такой анализ не проводится, поэтому в дальнейшем прогнозируется, строго говоря, лишь потребность авиакомпаний в воздушных судах для выполнения ожидаемых объемов транспортной работы.

Введем следующие условные обозначения:

$W(t)$ – пассажирооборот в году t , пкм/год;

$N(t)$ – списочная численность парка гражданской авиации;

m – пассажировместимость одного воздушного судна, мест;

v – средняя рейсовая скорость воздушного судна, км/ч;

h – среднегодовой налет на одно воздушное судно, летных часов в год;

k – коэффициент заполнения кресел.

Тогда число воздушных судов, необходимое для удовлетворения спроса на перевозки в году t , может быть оценено по следующей формуле:

$$(7) \quad N_{\text{потреб}}(t) = \frac{W(t)}{v \cdot \eta \cdot k \cdot m}.$$

Если $N_{\text{потреб}} > N(t)$, то в году t имеет место дефицит провозных мощностей. Если же, напротив, $N_{\text{потреб}} < N(t)$, то существующие провозные мощности избыточны. Поскольку существенное снижение коэффициента заполнения кресел и налета на воздушное судно приводит к убыткам авиакомпаний, они стремятся эксплуатировать авиатехнику максимально интенсивно. В связи с этим при избытке провозных мощностей авиакомпании, как правило, выводят из эксплуатации часть принадлежащих им воздушных судов. Следовательно, в году t эксплуатироваться будет лишь часть парка, численность которой $N_3(t)$ можно определить в общем случае следующим образом:

$$(8) \quad N_3(t) = \min\{N_{\text{потреб}}(t), N(t)\}.$$

Пусть T – назначенный ресурс изделия, выраженный в летних часах. Тогда эксплуатируемые изделия будут в среднем выработать за год долю ресурса, равную h/T . Следовательно, количество изделий, выработавших в году t свой ресурс и подлежащих списанию, может быть приближенно представлено в следующем виде (при равномерном распределении парка по возрасту):

$$(9) \quad q_{\text{спис}}(t) = \frac{\eta}{T} N_{\text{э}}(t).$$

Если авиакомпании не будут приобретать новых воздушных судов, численность их парка к следующему году сократится до $N(t) - q_{\text{спис}}(t)$. Необходимо оценить, достаточно ли этих провозных мощностей для удовлетворения ожидаемого спроса на перевозки. Обозначим $W_e(t+1)$ ожидаемый авиакомпаниями в году t уровень спроса на авиаперевозки в будущем году $t+1$. При неизменных значениях среднегодового налета h , рейсовой скорости n , пассажироместности m и коэффициента заполнения кресел k , потребная численность парка пропорциональна пассажирообороту. Тогда в будущем году $t+1$ потребуются следующее количество воздушных судов:

$$(10) \quad N(t+1) = \frac{W(t+1)}{\eta \cdot v \cdot k \cdot m} = \frac{W(t+1)}{W(t)} N_{\text{потреб}}(t).$$

Если это количество не превосходит численность парка, оставшегося к началу года $t+1$, т.е., $N_{\text{потреб}}(t) \leq N(t) - q_{\text{спис}}(t)$, то приобретения новых воздушных судов не требуется. Иначе для поддержания баланса провозных мощностей авиакомпаниям необходимо приобрести по меньшей мере $N_{\text{потреб}}(t+1) - N(t) + q_{\text{спис}}(t)$ воздушных судов. Таким образом, выражение для спроса на воздушные суда примет окончательный вид:

$$(11) \quad q_{\text{необх}}(t) = \max \left\{ 0; \frac{W_e(t+1)}{v \cdot \eta \cdot k \cdot m} + \frac{\eta}{T} N_{\text{э}}(t) - N(t) \right\}.$$

Если изначально не было ни дефицита, ни избытка провозных мощностей, тогда $N(t) = W(t)/(n \cdot h \cdot k \cdot m)$, $N_{\text{э}}(t) = N(t)$, и выражение упрощается:

$$(12) q_{\text{необх}}(t) = \max \left\{ 0; \left(\frac{W_e(t+1)}{W(t)} - 1 + \frac{h}{T} \right) \times N(t) \right\}.$$

Считаем, что мощности авиастроения достаточны для удовлетворения спроса на новые ВС, т.е. поставки ВС равны заказам.

Представленные выше уравнения (7)–(12) представляют собой упрощенную модель *баланса провозных мощностей* (далее БПМ) гражданской авиации, широко применяемую для прогнозирования совокупного спроса на авиатехнику, см. [1].

В расчете по описанной модели (1)–(12) использовались следующие данные, близкие к реальным для России: коэффициент оценки потерь при происшествии по отношению к месячному доходу $j = 1000$; коэффициент связи между доходом и стоимостью часа $b = 1/200$ (считаем, что в месяце 25 рабочих дней, в рабочем дне – 8 рабочих часов). Распределение населения по величине среднедушевых денежных доходов, основанное на официальных данных Росстата за 2009 г. [12], представлено на рис. 1.

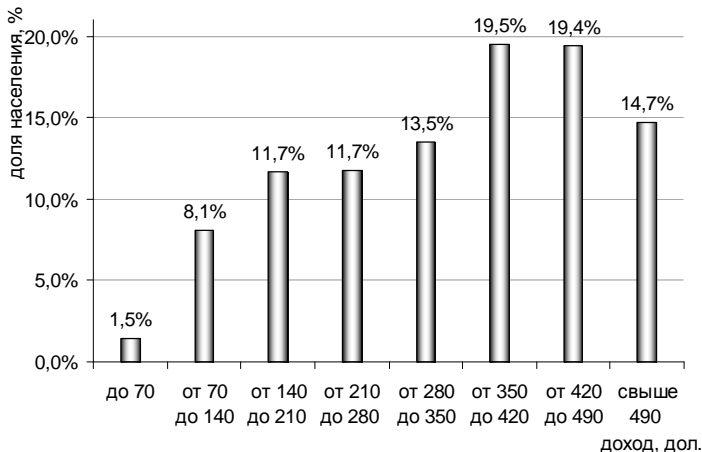


Рис. 1. Распределение населения по величине среднедушевых денежных доходов (по данным Росстата)

Согласно тем же данным, численность населения в данном году приблизительно принята равной 143 млн. чел., а прожиточ-

ный минимум – 5000 руб./месяц. Доля расходов на дальние поездки сверх прожиточного минимума $a = 5\%$. Рейсовая скорость самолета принята равной $n^{\text{авиа}} = 800$ км/ч; скорость железнодорожного транспорта $n^{\text{жд}} = 50$ км/ч, стоимость пассажирокилометра на авиатранспорте и железнодорожном транспорте соответственно $p^{\text{авиа}} = 0,12$ долл/пкм и $p^{\text{жд}} = 0,07$ долл/пкм. Значения авиатарифа были получены на основе статистической обработки данных [14] как отношение стоимости билета к дальности полета для разных направлений и авиакомпаний; рассматривался только эконом-класс с тарифом, не предусматривающем перенос даты или возврат билета. Были рассмотрены дальности полета от 1200 до 3500 км, средняя стоимость пассажирокилометра составила от 14 до 9 центов соответственно. Значения железнодорожного тарифа были получены на основе статистической обработки данных [15] как отношение стоимости билета к дальности поездки для разных направлений. Были рассмотрены дальности от 1200 до 3500 км, средняя стоимость пассажирокилометра составила от 9 до 5 центов соответственно. Оценка вероятности происшествия на самолете принята равной $y^{\text{авиа}} = 2 \cdot 10^{-8}$ пкм⁻¹, на железнодорожном транспорте $y^{\text{жд}} = 2 \cdot 10^{-9}$ пкм⁻¹. Пассажировместимость и среднегодовой налет самолета составляют, соответственно, $m = 150$ и $h = 3000$ л.ч., что соответствует типичным значениям для среднемагистрального самолета. Ожидания авиакомпаний относительно пассажирооборота считаются статическими, т.е. $W_e(t + 1) = W(t)$, и, в соответствии с (10), $N_{\text{потреб}}(t + 1) = N_{\text{потреб}}(t)$. Таким образом, предполагается, что авиакомпании, формируя заказы на новую авиатехнику, будут лишь стремиться восполнить текущий дефицит провозных мощностей.

На рис. 2. представлены полученные в данном примере относительные изменения спроса на авиаперевозки и на новые самолеты в зависимости от относительного изменения вероятности авиационного происшествия.

Видно, что изменение оценки вероятности происшествия оказывает сравнительно небольшое влияние на рынок авиаперевозок – пассажирооборот гражданской авиации падает всего лишь на 22% от первоначального уровня при увеличении оцен-

ки вероятности происшествия в 2,6 раза. Спрос же на самолеты падает на 100%, т.е. до нуля, уже при увеличении оценки вероятности происшествия в 1,5 раза. Скачки на данном графике связаны с тем, что приведенное на рис. 1 распределение населения по величине среднедушевых доходов дискретно.

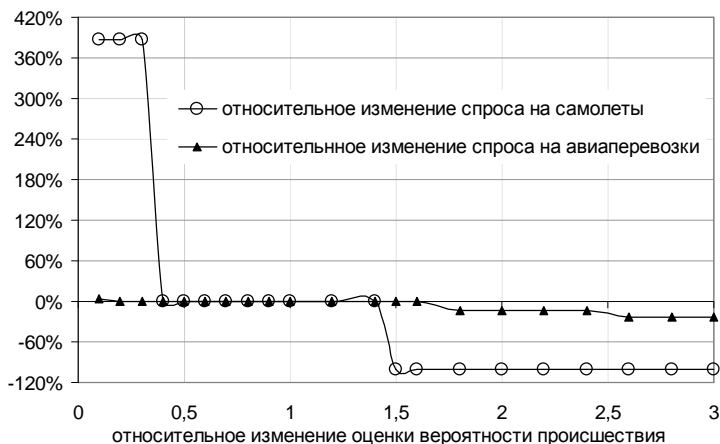


Рис. 2. Относительное изменение спроса авиаперевозки и потребности в воздушных судах в зависимости от относительного изменения оценки вероятности происшествия

Подчеркнем, что ожидаемые потери при попадании в авиационное происшествие входят в суммарную стоимость пассажиро-километра линейным образом. То есть пассажир считается нейтральным по отношению к риску. На самом деле это не так – вероятнее всего, он не расположен к риску, поэтому приведенные выше оценки спроса на авиаперевозки следует рассматривать как завышенные, оптимистические. Поэтому при снижении безопасности полетов спад спроса на авиаперевозки, а следовательно, и на самолеты будет более глубоким. По данным Национального совета США по безопасности на транспорте, *National Transportation Safety Board (NTSB)* [13], с 2000 по 2001 г. (т.е. в период, когда в гражданской авиации США произошли трагические события 9 сентября 2001 г.) годовое количество авиакатастроф в США выросло с 10 до 12, т.е. на 20%, что вы-

звало спад пассажирооборота за год, последовавший за данными событиями, на 9% относительно аналогичного периода до этих событий, см. [11]. Заметим, что согласно используемой здесь модели, при изменении оценки вероятности происшествия в 1,2 раза пассажирооборот вообще практически не должен измениться, в силу дискретности распределения населения по доходам. В то же время, в полном согласии с описанной моделью БПМ, хотя мировой пассажирооборот упал всего лишь на 3%, спрос на новые воздушные суда сократился гораздо сильнее – на 52%, см. рис. 3 (по официальным данным ведущих производителей магистральных пассажирских самолетов, обобщенным в [1]).

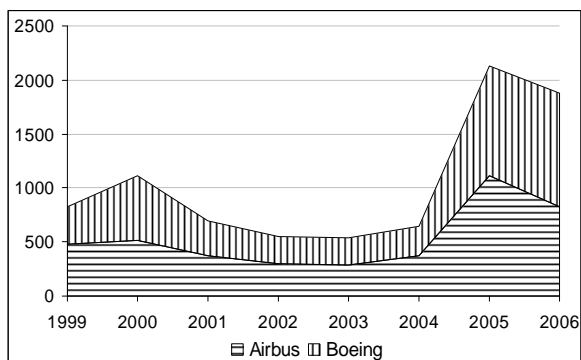


Рис. 3. Относительное изменение количества заказов на магистральные пассажирские самолеты в окрестности 2001 г.

Это подтверждает предположение о том, что в реальности спрос на авиаперевозки и новую авиатехнику более чувствителен к увеличению оценки вероятности авиационного происшествия, чем предсказывает предлагаемая здесь модель. Как показывает ее анализ и пример расчета, на авиастроителях увеличение оценки вероятности авиационного происшествия, как и предполагалось, отразится гораздо более глубоким спадом, чем на авиакомпаниях. При этом для авиационной промышленности спад выпуска чреват деградацией кадрового потенциала. В высокотехнологичных отраслях действует *эффект обучения*, который состоит в сокращении удельных трудозатрат с ростом накопленного опыта – по статистике [5, 10], удвоение выпуска

гражданских самолетов приводит к снижению удельных затрат на 15–20%. Однако это справедливо лишь при сохранении ритмичного выпуска изделий. В периоды вынужденного простоя действует обратный эффект – *забывания*. Согласно результатам статистических исследований, приведенным в [5], за год простоя самолетостроительного производства теряется около 40% накопленного опыта. Следовательно, авиастроители, как и предполагалось, гораздо больше, чем авиакомпании, заинтересованы в увеличении безопасности полетов.

3. Анализ заинтересованности авиакомпаний в ускоренной замене авиатехники на более безопасную

Однако даже если производители будут заинтересованы в разработке и освоении производства более безопасной авиатехники, необходимо оценить, насколько авиакомпании заинтересованы в ее приобретении, если у них в достаточном количестве имеется авиатехника предыдущего поколения, обладающая остатком ресурса и пригодная к эксплуатации. В работе [1] сформулированы условия, при которых становится целесообразным с экономической точки зрения досрочное списание современной авиатехники и ее замена на новую (более экономичную). Это имеет место, когда средняя ставка амортизации нового типа самолетов в расчете на летный час, определяемая как отношение цены самолета нового типа к его назначенному ресурсу, меньше, чем разность текущих эксплуатационных расходов (на авиатопливо, техобслуживание и ремонт и т.п.) старого и нового изделий. Однако в рассматриваемом случае это условие может не выполняться, хотя новые типы авиатехники, как правило, и безопаснее, и экономичнее старых. В то же время, при повышении безопасности полетов авиакомпании могут рассчитывать на повышение спроса на перевозки и выручки. Таким образом, чтобы оценить их заинтересованность в ускоренном обновлении парка авиатехники, необходимо рассмотреть не изменение удельных затрат на летный час или пассажиро-километр, а изменение совокупной прибыли авиакомпаний после замены старой техники на более безопасную.

Изменение прибыли авиакомпаний страны $\Delta\Pi$ выражается следующим образом:

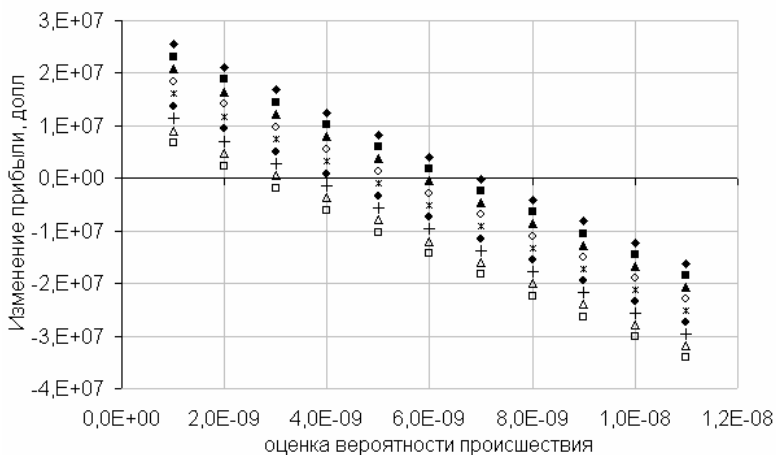
$$(13) \quad \Delta\Pi = \Pi^{\text{нов}} - \Pi^{\text{стар}} = W^{\text{нов}} (p^{\text{авиа}} - a^{\text{нов}} - p_{\text{ГСМ}} g^{\text{нов}} - c_{\text{ТОиР}}^{\text{нов}}) - \\ - W^{\text{стар}} (p^{\text{авиа}} - p_{\text{ГСМ}} g^{\text{стар}} - c_{\text{ТОиР}}^{\text{стар}}),$$

где $C_{\text{ТОиР}}^{\text{стар}}$ и $C_{\text{ТОиР}}^{\text{нов}}$ – стоимость технического обслуживания и ремонта (ТОиР) воздушных судов соответственно старого и нового типа; $W^{\text{стар}}$ и $W^{\text{нов}}$ – соответственно пассажирооборот до замены старой техники и после замены авиапарка на новые воздушные суда; $g^{\text{стар}}$ и $g^{\text{нов}}$ – удельный расход топлива, в расчете на пассажиро-километр, воздушных судов соответственно старого и нового типа; $p_{\text{ГСМ}}$ – цена авиатоплива (горюче-смазочных материалов, ГСМ); $p^{\text{авиа}}$ – тариф, в расчете на пассажиро-километр (здесь он считается постоянным, т.е. оптимизация ценовой политики авиакомпаний не рассматривается); $a^{\text{нов}}$ – средняя ставка амортизации в расчете на пассажиро-километр для нового типа самолетов (она может быть оценена как отношение стоимости самолета к его суммарному объему транспортной работы за весь расчетный срок службы).

В расчете модели использовались следующие значения параметров: удельный расход топлива $g^{\text{стар}} = 20$ г/пкм, цена ГСМ $p_{\text{ГСМ}} = 1000$ долл./тонн, амортизация нового самолета $a^{\text{нов}} = 0,01$ долл./пкм (оценка приведена в работе [1]), стоимость ТОиР $C_{\text{ТОиР}}^{\text{нов}} = C_{\text{ТОиР}}^{\text{стар}} = 0,01$ долл./пкм.

Для простоты не учитываются затраты на переобучение экипажей на новый тип воздушных судов, на получение нового сертификата эксплуатанта и т.п., поскольку они по порядку величины ниже статей расходов, перечисленных выше.

На рис. 4 изображены значения изменения прибыли авиакомпаний гражданской авиации России (пассажирооборот рассчитан на основе вышеописанной структурной модели). Проведены параметрические расчеты для различных уровней удельного расхода топлива нового типа авиатехники.



◆ $g^{noc} = 12,0$ ■ $g^{noc} = 13,0$ ▲ $g^{noc} = 14,0$ ◇ $g^{noc} = 15,0$ × $g^{noc} = 16,0$ ● $g^{noc} = 17,0$
 + $g^{noc} = 18,0$ △ $g^{noc} = 19,0$ □ $g^{noc} = 20,0$

Рис. 4. Изменение прибыли авиакомпании в зависимости от оценки вероятности происшествия и удельного расхода топлива в расчете на пассажиро-километр для нового типа самолетов

В тех случаях, когда изменение прибыли авиакомпаний не будет положительным, они не будут заинтересованы в ускоренной замене старой, менее безопасной авиатехники, и потребуются государственное вмешательство для стимулирования смены самолетов на более безопасные.

По причине исчерпания возможностей совершенствования конструкции современных гражданских самолетов и авиадвигателей, маловероятно существенное сокращение удельного расхода топлива новыми моделями ближнесреднемагистральных самолетов, которые планируется вывести на рынки в обозримом будущем – в 2020–2025 гг. Значительные улучшения, запланированные на более отдаленную перспективу, см. [8], уже требуют радикальных инноваций. В ближайшем же будущем можно рассчитывать на повышение топливной экономичности на 20–25% от нынешнего уровня, т.е. на снижение удельного расхода топлива с нынешних 20–22 г/пкм до 15–17 г/пкм. Легко

видеть, что без учета повышения безопасности полетов появление такого нового поколения изделий не вызовет морального устаревания современных самолетов и их массовой ускоренной замены: при нынешней цене авиатоплива ожидаемая экономия при переходе на новую технику составит, в лучшем случае, лишь 0,005 долл./пкм, тогда как ставка амортизации новых изделий окажется заведомо не ниже современного уровня 0,01 долл./пкм. Сопоставление указанных значений и $g^{\text{нов}}$ с графиком на рис. 4 показывает, что авиакомпании будут экономически заинтересованы в немедленном приобретении самолетов нового поколения только при условии, что вероятность попадания в авиационное происшествие сократится от нынешнего уровня до $4\text{--}5 \cdot 10^{-9}$ пкм⁻¹, т.е. в 4–5 раз. В принципе, такой уровень безопасности полетов уже достигнут в последние годы в гражданской авиации наиболее экономически развитых стран мира, т.е. он принципиально достижим даже при нынешнем техническом уровне авиационной техники и наземной инфраструктуры. Как отмечено выше, проведенный учеными и специалистами отрасли системный анализ сложившегося положения дел в сфере безопасности полетов в нашей стране (см., например, [3]) показал, что свыше 80% причин авиационных происшествий связано с человеческим фактором на стадии эксплуатации, технического обслуживания и ремонта авиатехники. Но это отнюдь не означает, что минимизировать влияние этих факторов авиационная промышленность совершенно не в состоянии. Проведенный анализ показывает, что авиационная промышленность экономически заинтересована в разработке авиатехники, практически не позволяющей нарушить правила ее безопасной эксплуатации, преднамеренно или непреднамеренно. Разумеется, не подвергается сомнению актуальность совершенствования институциональной среды в гражданской авиации, механизмов обеспечения дисциплины эксплуатации, совершенствования системы подготовки операторов (в том числе, опять-таки, при активном участии авиационной промышленности – в части разработки тренажерных технологий и тренажерной техники).

4. Заключение

1. Разработана упрощенная экономико-математическая модель зависимости спроса на авиаперевозки и новую авиатехнику от уровня риска авиационного происшествия, причем, в отличие от традиционных моделей подобного рода, в суммарную стоимость пассажиро-километра включено слагаемое, отражающее ожидаемые потери в случае происшествия. Аналитическое исследование модели и расчеты с использованием реалистичных исходных данных показали, что на авиастроителях увеличение субъективной оценки пассажирами вероятности авиационного происшествия отразится гораздо более глубоким спадом выпуска, чем на авиакомпаниях. В свою очередь, ущерб от краткосрочного спада выпуска авиатехники может быть велик из-за утраты накопленного опыта и деградации человеческого капитала. Следовательно, авиастроители больше, чем перевозчики, заинтересованы в повышении уровня безопасности полетов, даже за пределами своей зоны ответственности.

2. Определены условия, при которых для авиакомпаний становится целесообразным, с экономической точки зрения, досрочное списание современной авиатехники и ее замена на новую, более безопасную. Для этого было рассмотрено изменение совокупной прибыли авиакомпаний после замены старой техники на новую и учтено не только сокращение удельных затрат на пассажиро-километр, но и повышение спроса на авиаперевозки вследствие повышения доверия пассажиров. Расчеты с использованием реалистичных исходных данных показывают, что при реально достижимом для следующего поколения гражданских ближне- и среднемагистральных самолетов сокращении расхода топлива с нынешних 20–22 г/пкм до 15–17 г/пкм досрочная замена современной авиатехники станет экономически целесообразной, если вероятность авиационного происшествия сократится в 4–5 раз. Такой уровень безопасности полетов может быть достигнут при условии системных преобразований в авиатранспортной отрасли, а также благодаря новым конструктивно-технологическим решениям, нацеленным на минимизацию риска негативного проявления человеческого фактора. Авиационная промышленность экономически заинтересована в выработке таких решений.

Литература

1. КЛОЧКОВ В.В. *Управление инновационным развитием гражданского авиастроения: монография.* – М.: ГОУ ВПО МГУЛ, 2009. – 280 с.
2. ПОЗАМАНТИР Э.И. *Модели спроса на перевозки* // Экономико-математический энциклопедический словарь. – М.: Большая Российская Энциклопедия, 2003. – С. 506–508.
3. УВАРОВ Р.В. *Формирование организационно-экономического механизма обеспечения безопасности полетов:* Дис. канд. экон. наук. – М., 2005 – 167 с.
4. BEN-AKIVA M., LERMAN S.M. *Discrete Choice Analysis: Theory and Application to Travel Demand.* – MIT Press, Cambridge, Mass., 1985.
5. BENKARD C.L. *A Dynamic Analysis of the Market for Wide-bodied Commercial Aircraft* // Review of Economic Studies. – 2004. – Vol. 71, No. 3. – P. 581–611.
6. BORENSTEIN S., ZIMMERMAN M.B. *Market Incentives for Safe Commercial Airline Operations* // The American Economic Review. – 1988. – Vol. LXXVII, No. 5. – P. 913–936.
7. CUNNINGHAM L. F., YOUNG C.E., LEE M. *Perceptions of Airline Service Quality: Pre and Post 9/11* // Public Works Management & Policy. – 2004. – Vol. 9, No. 1. – P. 10–25.
8. *National Plan for Aeronautics Research and Development and Related Infrastructure.* – NASA, 2007. – 56 p.
9. SQUALLY J., SAAD M. *Accidents Airline Safety Perceptions and Consumer Demand* // Journal of Economics and Finance. – 2006. – Vol. 30, No. 3. – P. 297–305.
10. WRIGHT T.P. *Factors Affecting the Cost of Airplanes* // Journal of Aeronautical Sciences. – 1936. – Vol. 3. – P. 122–128.
11. <http://www.faa.gov/> (дата обращения 23.05.2011).
12. <http://www.gks.ru/> (дата обращения 23.05.2011).
13. <http://www.nts.gov/> (дата обращения 23.05.2011).
14. <http://poezdka.ru> (дата обращения 20.05.2011).
15. <http://rzd.ru/> (дата обращения 20.05.2011).

EFFICIENCY ANALYSIS OF AIRCRAFT SAFETY INCREASE ECONOMIC INCENTIVES

Ekaterina Varyukhina, Moscow Institute of Physics and Technology, Moscow, postgraduate student (*eyebrow@yandex.ru*).

Vladislav Klochkov, Institute of Control Sciences of RAS, Moscow, Doctor of Science, leading research fellow (*vlad_klochkov@mail.ru*).

Abstract: The economic-mathematical efficiency analysis of natural market mechanisms that stimulate the aircraft safety increase was carried out. It was shown that aircraft industry is more vulnerable in passenger traffic decrease than airlines after plane crashes so the aircraft industry is more interested in flight safety than airlines.

Keywords: flight safety, air traffic, aircraft, economic interest, incentives, economic-mathematical modeling.

*Статья представлена к публикации
членом редакционной коллегии В.Д. Богатыревым*

*3-я Российская конференция
с международным участием
«Технические и программные средства
систем управления, контроля и измерения» (УКИ-12)*

*ИПУ РАН, Москва, 16-19 апреля 2012 г.
<http://стт.ipu.ru>*

УДК 519

ББК 22.183.43 + 65в641

СВОЙСТВА ОПТИМАЛЬНОГО ДЕЛЕГИРОВАНИЯ УПРАВЛЕНИЯ В ОРГАНИЗАЦИИ¹

Мишин С. П.²

(Учреждение Российской академии наук
Институт проблем управления РАН, Москва)

Рассмотрена задача минимизации функции затрат выбором оптимального делегирования – распределения управленческого воздействия, оказываемого на производственную подсистему, между менеджерами управленческой подсистемы, с учетом их сложной взаимосвязи, заключающейся в дублировании управлений друг друга. Математически данная постановка эквивалентна задаче оптимального распределения заданного объема выпуска между взаимосвязанными предприятиями с многопродуктовыми функциями затрат. Исследованы общие свойства оптимального решения, позволяющие находить оптимальное делегирование для различных классов функций затрат. Показано, что ключевые свойства функции затрат отдельных менеджеров сохраняются и для функции затрат всей управленческой подсистемы при оптимальном делегировании.

Ключевые слова: оптимальное делегирование управления, менеджер, затраты, организация.

1. Введение

Иерархическая организация (основанная на асимметричном отношении элементов «начальник – подчиненный») характерна для

¹ Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (проект 10-07-00129).

² Сергей Петрович Мишин, кандидат физико-математических наук (smishin78@mail.ru).

самых различных сфер практической деятельности людей. Большая часть современной экономики состоит из иерархических организаций, которые позволяют повысить эффективность производства за счет разделения труда, порождая в то же время организационные издержки сложной системы управления с иерархической структурой (иерархией), состоящей из менеджеров, которые управляют производственной системой. Выделим производственную и управленческую подсистемы, изображенные на рис. 1.

Исполнители производственной подсистемы W (обозначаемые w_1, \dots, w_n) могут влиять на управленческую подсистему сколь угодно сложным образом. Например, затраты менеджеров могут зависеть от количества исполнителей n , от характеристик отдельных исполнителей, от взаимодействия в группах (множествах) исполнителей размера 2, 3 и т.д. В соответствии с системным подходом на рис. 1 влияние управленческой подсистемы на производственную ограничивается p видами управления – управленческого воздействия в объемах $x = (x^1, \dots, x^p)^T \geq 0$. В вектор x могут входить, например, такие виды управления, как обработка и передача информации, принятие решений, планирование и контроль, прием и увольнение, обучение и разъяснение, решение проблем и т.п.

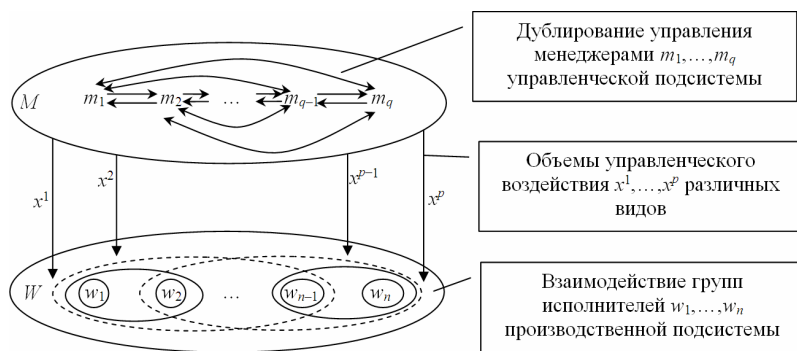


Рис. 1. Производственная и управленческая подсистемы

Подход с выделением аддитивных объемных характеристик управленческого воздействия является достаточно общим. Увеличивая размерность p , можно описать и прочие характеристики управленческого воздействия, например, качество (на-

сколько эффективно управленческое воздействие) и время (когда оказывается управленческое воздействие), а также конкретизировать, на каких исполнителей или их группы это воздействие оказывается.

Критерий эффективности всей организации можно представить как функцию выручки за вычетом функции затрат производственной подсистемы W и функции затрат управленческой подсистемы M , причем первые две функции зависят только от управления x , но не от деталей организации подсистемы M (поскольку ее влияние на подсистему W ограничивается компонентами вектора x):

$$(1) P(x, W, M) = S(x, W) - C_{prod}(x, W) - C_{manag}(x, W, M).$$

В результате обоснована общность следующего подхода. Фиксируется некоторая производственная система и *вектор управления x* (который называется *вектором управления производственной подсистемой*). На первом этапе решается задача об оптимальной управленческой подсистеме M (включая иерархию, полномочия менеджеров и т.п.), оказывающей заданное управленческое воздействие x с минимальными затратами. На втором этапе оптимизируется производственная подсистема и вектор управления x с учетом функции минимальных затрат на его реализацию управленческой подсистемой. Наиболее сложным является первый этап – *задача оптимизации управленческой подсистемы M , оказывающей заданное управленческое воздействие x с минимальными затратами*, называемая ниже *задачей об оптимальной организации*:

$$(2) C_{manag}^*(x) = \min_{M \in \mathbf{M}} C_{manag}(x, M),$$

где через \mathbf{M} обозначено *множество допустимых организаций*, а подсистема $M^*(x) \in \mathbf{M}$, доставляющая минимум (2), *оптимальной организацией, реализующей x* (фиксированный аргумент W для краткости опускается). На рис. 1 задачу (2) можно проиллюстрировать как построение минимально затратной управленческой подсистемы M внутри верхнего эллипса при условии реализации управления x .

Задача об оптимальной организации (2) может включать в себя оптимизацию разнообразных характеристик управленче-

ской подсистемы: количества и состава менеджеров, их взаимной подчиненности (вида иерархии), правил взаимодействия, стимулирования и т.п. В настоящей работе рассматривается лишь одна¹ из подзадач задачи об оптимальной организации (2): *задача об оптимальном делегировании в организации*: каждый из q менеджеров вносит свой вклад в управление x , то есть менеджерам m_1, \dots, m_q делегированы такие векторы управления y_1, \dots, y_q , что выполнено балансовое условие $y_1 + \dots + y_q = x$ и каждый менеджер оказывает неотрицательное управленческое воздействие ($y_1, \dots, y_q \geq 0$, то есть неотрицательны все компоненты векторов). Подобное аддитивное распределение управления между менеджерами возможно, поскольку рассматриваются лишь объемные виды управления (как указано выше, это не ограничивает общности). Для краткости распределение вектора x на сумму векторов y_1, \dots, y_q называется *делегированием*, а конкретный вектор y_i – *делегированием менеджера m_i* .

В настоящей работе классические методы непрерывной оптимизации применены для исследования задачи об оптимальном делегировании. В разделе 2 определен вид функции затрат управленческой подсистемы. В разделе 3 поставлена задача об оптимальном делегировании и обосновано, что результаты настоящей работы позволяют исследовать дальнейшие этапы оптимизации организации (иерархии, типов и состава менеджеров и т.п.), опираясь на свойства функции минимальных затрат, доказанные в последующих разделах в зависимости от свойств функций затрат отдельных менеджеров. В заключении сформулированы основные математические результаты работы.

Кроме того, в разделах 2 и 3 отмечено, что за последнее десятилетие произошел резкий рост количества работ, в которых проводится идентификация многопродуктовых функций затрат предприятия на выпуск набора продуктов в заданных объемах. Показано, что математический аппарат, предложенный в на-

¹ *Общее исследование задачи об оптимальной организации проведено в [3]. Там же изложена одна из частных моделей совместной оптимизации производственной и управленческой подсистем.*

стоящей работе, может использоваться для поиска оптимального распределения общего объема выпуска между предприятиями (отрасли, холдинга и т.п.), которые обмениваются продуктами производства по обобщенной модели многопродуктового межотраслевого баланса и описываются нелинейными многопродуктовыми функциями затрат.

2. Функция затрат управленческой подсистемы

Если управленческая подсистема состоит из единственного менеджера m , оказывающего весь объем управленческого воздействия x , то функцию затрат менеджера $C_m(x) = C_{manag}(x, M)$ (равную функции затрат управленческой подсистемы) можно моделировать по аналогии с затратами любого предприятия, выпускающего заданный объем «продукции» x (под «продукцией» менеджера понимается оказанное им управленческое воздействие). Для этого используются *многопродуктовые функции затрат*, которые начали применяться в эконометрических исследованиях в 1970-х для учета разнообразия производимых продуктов, количества заказчиков, размера территории и прочих аддитивных характеристик, влияющих на затраты.

При построении многопродуктовой функции затрат главной проблемой является недостаточность данных для достоверной статистической оценки большого количества параметров. Несмотря на разработку специальных методов «обхода» этой проблемы (например, Positive Mathematical Programming, см. [11]), до последнего десятилетия в основном удавалось статистически идентифицировать только иллюстративные двух-трехпродуктовые функции затрат. Ситуация кардинально изменилась в 2000-е годы в связи с качественным скачком уровня информатизации экономики, позволяющим накапливать достаточное количество статистических данных. Достаточно сказать, что поисковая система «Google» по запросу «multi-output cost function» выдает несколько миллионов ссылок (на 2011 год). Не претендуя на сколько-нибудь полный обзор работ по этой тематике, отметим, что в любой отрасли, характеризующейся возможностью выбора множества выпускаемых продуктов, можно

без труда найти прикладные исследования, в которых идентифицированы многопродуктовые функции затрат (например, коммунальное хозяйство [9], сельское хозяйство – [9], [8]; телекоммуникации и связь [14], [6]; транспорт – [15]; образование – [7]).

В настоящей работе предполагается лишь непрерывная дифференцируемость функции затрат $C_m(x)$ (существование и непрерывность функций предельных затрат). Конкретный вид функции не фиксируется, так же, как и свойства выпуклости, однородности, неотрицательности и монотонности, существенность которых явно оговаривается ниже.

Увеличение количества менеджеров приводит к усложнению управленческой подсистемы. Чем она сложнее, тем большую долю управления, реализуемого одними менеджерами, вынуждены дублировать другие менеджеры. В настоящей работе рассматривается модель с линейным дублированием, что не ограничивает общности, поскольку рассматриваются широкие классы функций затрат, возможно увеличение размерности управления и замена переменных.¹

Модель с линейным дублированием базируется на понятии коэффициентов дублирования. При единственном виде управления $p = 1$ коэффициент $d_{i,j}$ показывает, какую долю управления y_j , делегированного менеджеру m_j , вынужден повторять («дублировать») менеджер m_i , то есть задана матрица дублирования $D = (d_{i,j})$ размера $q \times q$. При нескольких видах управления $p > 1$ коэффициент $d_{i,j}^{k,l}$ показывает, какую долю управления вида l менеджера m_j дублирует менеджер m_i при управлении вида k . Соответственно, пара менеджеров m_i, m_j характеризуется матрицей $D_{i,j} = (d_{i,j}^{k,l})$ размера $p \times p$, а вся матрица дублирования $D = (D_{i,j})$ состоит из q^2 таких блоков, то есть D имеет размер $pq \times pq$ и состоит из элементов $d_{i,j}^{k,l}$, в которых нижние индексы

¹ В [3] доказано, что в рамках рассматриваемой модели можно описать и нелинейное дублирование, и любую функцию затрат, зависящую от составного вектора делегирования y .

определяют положение блока (i – во вертикали, j – по горизонтали), а верхние индексы определяют положение элемента в блоке (k – по вертикали, l – по горизонтали).

В результате вместо функции затрат $C_m(x)$ единственного менеджера рассматривается функция затрат управленческой подсистемы M : $C_{manag}(x, D, y) = C_{m_1}(\tilde{y}_1) + \dots + C_{m_q}(\tilde{y}_q)$, равная сумме функций затрат q менеджеров, зависящих от векторов нагрузки менеджеров $\tilde{y}_1, \dots, \tilde{y}_q$ с учетом дублирования менеджерами делегирований друг друга. Нагрузка вычисляется следующим образом: $\tilde{y} = Dy$, где вектор делегирования y составлен из векторов делегирования менеджеров y_1, \dots, y_q , а вектор \tilde{y} составлен из векторов нагрузки менеджеров $\tilde{y}_1, \dots, \tilde{y}_q$). Функция $C_{manag}(x, D, y)$ зависит от своих аргументов неявным образом, с учетом равенств $\tilde{y} = Dy$, $y_1 + \dots + y_q = x$, подобная запись используется ниже для удобства и единообразия изложения.

Обратная матрица D^{-1} , называемая матрицей продуктивности, представляет собой многопродуктовое обобщение матрицы межотраслевого баланса в леонтьевской модели. Отрицательные элементы D^{-1} , стоящие вне главной диагонали¹, характеризуют объемы различных видов управленческого воздействия различных менеджеров, которые дополнительно требуются для того, чтобы данный менеджер оказал единицу объема управленческого воздействия данного вида. Эта дополнительная нагрузка соответствует организационным издержкам сложной подсистемы управления, например, необходимости контроля подчиненных и отчетов перед начальниками. При $p = 1$ сумма \hat{d}_i элементов i -го столбца D^{-1} характеризует продуктивность менеджера m_i , то есть вклад единицы нагрузки в итоговое «чистое» управление x , оказываемое на производственную подсистему: $\hat{d}_1 \tilde{y}_1 + \dots + \hat{d}_q \tilde{y}_q = x$. В многомерном случае $\hat{D}_1 \tilde{y}_1 + \dots + \hat{D}_q \tilde{y}_q = x$,

¹ В рассматриваемой модели, в отличие от леонтьевской, внедиагональные элементы D^{-1} могут быть и положительными.

где $\hat{D}_j = D_{1,j}^{-1} + \dots + D_{q,j}^{-1}$ – матрица продуктивности менеджера m_j (сумма блоков D^{-1} размера $p \times p$).

Если под менеджерами (элементами подсистемы управления) понимать произвольные предприятия, каждое из которых описывается многопродуктовой функцией затрат, то матрица дублирования определяет обмен предприятий продуктами производства по обобщенной модели межотраслевого баланса (сколько сырья определенного вида, производимого одним предприятием, нужно для производства единицы продукции другого). Векторы нагрузки $\tilde{y}_1, \dots, \tilde{y}_q$ соответствуют суммарному выпуску предприятий, а векторы y_1, \dots, y_q – «чистому» выпуску, остающемуся после внутреннего потребления предприятиями продукции друг друга. Если при этом имеются r сугубо «внутренних» продуктов видов $p + 1, \dots, p + r$ (виды управления, не требующиеся производственной подсистемы), то их легко учесть в рамках модели, увеличивая размерность вектора x на r и полагая $x_{p+1} = 0, \dots, x_{p+r} = 0$. В результате, рассматриваемая функция $C_{\text{manag}}^*(x, D, y) = C_{m_1}(\tilde{y}_1) + \dots + C_{m_q}(\tilde{y}_q)$ определяет затраты всех предприятий (отрасли, холдинга и т.п.) с учетом сложного обмена продуктами производства по обобщенной модели многопродуктового межотраслевого баланса и сложных нелинейных многопродуктовых функций затрат.

3. Задача об оптимальном делегировании

С учетом вида функции затрат управленческой подсистемы, определенной в предыдущем разделе, задача об оптимальном делегировании имеет вид:

$$(3) \quad C_{\text{manag}}^*(x, D) = \min_{y_1 \geq 0, \dots, y_q \geq 0, y_1 + \dots + y_q = x} C_{\text{manag}}(x, D, y) = \min_{y_1 \geq 0, \dots, y_q \geq 0, y_1 + \dots + y_q = x} \sum_{i=1}^q C_{m_i}(D_i y).$$

Результатом решения задачи (3) является функция $C_{\text{manag}}^*(x, D)$, характеризующая минимальные затраты на оказание управленческого воздействия x при условии его оптимального распределения между менеджерами с функциями затрат

$C_{m_1}(\cdot), \dots, C_{m_r}(\cdot)$, входящими в управленческую подсистему с матрицей дублирования D . Всевозможные характеристики управленческой подсистемы и отдельных менеджеров (типы менеджеров, иерархия менеджеров, порядок взаимодействия менеджеров, стимулирование менеджеров и т.п.) могут гибко моделироваться с помощью матрицы дублирования и параметров функции затрат.¹ Для оптимизации вышеупомянутых характеристик организации необходима дальнейшая минимизация функции $C_{\text{manag}}^*(x, D)$, например, по элементам матрицы D . Для подобной оптимизации необходимо опираться на общие свойства функции $C_{\text{manag}}^*(x, D)$ (основным из которых является непрерывная дифференцируемость, а дополнительными – выпуклость, однородность и монотонность), проверке которых и посвящена настоящая работа. Кроме технических свойств, необходимых для оптимизации, результаты настоящей работы устанавливают взаимосвязь между свойствами функций затрат отдельных менеджеров и всей управленческой подсистемы при условии оптимального делегирования.

В терминах оптимального распределения объемов производства (см. раздел 2), результаты работы позволяют установить взаимосвязь между функциями затрат отдельных предприятий (взаимодействующих по модели многопродуктового межотраслевого баланса) и функцией затрат всей отрасли (объединения, холдинга и т.п.). Это дает возможность далее оптимизировать структуру отрасли (определяющую матрицу дублирования D и параметры функции затрат), минимизируя функцию $C_{\text{manag}}^*(x, D)$ с учетом ее свойств, доказанных в настоящей работе.

4. Общие свойства оптимального делегирования

Задача минимизации затрат (3) решается на ограниченном ($0 \leq y_i^k \leq x^k$) и замкнутом множестве допустимых делегирований. Поэтому для того, чтобы минимум (3) достигался, достаточно

¹ См. соответствующие модели в [3].

непрерывности функций затрат $C_{m_1}(\cdot), \dots, C_{m_q}(\cdot)$ (теорема Вейерштрасса, [5]). Следовательно, задача об оптимальном делегировании имеет решение, поскольку рассматриваются непрерывно-дифференцируемые функции затрат менеджеров, которые очевидно непрерывны.

С учетом покоординатных ограничений $x^k - y_1^k - \dots - y_q^k = 0$, $-y_i^k \leq 0$ запишем лагранжиан задачи (3):

$$(4) \quad L = I_0 \sum_{i=1}^q C_{m_i}(D_i y) + \sum_{k=1}^p I^k (x^k - y_1^k - \dots - y_q^k) + \sum_{k=1}^p \sum_{i=1}^q q_i^k (-y_i^k),$$

где коэффициенты Лагранжа $I_0 \geq 0$, $I = (I^1, \dots, I^p)^T \in R^p$, $q_i^k \geq 0$, $q_i = (q_i^1, \dots, q_i^p)^T \in R_+^p$. Вектор частных производных функции затрат $C_{m_i}(D_i y)$ менеджера m_i по делегированию менеджера m_j :

$$(5) \quad (\partial C_{m_i}(D_i y) / \partial y_j^1, \dots, \partial C_{m_i}(D_i y) / \partial y_j^p)^T = D_{i,j}^T \nabla C_{m_i}(\tilde{y}_i),$$

где в правой части равенства стоит вектор частных производных по переменным $\tilde{y}_i^1, \dots, \tilde{y}_i^p$, то есть градиент функции затрат $C_{m_i}(\cdot)$ менеджера m_i при нагрузке $\tilde{y}_i = D_i y$.¹ pq условий равенства нулю частных производных лагранжиана (4) примут вид:

$$(6) \quad I_0 \sum_{i=1, q}^- D_{i,j}^T \nabla C_{m_i}(\tilde{y}_i) = I + q_j \text{ для всех } j = \overline{1, q}.$$

Обозначим через $g = (\nabla C_{m_1}(\tilde{y}_1)^T, \dots, \nabla C_{m_q}(\tilde{y}_q)^T)^T$ вектор-столбец, составленный из стоящих «друг под другом» векторов градиентов; через $\bar{I} = (I^1, \dots, I^p)^T$ – вектор, составленный из q повторяющихся векторов λ , $q = (q_1^T, \dots, q_q^T)^T$ – вектор, составленный из компонент векторов $\theta_1, \dots, \theta_q$. Условия равенства нулю производных лагранжиана (6) в компактной форме примут вид:

$$(7) \quad I_0 D^T g = \bar{I} + q.$$

¹ Формула (5) преобразования градиента при линейной замене переменных широко известна (см., например, [13]): $\partial C_{m_i}(D_i y) / \partial y_j^l = \sum_{k=1, p}^- d_{i,j}^{k,l} \partial C_{m_i}(D_i y) / \partial x^k = (d_{i,j}^{l, \cdot} \nabla C_{m_i}(\tilde{y}_i))$, где $d_{i,j}^{l, \cdot}$ – l -й столбец матрицы $D_{i,j}$. Таким образом, левая часть (5) есть скалярное произведение столбцов $D_{i,j}$ на $\nabla C_{m_i}(\tilde{y}_i)$, то есть $D_{i,j}^T \nabla C_{m_i}(\tilde{y}_i)$.

Делегирование допустимо, если $y_1 \geq 0, \dots, y_q \geq 0$ и выполнены балансовые условия:

$$(8) \quad y_1 + \dots + y_q = x.$$

Кроме того, выпишем условия дополняющей нежесткости, связывающие компоненту делегирования y_i^k с соответствующим ей коэффициентом Лагранжа $q_i^k \geq 0$:

$$(9) \quad q_i^k y_i^k = 0 \text{ для всех } i = \overline{1, q}, k = \overline{1, p}.$$

Критической точкой назовем делегирование и коэффициенты Лагранжа, удовлетворяющие условию (7) равенства нулю производных лагранжиана (или в развернутой форме – условию (6)), условию допустимости делегирования (8) и условию дополняющей нежесткости (9). Достаточно искать оптимальное делегирование среди критических точек, что подтверждается следующей классической теоремой, справедливой для непрерывно дифференцируемых целевых функций.

Теорема 1 (Каруша-Куна-Таккера,¹ см., например, [1]). Если y – оптимальное правило делегирования, доставляющее минимум затрат в задаче (3), то найдутся такие коэффициенты Лагранжа $l_0 \geq 0$, $l \in R^p$, $q_i \in R_+^p$, не все равные нулю, которые совместно с делегированием y являются критической точкой (выполнены условия (7), (8), (9)).

Итак, при поиске оптимального делегирования можно ограничиться решениями системы уравнений (7), (8), (9) при ограничении неотрицательности y и θ . Решения со строго положительными компонентами вектора y будем называть *внутренними*, прочие решения – *краевыми*. Из условий (9) следует, что для любого внутреннего решения $\theta = 0$. Для краевого решения с $y_i^k = 0$ коэффициент Лагранжа q_i^k может быть положительным, в этом случае ограничение-неравенство $y_i^k \geq 0$ называется *напряженным*. Для внутреннего решения все ограничения-неравенства будут ненапряженными.

¹ Иногда эта теорема еще называется теоремой Каруша-Джона.

Коэффициент λ_0 может быть равен нулю или единице (при прочих $\lambda_0 > 0$ коэффициенты Лагранжа в (4) можно разделить на λ_0 и перейти к новым переменным, что, очевидно, не повлияет на выполнение условий (7), (8), (9)). В случае $\lambda_0 = 1$ имеет место так называемый *регулярный случай*, при $\lambda_0 = 0$ вырожденный. Линейность ограничений является достаточным условием регулярности оптимизационной задачи, то есть в теореме 1 можно считать $\lambda_0 = 1$. В рассматриваемой задаче этот факт легко доказать непосредственно.

Следствие 1. *В условиях теоремы 1 достаточно рассмотреть только регулярный случай $I_0 = 1$, то есть найдутся такие коэффициенты Лагранжа $I \in R^p$, $q_i \in R_+^p$, которые совместно с делегированием u являются критической точкой.*

Доказательство. В вырожденном случае $\lambda_0 = 0$ из (6) имеем $q_1^k = \dots = q_q^k = -I^k$ для всех $k = \overline{1, p}$. При $I_k \neq 0$ условия дополняющей нежесткости (9) дают $y_1^k = \dots = y_q^k = 0$, что с учетом условия баланса даст $x^k = 0$. Поэтому в случае $x^k > 0$ имеем $q_1^k = \dots = q_q^k = -I^k = 0$. Таким образом, при $x > 0$ (требуется реализация положительного объема управленческого воздействия любого вида) из $\lambda_0 = 0$ следует равенство нулю всех коэффициентов Лагранжа, что противоречит условиям теоремы 1. Поэтому для случая $x > 0$ следствие доказано.

В общем случае считаем $x^1 > 0, \dots, x^{p_0} > 0, x^{p_0+1} = 0, \dots, x^p = 0$, что не ограничивает общности, поскольку виды управления можно перенумеровать так, чтобы нулевые компоненты вектора x были последними. Тогда $y_1^l = \dots = y_q^l = 0$ для всех $l > p_0$. Рассмотрим задачу (3) с оставшимися переменными $y_i^k \geq 0, \dots, y_q^k \geq 0$ для всех $k \leq p_0$, то есть рассмотрим p_0 -мерные векторы неизвестных $y_1 + \dots + y_q = (x^1, \dots, x^{p_0})^T$. Рассмотрим векторное уравнение (6):

$$(*) \quad I_0 \sum_{i=1}^q D_{i,j}^T \nabla C_{m_i}(\tilde{y}_i) = I + q_j \text{ для всех } j = \overline{1, p}.$$

В (*) должны выполняться только первые p_0 скалярных уравнений, поскольку в (5) дифференцирование ведется только по $y_j^1, \dots, y_j^{p_0}$. При этом в (6) «задействованы» только первые p_0 строк матрицы $D_{i,j}^T$, то есть только первые p_0 столбцов матрицы $D_{i,j}$ (остальные столбцы не играют роли в функции затрат, поскольку умножаются на нулевое делегирование).

Для рассмотренной задачи запишем любое оптимальное решение $y_1^k \geq 0, \dots, y_q^k \geq 0, k \leq p_0, y_1^k + \dots + y_q^k = x^k$ обеспечивающее минимум затрат. Для него также верна теорема 1, причем с учетом $x^1 > 0, \dots, x^{p_0} > 0$ по доказанному выше $\lambda_0 = 1$ и найдутся такие $I^1, \dots, I^{p_0}, q_i^1 \geq 0, \dots, q_i^{p_0} \geq 0$, что при каждом j выполнены p_0 первых уравнений (*) и $q_i^k y_i^k = 0$ для всех $i = \overline{1, q}, k = \overline{1, p_0}$. Продолжим это решение. Для этого через $d_{i,j}^l$ обозначим l -й столбец матрицы $D_{i,j}$ (или l -ю строку $D_{i,j}^T$), $l = \overline{p_0 + 1, p}$. l -е уравнение (*)

имеет вид
$$\sum_{i=1}^q (d_{i,j}^l, \nabla C_{m_i}(\tilde{y}_i)) = I^l + q_j^l.$$
 Положим

$I^l = \min_{1 \leq j \leq q} \sum_{i=1}^q (d_{i,j}^l, \nabla C_{m_i}(\tilde{y}_i)), q_j^l = \sum_{i=1}^q (d_{i,j}^l, \nabla C_{m_i}(\tilde{y}_i)) - I^l$, что при каждом $j = \overline{1, q}$ обеспечит выполнение всех p уравнений (*) и неравенство $q_j^l \geq 0$. В силу $y_1^l = \dots = y_q^l = 0$ выполнено $q_i^l y_i^l = 0$, а также условие баланса $y_1^l + \dots + y_q^l = x^l = 0$.

Найденными решениями вида $y_i = (y_i^1, \dots, y_i^{p_0}, 0, \dots, 0)^T, i = \overline{1, q}$ исчерпываются все оптимальные решения исходной задачи (3). Для любого такого решения при $\lambda_0 = 1$ найдены коэффициенты Лагранжа $I \in R^p, q_i \in R_+^p$, обеспечивающие выполнение условий (6) и условий дополняющей нежесткости $q_i^k y_i^k = 0$ для всех $i = \overline{1, q}, k = \overline{1, p}$. Следствие доказано. ■¹

¹ Символ «■» здесь и далее обозначает окончание доказательства.

Ниже без дополнительных оговорок будем считать, что $I_0 = 1$. В результате для определения pq неизвестных y_i^k , p неизвестных I^1, \dots, I^p и pq неизвестных q_i^k теорема 1 и следствие 1 дают соответствующее количество уравнений: pq условий равенства нулю производных лагранжиана (7), p балансовых ограничений (8) и pq условий дополняющей нежесткости (9). Совпадение числа уравнений и числа неизвестных в общем случае не гарантирует единственности делегирования. В случае нескольких решений требуется еще минимизация затрат по всем делегированиям, для которых нашлась критическая точка.

Докажем следствие, гарантирующее при $x^k > 0$ единственность коэффициентов Лагранжа, определяемых по заданному оптимальному делегированию.

Следствие 2. Если y – оптимальное правило делегирования, доставляющее минимум затрат в задаче (3), то при $x^k > 0$ однозначно определяются коэффициенты Лагранжа $I^k = \min_j \sum_i (d_{i,j}^{:,k}, \nabla C_{m_i}(\tilde{y}_i))$,¹ $q_j^k = \sum_i (d_{i,j}^{:,k}, \nabla C_{m_i}(\tilde{y}_i)) - I^k$, которые совместно с делегированием y являются критической точкой. При $x^k = 0$ можно выбрать любое $I^k \leq \min_j \sum_i (d_{i,j}^{:,k}, \nabla C_{m_i}(\tilde{y}_i))$, после чего q_j^k определяется аналогичным образом.

Доказательство. Из уравнения (6) (детализирующего уравнения критической точки (7)), при $q_j^k = 0$ получаем $I^k = \sum_i (d_{i,j}^{:,k}, \nabla C_{m_i}(\tilde{y}_i))$, а при $q_j^k > 0$ получаем $I^k < \sum_i (d_{i,j}^{:,k}, \nabla C_{m_i}(\tilde{y}_i))$. При $x^k > 0$ найдется j , для которого $y_j^k > 0$, поэтому $q_j^k = 0$ и коэффициент Лагранжа $I^k = \min_j \sum_i (d_{i,j}^{:,k}, \nabla C_{m_i}(\tilde{y}_i))$ однозначно определяется по оптимальному делегированию y . При $x^k = 0$ выполнено $I^k \leq \min_j \sum_i (d_{i,j}^{:,k}, \nabla C_{m_i}(\tilde{y}_i))$ (иначе нарушается условие $q \geq 0$), однако $y_1^k = \dots = y_q^k = 0$, поэтому λ^k можно сделать

¹ Напомним, что нагрузка вычисляется по формуле $\tilde{y} = Dy$, через $d_{i,j}^{:,k}$ обозначается k -й столбец матрицы $D_{i,j}$.

произвольно малым, соответственно увеличивая q_1^k, \dots, q_q^k (условия дополняющей нежесткости (9) при этом удовлетворены). Поэтому можно выбрать любое $I^k \leq \min_j \sum_i (d_{i,j}^k, \nabla C_{m_i}(\tilde{y}_i))$. ■

5. Непрерывность и дифференцируемость

В результате решения задачи об оптимальном делегировании (3) вычисляется функция оптимальных затрат $C_{manag}^*(x, D)$. Для дальнейшей оптимизации управленческой подсистемы по типам менеджеров, иерархии, порядку взаимодействия менеджеров, их стимулированию и т.п. (см. раздел 3) необходимо оптимизировать $C_{manag}^*(x, D)$ по параметрам x, D . Для решения этой задачи методами непрерывной оптимизации необходимо проверить непрерывность и непрерывную дифференцируемость функции $C_{manag}^*(x, D)$, чему и посвящен настоящий раздел.

Обозначим через $y(x, D)$ некоторое оптимальное решение задачи (3), принадлежащее множеству всех оптимальных решений: $y(x, D) \in Y^*(x, D)$. Функция $C_{manag}^*(x, D) = C_{manag}(x, D, y(x, D))$ не зависит от того, какое именно оптимальное решение $y(x, D) \in Y^*(x, D)$ выбрано при каждом значении параметров x и D . По следствию 1 из теоремы 1 найдутся коэффициенты $\lambda(x, D), \theta(x, D)$, которые совместно с $y(x, D)$ обеспечивают выполнение условий (7), (8), (9) (являются критической точкой).

Для исследования вопроса о непрерывной дифференцируемости $C_{manag}^*(x, D)$ по x и D нельзя напрямую применять известную из экономической литературы теорему об огибающей, поскольку она требует не только единственности, но и непрерывной дифференцируемости решения $y(x, D)$. В общем случае множество $Y^*(x, D)$ может содержать более одного решения, например, при нестрогих выпуклых функциях затрат. Кроме того, даже в случае строго выпуклых функций затрат, гарантирующего единственное решение (см. ниже), функция $y(x, D)$ может быть недифференцируемой (производная $y(x, D)$ по коэффициенту дублирования может терпеть разрыв при смене внутренне-

го решения краевым, хотя функция $C_{\text{manag}}^*(x, D)$ остается непрерывно дифференцируемой).¹

Поэтому при доказательстве результатов настоящего раздела использована более общая версия теоремы об огибающей, доказанная сравнительно недавно в статье [12].²

Вопрос непрерывной дифференцируемости по коэффициентам дублирования решается следующей леммой.

Лемма 1. *Левосторонняя и правосторонняя производные функции $C_{\text{manag}}^*(x, D)$ по $d_{i,j}^{k,l}$ существуют, совпадают почти всюду и равны соответственно минимуму и максимуму $y_j' \partial C_{m_i}(D_i y) / \partial x^k$ по всем $y(x, D) \in Y^*(x, D)$. Если левосторонняя и правосторонняя производные совпадают (в частности, если функции затрат менеджеров выпуклы или $y(x, D)$ единственно), то функция $C_{\text{manag}}^*(x, D)$ непрерывно дифференцируема по $d_{i,j}^{k,l}$ и $\partial C_{\text{manag}}^*(x, D) / \partial d_{i,j}^{k,l} = y_j' \partial C_{m_i}(D_i y) / \partial x^k$.*

Доказательство. Задача об оптимальном делегировании (3) представляет собой задачу минимизации на непустом выпуклом компакте $(y_1 + \dots + y_q = x, y_1 \geq 0, \dots, y_q \geq 0)$ целевой функции $C_{\text{manag}}(x, D, y) = \sum_{i=1,q} C_{m_i}(D_i y)$. Очевидно, что эта функция непрерывно дифференцируема по любому элементу матрицы D , поскольку функция каждого менеджера непрерывно дифференцируема и зависит от линейной комбинации элементов D . В статье [12] доказано, что непрерывная дифференцируемость целевой функции по параметрам D и по переменным y , принадлежащим непустому выпуклому компакту, достаточна для дифференцируемости функции $C_{\text{manag}}^*(x, D)$ по $d_{i,j}^{k,l}$ почти всюду (для существования производной во всех точках, за исключением, может быть, множества точек нулевой меры). Кроме того, во всех без исключения точках существует левосторонняя и право-

¹ Подробнее см. [3].

² Там же имеются ссылки на более ранние математические работы, в которых доказаны схожие теоремы.

сторонняя производные, равные соответственно минимуму и максимуму по всем $y(x, D) \in Y^*(x, D)$ производной целевой функции $\partial C_{\text{manag}}(x, D, y) / \partial d_{i,j}^{k,l}$. В выражении $C_{\text{manag}}(x, D, y) = \sum_{i=l,q} C_{m_i}(D_i, y)$ только k -й аргумент функции $C_{m_i}(D_i, y)$ зависит от $d_{i,j}^{k,l}$, поэтому по правилу вычисления сложной функции $\partial C_{\text{manag}}(x, D, y) / \partial d_{i,j}^{k,l} = y_j^l \partial C_{m_i}(D_i, y) / \partial x^k$, что и доказывает первую часть леммы.

Если левосторонняя и правосторонняя производные совпадают всюду (значение $y_j^l \partial C_{m_i}(D_i, y) / \partial x^k$ одинаково при всех $y(x, D) \in Y^*(x, D)$), то функция $C_{\text{manag}}^*(x, D)$ дифференцируема по элементам матрицы дублирования $d_{i,j}^{k,l}$ – производная $\partial C_{\text{manag}}^*(x, D) / \partial d_{i,j}^{k,l} = y_j^l \partial C_{m_i}(D_i, y) / \partial x^k$ существует во всех точках. Докажем непрерывность производной по всем параметрам (в [12] показана непрерывность только по параметру $d_{i,j}^{k,l}$, по которому ведется дифференцирование). По теореме Бержа (см., например, [2]) многозначное отображение $Y^*(x, D)$ полунепрерывно сверху: для любого $\epsilon > 0$ найдется достаточно малая окрестность точки x, D , такая, что для любой точки \hat{x}, \hat{D} из этой окрестности множество $Y^*(\hat{x}, \hat{D})$ будет вложено в ϵ -расширение множества $Y^*(x, D)$ (то есть во множество точек, лежащих на расстоянии не более ϵ от множества $Y^*(x, D)$ в смысле некоторой метрики). Поэтому найдутся точки $y(x, D) \in Y^*(x, D)$ и $y(\hat{x}, \hat{D}) \in Y^*(\hat{x}, \hat{D})$, расстояние между которыми не превышает ϵ . В силу непрерывности $\partial C_{m_i}(\cdot) / \partial x^k$ при достаточно малом ϵ значения $y_j^l(x, D) \partial C_{m_i}(D_i, y(x, D)) / \partial x^k$ и $y_j^l(\hat{x}, \hat{D}) \partial C_{m_i}(\hat{D}_i, y(\hat{x}, \hat{D})) / \partial x^k$ произвольно мало отличаются друг от друга, что и доказывает непрерывность производной $\partial C_{\text{manag}}^*(x, D) / \partial d_{i,j}^{k,l}$.

Если решение задачи об оптимальном делегировании $y(x, D)$ единственно, то $Y^*(x, D)$ состоит из одной точки, левосторонняя и правосторонняя производные совпадают, поскольку минимум $y_j^l \partial C_{m_i}(D_i, y) / \partial x^k$ по $y(x, D) \in Y^*(x, D)$ равен максимуму

му. Однако и в случае множественности решений минимум и максимум могут совпадать. Например, при выпуклых функциях затрат менеджеров целевая функция задачи об оптимальном делегировании $C_{\text{manag}}(x, D, y) = \sum_{i=1, q} C_{m_i}(D_i, y)$ также будет выпуклой (каждое слагаемое выпукло, поскольку зависит от линейной комбинации переменных, сумма выпуклых функций выпукла, см., например, [13]). Для задачи минимизации выпуклой функции на компакте в работе [12] доказано,¹ что функция оптимальных затрат $C_{\text{manag}}^*(x, D)$ дифференцируема по $d_{i,j}^{k,l}$, поэтому левосторонняя и правосторонняя производные совпадают (хотя решение $y(x, D) \in Y^*(x, D)$ может быть множественным при нестрогой выпуклости). ■

Следующая лемма 2 использует результат леммы 1 для доказательства непрерывности функции $C_{\text{manag}}^*(x, D)$ по элементам матрицы дублирования D . Непрерывность по x доказывается непосредственно.

Лемма 2. *Функция $C_{\text{manag}}^*(x, D)$ непрерывна по параметрам x и D .*

Доказательство. Согласно лемме 1 существуют левосторонняя и правосторонняя производные функции $C_{\text{manag}}^*(x, D)$ по $d_{i,j}^{k,l}$, равные соответственно минимуму и максимуму $y_j^l \partial C_{m_i}(D_i, y) / \partial x^k$ по всем $y(x, D) \in Y^*(x, D)$. Эти односторонние производные конечны, поскольку частные производные $\partial C_{m_i}(\cdot) / \partial x^k$ ограничены, как и любая непрерывная функция на компакте $y_1 + \dots + y_q = x, y_1 \geq 0, \dots, y_q \geq 0$, на котором решается задача оптимизации (3). Если бы функция $C_{\text{manag}}^*(x, D)$ не была непрерывной слева (справа) в точке $d_{i,j}^{k,l}$, то левосторонняя (правосторонняя) частная производная не существовала бы или

¹ Точнее, это утверждение доказано в [12] для задачи максимизации вогнутой функции, эквивалентной задаче минимизации выпуклой функции (после умножения на -1).

была бы бесконечной. Отсюда следует непрерывность функции $C_{manag}^*(x, D)$ по элементам матрицы дублирования D .

Докажем непрерывность $C_{manag}^*(x, D)$ по x непосредственно.

Предположим противное: найдется такое $\epsilon > 0$, что для любого $d > 0$ существует \hat{x}_d , лежащее на расстоянии не более d от x (в смысле некоторой метрики), и $|C_{manag}^*(x, D) - C_{manag}^*(\hat{x}_d, D)| > \epsilon$.

Рассмотрим $y(x, D) \in Y^*(x, D)$ и определим делегирование $\hat{y}_i^k = y_i^k(x, D) \cdot \hat{x}^k / x^k$ при $x^k > 0$, $\hat{y}_i^k = \hat{x}^k / q$ при $x^k = 0$. Выполнено $\hat{y}_1 + \dots + \hat{y}_q = \hat{x}$, и при достаточно малом d точка \hat{y} лежит сколь

угодно близко к точке $y(x, D)$, поэтому в силу непрерывности целевой функции $|C_{manag}(\hat{x}, D, \hat{y}) - C_{manag}(x, D, y(x, D))| < \epsilon$ или

$|C_{manag}(\hat{x}, D, \hat{y}) - C_{manag}^*(x, D)| < \epsilon$. Если $C_{manag}^*(\hat{x}, D) > C_{manag}^*(x, D) + \epsilon$,

то получаем противоречие с оптимальностью $C_{manag}^*(\hat{x}, D)$, поскольку затраты $C_{manag}(\hat{x}, D, \hat{y})$ меньше $C_{manag}^*(\hat{x}, D)$. Если

$C_{manag}^*(x, D) > C_{manag}^*(\hat{x}, D) + \epsilon$, то рассмотрим точку $y(\hat{x}, D) \in Y^*(\hat{x}, D)$

и аналогично определим делегирование y , $y_1 + \dots + y_q = x$, лежащее сколь угодно близко к точке $y(\hat{x}, D)$, поэтому

$|C_{manag}(\hat{x}, D, y(\hat{x}, D)) - C_{manag}(x, D, y)| < \epsilon$, $|C_{manag}^*(\hat{x}, D) - C_{manag}(x, D, y)| < \epsilon$.

Получаем противоречие с оптимальностью $C_{manag}^*(x, D)$, поскольку затраты $C_{manag}(x, D, y)$ меньше $C_{manag}^*(x, D)$. ■

В случае выпуклых функций затрат менеджеров лемма 1 исчерпывающим образом решает вопрос о непрерывной дифференцируемости функции оптимальных затрат $C_{manag}^*(x, D)$ по элементам матрицы дублирования. Для невыпуклых функций необходима дополнительная проверка совпадения левосторонней и правосторонней производной (например, проверка единственности оптимального делегирования $y(x, D)$). Кроме того заметим, что в статье [12] доказан следующий факт: если $d_{i,j}^{k,l}$ – точка локального минимума функции $C_{manag}^*(x, D)$ по $d_{i,j}^{k,l}$, то

левосторонняя и правосторонняя производная по $d_{i,j}^{k,l}$ в этой точке совпадают и равны нулю. Это может быть полезно при дальнейшей минимизации затрат организации $C_{manag}^*(x, D)$ по параметрам дублирования (которые зависят от иерархии, порядка взаимодействия менеджеров и т.п.), поскольку в точке оптимума производная существует и равна нулю при любых рассматриваемых функциях затрат.

Исследуем непрерывную дифференцируемость по суммарному управленческому воздействию x , которое оказывается на производственную подсистему. Если решение задачи об оптимальном делегировании (3) $y(x, D) \in Y^*(x, D)$ единственно и непрерывно дифференцируемо по компонентам управления x , то функция $C_{manag}^*(x, D) = C_{manag}(x, D, y(x, D))$ также непрерывно дифференцируема. Если при этом решение внутреннее, то все ограничения-неравенства ненапряженные и ими можно пренебречь ($\theta = 0$). Для задачи минимизации с ограничениями-равенствами, применяя правило дифференцирования сложной функции, можно показать (см., например, [1]):

$$(10) \quad \partial C_{manag}^*(x, D) / \partial x^k = I^k(x, D),$$

где $\lambda(x, D)$ – множители Лагранжа, которые совместно с $y(x, D)$ являются критической точкой лагранжиана, то есть удовлетворяют условиям (7), (8), (9). Таким образом, коэффициент Лагранжа $\lambda^k(x, D)$ можно интерпретировать как предельные затраты на реализацию управления вида k .

Приведенное выше обоснование выражения (10) предполагает выполненным весьма обременительное требование единственности и непрерывной дифференцируемости внутреннего решения $y(x, D)$. Без этого требования в случае выпуклых функций затрат вопрос непрерывной дифференцируемости по x решается следующей леммой.

Лемма 3. *Если функции затрат менеджеров выпуклы, то левосторонняя (при $x^k > 0$) и правосторонняя производные функции $C_{manag}^*(x, D)$ по x^k существуют, совпадают почти всюду*

и равны соответственно минимуму и максимуму $I^k(x, D) = \min_j \sum_i (d_{i,j}^{.k}, \nabla C_{m_i}(\tilde{y}_i))$ по всем $y(x, D) \in Y^*(x, D)$.¹

Если при $x^k > 0$ левосторонняя и правосторонняя производные по x^k совпадают, то функция $C_{\text{manag}}^*(x, D)$ непрерывно дифференцируема по x^k и $\partial C_{\text{manag}}^*(x, D) / \partial x^k = I^k(x, D)$ при всех $x^k \geq 0$. Условие совпадения односторонних производных выполнено при единственном $y(x, D)$, в частности, если функции затрат менеджеров строго выпуклы и $\det D \neq 0$.

Доказательство. При выпуклых функциях затрат менеджеров целевая функция задачи об оптимальном делегировании $C_{\text{manag}}(x, D, y) = \sum_{i=1, n} C_{m_i}(D_i, y)$ также будет выпуклой (каждое слагаемое выпукло, поскольку зависит от линейной комбинации переменных, сумма выпуклых функций выпукла, см., например, [13]). Для задачи минимизации выпуклой функции на компакте в работе [12] доказано², что при $x^k > 0$ функция $C_{\text{manag}}^*(x, D)$ дифференцируема по x^k почти всюду (производная существует во всех внутренних точках, за исключением, может быть, множества точек нулевой меры). Кроме того, во всех без исключения точках существует левосторонняя (при $x^k > 0$) и правосто-

¹ Напомним, что нагрузка вычисляется по формуле $\tilde{y} = Dy$, через $d_{i,j}^{.k}$ обозначается k -й столбец матрицы $D_{i,j}$. Оговорка существования левосторонней производной лишь при $x^k > 0$ необходима, поскольку слева от $x^k = 0$ задача об оптимальной организации не определена.

² Точнее, в [12] рассматривается задача максимизации вогнутой функции $f(y)$ при наличии ограничений вида $g(y) \geq 0$ с вогнутой функцией $g(\cdot)$ при выполнении условия Слейтера (существование допустимой точки, в которой все ограничения выполнены строго). Очевидно, что это аналогично задаче минимизации выпуклой функции $-f(y)$ с выпуклыми ограничениями $-g(y) \leq 0$. Условие Слейтера используется только для того, чтобы, сославшись на Rockafellar R.T. (1970) [13], показать предствимость оптимума в виде седловой точки лагранжиана. Однако в [13] показано выполнение этого свойства и в случае линейных ограничений-равенств совместно с выпуклыми ограничениями-неравенствами, что соответствует рассматриваемой задаче.

ронная производные, равные соответственно минимуму и максимуму производной лагранжиана $\partial L / \partial x^k = I^k(x, D)$ (см. выражение (4)) по всем оптимальным делегированиям $y(x, D) \in Y^*(x, D)$ и по во всем $\lambda(x, D)$, которые совместно с $y(x, D)$ являются критической точкой. По следствию 2 из теоремы 1 $I^k(x, D) = \min_j \sum_i (d_{i,j}^{:,k}, \nabla C_{m_i}(\tilde{y}_i))$ определяется однозначно при $x^k > 0$, а при $x^k = 0$ максимум $\lambda^k(x, D)$, необходимый для вычисления правосторонней производной, также равен $\min_j \sum_i (d_{i,j}^{:,k}, \nabla C_{m_i}(\tilde{y}_i))$. Первая часть леммы доказана.

Если левосторонняя и правосторонняя производные совпадают при всех $x^k > 0$ (значение $I^k(x, D) = \min_j \sum_i (d_{i,j}^{:,k}, \nabla C_{m_i}(\tilde{y}_i))$ одинаково при всех $y(x, D) \in Y^*(x, D)$), то функция $C_{manag}^*(x, D)$ дифференцируема – производная $\partial C_{manag}^*(x, D) / \partial x^k = I^k(x, D)$ существует во всех точках. Докажем непрерывность производной по всем параметрам. Аналогично доказательству леммы 1 для любой точки x, D и любого $\epsilon > 0$ найдется достаточно малая окрестность, такая, что для любой точки \hat{x}, \hat{D} из этой окрестности найдутся точки $y(x, D) \in Y^*(x, D)$ и $y(\hat{x}, \hat{D}) \in Y^*(\hat{x}, \hat{D})$, расстояние между которыми не превышает ϵ . В силу непрерывности функции $\min_j \sum_i (d_{i,j}^{:,k}, \nabla C_{m_i}(\tilde{y}_i))$ при достаточно малом ϵ значения $I^k(x, D)$ и $I^k(\hat{x}, \hat{D})$ произвольно мало отличаются друг от друга, что и доказывает непрерывность производной $\partial C_{manag}^*(x, D) / \partial x^k$. Правосторонняя производная определяется выражением $\min_j \sum_i (d_{i,j}^{:,k}, \nabla C_{m_i}(\tilde{y}_i))$ и при $x^k = 0$, поэтому ее значение в нуле совпадает с пределом производной при $x^k \rightarrow 0$. Таким образом, производная $\partial C_{manag}^*(x, D) / \partial x^k = I^k(x, D)$ непрерывно дифференцируема при всех $x^k \geq 0$.

Если оптимальное делегирование единственно, то $Y^*(x, D)$ состоит из одной точки, левосторонняя и правосторонняя производные совпадают, поскольку минимум

$I^k(x, D) = \min_j \sum_i (d_{i,j}^{i,k}, \nabla C_{m_i}(\tilde{y}_i))$ по $y(x, D) \in Y^*(x, D)$ равен максимуму. Если функции затрат менеджеров строго выпуклы и $\det D \neq 0$ (матрица D невырождена), то оптимальное делегирование y единственно (см. следствие из теоремы 2 ниже). ■

Лемма 3 показывает, что в выпуклом случае почти во всех точках справедливо выражение (10), поэтому функцию $I^k(x, D) = \min_j \sum_i (d_{i,j}^{i,k}, \nabla C_{m_i}(\tilde{y}_i))$ можно интерпретировать как предельные затраты на реализацию управления вида k . В отдельных точках (множество которых имеет нулевую меру) производная (10) может не существовать, поскольку при различных оптимальных делегированиях $y(x, D)$ возможны различные значения коэффициента $\lambda^k(x, D)$, минимум и максимум которых соответственно определяют левостороннюю и правостороннюю производную. Единственность $y(x, D)$ (в частности, строгая выпуклость функций затрат менеджеров и невырожденность матрицы дублирования) гарантирует непрерывную дифференцируемость. В случае нестрогой выпуклости и множественного решения $y(x, D)$ необходимо дополнительно проверять совпадение левосторонней и правосторонней производных.

Таким образом, лемма 2 показывает, что после решения задачи об оптимальном делегировании при дальнейшей оптимизации организации $C_{manag}^*(x) = \min_{M \in \mathbf{M}} C_{manag}^*(x, D)$ можно пользоваться непрерывностью функции минимальных затрат $C_{manag}^*(x, D)$ по параметрам D и x . Выпуклость функций затрат менеджеров достаточна для непрерывной дифференцируемости $C_{manag}^*(x, D)$ по элементам D , в общем случае лемма 1 дает конструктивный способ расчета односторонних производных и проверки непрерывной дифференцируемости. Строгая выпуклость функций затрат менеджеров и невырожденность матрицы дублирования достаточны для непрерывной дифференцируемости $C_{manag}^*(x, D)$ по x^1, \dots, x^p , для выпуклых функций и любой D лемма 3 дает конструктивный способ расчета односторонних производных и проверки непрерывной дифференцируемости.

6. Однородные функции затрат

Запишем функцию затрат менеджера $C_m(x) = C_m^v(x) + C_m^f$ в виде суммы функции переменных затрат $C_m^v(x)$ и величины постоянных затрат C_m^f , где $C_m^v(0) = 0$. Функцию затрат менеджера $C_m(x)$ назовем *однородной*,¹ если для любого $x \in R_+^p$ и любого $\xi \geq 0$ выполнено

$$(11) C_m^v(\xi x) = f(\xi) C_m^v(x),$$

то есть увеличение всех компонент произвольного управления x в ξ раз приведет к росту переменных затрат в $f(\xi)$ раз независимо от x . Функция $f(\xi)$ описывает зависимость от масштаба.

Известно, что $f(\xi) = \xi^\gamma$ для некоторой степени однородности $\gamma \in (-\infty; +\infty)$, $C_m^v(\xi x) = \xi^\gamma C_m^v(x)$; частные производные функции затрат однородны степени $\gamma - 1$, $\nabla C_m^v(\xi x) = \xi^{\gamma-1} \nabla C_m^v(x)$,² поэтому для выполнения условия непрерывной дифференцируемости рассматриваем только $\gamma \geq 1$, что соответствует выпуклости затрат по масштабу; однородность (11) обеспечивает постоянную эластичность переменных затрат по масштабу (при изменении всех компонент x на 1 % переменные затраты изменятся на γ %); для одномерного случая ($p = 1$) функция затрат однородна тогда и только тогда, когда ее можно представить в виде степенной функции вида:

$$(12) C_m(x) = x^\gamma / \gamma + C_m^f.$$

В настоящем разделе предположим, что функции переменных затрат всех менеджеров однородны и имеют одинаковую степень $\gamma \geq 1$, то есть одинаково изменяются при изменении масштаба управления. Тогда в задаче об оптимальном делегировании (3) переменная часть целевой функции $C_{manag}(x, D, y)$

¹ Строго говоря, следовало бы назвать ее однородной с точностью до константы постоянных затрат, однако для краткости в этом смысле ниже используется термин «однородная».

² Подробные выкладки приведены, например, в [3].

также однородна $C_{manag}(x, D, y) = x^g C_{manag}^v(x, D, y) + C_{manag}^f$, что позволяет доказать следующую лемму.

Лемма 4. Если функции переменных затрат всех менеджеров однородны одинаковой степени $\gamma \geq 1$, то:

1. критическое делегирование $y(x, D)$ однородно степени 1 по x , а соответствующие коэффициенты Лагранжа¹ $\lambda(x, D)$, $\theta(x, D)$ однородны степени $\gamma - 1$ по x ;²

2. делегирование $y(x, D)$ оптимально при суммарном управлении x тогда и только тогда, когда для любого $\zeta > 0$ делегирование $\zeta y(x, D)$ оптимально при суммарном управлении ζx ;

3. при любом критическом делегировании y и соответствующих λ выполнено:

$$C_{manag}(x, D, y) = (I^1 x^1 + \dots + I^p x^p) / g + \sum_{i=1, q} C_{m_i}^f;$$

4. функция оптимальных затрат $C_{manag}^*(x, D)$ однородна степени γ по x :

$$C_{manag}^*(x, D) = C_{manag}^{*,v}(x, D) + C_{manag}^{*,f} = x^g C_{manag}^{*,v}(x, D) + C_{manag}^{*,f}.$$

Доказательство. По теореме Эйлера для однородных функций затрат менеджеров выполнено $(\nabla C_{m_i}^v(x), x) = g C_{m_i}^v(x)$ (см., например, [5]), поэтому:

$$C_{m_i}^v(\tilde{y}_i) = \frac{1}{g} (\tilde{y}_i^1 \frac{\partial C_{m_i}(\tilde{y}_i)}{\partial \tilde{y}_i^1} + \dots + \tilde{y}_i^p \frac{\partial C_{m_i}(\tilde{y}_i)}{\partial \tilde{y}_i^p}) = \frac{1}{g} (\tilde{y}_i, \nabla C_{m_i}(\tilde{y}_i)),$$

$$C_{manag}(x, D, y) = \tilde{y}^T g / g + \sum_{i=1, q} C_{m_i}^f,$$

где g – «составной» вектор градиентов затрат всех q менеджеров при нагрузке \tilde{y} (см. выражение (7)). Докажем свойство 3. Пусть

¹ Напомним, что по следствию 2 из теоремы 1 при $x^k > 0$ коэффициенты $\lambda^k(x, D)$, $q_j^k(x, D)$ однозначно определяются по $y(x, D)$, а при $x^k = 0$ однозначно определяется максимум $\lambda^k(x, D)$ и минимум $q_j^k(x, D)$.

² Под однородностью критической точки понимается, что если векторы $y(x, D)$, $\lambda(x, D)$, $\theta(x, D)$ являются критической точкой при управлении x , то для любого $\zeta > 0$ векторы $\zeta y(x, D)$, $x^{g-1} \lambda(x, D)$, $x^{g-1} \theta(x, D)$ являются критической точкой при управлении ζx .

u и λ являются критической точкой, тогда с учетом $\tilde{y} = Du$ и уравнений критической точки (7) справедливо равенство $\tilde{y}^T g = (Du)^T g = y^T D^T g = y^T (\bar{I} + q)$. В силу условий дополняющей нежесткости (9) имеем $y^T q = 0$, в силу балансовых условий (8) имеем $y^T \bar{I} = \sum_{j,k} y_j^k I^k = \sum_k I^k \sum_j y_j^k = \sum_k I^k x^k$, что и доказывает равенство свойства 3.

Докажем свойство 1. Предположим, что $y(x, D)$, $\lambda(x, D)$, $\theta(x, D)$ – критическая точка в задаче об оптимальном делегировании (3) с некоторым суммарным управлением x . Рассмотрим ту же задачу, но с вектором x , умноженным на произвольную величину $\zeta \geq 0$: $x' = \zeta x$. Покажем, что в этой задаче точка $y' = \zeta y$, $I' = \zeta^{g-1} I$, $q' = \zeta^{g-1} q$ также будет критической,¹ проверив условия (7), (8), (9). Балансовые условия (8) выполнены по построению в силу $y' = \zeta y$, аналогично выполнены и условия дополняющей нежесткости (9), поскольку нулевые компоненты y, q останутся нулевыми и в y', q' . Кроме того, $\tilde{y}' = \zeta \tilde{y}$, $\nabla C_{m_i}(\zeta x) = \zeta^{g-1} \nabla C_{m_i}(x)$, поэтому $g(\tilde{y}') = \zeta^{g-1} g(\tilde{y})$. В результате равенство (7) $x'^{g-1} D^T g(\tilde{y}') = \bar{I}' + q' = \zeta^{g-1} \bar{I} + \zeta^{g-1} q$ выполнено, поскольку в исходной критической точке $D^T g(\tilde{y}) = \bar{I} + q$. Свойство 1 доказано.

Докажем свойство 2. Пусть делегирование $y = y(x, D)$ оптимально при суммарном управлении x . Рассмотрим произвольное $\zeta > 0$ и предположим, что делегирование ζy неоптимально при суммарном управлении ζx . Тогда для некоторого делегирования $\hat{y} \neq \zeta y$ выполнено $C_{manag}(\zeta x, D, \hat{y}) < C_{manag}(\zeta x, D, \zeta y)$. Отсюда, игнорируя постоянные затраты и учитывая однородность переменных затрат целевой функции $C_{manag}(x, D, y)$, получим: $x^g C_{manag}^v(x, D, \hat{y}/x) = C_{manag}^v(\zeta x, D, \hat{y}) < C_{manag}^v(\zeta x, D, \zeta y) = x^g C_{manag}^v(x, D, y)$, что противоречит оптимальности y (делегирующее \hat{y}/x имеет

¹ При $\zeta = 0$ и $g = 1$ выражение ζ^{g-1} считается равным 1 по определению.

меньшие затраты). Поэтому из оптимальности делегирования y при суммарном управлении x для любого $\zeta > 0$ следует оптимальность делегирования ζy при суммарном управлении ζx . Обратно, рассмотрев делегирование $y' = \chi y$, оптимальное при суммарном управлении $x' = \chi x$, и $x' = 1/x$, аналогично докажем оптимальность делегирования y при суммарном управлении x . Свойство 2 доказано.

Осталось доказать свойство 4. Зафиксируем произвольное суммарное управление x и соответствующее оптимальное делегирование y . Тогда по свойству 2 при $\zeta > 0$ делегирование ζy оптимально при суммарном управлении ζx , при $\zeta = 0$ это также верно, поскольку для $x = 0$ оптимально единственное возможное делегирование $y = 0$. С учетом однородности $C_{manag}(x, D, y)$ имеем $C_{manag}^*(\chi x, D) = C_{manag}(\chi x, D, \chi y) = \chi^g C_{manag}^v(x, D, y) + C_{manag}^f$. Обозначив $C_{manag}^{*,f} = C_{manag}^f$, $C_{manag}^{*,v}(\chi x, D) \equiv C_{manag}^v(\chi x, D, \chi y)$ при любом ζ , $C_{manag}^*(\chi x, D) = C_{manag}^{*,v}(\chi x, D) + C_{manag}^{*,f} = \chi^g C_{manag}^{*,v}(x, D) + C_{manag}^{*,f}$, что и доказывает лемму. ■

Лемма 4 показывает, что все критические точки (включая и делегирование и коэффициенты Лагранжа) являются однородными функциями суммарного управления x (свойство 1). То же верно и для оптимальных критических точек (свойство 2), что позволяет доказать однородность функции оптимальных затрат $C_{manag}^*(x, D)$ по x (свойство 4).

В частном случае эквивалентность свойств 3 и 4 следует из теоремы Эйлера. Для строго выпуклых функций затрат менеджеров и невырожденной матрицы D коэффициент Лагранжа λ^k равен предельным затратам $I^k(x, D) = \partial C_{manag}^*(x, D) / \partial x^k$ и непрерывен (см. лемму 3). По теореме Эйлера (см. [5]) однородность функции $C_{manag}^*(x, D)$ эквивалентна ее записи в виде $C_{manag}^*(x, D) = (\nabla C_{manag}^*(x, D), x) / g + C_{manag}^{*,f}$, то есть в виде равенства свойства 3. Лемма 4 показывает, что равенство свойства 3 справедливо и в общем случае, независимо ни от выпуклости, ни от оптимальности рассматриваемого критического делегирования y, λ .

7. Выпуклые функции затрат

Выпуклость затрат сотрудника по нагрузке подтверждается статистическими исследованиями, например, в [4] показан строго выпуклый рост желательного вознаграждения (зарплаты) в зависимости от отрабатываемого количества часов. Поэтому в одномерном случае ($p = 1$) или при линейном росте масштаба управления ξx , $x \neq 0$ функция затрат менеджера $f_x(\xi) = C_m(\xi x)$ при $\xi \geq 0$ обычно растет строго выпукло: для любых $\xi_1, \xi_2 \geq 0$, $\xi_1 \neq \xi_2$ и любого $\theta \in (0; 1)$ выполнено неравенство $f_x(q\xi_1 + (1-q)\xi_2) < qf_x(\xi_1) + (1-q)f_x(\xi_2)$, то есть график функции $f_x(\xi)$ лежит ниже прямой, соединяющей любые две точки этого графика. Из строгой выпуклости следует рост предельных затрат – каждый лишний час отработанного времени требует все больших усилий.

В многомерном случае функцию затрат менеджера назовем *выпуклой*, если для любых $q \in (0; 1)$, $x_1 \in R_+^p$, $x_2 \in R_+^p$ выполнено:

$$(13) \quad C_m(qx_1 + (1-q)x_2) \leq qC_m(x_1) + (1-q)C_m(x_2),$$

функцию затрат менеджера назовем *строго выпуклой*, если неравенство (13) выполнено строго при любых векторах $x_1 \neq x_2$.

Содержательно выпуклость (13) означает аспециализацию¹, например, выгоднее поручить каждому из двух менеджеров планирование и контроль половины исполнителей, чем поручать одному менеджеру планирование, а второму – контроль исполнения этих планов.

В случае выпуклых функций затрат менеджеров целевая функция $C_{manag}(x, D, y) = \sum_i C_m(D_i y)$ задачи об оптимальном делегировании (3) также выпукла по y , поскольку выпуклая функция от линейной комбинации аргументов выпукла, сумма выпуклых функций выпукла, см., например, [13]. Классический аппарат выпуклого анализа [13] позволяет показать, что для выпуклой целевой функции затрат не только оптимальное решение задачи (3) будет критической точкой, но и наоборот –

¹ Подробнее см. [3].

любая критическая точка (с $I_0 = 1$) будет оптимальным решением. В результате теорема 1 примет следующий вид.

Теорема 2 (Куна-Таккера, см., например, [13]). *Если функции затрат менеджеров выпуклы, то y – оптимальное делегирование, доставляющее минимум в задаче (3), тогда и только тогда, когда найдутся такие коэффициенты Лагранжа $I_0 = 1$, $I \in R^p$, $q_i \in R_+^p$, которые совместно с делегированием y являются критической точкой (выполнены условия (7), (8), (9)).*

Докажем следствие, обеспечивающее строгую выпуклость целевой функции и единственность критической точки.

Следствие. *Если функции затрат менеджеров строго выпуклы и $\det D \neq 0$, то целевая функция $C_{\text{manag}}(x, D, y)$ строго выпукла по y , оптимальное делегирование y единственно и только для него найдутся такие коэффициенты Лагранжа $I_0 = 1$, $I \in R^p$, $q_i \in R_+^p$, которые совместно с делегированием y являются критической точкой (выполнены условия (7), (8), (9)).*

Доказательство. Докажем строгую выпуклость по y целевой функции $C_{\text{manag}}(x, D, y) = \sum_i C_{m_i}(D_i y)$ задачи об оптимальном делегировании (3). Рассмотрим произвольные делегирования y, z , $y \neq z$ и любое $q \in (0; 1)$. Поскольку выпуклая функция линейной комбинации аргументов выпукла (Rockafellar R.T. (1970) [13]) неравенство $C_{m_i}(D_i\{qy + (1-q)z\}) \leq qC_{m_i}(D_i y) + (1-q)C_{m_i}(D_i z)$ выполнено для всех $i = \overline{1, q}$. В силу $\det D \neq 0$ и $y \neq z$ выполнено $D(y - z) \neq 0$. Следовательно, $D_i y \neq D_i z$ хотя бы для одного i , поэтому из строгой выпуклости $C_{m_i}(\cdot)$ следует $C_{m_i}(qD_i y + (1-q)D_i z) < qC_{m_i}(D_i y) + (1-q)C_{m_i}(D_i z)$, что и доказывает $C_{\text{manag}}(x(y+z), D, qy + (1-q)z) < qC_{\text{manag}}(x(y), D, y) + (1-q)C_{\text{manag}}(x(z), D, z)$.

Легко показать (см., например, [13]), что в случае строго выпуклой целевой функции задача минимизации (3) имеет единственное решение (при наличии двух решений значение функции на соединяющем их отрезке не может быть меньше минимума, что противоречит строгой выпуклости). По теореме 2 любая критическая точка является оптимальным решением

задачи (3), поэтому во всех критических точках вектор оптимального делегирования y одинаков, и только для этого вектора найдутся коэффициенты Лагранжа $I_0 = 1$, $I \in R^p$, $q_i \in R_+^p$, которые совместно с делегированием y являются критической точкой, то есть выполнены условия (7), (8), (9). ■

Докажем лемму, определяющую свойства функции оптимальных затрат $C_{manag}^*(x, D)$.

Лемма 5. *Если функции затрат менеджеров выпуклы, то функция оптимальных затрат $C_{manag}^*(x, D)$ выпукла по x . Если функции затрат менеджеров строго выпуклы и $\det D \neq 0$, то функция оптимальных затрат $C_{manag}^*(x, D)$ строго выпукла по x .*

Доказательство. Рассмотрим два произвольных вектора $x', x'' \in R_+^p$. Пусть y', y'' – соответствующие оптимальные делегирования: $C_{manag}^*(x', D) = C_{manag}(x', D, y')$, $C_{manag}^*(x'', D) = C_{manag}(x'', D, y'')$. Тогда для любого $\theta \in (0; 1)$ делегирование $y = qy' + (1 - q)y''$ является допустимым в задаче минимизации $C_{manag}(x, D, y)$ с суммарным управлением $x = qx' + (1 - q)x''$. Отсюда с учетом выпуклости целевой функции $C_{manag}(x, D, y)$ по y :

$$\begin{aligned} C_{manag}^*(qx' + (1 - q)x'', D) &= C_{manag}^*(x, D) \leq C_{manag}(x, D, y) = \\ C_{manag}(qx' + (1 - q)x'', D, qy' + (1 - q)y'') &\leq qC_{manag}(x', D, y') + \\ + (1 - q)C_{manag}(x'', D, y'') &= qC_{manag}^*(x', D) + (1 - q)C_{manag}^*(x'', D). \end{aligned}$$

В итоге доказано, что выпуклость функций затрат менеджеров влечет выпуклость функции минимальных затрат $C_{manag}^*(x, D)$ по суммарному управлению x . Если функции затрат менеджеров строго выпуклы и $\det D \neq 0$, то по следствию из теоремы 2 целевая функция $C_{manag}(x, D, y)$ строго выпукла по y , при $x' \neq x''$ выполнено $y' \neq y''$, потому доказанное выше неравенство будет строгим. Поэтому функция оптимальных затрат $C_{manag}^*(x, D)$ строго выпукла по x . Лемма доказана. ■

Таким образом, теорема 2 гарантирует, что в выпуклом случае решение системы уравнений (7), (8), (9) дает оптимальное делегирование. Следствие гарантирует, что в строго выпук-

лом случае при невырожденной матрице дублирования решение системы уравнений дает единственное оптимальное делегирование. Лемма 5 позволяет доказать, что выпуклость функций затрат менеджеров влечет выпуклость функции оптимальных затрат $C_{manag}^*(x, D)$ по x (а строго совместно с невырожденностью матрицы D влечет строгую выпуклость $C_{manag}^*(x, D)$).

8. Монотонные функции затрат

Предположим, что функции затрат всех менеджеров монотонны, то есть затраты не убывают при росте любой из компонент нагрузки. Тогда в задаче об оптимальном делегировании (3) целевая функция $C_{manag}(x, D, y) = \sum_i C_{m_i}(D_i y)$ монотонна и по компонентам делегирования y , и по коэффициентам дублирования (неотрицательным элементам матрицы D). Монотонность целевой функции позволяет доказать следующую лемму.

Лемма 6. Если функции затрат менеджеров монотонны, то:

1. для любого оптимального делегирования y найдутся такие коэффициенты Лагранжа $I^1 \geq 0, \dots, I^p \geq 0, q_i \in R_+^p$, которые совместно с делегированием y являются критической точкой (удовлетворяют условиям (7), (8), (9));

2. функция оптимальных затрат $C_{manag}^*(x, D)$ монотонна как по компонентам управления x , так и по элементам матрицы дублирования D ;

3. для строго монотонных функций затрат менеджеров и матрицы D без нулевых столбцов функция оптимальных затрат $C_{manag}^*(x, D)$ строго монотонна по компонентам x .

Доказательство. По следствию 2 из теоремы 1 при $x^k > 0$ выполнено $I^k = \min_j \sum_i (d_{i,j}^{k,l}, \nabla C_{m_i}(\tilde{y}_i))$, а при $x^k = 0$ можно выбрать такое же I^k . Монотонность функций затрат менеджера m_i влечет неотрицательность компонент градиента $\nabla C_{m_i}(\tilde{y}_i)$. В результате с учетом неотрицательности $d_{i,j}^{k,l}$ имеем $I^k \geq 0$. По-

этому в следствии 1 из теоремы 1 можно считать $I^1 \geq 0, \dots, I^p \geq 0$, что и доказывает свойство 1.

Для доказательства свойства 2 рассмотрим $x \leq \hat{x}$, $D \leq \hat{D}$, соответствующие оптимальные затраты $C_{manag}^*(x, D)$, $C_{manag}^*(\hat{x}, \hat{D})$ и оптимальное делегирование \hat{y} , для которого $C_{manag}(\hat{x}, \hat{D}, \hat{y}) = C_{manag}^*(\hat{x}, \hat{D})$. Построим вектор делегирования z по следующему правилу: $z_i^k = \hat{y}_i^k \cdot x^k / \hat{x}^k$ при $\hat{x}^k > 0$ и $z_i^k = 0$ при $\hat{x}^k = 0$. Очевидно $z \leq \hat{y}$, $z_1 + \dots + z_q = x$, то есть z – допустимое делегирование с суммарным управлением x . В силу $D \leq \hat{D}$ и неотрицательности элементов матриц дублирования имеем $Dz \leq \hat{D}\hat{y}$, поэтому нагрузки менеджеров при делегировании z с D не превышают нагрузок при делегировании \hat{y} с \hat{D} . В силу монотонности функций затрат менеджеров выполнено неравенство $C_{manag}(x, D, z) \leq C_{manag}(\hat{x}, \hat{D}, \hat{y}) = C_{manag}^*(\hat{x}, \hat{D})$. В силу $C_{manag}^*(x, D) \leq C_{manag}(x, D, z)$ имеем $C_{manag}^*(x, D) \leq C_{manag}^*(\hat{x}, \hat{D})$, что и доказывает свойство 2.

Для доказательства свойства 3 рассмотрим такие $x \leq \hat{x}$, что для некоторого k выполнено $x^k < \hat{x}^k$, $D = \hat{D}$, и сохраним остальные обозначения из доказательства свойства 2. Тогда в силу $\hat{x}^k > 0$ для некоторого i выполнено $\hat{y}_i^k > 0$, поэтому $z_i^k < \hat{y}_i^k$. Поскольку в векторном неравенстве $Dz \leq \hat{D}\hat{y}$ компоненты z_i^k и \hat{y}_i^k умножаются на один и тот же ненулевой столбец, хотя бы одно скалярное неравенство выполнено строго, то есть хотя бы у одного менеджера хотя бы один вид нагрузки строго меньше при делегировании z , чем при делегировании \hat{y} , а остальные нагрузки не больше. В силу строгой монотонности функций затрат менеджеров выполнено $C_{manag}(x, D, z) < C_{manag}(\hat{x}, \hat{D}, \hat{y})$, откуда $C_{manag}^*(x, D) < C_{manag}^*(\hat{x}, \hat{D})$. Лемма доказана. ■

Лемма 6 доказывает, что для монотонных функций затрат можно рассматривать только критические точки с неотрицательными коэффициентами Лагранжа I (свойство 1), функция

оптимальных затрат $C_{manag}^*(x, D)$ монотонна по параметрам x и D (свойство 2), а строгая монотонность и отсутствие нулевых столбцов в матрице дублирования приводит к строгой монотонности $C_{manag}^*(x, D)$ по суммарному управлению x .

9. Заключение

В целом в настоящей работе доказано, что *функция минимальных суммарных затрат управленческой подсистемы* (при оптимальном делегировании) будет зависеть от параметров (управления x и коэффициентов дублирования) непрерывно, а во многих случаях и непрерывно дифференцируемо по параметрам. Также доказано, что свойства однородности, выпуклости, монотонности функций затрат менеджеров, зависящих от управленческой нагрузки, приводят соответственно к тем же свойствам функции минимальных затрат, зависящей от суммарного управленческого воздействия, оказываемого всей управленческой подсистемой.

Таким образом, предполагая весьма общие, хорошо известные в экономике свойства функций затрат отдельных менеджеров, можно устанавливать наличие этих свойств для функции затрат сложной управленческой подсистемы, состоящей из множества взаимодействующих друг с другом менеджеров.

Полученные результаты позволяют применить методы непрерывной оптимизации и для поиска оптимального делегирования, и для дальнейшей минимизации затрат управленческой подсистемы выбором иерархии, типов и состава менеджеров, стимулирования и т. п.

Кроме того, математический аппарат, предложенный в настоящей работе, может использоваться для поиска оптимального распределения общего объема выпуска между предприятиями (отрасли, холдинга и т.п.), которые обмениваются продуктами производства по обобщенной модели многопродуктового межотраслевого баланса и описываются нелинейными многопродуктовыми функциями затрат.

Литература

1. БОСС В. *Лекции по математике: оптимизация*. Том 7. М.: КомКнига, 2007.
2. БУСЫГИН В.П., ЖЕЛОБОДЬКО Е.В., ЦЫПЛАКОВ А.А. *Микроэкономика – третий уровень*. Новосибирск: СО РАН, 2003.
3. МИШИН С.П. *Модели и методы оптимизации иерархических организаций*: Диссертация на соискание ученой степени доктора технических наук. М.: ИПУ РАН, 2011.
4. НОВИКОВ Д.А. *Экспериментальное исследование индивидуальных стратегий предложения труда*. М.: Эгвес, 2010.
5. ФИХТЕНГОЛЬЦ Г.М. *Курс дифференциального и интегрального исчисления (в 3-х томах)*. М.: Физматлит, 2001.
6. BLOCH H., MADDEN G. AND SAVAGE S.J. *Economies of Scale and Scope in Australian Telecommunications* // Review of Industrial Organization. 2001. 18. P. 219 – 227.
7. COHN E., RHINE S., SANTOS M. *Institutions of Higher Education as Multi-Product Firms: Economies of Scale and Scope* // The Review of Economics and Statistics. 1989. Vol. 71(2). P. 284 – 290.
8. DUONG T.M., MCLAREN N.K., ZHAO X. *Multi-output Broadacre Agricultural Production: Estimating a Cost Function Using Quasi-Micro Farm Level Data from Australia* / Proceedings of AARES 52nd Annual Conference, February 5-8, 2008, Canberra, Australia (<http://purl.umn.edu/6009>).
9. POLOME P., FERNAGUT P.P., HARMIGNIE B., O. DE FRAHAN, BRUNO H. *Multi-input Multi-output Farm-level Cost Function: A Comparison of Least Squares and Entropy Estimators* / Proceedings of European Association of Agricultural Economists, International Congress, August 23-27, 2005, Copenhagen, Denmark (<http://purl.umn.edu/24727>).
10. FILIPPINI M., FARSI M. *Cost Efficiency and Scope Economies in Multi-output Utilities in Switzerland* / Study on behalf of the State Secretariat for Economic Affairs SECO. Berne, 2008.
11. HOWITT R.E. *Positive mathematical programming* // American Journal of Agricultural Economics. 1995. 77(2). P. 329 – 342.

12. MILGROM P., SEGAL I. *Envelope Theorems for Arbitrary Choice Sets* // *Econometrica*. 2002. Vol. 70. No.2. P. 583 – 601.
13. ROCKAFELLAR R.T. *Convex Analysis*. N.J., Princeton: Princeton University Press, 1970.
14. ROLLER L.H. *Proper Quadratic Cost Functions with an Application to the Bell System* // *Review of Economics and Statistics*. 1990. 72. P. 202 – 210.
15. TOVAR B., JARA-DIAZ S.R., TRUJILLO L. *A Multioutput Cost Function for Port Terminals: Some Guidelines for Regulation*. Paper provided by The World Bank in its series Policy Research Working Paper Series with number 3151. 2003 (http://www-wds.worldbank.org/servlet/WDSContentServer/WDSP/IB/2003/10/27/000160016_20031027124418/Rendered/PDF/wps3151.pdf).

PROPERTIES OF OPTIMAL DELEGATION OF AUTHORITY IN FIRMS

Sergei Mishin, Institute of Control Sciences of RAS, Moscow, Cand.Sc. (smishin78@mail.ru).

Abstract: We minimize a cost function by choosing an optimal delegation of authority (DoA). DoA means distribution of management actions needed for a production subsystem between managers with complicated duplication of each others' efforts. Formally this problem is equivalent to optimization of production volume distribution between firms with complicated exchange of products and multi-output cost functions. We have proved general properties of optimal solution which allow finding out optimal DoA for several classes of cost function. Key properties of managers' cost functions hold true for a cost function of a whole management subsystem with optimal DoA.

Keywords: optimal delegation of authority, manager, cost, firm.

Статья представлена к публикации членом редакционной коллегии М.В. Губко

УДК 519.254 + 004.93'14
ББК 3.32.965.32.965.9

ИНДЕКСНЫЕ ПАЕВЫЕ ИНВЕСТИЦИОННЫЕ ФОНДЫ (АНАЛИЗ ДОХОДНОСТИ)¹

Спиро А. Г.², Дорофеев Ю. А.³

(Учреждение Российской академии наук
Институт проблем управления РАН, Москва)

Alperovich Ed.⁴

(Polytechnic Institute of New York University, New York, USA)

Проведен анализ доходности инвестора открытого индексного паевого инвестиционного фонда, показано, что активы, не входящие в расчёт индекса, могут существенно влиять на доход инвестора.

Ключевые слова: паевой инвестиционный фонд, структура инвестиционного портфеля фонда, ошибка следования, индексная часть фонда, стоимость чистых активов.

1. Введение

С развитием финансовых рынков и привлекательности их для сохранения и приумножения своих денежных средств (о чём настойчиво «вещает» реклама во всём мире, в том числе и в России) всё большее число людей размещают свои деньги в фонды, которые объединяют эти средства и используют их для инвестирования.

¹ Работа выполнена при поддержке РФФИ, проект 10-07-00210.

² Арнольд Григорьевич Спиро, кандидат технических наук, старший научный сотрудник (spiro@ipu.ru).

³ Юлия Александровна Дорофеев, научный сотрудник (dorofeyuk_julia@mail.ru).

⁴ Eduard Alperovich, Ph.D., Professor (alpedu@yahoo.com).

В США это взаимные фонды (*mutual funds*), в России это **Паевые Инвестиционные Фонды** (ПИФ). Частный инвестор, покупая долю (пай) в фонде, становится одним из владельцев той части активов фонда, которая зависит от количества паёв, приобретённых инвестором. Цена пая зависит от общей стоимости активов фонда и определяется как **сумма чистых активов** (СЧА) фонда (*net asset value*), разделенная на число паёв [6, 8]. Если пай можно приобрести в любой момент времени, то фонд называется открытым (*open-end fund*), в противном случае он является закрытым (*closed-end fund*). Как правило, закрытые фонды не являются фондами взаимного вложения и средства собирают один раз посредством начального публичного предложения (*Initial Public Offering*). В США широко распространены фонды, работающие с акциями, входящими в расчёт индексов *S&P 500*, *DJI* (Доу-Джонса), *NASDAQ* и других.

Фонды, акции которых свободно торгуются на биржах США, объединены в группу *ETF – Equity traded fund* [3].

В России фонды, работающие с акциями, входящими в расчёт индексов, являются **Открытыми Индексными Паевыми Инвестиционными Фондами** (ОИПИФ). Они рассчитаны в основном на работу с мелкими инвесторами. Законодательства по ОИПИФ в РФ пока нет. На сегодняшний день основным нормативно-правовым актом, регулирующим деятельность ПИФов, является Федеральный закон «Об инвестиционных фондах» от 29.11.2001г. № 156-ФЗ, однако в этом законе ОИПИФ даже не упоминаются. В настоящее время готовится поправка к закону, в которую войдут также и правовые основы работы ОИПИФ. На основании этой поправки в РФ будут созданы и начнут функционировать аналоги фондов группы *ETF* [1, 3].

Поскольку нигде не оговорено соотношение активов, копирующих структуру индекса и активов, не связанных с ним, выбор инвестиционного портфеля относится к компетенции управляющей компании (УК) фонда. Обычно индексным называют фонд, в котором значительная часть его средств идёт на копирование индекса. Объём таких средств, в принципе, должен оговариваться в инвестиционной декларации фонда.

Рынки акций в долгосрочной перспективе растут (в отсутствии затяжных финансовых кризисов), поэтому УК практически невозможно переиграть фондовый индекс по доходности на больших промежутках времени. Однако в краткосрочной перспективе возможны ощутимые колебания стоимости активов как в сторону увеличения, так и в сторону уменьшения, поэтому менеджеры фондов и оставляют часть средств для игры «на коротких дистанциях», чтобы компенсировать потери в случае падения рынка.

Отметим ряд особенностей, отличающих индексный ПИФ от других типов ПИФов:

1. Вся (или значительная часть) активов индексного ПИФа копирует в своём портфеле структуру того индекса, которому он соответствует. Например, декларируется, что в индексном фонде ММВБ не менее 85% его портфеля в течение двух третей рабочих дней в месяце составляют именно те бумаги, по которым этот индекс рассчитывается [4].
2. Управление индексным ПИФом частично или полностью является пассивным (не требует активных действий менеджеров УК). Поскольку структура части инвестиционного портфеля определена, то для работы с этой частью не требуется квалифицированный персонал, что существенно удешевляет процесс управления, который ещё часто удается автоматизировать, то есть свести к минимуму затраты. Поэтому ежегодное вознаграждение персоналу УК фонда, обычно, невелико.
3. В связи с тем, что структура индекса периодически меняется, за состоянием индексной части инвестиционного портфеля необходимо следить, обновляя её содержимое и поддерживая соотношение различных акций таким образом, чтобы это соотношение совпадало со структурой индекса.
4. Некоторую свободу действий менеджерам УК фонда даёт то, что в РФ при расчёте СЧА используется «признаваемая котировка», то есть средневзвешенная цена акции за день. Такая цена отличается от цены покупки. Поэтому возможны некоторые вариации с определением числа паёв для вновь вступающих в ОИПИФ. Кроме того, есть некоторая задержка по

времени приобретения акций, входящих в расчет индекса, что приводит к несовпадению расчетных значений изменения индекса и паёв фонда.

5. Обычно критерием «правильного управления» ОИПИФ является близость совпадения динамики изменения цен паёв и динамики индекса в долгосрочной перспективе. Существует коэффициент, обычно называемый «ошибка следования» (tracking error – TE), который показывает, насколько близко данный фонд повторяет динамику индекса. Чем меньше TE , тем меньше отклонение и тем лучше ОИПИФ. В идеале TE должна равняться нулю.

Аналогичным является параметр среднеквадратичного отклонения (R -squared или R^2), характеризующий точность повторения индекса фондом. Иногда этот параметр называют коэффициентом детерминации. Коэффициент R^2 показывает степень связи фонда с выбранным индексом. Показатель обычно выражается в процентах и может находиться в пределах от 0% (связь отсутствует) до 100% (полная функциональная связь). Коэффициент равный 90% означает, что 10% колебаний доходности фонда не связаны с индексом, а являются результатом работы УК. Естественно, чем коэффициент больше, тем лучшие показатели индексного фонда и тем меньше влияние человеческого фактора [9].

В литературе не уделено должного внимания функционированию индексных паевых фондов России, не считая рекламных материалов для привлечения средств частных инвесторов, которые не отражают объективную картину состояния фондов. Как правило, ОИПИФ публикует два показателя: стоимость чистых активов и цену пая на текущую дату. Однако этих данных недостаточно для принятия решения об инвестировании средств в ОИПИФ.

Основной целью настоящей работы является оценка доходности пайщика ОИПИФ и разработка показателей, позволяющих эффективно выбирать конкретный фонд для максимизации доходности инвестирования. Ниже будут рассматриваться лишь те индексные фонды, в которых индексная часть больше всех других активов фонда.

2. Связь стоимости чистых активов (СЧА) с индексом

Индекс ММВБ обычно рассчитывается следующим образом [5]:

$$(1) I(t_i) = \left(\sum_{j=1}^N k_j A_j(t_i) \right) / V \sum_{l=1}^L Q_l^b P_l^b,$$

где k_j – весовой коэффициент (коэффициент капитализации); V – поправочный коэффициент, корректирующий значения индекса при изменении базы его расчета; а также:

- на момент t_i расчёта индекса: N – количество различных видов акций (компаний), входящих в базу расчёта индекса; $A_j(t_i) = Q_j(t_i)P_j(t_i)$ – стоимости акций j -го вида, где: $Q_j(t_i)$ – количество акций j -го вида; $P_j(t_i)$ – средневзвешенная цена акции j -го вида;

- на момент определения первого значения индекса: Q_l^b – количество акций l -го вида; P_l^b – средневзвешенная цена акции l -го вида; L – количество акций, входящих в базу расчёта индекса;

При этом средневзвешенная цена любой акции P определяется соотношением

$$(2) P = \left(\sum_{k=1}^K p_k q_k \right) / \sum_{k=1}^K q_k,$$

где p_k – цена акции в k -ой сделке в рублях; q_k – объем k -ой сделки в штуках; K – количество заключённых сделок в пределах анализируемой сессии.

Выражение (1) можно представить следующим образом:

$$(3) I(t_i) = I \sum_{j=1}^N k_j A_j(t_i),$$

где нормирующий коэффициент λ равен $I = 1/V \sum_{l=1}^L Q_l^b P_l^b$, – все обозначения из (1).

Рассмотрим теперь открытые индексные паевые инвестиционные фонды (ОИПИФ). Для индексного фонда справедливо соотношение:

$$(4) \sum_{j=1}^N k_j B_j(t_i) + \sum_{m=1}^M k_m D_m(t_i) = n(t_i) C(t_i),$$

где: $\sum_{j=1}^N k_j B_j(t_i)$ – индексная часть фонда; $\sum_{m=1}^M k_m D_m(t_i)$ – неиндексная часть фонда, используемая УК фонда по своему усмотрению; $n(t_i)$ – общее число паёв, а $C(t_i)$ – цена пая в текущий момент времени t_i .

Если стоимость активов фонда отличается от СЧА, то реальная цена пая $C_R(t_i)$ будет:

$$(5) C_R(t_i) = \frac{1}{n(t_i)} \left[\sum_{j=1}^N k_j B_j(t_i) + e(t_i) \right],$$

где $e(t_i)$ – зависящие от времени активы фонда, не входящие в индексную часть фонда (индексная часть фонда – набор акций, соответствующих индексу). Представим $e(t_i)$ в виде трёх составляющих $e(t_i) = e_1(t_i) + e_2(t_i) + e_3(t_i)$, где:

$e_1(t)$ – активы (резервные депозиты, денежные средства от не основной деятельности, дебиторская задолженность и пр.) и пассивы, а также – кредиторская задолженность, резервы предстоящих расходов и платежей;

$e_2(t)$ – вознаграждение УК, расходуемое на некоторое постоянное имущество фонда, оплата менеджеров и т.п.;

$e_3(t)$ – стоимость активов неиндексной части ПИФа, то есть

$$e_3(t) = \sum_{m=1}^M k_m D_m(t_i).$$

3. Связь неиндексной части фонда со стоимостью чистых активов

Рассмотрим для примера индексный фонд ММВБ. Поскольку индекс ММВБ определяется взвешенной по рыночной капитализации суммой активов 30 компаний, то величина приращения индекса ΔI для последовательных моментов времени t_i и t_{i+1} с учётом (1) запишется в виде:

$$(6) \Delta I = [I(t_{i+1}) - I(t_i)] / I(t_i) =$$

$$= [\sum_{j=1}^{30} k_j A_j(t_{i+1}) - \sum_{j=1}^{30} k_j A_j(t_i)] / \sum_{j=1}^{30} k_j A_j(t_i),$$

где все обозначения из (1).

По правилам формирования индексной части ПИФа, количество каждого вида акций должно быть пропорционально акциям индекса, т.е. $B_j(t_i) = \alpha A_j(t_i)$, поэтому для индексного фонда, с учётом (4) и (5) получаем:

$$(7) \Delta I_\phi = \frac{[\sum_{j=1}^{30} \alpha k_j A_j(t_{i+1}) + e(t_{i+1})] - [\sum_{j=1}^{30} \alpha k_j A_j(t_i) + e(t_i)]}{[\sum_{j=1}^{30} \alpha k_j A_j(t_i) + e(t_i)]},$$

После преобразований выражения (7), получим:

$$(8) \Delta I_\phi = \frac{\sum_{j=1}^{30} \alpha k_j A_j(t_{i+1}) - \sum_{j=1}^{30} \alpha k_j A_j(t_i)}{\sum_{j=1}^{30} \alpha k_j A_j(t_i) + e(t_i)} + \frac{e(t_{i+1}) - e(t_i)}{\sum_{j=1}^{30} \alpha k_j A_j(t_i) + e(t_i)}$$

Величина $e(t_i)$ для индексного фонда ММВБ существенно меньше $\sum_{j=1}^{30} \alpha k_j A_j(t_i)$. Следовательно, с учётом выражения (8)

оценка сверху величины приращения ΔI_ϕ имеет вид:

$$\Delta I_\phi = \Delta I + [e(t_{i+1}) - e(t_i)] / [\sum_{j=1}^{30} \alpha k_j A_j(t_i) + e(t_i)].$$

Если пренебречь «прочими расходами» ($\varepsilon_1(t_i) + \varepsilon_2(t_i) \ll \varepsilon_3(t_i)$), то ΔI_ϕ можно записать в виде:

$$(9) \Delta I_\phi \approx \Delta I + \Delta e / [\sum_{j=1}^{30} k_j B_j(t_i) + \sum_{m=1}^M k_m D_m(t_i)],$$

где $\Delta e = [e(t_{i+1}) - e(t_i)] \approx$

$$\approx e_3(t_{i+1}) - e_3(t_i) = \sum_{m=1}^M k_m D_m(t_{i+1}) - \sum_{m=1}^M k_m D_m(t_i).$$

То есть второе слагаемое в (9) приблизительно равно отношению разности не вложенных в индекс активов за $(i+1)$ -ый и i -ый моменты времени, к СЧА в i -ый момент времени.

Другими словами: относительное изменение активов фонда приближённо (оценка сверху) равно изменению индекса плюс относительному изменению активов, не вложенных в акции индекса (неиндексная часть).

Изменение индекса – это величина легко оцениваемая, поскольку графики и значения индекса широко публикуются. Величина же Δe и её большая часть $\Delta e_3(t)$ – изменение стоимости активов неиндексной части ОИПИФа, нигде не публикуются.

Как правило, вкладчики выбирают именно индексный фонд для инвестирования на длительные промежутки времени, порядка 5-10 лет. Число таких пайщиков быстро растёт. Все они полагаются на «очевидный» постулат – рынки акций в долгосрочной перспективе растут, и УК невозможно переиграть фондовый индекс по доходности на длинном промежутке времени. Кроме того, индексные паевые фонды легкодоступны и, предполагается, что в них отсутствует «человеческий фактор». Всё это и привлекает людей со средним достатком, не имеющих достаточных экономических знаний для работы на бирже. Однако все рекламные «разговоры» о полной прозрачности ОИПИФ являются пустыми словами, поскольку управляющая фондом компания может в любой момент совершить ошибку при использовании неиндексной части ОИПИФа и нанести ущерб пайщикам (иногда весьма существенный). Таким образом, пайщиков сознательно вводят в заблуждение. Пока информация о том, что ОИПИФ содержит в своем составе значительную неиндексную часть, не будет публиковаться (включая её величину), говорить о прозрачности таких фондов и оценивать их качество чрезвычайно сложно. Это под силу только биржевым аналитикам высокого класса, владеющим полной информацией о состоянии фонда.

Публикуемая же информация о том, что оценочная стоимость ценных бумаг, по которым рассчитывается индекс, должна составлять (например, для ММВБ) не менее 85% стоимости активов, не имеет под собой никакой правовой базы и ничего не говорит о реальной величине неиндексной части конкретного ОИПИФ [2]. Оценка же качества индексного паевого фонда

лишь по публикуемым графикам субъективна, и будет зависеть от квалификации аналитика и выбранного временного интервала.

4. Доходность ПИФа

Как известно, [1] ПИФы могут быть с так называемой передней или задней нагрузкой. Передняя нагрузка – это вычеты на нужды фонда при вступлении в ПИФ, задняя нагрузка – это вычеты при выходе из ПИФа – при продаже (возврате) пая. Без ограничения общности далее рассматривается второй вариант нагрузки. При продаже доходность (чистая прибыль) ПИФа для частного инвестора может быть представлена следующим образом:

$$(10) H(t_i) = [1 - m_1(t_i)]k(t_i)S - m_2(t_i)(k(t_i)S - m_1(t_i)k(t_i)S - S),$$

где: S – депозит; $m_1(t_i)$ – коэффициент скидки (величина скидки зависит от времени владения паем до его продажи); $m_2(t_i)$ – коэффициент, зависящий от величины подоходного налога; $k(t_i) = C(t_i)/C(t_0)$ – коэффициент доходности, где $C(t_0)$ – цена пая на дату приобретения пая в ПИФе, $C(t_i)$ – цена пая на момент t_i выхода из ПИФа (продажи пая). Величину $H_{\delta}(t_i) = (k(t_i)S - m_1(t_i)k(t_i)S - S)$ обычно называют балансовой или экономической прибылью, она является налогооблагаемой базой для операций по покупке-продаже паёв. Тогда условием положительной доходности операции является $H_{\delta}(t_i) > 0$, после несложных преобразований получаем простое условие для положительной доходности:

$$(11) k(t_i) > 1/(1 - m_1).$$

Другими словами, условие (11) выполняется, если цена пая в момент времени t_i будет в $1/(1 - m_1)$ раз больше, чем на дату t_0 покупки пая. Так, например, для ОИПИФ «Райффайзен – Индекс ММВБ» (УК Райффайзен-Капитал) величина $m_1 = 0,02$, тогда $1/(1 - m_1) = 1,0204$, т. е. цена продажи пая должна быть больше цены покупки на 2,05%.

Иногда для характеристики динамики соответствующего финансового рынка используется величина

$$\delta H_{\delta}(t_{i+k}) = [H_{\delta}(t_{i+k}) - H_{\delta}(t_i)] / H_{\delta}(t_i)$$

относительного изменения прибыли на интервале времени $(t_{i+k} - t_i)$.

Как упоминалось выше, не вся сумма денег пайщиков фонда идет на покупку ценных бумаг в соответствии с индексом. Используя формулы (7)-(9) и (10), (11) можно получить условие прибыльности (положительной доходности) для пайщика в условиях, когда неиндексная часть фонда (так называемые, свободные средства) составляет заметную часть средств фонда:

$$(12) \Delta_r(t_i) + \Delta_p(t_i) > m_1 / (1 - m_1),$$

где $\Delta_r(t_i) = \Delta C_r(t_i) / C_r(t_0)$ – относительное приращение индексной части стоимости пая с момента t_0 покупки пая к моменту t_i его продажи; аналогично определяется величина $\Delta_p(t_i)$, характеризующая относительное приращение неиндексной части стоимости пая. Здесь $\Delta C_r(t_i) = C_r(t_i) - C_r(t_0)$, величина $\Delta C_p(t_i)$ определяется аналогично.

Из (12) непосредственно следует, что условие прибыльности пайщика в ОИПИФ существенным образом зависит от эффективности работы управляющей компании (УК) с неиндексной частью фонда. К сожалению, практика работы с неиндексной долей фонда не ясна, что не позволяет делать однозначные выводы об увеличении прибыли фонда. В связи с отсутствием законодательной базы по ОИПИФ можно предположить следующее развитие ситуации. При успешных действиях менеджеров УК (прибыль по неиндексной части выше, чем по индексной), вся дополнительная прибыль или определенный её процент под разными предлогами изымается из дальнейшего оборота (непредвиденные расходы, поощрение менеджеров, которые провели успешную операцию и т.д.). В свою очередь, при неуспешных действиях менеджеров УК может наблюдаться обратная картина – дефицит прибыли может покрываться за счёт пайщиков.

Эффективность работы УК фонда можно оценить, например, в процентах для одного и того же момента времени по относительному изменению величины пая и индекса. Если изменение пая значительно меньше изменения индекса, то возможны следующие сценарии:

1. Индексная часть активов фактически не соответствует набору акций для расчета индекса;
2. Неиндексная часть активов соизмерима с индексной частью и УК в процессе работы понесла большой убыток;
3. По каким-либо причинам сократилась стоимость чистых активов фонда.

Когда относительное изменение пая значительно больше изменения индекса, возникает обратная картина:

1. Индексная часть полностью соответствует набору акций для расчета индекса;
2. Неиндексная часть активов соизмерима с индексной частью и УК в процессе работы получила большую прибыль;
3. По каким-либо причинам увеличилась стоимость чистых активов фонда.

В качестве примера рассмотрим данные по ОИПИФ «Райффайзен – Индекс ММВБ», УК Райффайзен-Капитал с 14.04.2011 по 13.05.2011 г. На рис. 1 приведены изменения (в процентах) индекса ММВБ и пая в фонде. Следует отметить, что на этом же временном интервале инвестор будет иметь как прибыль, так и убытки (рис 2). Это зависит от величины коэффициента скидки m_1 (для указанного ОИПИФ $m_1 = 0,02$).

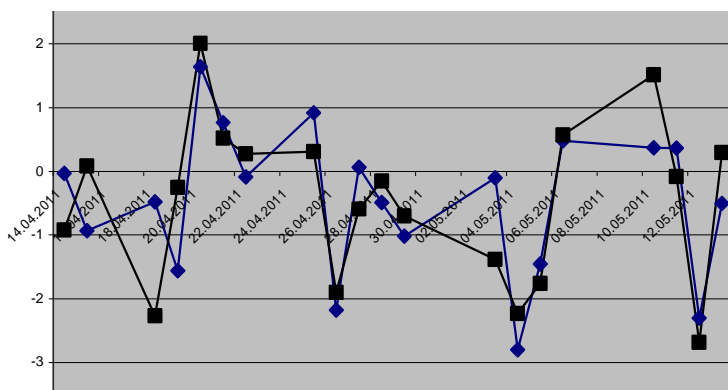


Рис. 1. Изменения в %: индекса (ромб), величины пая (квадрат)

Из рис. 2 следует, что и для малых сроков инвестирования при достаточной квалификации инвестора (или менеджера УК)

можно получить вполне приемлемую доходность (за 14 дней – 5%). Однако в подавляющем числе случаев паи ОИПИФ на финансовых рынках являются инструментами долгосрочного инвестирования.

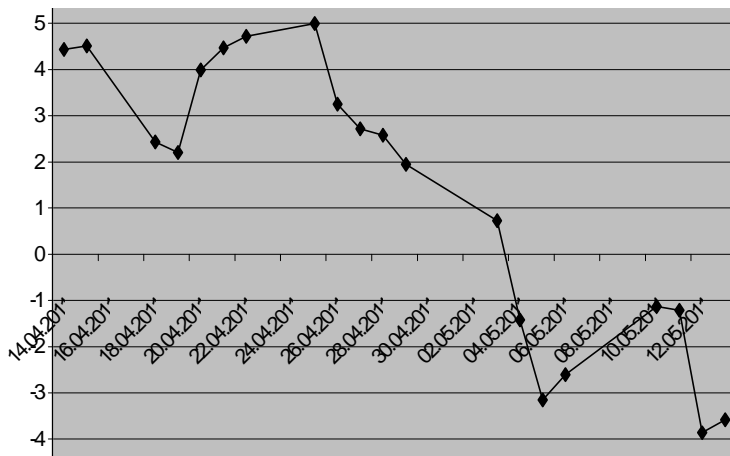


Рис. 2. Доходность инвестора в %

5. Заключение

Для выбора конкретного индексного фонда с целью долгосрочного инвестирования необходимо на момент инвестирования иметь информацию о стоимости неиндексной части активов фондов-претендентов. Наличие этой информации расширяет возможности традиционных методов анализа, основанных на использовании коэффициентов Шарпа, Сортино, b , а также на анализе волатильности, VaR, R2 и т.п. [7].

Литература

1. ЗАКАРЯН И. О. *Практический интернет-трейдинг (как работать на рынках акций, фьючерсов, опционов и FOREX)*. – М.: Smart BOOK, 2006. – 377 с.
2. Интернет-конференции по ПИФ: сайт. – URL: <http://pif.investfunds.ru> (дата обращения: 12.07.2011).

3. КАПИТАН М. Е., БАРАНОВСКИЙ Д. С. *Паевые фонды: современный подход к управлению деньгами*. – СПб.: Питер, 2007. – 256 с.
4. МАЗУНИН А., ЛАДЫГИН Д. *ПИФы обнаружили клиентов*. // Коммерсантъ. – №. 223(4523) от 02.12.2010.
5. Положение о сводных фондовых индексах Московской межбанковской валютной биржи: сайт. – URL: <http://www.dinform.ru> (дата обращения: 09.07.2011).
6. СОЛАБУТО Н. В. *Секреты инвестирования в ПИФ*. – СПб.: Питер, 2007. – 144 с.
7. *Сравнительная таблица ПИФов*. // Информационный ресурс Investfunds: сайт. – URL: <http://pif.investfunds.ru/1008> (дата обращения: 17.07.2011).
8. *Что такое СЧА? (Стоимость чистых активов)*. // О ПИФах для пайщиков: сайт. – URL: <http://www.invest-pif.ru> (дата обращения: 15.07.2011).
9. CHERRY N. *Блог о том, как достичь финансовой независимости и перестать работать за деньги*: сайт. – URL: <http://www.myrichway.ru> (дата обращения: 18.07.2011).

INDEX MUTUAL FUNDS (COST-BENEFIT ANALYSIS)

Arnold Spiro, Institute of Control Sciences of RAS, Moscow, PhD in technical sciences, senior research assistant (spiro@ipu.ru).

Julia Dorofeyuk, Institute of Control Sciences of RAS, Moscow, research assistant (dorofeyuk_julia@mail.ru).

Eduard Alperovich, Polytechnic Institute of New York University, New York, Ph.D., Professor (alpedu@yahoo.com).

Abstract: The analysis of open-end index mutual fund investor's profitability was conducted. It is shown, that the assets, which are not included in the index, can significantly affect the investor's income.

Keywords: mutual fund, structure of investment portfolio, tracking error, index part of fund, net asset value.

Статья представлена к публикации членом редакционной коллегии А.С. Манделем

УДК 332.145
ББК 65в6

ОБ ОДНОМ ПОДХОДЕ К РАЗРАБОТКЕ СОЦИАЛЬНО ОРИЕНТИРОВАННЫХ ТАРИФОВ НА ЭЛЕКТРОЭНЕРГИЮ ДЛЯ НАСЕЛЕНИЯ

Эйфельд А. А.¹

(Волгоградский государственный университет, Волгоград)

Статья посвящена моделированию социально-ориентированного блочного тарифа на электроэнергию для населения. Приведен механизм расчета данного тарифа на основе статистических данных г. Волгограда и Волгоградской области. Проведены численные исследования свойств данного тарифа.

Ключевые слова: двухставочный тариф, тарифы на электроэнергию, моделирование тарифов, социальная норма потребления, бытовые потребители.

Либерализация электроэнергетических рынков и современные социально-экономические условия жизни определяют необходимость разработки новых, более эффективных моделей тарифных систем для населения, учитывающих множество факторов. Величина тарифов на электроэнергию для бытовых клиентов в России приближается к европейским значениям, однако уровень платежеспособности и уровень жизни остаются несравнимо разными. Более того, в России на протяжении долгого периода наблюдается процесс высокой дифференциации населения по доходам. Об этом свидетельствует динамика значений коэффициента фондов, который характеризует степень расслоения общества и показывает отношение среднего уровня доходов 10% самых богатых граждан к среднему уровню доходов 10% самых бедных. Чем выше данный показатель, тем

¹ Анастасия Александровна Эйфельд, ассистент
(aeisfeld@yandex.ru).

выше дифференциация населения по доходам. По рекомендации ООН он не должен превышать 8–10, иначе ситуация в стране чревата социальными катаклизмами. В странах Европы этот коэффициент равен 6–10 и ниже, для России динамика его изменения представлена данными таблицы 1.

Таблица 1. Динамика изменения коэффициента фондов в России за период с 2004 по 2010 гг [12].

2004	2005	2006	2007	2008	2009	2010
15,2	15,2	16,0	16,8	16,8	16,7	15,8

Таким образом, по официальным данным Федеральной службы государственной статистики доходы 10% самых богатых россиян превышают доходы 10% самых бедных в 15–16 раз. При этом не учитываются «теневые» доходы граждан. Тогда, вероятно, данный показатель окажется ещё выше.

Данное исследование продиктовано необходимостью разработки новой, социально-ориентированной системы тарифов. Среди авторов, работы которых посвящены методам формирования тарифов на электроэнергию, следует отметить работы Богачковой Л.Ю. [1], Забелло Е.П. [3], Зайцевой Ю.В. [5], Орлова В.С. [6], Шагова А.В. [7], Brown S. [9].

Модель «тарифного меню» для населения как механизм согласования интересов участников рынка и реализации социальной ответственности предложена в работе [1]. Конкретизация этого тарифа с учетом статистических данных о спросе на электроэнергию со стороны населения Волгоградской области получена в работе [4]. Несмотря на преимущество этой модели, состоящее в возможности согласования интересов всех участников рынка, её применение на практике затруднено в связи с неопределенностью методики определения тарифных ставок, которые подбирались, по сути, эвристически.

В работе [8] описана модель блочного тарифа, целью построения которого является социальная защита наименее обеспеченных групп населения. Такой тариф перекладывает нагрузку по оплате электроэнергии на группу населения с высоким уровнем электропотребления. Предполагается, что данная кате-

гория населения имеет более высокий уровень доходов. Первая ставка тарифа за объем потребления, равный социальной норме, низкая. Вторая ставка выбирается из условия самоокупаемости энергокомпании. Дальнейшее развитие и анализ разработанная модель получила в настоящей статье.

Предполагается, что спрос потребителя на электроэнергию зависит от цены (P), дохода (I) и, возможно, ряда других факторов. Для построения двухставочного тарифа существенны только факторы цены и дохода, поэтому для функции спроса будем использовать обозначение $Q = Q(P, I)$. Пусть минимальный среднедушевой доход составляет I^{min} , тогда $Q = Q(P, I^{min})$ – функция спроса потребителей с минимальным среднедушевым доходом. Будем считать, что значение социальной нормы потребления Q_{soc} задано. Тогда определим первую социальную ставку как $P_{soc} = P(Q_{soc}, I^{min})$, где $P(Q, I^{min})$ – обратная функция спроса для потребителя с минимальным доходом. В этом случае все потребители приобретут электроэнергию в объеме не меньше социальной нормы Q_{soc} . Обозначим вторую ставку тарифа через $P_1 > P_{soc}$. Эту ставку будем выбирать из условия безубыточности энергокомпании. Самые низкодоходные потребители остановятся на уровне потребления Q_{soc} . Это потребители с доходом ниже, чем предельный доход I^{pred} . Значение I^{pred} определяется как корень уравнения $Q_{soc} = Q(P_1, I^{pred})$. Потребители с доходом $I > I^{pred}$ будут покупать электроэнергии в количестве больше Q_{soc} , остальные – только Q_{soc} .

Графическое изображение двухставочного тарифа с социальной нормой потребления, предложенного потребителю, представлено на рис. 1.

Предположим, что предельные издержки энергокомпании составляют c рублей, а постоянные издержки – FC рублей (в расчете на одного потребителя). Доход потребителей будем рассматривать как случайную величину с известной функцией распределения $F(x)$ и плотностью распределения $f(x)$.

Излишек производителя, полученный энергокомпанией при обслуживании «бедного» потребителя со значением дохода $I \leq I^{pred}$, равен

$$\Pi(P_{soc}, P_1, I) = (P_{soc} - mc) Q_{soc}.$$

Излишек, полученный при обслуживании «богатого» потребителя со значением дохода $I > I^{pred}$, равен

$$\Pi(P_{soc}, P_1, I) = (P_{soc} - mc) Q_{soc} + (P_1 - mc) (Q(P_1, I) - Q_{soc}).$$

Средний излишек производителя при двухставочном тарифе (P_{soc}, P_1) составит величину

$$\Pi(P_{soc}, P_1) = \int_{I^{min}}^{\infty} \Pi(P_{soc}, P_1, I) f(I) dI.$$

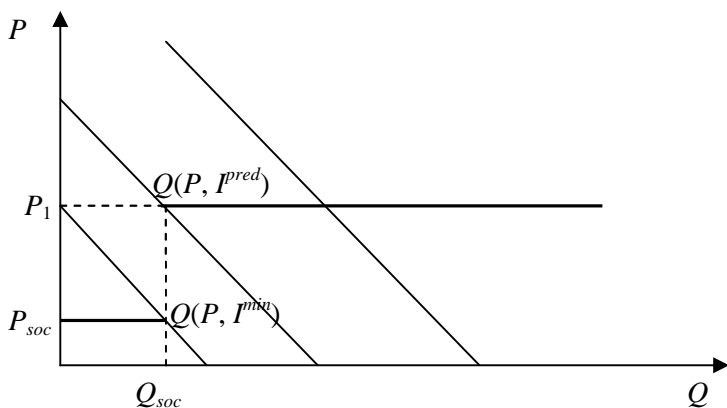


Рис. 1. Двухставочный тариф с социальной нормой потребления

Приравняв средний излишек энергокомпании к постоянным издержкам, получим условие нулевой прибыли:

$$Q_{soc} (P_{soc} - mc) + (P_1 - mc) \int_{I^{pred}}^{\infty} (Q(P_1, I) - Q_{soc}) \cdot f(I) dI = FC.$$

Найдем вторую ставку тарифа P_1 и предельный доход I^{pred} , решив систему уравнений:

$$(1) \quad \begin{cases} Q(P_1, I^{pred}) = Q_{soc}, \\ Q_{soc} (P_{soc} - mc) + (P_1 - mc) \int_{I^{pred}}^{\infty} (Q(P_1, I) - Q_{soc}) \cdot f(I) dI = FC. \end{cases}$$

Для решения системы (1) была написана программа в среде Lazarus¹. Входными параметрами явились социальная норма потребления Q_{soc} (кВт-ч) и минимальный доход потребителя I^{min} (руб. в месяц), которые задаются пользователем в активном окне программы.

Активное окно программы представлено на рис. 2.

Введите значения:	
Социальная норма, кВт-ч	<input type="text"/>
Минимальный доход потребителя, руб. в месяц	<input type="text"/>
<input type="button" value="Вычислить"/>	
Предельный доход, руб. в месяц	<input type="text"/>
Социальная ставка, руб./кВт-ч	<input type="text"/>
Экономически обоснованная ставка, руб./кВт-ч	<input type="text"/>
Доля потребителей, приобретающих только социальную норму	<input type="text"/>
Излишек потребителя при одноставочном тарифе	<input type="text"/>
Излишек потребителя при двухставочном тарифе	<input type="text"/>

Рис. 2. Активное окно программы для вычисления значений предельного дохода и двухставочного тарифа

В качестве значения минимального дохода в расчете на душу населения примем величину прожиточного минимума. В соответствии с постановлением Администрации Волгоградской области №146-п величина прожиточного минимума по Волгоградской области с 05.05.10 г. составляет 5185 рублей.

Выходными параметрами являются две ставки тарифа. Первая – низкая социальная – определяется как $P_{soc} = P(Q_{soc}, I^{min})$,

¹ Lazarus – свободная среда разработки программного обеспечения для компилятора Free Pascal Compiler. Интегрированная среда разработки предоставляет возможность кроссплатформенной разработки приложений в Delphi-подобном окружении.

т.е. значение обратной функции спроса для потребителя с минимальным доходом. Вторая – высокая P_1 как решение системы (1). Также в программе вычисляется значение предельного дохода I^{pred} , доля потребителей, потребляющих электроэнергию в пределах социальной нормы, и излишки потребителей при традиционном одноставочном тарифе и предлагаемом двухставочном.

Для построения модели спроса на электроэнергию в исследовании были использованы помесечные статистические данные о полезном отпуске электроэнергии населению (Q), цене на электроэнергию (P), а также доходах потребителя (I) за период с января 2008 г. по июнь 2010 г. для г. Волгограда и Волгоградской области [10, 11, 12]. В пакете *Statistica* с помощью МНК были получены параметры регрессионного уравнения степенной множественной регрессии спроса на электроэнергию бытового потребителя в зависимости от цены, личного располагаемого дохода и потребления электроэнергии в предыдущий период времени:

$$Q = (P)^{-0,08} (I)^{0,49} (Q_{t-1})^{0,62} e^{-3,05}.$$

Для моделирования функции издержек энергокомпании были использованы поквартальные статистические данные о фактических объемах полезного отпуска электроэнергии населению гарантирующим поставщикам г. Волгограда ОАО «Волгоградэнергообеспечение» и валовых издержках за период с января 2008 г. по июнь 2010 г. В результате вычислений в пакете *Statistica* получается функция издержек следующего вида:

$$C = 0,23 \cdot Q + 110,40.$$

Значит, увеличение объема выпуска электроэнергии на 1 кВт·ч в краткосрочном периоде обойдется компании в 23 коп. Кроме того, независимо от объема выпуска компания тратит на поддержание мощностей в рабочем состоянии в расчете на одного человека 110,4 руб. в месяц.

Предполагалось, что среднедушевой доход имеет логнормальное распределение, параметры которого оценивались на основе статистических данных за 2010 г. Была получена оценка плотности распределения:

$$f(x) = \frac{1}{\sqrt{2 \cdot p \cdot 0,15 \cdot x}} \cdot \exp\left(-\frac{(\ln x - 9,46)^2}{0,043}\right), \quad x > 0.$$

Для расчета ставок блочного тарифа на электроэнергию с социальной нормой потребления необходимо изначально определить размер данной нормы. По данным о потреблении электроэнергии домохозяйствами г. Волгограда и Волгоградской области был построен график распределения доли домохозяйств в зависимости от количества потребляемой электроэнергии (рис. 3).

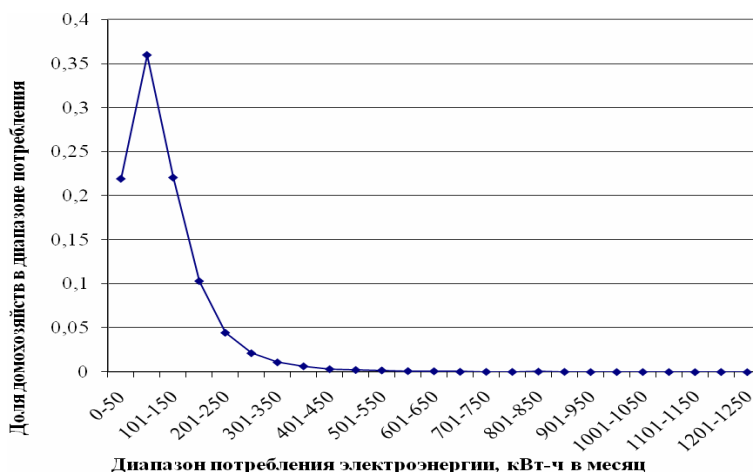


Рис. 3. Распределение доли домохозяйств в зависимости от количества потребляемой электроэнергии

Анализ данного графика показывает, что ежемесячное потребление электроэнергии наибольшей доли домохозяйств приходится на диапазон потребления 101–150 кВт·ч.

Домохозяйство — это экономическая единица в составе одного или нескольких лиц. Распределение численности домохозяйств в Волгоградской области по данным переписи населения 2002 г. представлено в таблице 2.

Таблица 2. Распределение жилых помещений по числу проживающих

Состав одного домохозяйства	
количество человек	%
1 человек	19,09
2 человека	29,06
3 человека	24,28
4 человека	18,23
5 и более человек	9,34

Так как большинство домохозяйств состоит из двух-трех человек, то для расчетов примем значение социальной нормы потребления электроэнергии в размере $Q_{soc} = 40$ кВт-ч в расчете на одного потребителя.

Результатом выполнения программы с заданными параметрами $Q_{soc} = 40$ кВт-ч и $I^{min} = 5185$ руб. является следующий двухставочный тариф на электроэнергию:

$$\begin{cases} P_{soc} = 0,74 & \text{за первые } Q \leq 40 \text{ кВт-ч,} \\ P_1 = 7,84 & \text{за остальные } Q > 40 \text{ кВт-ч.} \end{cases}$$

При этом значение предельного дохода $I^{pred} = 7503,06$ руб. в месяц. Используя оценку плотности распределения по доходам, определим процент потребителей с доходом ниже предельного – 0,14%. Таким образом, 0,14% потребителей электроэнергии останутся на уровне потребления в 40 кВт-ч, остальные будут приобретать большее количество электроэнергии.

Расходы потребителя при данном двухставочном тарифе вычисляются по формуле:

$$R(Q) = \begin{cases} 0,74 \cdot Q, & \text{если } Q \leq 40; \\ 0,74 \cdot 40 + 7,84 \cdot (Q - 40), & \text{если } Q > 40. \end{cases}$$

Например, если потребитель приобретет в месяц только 40 кВт-ч электроэнергии, то его расходы составят 29,6 рубля. Заметим, что при действующем в настоящий момент в Волгоградской области одноставочном тарифе 2,53 рубля за кВт-ч, расходы этого потребителя составляют 101,2 рубля. Таким образом, переход к предложенному двухставочному тарифу

уменьшает расходы данного потребителя на 71,6 рубля. Причем это уменьшение в расходах произойдет не за счет государственного бюджета, а за счет «богатых» потребителей. Например, потребитель, приобретающий 80 кВт-ч электроэнергии, заплатит за них по двухставочному тарифу 343,2 рубля. При действующем одноставочном тарифе 2,53 рубля за кВт-ч расходы этого потребителя составляют 202,4 рубля. Таким образом, расходы «богатого» потребителя увеличились на 140,8 рубля.

В программе рассчитываются также излишки потребителей при предложенном двухставочном тарифе (P_{soc}, P_1) и одноставочном тарифе P_{odn} , действующем в настоящий момент для населения г. Волгограда. Потребительский излишек – это разница между тем максимумом, который потребитель готов заплатить за каждый киловатт-час электроэнергии, и ценой этого киловатт-часа в соответствии с тарифом.

Средний потребительский излишек при одноставочном тарифе P_{odn} вычисляется по формуле:

$$S(P_{odn}) = \int_{I^{\min}}^{\infty} \left(\int_0^{Q(P_{odn}, I)} (P(Q, I) - P_{odn}) dQ \right) \cdot f(I) dI .$$

Излишек потребителя со значением дохода I при одноставочном тарифе P_{odn} представляет собой площадь заштрихованного треугольника, изображенного на рис. 4.

При действующем в 2010 г. тарифе на электроэнергию на территории г. Волгограда и Волгоградской области $P_{odn} = 2,30$ руб./кВт-час средний потребительский излишек составляет $S(P_{odn}) = 2276,36$.

Излишек для потребителя с доходом $I \leq I^{pred}$ рассчитывается как

$$S^I(I, P_{soc}) = \int_0^{Q_{soc}} (P(Q, I) - P_{soc}) dQ .$$

Излишек для потребителя с доходом $I > I^{pred}$ рассчитывается как

$$S^{II}(I, P_{soc}, P_1) = \int_0^{Q_{soc}} (P(Q, I) - P_{soc}) dQ + \int_{Q_{soc}}^{Q(P_1, I)} (P(Q, I) - P_1) dQ .$$

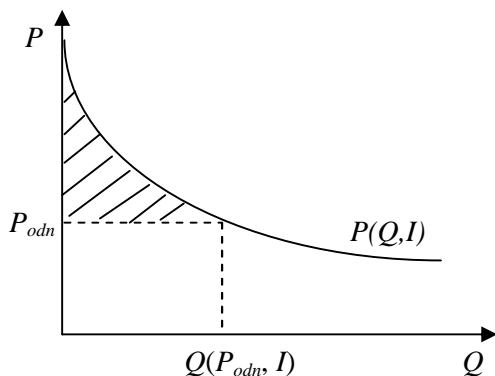


Рис. 4. Излишек потребителя со значением дохода I при одноставочном тарифе P_{odn}

Излишек потребителя со значением дохода I при применении социально-ориентированной модели двухставочного тарифа (P_{soc}, P_1) представляет собой площадь заштрихованной фигуры, изображенной на рис. 5.

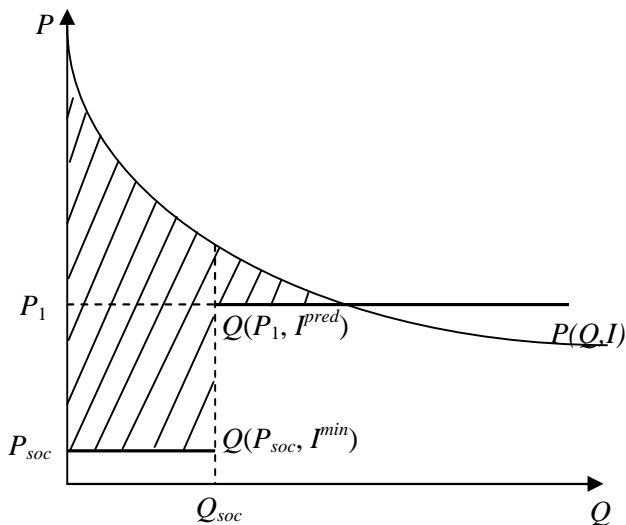


Рис. 5. Излишек потребителя со значением дохода I при социально-ориентированном двухставочном тарифе (P_{soc}, P_1)

Средний потребительский излишек покупателя при применении двухставочного тарифа (P_{soc}, P_1) рассчитывается по формуле:

$$S(P_{soc}, P_1) = \int_{I^{\min}}^{I^{pred}} S^I(I, P_{soc}) \cdot f(I) dI + \int_{I^{pred}}^{\infty} S^{II}(I, P_{soc}, P_1) \cdot f(I) dI.$$

Подставляя известные значение социального тарифа $P_{soc} = 0,74$ руб./кВт-ч, социальной нормы потребления $Q_{soc} = 40$ кВт-ч, предельного дохода $I^{pred} = 7503,06$ руб. и экономически-обоснованной цены $P_1 = 7,84$ руб./кВт-ч, получим следующее значение среднего потребительского излишка при двухставочном тарифе (P_{soc}, P_1): $S(P_{soc}, P_1) = 2305,31$.

Пусть S – выигрыш в среднем потребительском излишке при переходе от одноставочного к двухставочному тарифу: $S = S(P_{soc}, P_1) - S(P_{odn})$. Тогда при $S(P_{odn}) = 2276,36$ и $S(P_{soc}, P_1) = 2305,31$ выигрыш в среднем потребительском излишке при переходе от одноставочного тарифа к двухставочному составит $S = 2305,31 - 2276,36 = 28,95$.

В таблице 3 представлены результаты моделирования социального тарифа P_{soc} , экономически-обоснованной цены на электроэнергию P_1 , предельного дохода I^{pred} и выигрыша в среднем потребительском излишке при различных начальных значениях социальной нормы потребления Q_{soc} .

Регулирующие органы, используя данную методику формирования тарифов на электроэнергию для бытовых потребителей, могут утверждать ставки двухставочного тарифа для населения, основываясь, например, на изменении значения выигрыша в потребительском излишке. С учетом данного показателя наиболее предпочтительным является следующий тариф для населения: потребитель оплачивает социальную цену $P_{soc} = 2,03$ руб./кВт-ч за первые $Q_{soc} = 37$ кВт-ч потребленной электроэнергии, а за остальные, потребленные сверх этой нормы киловатт-часы, цена составит $P_1 = 2,63$ руб./кВт-ч.

Для получения льгот на оплату жилищно-коммунальных услуг в настоящий момент малоимущие граждане должны каждые полгода предоставлять в органы социальной защиты ряд документов, в частности, справки о доходах каждого члена

семьи. Органы социальной защиты работают с предоставленными документами, работа оплачивается за счет государственного бюджета. Предложенный двухставочный тариф полностью упраздняет работу органов социальной защиты, связанную с назначением субсидий на оплату электроэнергии. Потребители сами дифференцируют себя на «богатых» и «бедных», регулируя уровень электропотребления.

Таблица 3. Результаты расчета блочного тарифа и выигрыша в среднем потребительском излишке при применении различных значений социальной нормы потребления¹

Q_{soc} , кВт-ч	P_{soc} , руб./кВт-ч	P_1 , руб./кВт-ч	предельный доход I^{pred} , руб.	процент населения с доходом ниже I^{pred}	S , руб.
36	2,54	1,08	5182,6	0,00	43,23
37	2,03	2,63	5367,7	0,00	51,08
38	1,43	4,23	6123,6	0,00	45,27
39	1,02	6,02	6832,4	0,01	36,54
40	0,74	7,84	7503,1	0,14	29,12
41	0,53	10,13	8220,0	0,98	18,24
42	0,39	13,49	9034,9	4,73	2,54

Следует подчеркнуть, что предлагаемая модель тарифа не перекладывает бремя перекрестного субсидирования малообеспеченных слоев населения на более богатые слои населения, поскольку каждый потребитель самостоятельно выбирает тот объем электроэнергии, который он будет потреблять. Обеспеченные потребители также имеют возможность получить часть электроэнергии по дешевому (социальному) тарифу. Их никто не заставляет потреблять больше – это их выбор.

¹ При расчетах полагалось, что $Q_{t-1} = 60$ кВт-ч.

Предлагаемая модель позволяет сформировать тарифную систему таким образом, чтобы защитить низкодходную категорию потребителей электроэнергии от резкого повышения цены на электроэнергию в процессе доведения тарифов до экономически обоснованного уровня. Наряду с этим данная модель обеспечивает безубыточность поставщиков электроэнергии. Внедрение предлагаемой модели двухставочного тарифа позволит повысить финансовую устойчивость и инвестиционную привлекательность энергокомпаний, а также будет способствовать постепенной ликвидации перекрестного субсидирования населения промышленными потребителями в процессе реформирования электроэнергетики.

Литература

1. БОГАЧКОВА Л.Ю., ЗАЙЦЕВА Ю.В. *О развитии системы цен в электроэнергетике: моделирование тарифного меню для населения* // Управление большими системами. – 2006. – Вып. 12-13. – С. 32–45.
2. БОГАЧКОВА Л.Ю. *Совершенствование управления отраслями российской энергетики: теоретические предпосылки, практика, моделирование.* – Волгоград: Волгоградское научное издательство, 2007. – 427 с.
3. ЗАБЕЛЛО Е. П. *О тарифной политике в электроэнергетике на современном этапе и на ближайшую перспективу* // Промышленная энергетика. – 2005. – №11. – С. 2–6.
4. ЗАЙЦЕВА Е.Е. *Моделирование ценового механизма согласования интересов субъектов розничного рынка электроэнергии* // Научно-технические ведомости СПбГПУ. Экономические науки. – 2007. – №4. – С. 231–233.
5. ЗАЙЦЕВА Ю.В. *Математические модели ценообразования в естественной монополии.* – ВолГУ. - Волгоград: Изд-во ВолГУ, 2006. – 117 с.
6. ОРЛОВ В.С., ПАПКОВ Б.В., ЕРШОВ Е.П., КОПАЛОВ Л.Н. *Анализ электропотребления и тарифов для бытовых потребителей* // Промышленная энергетика. – 1997. – №6. – С. 8–10.

7. ШАГОВ А.В., КОЛБИН В.В. *О некоторых подходах к управлению тарифной политикой в топливно-энергетическом комплексе региона* // Экономика и математические методы. – 2005. – № 41:1. – С. 54–64.
8. ЭЙСФЕЛЬД А.А. *Моделирование двухставочного тарифа на электроэнергию с учетом социальной нормы потребления на примере населения г. Волгограда* // Управление большими системами. – 2010. – Вып. 28. – С. 197–210.
9. BROWN S., SIBLEY D. *The theory of public utility pricing*. Ch. 3, 4, 5. – Cambridge University Press, 1986. – P. 26–129.
10. <http://www.energosal34.ru> – *Официальный сайт ОАО «Волгоградэнергообит»* (дата обращения 15.10.2010).
11. <http://www.volgastat.ru> – *Официальный сайт Территориального органа Федеральной службы государственной статистики по Волгоградской области* (дата обращения 01.10.2010).
12. <http://www.gks.ru> – *Официальный сайт Федеральной службы государственной статистики* (дата обращения 12.10.2010).

ABOUT ONE APPROACH TO WORKING OUT SOCIALLY FOCUSED DOMESTIC ELECTRICITY RATES

Anastasia Eisfeld, Volgograd State University, assistant (aeisfeld@yandex.ru).

Abstract: The article is devoted to modeling socially-focused blocked domestic electricity rate scales. The blocked rate is calculated on the basis of Volgograd and the Volgograd region statistics. Properties of the proposed tariff scheme are numerically analyzed.

Keywords: two-rate tariff, tariffs for the electric power, modeling of tariffs, social norm of consumption, domestic consumers.

Статья представлена к публикации членом редакционной коллегии А.А. Ворониным

УДК 519.711.2:614.4

ББК в.6.3.5

МОДЕЛЬ РАСПРОСТРАНЕНИЯ ВИЧ-ИНФЕКЦИИ В РЕЗУЛЬТАТЕ СОЦИАЛЬНОЙ ДЕЗАДАПТАЦИИ

Романюха А. А.¹

(Учреждение Российской Академии Наук Институт вычислительной математики, Москва)

Носова Е. А.²

(Федеральное государственное учреждение Научно-исследовательский институт организации и информатизации здравоохранения Минздравсоцразвития, Москва)

Цель исследования состоит в разработке методов для оценки эффективности мер по контролю распространения вируса иммунодефицита человека (ВИЧ) на территории России. Региональные различия по уровню распространенности ВИЧ-инфекции тесно связаны с различиями в распространённости алкогольной и наркотической зависимости, которые можно рассматривать как характеристики социальной дезадаптации населения. Учет влияния этих процессов на формирование групп с высоким риском инфицирования ВИЧ позволяет описать различия регионов России по масштабу эпидемии и оценить эффективность мер противодействия.

В работе рассмотрены модель взаимодействия процессов социальной дезадаптации и распространения ВИЧ в неоднородной популяции и задача идентификации параметров этой модели по данным регионов России.

Ключевые слова: модель, социальная дезадаптация, ВИЧ-инфекция.

¹ Алексей Алексеевич Романюха, доктор физико-математических наук, профессор (eburg@im.ras.ru).

² Екатерина Александровна Носова, аспирант (nosova@mednet.ru).

Введение

Распространение вируса иммунодефицита человека (ВИЧ) является одной из актуальных проблем современного здравоохранения. Это связано с такими особенностями инфекции, как невозможность элиминировать вирус из инфицированного организма, предотвратить развитие синдрома приобретенного иммунодефицита (СПИД), высокая стоимость поддерживающей терапии.

В связи с этим возникает задача оценки эффективности мер по противодействию распространению ВИЧ-инфекции. Разработке единой методики ее решения препятствует неравномерность распространения ВИЧ-инфекции как по территории Земли в целом, так и внутри отдельных государств [18]. Причины данного явления могут быть различными. Так, например, в США заболеваемость СПИДом по штатам достоверно коррелирует с расовым составом населения [2]. В России не удается установить факторы неоднородности аналогичным методом.

Особенности передачи ВИЧ обуславливают значительную изменчивость вероятности инфицирования за счет изменения поведения. Для инфекций, обладающих таким свойством, характерно объединение индивидов в группы, замкнутые относительно остальной популяции вследствие поведенческих особенностей. Поэтому при выполнении определенных условий инфекция может быть локализована внутри таких групп. Важно отметить, что скорость распространения вируса внутри групп гораздо выше, чем между группами, что может приводить к скачкообразным изменениям скорости распространения инфекции, чередующимся с периодами с постоянными уровнями инфицирования.

С целью прояснить особенности этого явления в распространении ВИЧ по территории Российской Федерации в данной работе предложена модель, описывающая не только процессы передачи вируса, но и формирования групп с различной вероятностью инфицирования.

1. Постановка задачи

1.1. ПРОБЛЕМА ПРОГНОЗИРОВАНИЯ РАСПРОСТРАНЕНИЯ ВИЧ НА ТЕРРИТОРИИ РОССИИ

Рекомендации ведущих зарубежных организаций (ВОЗ, ООН) по построению программ противодействия распространению ВИЧ-инфекции на территории стран СНГ, как правило, опираются на прогнозы и оценки, полученные с помощью математических моделей, идентифицированных на данных стран Африки. Высокий уровень заболеваемости и распространенности в данных двух группах территорий [18], а также хорошо отработанная на данных африканских исследований модель распространения заболевания позволяют международным экспертам распространять результаты на другие развивающиеся страны.

Проведение таких, казалось бы, адаптированных к местным условиям мероприятий на территории стран СНГ демонстрирует на практике весьма низкую отдачу [5, 8]. Мы предполагаем, что это вызвано тем, что механизмы формирования групп повышенного риска инфицирования ВИЧ на территории России и стран Африки существенно различаются.

Анализ данных [21] позволяет утверждать, что важную роль в распространении ВИЧ на территории России играет социальная дезадаптация населения. Социальная дезадаптация – неспособность индивида к социальной среде. Проявлениями социальной дезадаптации в индивидуальном поведении являются поступки, запрещаемые нормами права, нравственности, правилами общежития. Сюда входят разнообразные виды девиантного поведения: алкоголизм, наркомания, аморальное поведение, нарушение социальных норм.

Индивидуальный уровень социальной дезадаптации может значительно варьироваться и меняться со временем и должен влиять на эпидемиологические характеристики индивида: скорость заражения, диагностики заболевания, качество лечения. Измерить его можно, например, оценив количество ресурсов, необходимое на минимизацию асоциального поведения индивида. В этом смысле проявлением максимальной дезадаптации является развитие наркологической патологии – заболевания,

вызванного злоупотреблениями психоактивных веществ: алкоголя, различных видов наркотиков и ненаркотических токсинов, которые при введении в организм человека изменяют его поведение. В таком состоянии увеличивается риск инфицирования ВИЧ.

Социально дезадаптированные индивиды составляют основную часть резервуара ВИЧ-инфекции и ряда других социально-опасных заболеваний на начальных этапах распространения (рис. 1). В связи с этим остро становится вопрос об оценке риска инфицирования социально адаптированного населения.

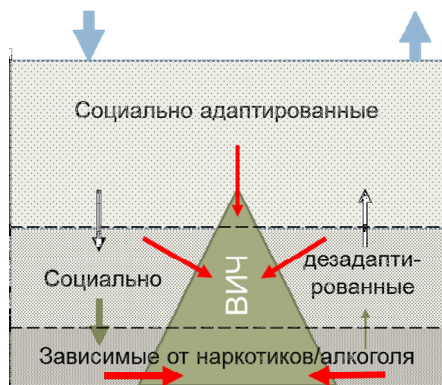


Рис. 1. Схематическое представление социальной дезадаптации и ее роли в распространении ВИЧ-инфекции. Стрелки – потоки индивидов.

1.2. ПРОБЛЕМА ОЦЕНКИ ЭФФЕКТИВНОСТИ КОНТРОЛЯ РАСПРОСТРАНЕНИЯ ВИЧ.

Одна из трудностей, возникающих перед исследователями распространения ВИЧ-инфекции, заключается в отсутствии единого понимания целей контроля данного процесса.

Так, например, в Декларации Тысячелетия ООН от 2000 г, с одной стороны, стоит задача остановить распространение заболевания, решением которой является скорейшая элиминация вируса из популяции и защита здорового населения от возможного инфицирования, а с другой – увеличение продолжительности и повышение качества жизни ВИЧ-инфицированных инди-

видов, их социализации, что ведет к увеличению времени жизни вируса в популяции. Для согласования таких противоречивых целей требуется комплексный критерий эффективности.

Согласно обзору [25] за период 1984-2000 г было опубликовано 38 работ, посвященных вопросам эффективности вмешательств направленных на борьбу с распространением ВИЧ-инфекции. Среди них большинство (23 работы) касаются ситуации в Африке и лишь 3 статьи рассматривают мероприятия в Восточной Европе. В качестве показателя, характеризующего эффективность, используется число предотвращенных случаев инфекции. В каждой статье рассмотрена экономическая эффективность отдельно взятых мероприятий, а не всего комплекса в целом. Такой подход не учитывает возможности синергии негативных последствий вмешательств, преследующих разные цели.

Например, две сходные по внешним признакам меры воздействия: снижение социальной дезадаптации неинфицированных и инфицированных ВИЧ-индивидов могут оказывать различное влияние на распространение вируса.

При воздействии на когорту, свободную от ВИЧ, снижается размер потенциального резервуара инфекции при сохранении достаточно высокого уровня инфекционности. При социальной адаптации уже инфицированных индивидов снижается инфекционность, при этом размер резервуара продолжает расти, к тому же повышается скорость проникновения вируса в социально адаптированную популяцию за счет поступления в нее инфицированных индивидов.

Влияние описанных процессов на заболеваемость и распространенность ВИЧ-инфекции оказывается различным (рис. 2). При достижении распространенностью ВИЧ некоторого критического значения фактор размера инфекционного резервуара становится доминирующим над фактором инфекционности и соотношение эффективности по кумулятивному числу предотвращенных случаев стратегий меняется.

В то же время, с точки зрения экономической эффективности работа по адаптации инфицированных индивидов сопряжена с меньшими затратами ввиду охвата меньшей популяции и выглядит более выгодной в условиях ограниченного бюджета.

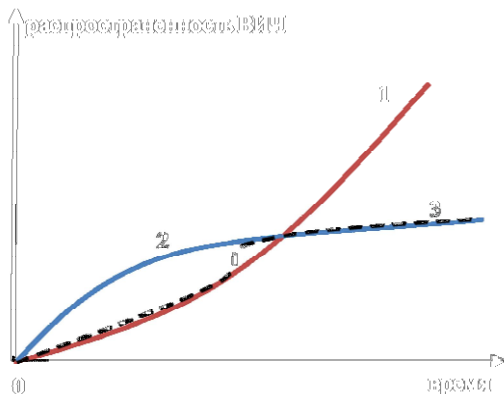


Рис. 2. Схематическое представление результата воздействия различных стратегий на распространение ВИЧ.

1 – «адаптация инфицированных индивидов»,
2 – «адаптация неинфицированных индивидов», пунктиром отмечена оптимальная стратегия.

Формулировка комплексного критерия эффективности позволила бы давать более достоверные оценки результатам деятельности системы здравоохранения по противодействию ВИЧ.

1.3. ЦЕЛИ И МЕТОДЫ

В связи с описанными выше проблемами перед исследователями стоит ряд целей. Во-первых, прояснить механизм формирования заболеваемости ВИЧ-инфекцией в результате социальной дезадаптации населения, а также оценить количественные характеристики связи этих явлений. Во-вторых, необходимо оценить эффект, оказываемый на распространение ВИЧ существующими программами противодействия. Все это позволит сформировать превентивную стратегию, направленную на оптимизацию компромиссного критерия.

Для достижения этих целей был проведен анализ данных, результаты которого были опубликованы ранее [21]. В настоящей работе представлен второй шаг исследования – построение модели эпидемической динамики ВИЧ-инфекции в неоднородной популяции, которая позволила бы отслеживать и прогнози-

ровать явления фазовых переходов эпидемии, исходя из формирования групп риска в результате социальной дезадаптации индивидов.

1.4. ПРЕДПОЛОЖЕНИЯ

Источником и резервуаром ВИЧ является человек на всех стадиях заболевания. Возбудителем является вирус, передача которого осуществляется по одному из трех путей: при половых контактах, парентерально (через кровь), вертикально (от инфицированной матери к ребенку). При этом каждый тип контакта отличается своей характерной частотой передачи. (табл. 1)

Таблица 1. Частота инфицирования [9]

Тип контакта	Частота
При уколе	0,003
При совместной инъекции	0,7-0,9
При переливании инфицированных кровепродуктов	0,9-1
При незащищенном сексуальном контакте	0,0003-0,005
При родах	0,3-0,5
При грудном вскармливании	0,13-0,3

По клиническим и эпидемиологическим характеристикам следует выделить среди распространителей ВИЧ группу активных распространителей – индивидов, чей вклад в эпидемическую динамику заболевания является основным. Для случая эпидемии ВИЧ на территории России возраст активных распространителей находится в интервале 15-49 лет [6]. Большинство случаев ВИЧ-инфекции, приходящееся на возраст до 15 лет, связано с вертикальной передачей вируса. Детская популяция представлена сексуально неактивными индивидами и отличается практически полным отсутствием горизонтального (между индивидами, не являющимися родителем и потомком) распространения заболевания. Инфицирование детского населения является следствием процессов, происходящих во взрослой популяции.

Поведение, при котором индивид не прибегает к мерам защиты от инфицирования, называется *рискованным*. Активные распространители ВИЧ неоднородны по склонности к рискованному поведению и частотам контактов.

Для ВИЧ-инфекции, как и для других заболеваний, передаваемых половым путем, характерно наличие «ядра суперраспространителей» [1]. *Ядерная группа (core-group)* — множество индивидов, уровень рискованного поведения которых достаточен для порождения жизнеспособных цепей передачи заболевания в популяции, и отсутствие которых приводит к полному исчезновению инфекции.

Изоляция ядра от остальной популяции относительна. Существуют индивиды, вступающие в рискованные контакты с представителями обеих групп. Они образуют группу-мост (*bridge-group*) [14], через которую возбудитель передается из ядра в основную популяцию (рис. 3).

Группы населения, находящиеся в условиях повышенного риска инфицирования ВИЧ ввиду социально-экономических или поведенческих причин, называются *группами риска*.

Для ВИЧ-инфекции принято выделять следующие группы риска: мужчины би- и гомосексуальной ориентации, работники коммерческого секса и потребители инъекционных наркотиков – представители перечисленных категорий формируют ядерную группу. Больные, зависимые от переливаний крови, вахтовые рабочие, водители дальнобойщики, дети-сироты могут взаимодействовать с ядерной группой, однако в ее отсутствие не способны создавать устойчивые цепи передачи вируса. Они образуют мост между ядром и основной популяцией.

Частота проникновения вируса за пределы ядерной группы определяется различными факторами: проницаемость границы для взаимодействий, размер группы-моста – которые в свою очередь определяются социально-экономическими, географическими условиями каждой конкретной территории. Региональная неоднородность масштабов течения эпидемии ВИЧ, наблюдаемая во всем мире, тесно связана с величиной данной частоты.

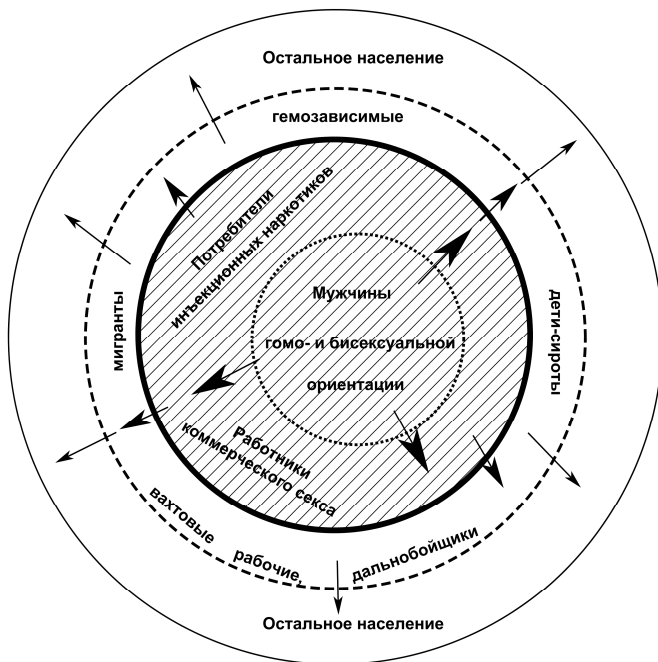


Рис. 3. Структура популяции относительно риска распространения ВИЧ. Штриховкой обозначена ядерная группа, прерывистой границей – группа-мост

UNAIDS¹ предлагает следующий сценарий развития эпидемии ВИЧ-инфекции и СПИДа. При проникновении вируса на новую не освоенную им территорию в течение некоторого времени наблюдается низкоуровневая эпидемия (low-level), при которой инфекция распространяется преимущественно половым путем в относительно небольшой и замкнутой группе индивидов с наивысшим уровнем рискованного поведения. Затем, по мере распространения вируса в ядерной группе начинается концентрированная эпидемия (concentrated), при которой абсолютная распространенность не превышает совокупной числен-

¹ UNAIDS – Объединенная международная программа противодействия распространению ВИЧ/СПИД ООН.

ности групп риска и уязвимого населения. В случае недостаточно эффективных мер противодействия эпидемия может достичь генерализованной (generalized) фазы, при которой ВИЧ прочно обосновывается во всей популяции.

Продолжительность и выраженность фаз связаны со временем проникновения вируса на территорию, а также проницаемостью границ ядерной и связующей группы. Последняя обусловлена традициями, социально-экономическими условиями и качеством организации медицинской службы каждого конкретного региона.

Согласно данным отечественной медицинской статистики в период 2000-2002 год на территории России наблюдалась «вспышка эпидемии ВИЧ/СПИДа» (рис. 4) – резкое увеличение распространенности и заболеваемости ВИЧ-инфекцией. До этого времени в течение 10 лет с момента выявления первого случая данного заболевания на территории России в 1987 г. практически 100% выявленных новых случаев приходились на половой путь передачи в результате гомосексуальных контактов. Редкие случаи парентеральной передачи вируса были связаны с использованием загрязненного оборудования при медицинских манипуляциях [6].

По мнению экспертов [3] явление, наблюдавшееся в 2000-2002 гг, обусловлено проникновением вируса в популяцию потребителей наркотиков и освоением ядерной группы. Подтверждает это и смена преобладающего способа передачи в выявленной заболеваемости. До 2000 года половой путь составлял до 100% случаев в выявленной заболеваемости.

Таким образом, начиная с 2000 г в России наблюдается концентрированная эпидемия. Вопрос о возможности перехода процесса в генерализованную фазу пока остается открытым. Сигналами к этому может служить повторная смена преобладания путей передачи. Подобное явление уже наблюдается в некоторых регионах России.

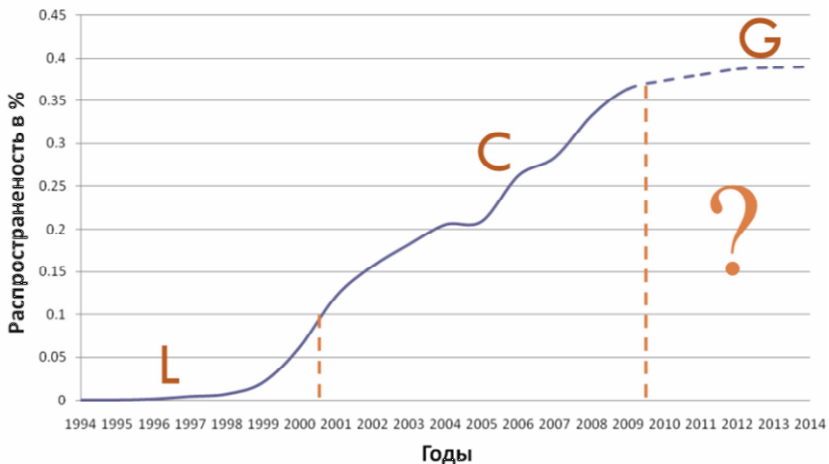


Рис. 4. Распространённость (в %) ВИЧ-инфекции и СПИДа среди населения Российской Федерации в 1994-2009 гг. и прогноз UNAIDS. Буквами обозначены фазы эпидемии: L – низкоуровневая, C – концентрированная, G – генерализованная [6]

2. Обзор литературы

2.1. НАПРАВЛЕНИЯ РАЗВИТИЯ МОДЕЛЕЙ ДИНАМИКИ И КОНТРОЛЯ ВИЧ-ИНФЕКЦИИ

Множество работ, доступных на сегодняшний день под ключевыми словами «математическая модель, ВИЧ/СПИД», впечатляет разнообразием методов и целей исследования и продолжает активно пополняться. Располагая работы в хронологическом порядке можно обнаружить, что развитие в данной предметной области происходило одновременно по нескольким направлениям.

Следует выделить три основные задачи: исследование патогенеза в рамках одного организма, прогнозирование развития эпидемической ситуации, прогнозирование последствий в экономике и социальной сфере.

Решение последней задачи является одним из важнейших направлений при принятии управленческих решений. Как правило, для прогнозирования последствий специалисты прибегают к корректировке существующих экономико-демографических

моделей экспертными оценками влияния эпидемии ВИЧ/СПИД. Данное направление довольно подробно рассмотрено в работе Денисова и Сакевич [4].

Задача моделирования патогенеза ВИЧ-инфекции в рамках одного организма направлена на изучение биологических процессов взаимодействия вируса с организмом хозяина. Первые работы по данному направлению появились несколько позже эпидемиологических моделей (в 1990 г) поскольку на момент открытия вируса исследователи обладали первыми статистическими данными о заболеваемости и распространенности нового заболевания, а на изучение свойств инфекционного агента требовалось определенное время.

Основные вопросы, которые исследуют с помощью иммунологических моделей: естественное течение ВИЧ-инфекции, роль терапии и поиск вакцины от вируса иммунодефицита человека.

Возникновение и развитие иммунологических моделей связано, главным образом с работами Перельсона и Нельсона [20], Новак и Мэй [22], Снедекор [24] и ряда других исследователей.

Первые модели эпидемиологической динамики вируса иммунодефицита человека в популяции носителей появились в 1983 г. Их разработка мотивировалась нуждами страховых компаний [15].

Эпидемиологические модели в качестве главной задачи содержат прогнозирование основных эпидемических показателей распространения ВИЧ-инфекции в популяции. Данный класс работ отличается многообразием методов и направлений исследования, речь о которых пойдет далее.

2.2. ПОПУЛЯЦИОННЫЕ МОДЕЛИ ЭПИДЕМИЧЕСКОЙ ДИНАМИКИ ВИЧ-ИНФЕКЦИИ

Популяционный (компаратментный или осредненный) подход основан на редукции исследования сообщества индивидов к рассмотрению классов зависимости отношения индивидов к инфекции: восприимчивые, инфекционные, иммунные и др. Дальнейшая работа осуществляется только в терминах состояния этих классов, детерминированным или стохастическим способом. В детерминированных моделях динамика состояний

классов описывается с помощью аппарата интегро-дифференциальных уравнений. Стохастический подход осуществляется с привлечением цепей Маркова.

Наиболее подходящую классификацию популяционных моделей эпидемической динамики ВИЧ/СПИД приводит в своем обзоре Хаберман [15]. Дальнейшее развитие эта классификация получила в работах Хаймана и Ли [17]. Критерием классификации моделей является учет неоднородности популяции.

К классу агрегированных моделей относятся работы, в которых предполагается однородность исследуемой популяции. Такие модели применяются при изучении динамики инфекционности индивидов в ходе естественного течения ВИЧ-инфекции и под влиянием лечения. Сокращенно их принято называть SP-модели (stage progression).

Развитие детерминированных агрегированных моделей берет начало в работе Андерсона и Мэя [9]. Изучением данного класса моделей занимались в разное время Кастильо-Чавез [13] с учениками и ряд других исследователей.

В полной мере стохастические агрегированные модели исследованы Лонджини, Кларком и Саттен с соавторами [19].

К следующей группе следует отнести модели, в которых прогрессия заболевания от инфицирования ВИЧ к СПИДу не учтена. Главной задачей таких работ является исследование роли гетерогенности популяций источников и вирусов, поэтому они получили название структурных или DI (differential infectivity) и DS (differential susceptibility) в зависимости от исследования гетерогенности инфекционности или восприимчивости носителей соответственно.

Исследование дискретных моделей начиналось также с работ Андерсона и Мэя и Кастильо-Чавеза. Вопросами пространственной неоднородности на примере южноафриканской провинции Квазулу-Натал занимались Блоуэр [26] и Смит. Большая часть работ в данном направлении посвящена теоретическому изучению свойств моделей.

Для решения более реалистичных задач используют комбинированные DISP и DSSP модели.

Исследование распространения ВИЧ в популяции с учетом изменения инфекционности индивидов в процессе старения и

прогрессии заболевания подробно рассмотрено в работах Лонджини.

Роль качества выявления инфицированных индивидов с учетом стадии заболевания рассмотрена в работах Хаймана, Ли и Стэнли, Алена [7], Аразозы [11].

Модели возможного развития эпидемии ВИЧ/СПИД в условиях существования вакцины берут начало от ранних иммунологических моделей Новака, МакЛин и Блоуэр. Первой в 1991 г была предложена модель Андерсона-Мэя-Гупты [10]. Основной же вклад в развитие данного направления внесли Блоуэр, Гумель, Дэйвенпорт и их последователи.

В данную классификацию не укладывается обширная группа работ по моделированию взаимодействия ядерной группы с основной популяцией с использованием осредненного подхода.

Главной их особенностью является рассмотрение усредненной динамики групп не отдельных индивидов, а кластеров (партнерств с различным числом участников, семейных пар).

Модели взаимодействия ядерной группы с основной популяцией были исследованы коллективом под руководством Сиеха [16] и некоторыми отдельными авторами. Каждая модель разрабатывалась для случая отдельного государства и в качестве ядерной группы рассматривает представителей только одной какой-либо связки «группа риска-уязвимая группа», отражающей доминирующий путь передачи. Попытки распространения результатов и метода оценки параметров на территории, где выраженность путей передачи не настолько сильна, потерпели неудачу. Данные работы оказались недостаточно хороши и в прогностическом плане, так как не учитывали процессы формирования ядра суперраспространителей и возможного влияния на них мероприятий противодействия.

2.3. ИМИТАЦИОННЫЕ МОДЕЛИ ЭПИДЕМИОЛОГИИ ВИЧ.

Модели, предназначенные для изучения интегральных характеристик популяции как результата множества локальных взаимодействий ее отдельных членов – агентов – следует называть имитационными. Первые попытки имитационного моделирования распространения ВИЧ/СПИД появились в 2000 г. Среди

множества имитационных моделей по способу управления параметрами удается выделить три основных класса.

В мультиагентных моделях центральным объектом является действующая единица популяции – агент, свойства которого и становятся ведущими в динамике системы. Агентом может являться как отдельный индивид, так и группа индивидов, объединенных по некоторому признаку.

При реализации мультиагентных моделей устанавливается множество агентов, их основные характеристики и способы их измерения с некоторым заданным начальным распределением, а также правила взаимодействия, в результате которых осуществляется эффективный контакт. При исследовании мультиагентных моделей прибегают к использованию аппарата стохастических процессов и теории вероятностей. Исследование данного вида моделей эпидемической динамики ВИЧ-инфекции тесно связано с именем Бертрана Увера [12].

При сетевом моделировании агенты как таковые отсутствуют, на первый план выходят характеристики связей между отдельными частями системы, по которым осуществляется взаимодействие. В связи с данным направлением широкое распространение получили комплексные сети – модели сетей со сложной топологией. В данном направлении основным математическим аппаратом является теория графов и другие разделы дискретной математики.

Исследование распространения ВИЧ в комплексных сетях, а также разработка специализированного программного обеспечения для реализации данного вида моделей проводится Слот и Бухановским [23]. Моделирование сети агентов представляет собой попытку объединения двух первых технологий. Это самое молодое направление в имитационном моделировании задач эпидемической динамики ВИЧ-инфекции в популяции носителей также берет начало в работах Слот и Бухановского.

Одним из основных недостатков имитационных моделей является невозможность получения решения задачи в аналитическом виде и необходимость исследовать выходные данные вероятностными методами.

3. Социально-эпидемиологическая модель распространения ВИЧ-инфекции

Классическая модель [1] описания инфекционных объектов на популяционном уровне для случая системы «хозяин-микроразит» подразумевает разделение общей популяции хозяев на три категории: восприимчивых, инфицированных и иммунных индивидов. Для ВИЧ-инфекции сообщения о случаях выздоровления с последующей иммунизацией носят характер редких данных с сомнительной достоверностью. Поэтому в нашей модели популяцию из $N(t)$ индивидов в любой момент времени t можно представить как совокупность двух групп индивидов: инфицированных $I(t)$ восприимчивых к ВИЧ $S(t)$.

$$(1) \quad N(t) = I(t) + S(t).$$

Как было упомянуто в обзоре литературы, существующие модели эпидемической динамики оперируют лишь отдельными частями ядерной группы и рассматривают распространение вируса только одним путем. Из постановки задачи следует, что фазовые переходы эпидемии ВИЧ/СПИДа представляют собой выход вируса за пределы рассматриваемых групп и связаны со сменой преобладающего пути передачи. В предлагаемой модели популяция носителей структурирована по отношению к риску развития наркомании или алкоголизма (табл. 2).

Таблица 2. Переменные модели – численности популяций

Развитие наркологической патологии	Восприимчивые	Инфицированные
Социально адаптированные	S_G	I_G
Повышенный риск зависимости	S_S	I_S
Хронический алкоголизм	S_A	I_A
Наркомания	S_D	I_D

В группы с индексом G были включены социально адаптированные индивиды, с S — лица, для которых факт наличия наркологического заболевания в острой фазе не установлен медицинским учреждением, но при этом влияние факторов риска превалирует над факторами защиты. В группы с индексом

A были отнесены индивиды с установленным диагнозом «хронический алкоголизм», а с *D* — «наркомания».

Предполагается, что выявление наркоманий и алкоголизма построено таким образом, что наркотическая зависимость имеет приоритет при постановке на учёт, как более тяжёлая в лечении, а постановка на учёт по двум патологиям невозможна, т. е. группы *D* и *A* не пересекаются.

Состояние алкогольной и наркотической зависимости не является необратимым. После постановки на учёт в наркологическом диспансере индивид находится под наблюдением врача соответствующей специализации, который предлагает пациенту программу лечения. В результате может наступить стойкая ремиссия, которая повлечет за собой условное выздоровление индивида. В группы *A* и *D* входят только те индивиды, для которых ремиссия имеет среднюю продолжительность, недостаточную для объявления его здоровым с точки зрения системы здравоохранения. Индивиды в продолжительной (3-5 лет) ремиссии в меньшей степени социально дезадаптированы, но сохраняют повышенный риск развития наркологической патологии. Поэтому таких индивидов мы отнесли к когорте *S*.

Модель представляет собой систему восьми обыкновенных дифференциальных уравнений, каждое из которых описывает динамику численности в одной из восьми групп населения. На рис. 5 приведена схема переходов индивида между состояниями, соответствующая уравнениям модели.

Приток новых индивидов в популяцию возможен в результате дожития родившихся до возраста 15 лет и миграции из других территорий. При этом в модели не учтена возможность «приноса» инфекции мигрантами. Данные Федерального научно-методического центра по борьбе с распространением ВИЧ/СПИД указывают на незначительный вклад миграции в эпидемическую динамику.

Уменьшение численности популяции происходит в результате смерти индивидов или выбытия в места лишения свободы, для социально дезадаптированных.

При построении модели мы предположили, что процесс социальной дезадаптации в большинстве случаев можно описать как последовательные переходы между состояниями с относи-

тельно близкими уровнями социальной адаптации. Первоначально адаптированный индивид может оказаться под влиянием факторов риска развития наркологической патологии. Только находясь в этом состоянии индивид имеет возможность приобрести диагноз «хронический алкоголизм» или «наркомания». Вместе с тем индивид может преодолеть влияние этих факторов и вернуться в группу социально адаптированных индивидов.

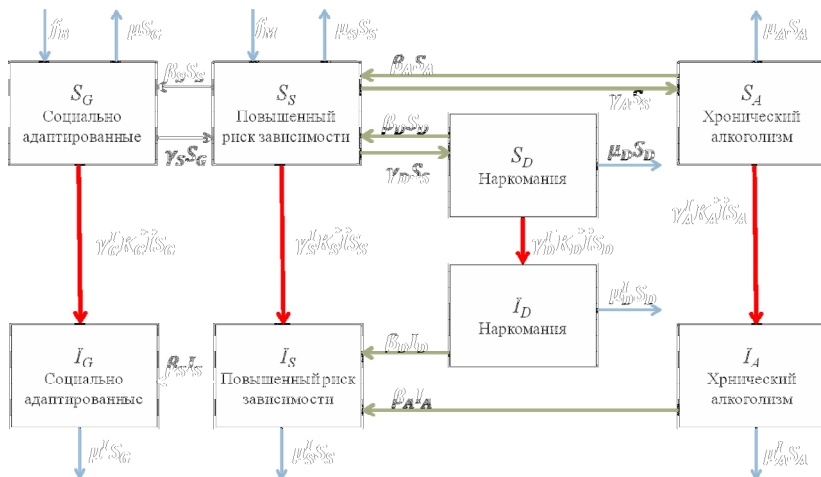


Рис. 5. Схема переходов состояний индивидов в модели, верхний слой состояний соответствует подмодели социальной дезадаптации

Индивид из любой группы может быть инфицирован ВИЧ. Внутри каждой из групп индивиды неразличимы по инфекционным свойствам: трансмиссивности вируса, частотам контактов, соотношению доли половых и парентеральный взаимодействий. Различия обнаруживаются между представителями групп с разными уровнями социальной дезадаптации. В нашей модели предусмотрено влияние склонности индивидов к взаимодействию с представителями своей и других групп.

Слагаемое инфицирования в модели состоит из двух компонент: матрицы вероятностей формирования пар K и вектора вероятностей эффективных контактов γ^I . Такой способ описания

процесса инфицирования впервые был предложен в [16] и нашел широкое распространение.

В России среди ВИЧ-инфицированных из групп риска проводится значительное число мероприятий по их социальной адаптации и поддержке, и может приводить к значительным потокам индивидов из групп I_A и I_D в I_S и далее в I_G . Данный процесс является переносом инфекции из ядра в основную популяцию и поэтому представляет особый интерес для моделирования.

4. Результаты

4.1 УСЛОВИЯ СУЩЕСТВОВАНИЯ РАВНОВЕСИЯ

Классическая модель распространения инфекций, передаваемых половым путем в неоднородной популяции [6], предполагает, что время с момента инфицирования до наступления законченного случая заболевания (выздоровления или смерти больного), а также время между инфицированиями достаточно мало для изменения стратегии поведения на индивидуальном уровне. Каждой из n групп индивидов, объединенных по уровню риска заражения, ставится в соответствие пара дифференциальных уравнений, описывающих динамику числа инфицированных и восприимчивых индивидов в группе.

$$(2) \begin{cases} \frac{dS_i}{dt} = f_i - m_i^S S_i - g_i \sum_{j=1}^n k_{ij} I_j S_i \\ \frac{dI_i}{dt} = g_i \sum_{j=1}^n k_{ij} I_j S_i - m_i^I I_i \end{cases} \quad i = \overline{1, n}$$

Данному классу моделей посвящено достаточно большое количество работ, приведено много аналитических результатов.

В случае ВИЧ-инфекции, исходя из характерных времен заболевания и среднего возраста эффективных носителей, целесообразно учесть в модели возможности перехода индивидов между группами как в инфицированном состоянии, так и в состоянии восприимчивости к вирусу. Предлагаемая модель, как дифференциальная задача, принадлежит к более обширному и

малоизученному в смысле приложения к задачам распространения ИППП классу.

$$(3) \begin{cases} \frac{dS_i}{dt} = f_i + \sum_{j=1}^n a_{ij}^S S_j - g_i \sum_{j=1}^n k_{ij} I_j S_i \\ \frac{dI_i}{dt} = g_i \sum_{j=1}^n k_{ij} I_j S_i + \sum_{j=1}^n a_{ij}^I I_j \end{cases} \quad i = \overline{1, n}$$

Коэффициенты a , соответствующие неинфекционной динамике в популяции, удовлетворяют следующим требованиям ($\forall i = \overline{1, n}, k \in \{I, S\}$):

- 1) $a_{ii}^k \leq 0$,
- 2) $\forall j \neq i \quad a_{ij}^k \geq 0$,
- 3) $a_{ii}^k = -\left(\mu_i + \sum_{j \neq i} a_{ji}^k\right)$, где μ_i – константа скорости оттока индивидов из популяции в группе с индексом i .

Данная система имеет, по меньшей мере, одно положение равновесия – в отсутствие инфекции. Чтобы сформулировать условие устойчивости для него введем вспомогательную величину:

$$R_i = \frac{g_i k_{ii} N_i}{|a_{ii}^S + a_{ii}^I|}.$$

Её физический смысл близок к понятию базовой скорости репродукции: произведение числа индивидов, инфицированных в единицу времени одним инфицированным индивидом в своей группе и среднего времени пребывания в группе с индексом i .

Утверждение 1. Если в задаче (3) выполнены условия (1)-(3) и найдется $p < n$ групп, для которых выполнено условие $\frac{1}{p} \sum_{i=1}^p R_i \leq 1$, и при этом $\frac{1}{n} \sum_{i=1}^n R_i < 1$, то положение равновесия системы при $I = 0$ является устойчивым по Ляпунову.

4.2. ОЦЕНКА ПАРАМЕТРОВ

Параметры демографических процессов – потоки и константы скорости – могут быть оценены исходя из данных, публикуемых Федеральной службой статистики и Минздравсоцразвития по отдельным субъектам Российской Федерации.

Параметры процессов, отвечающих за распространение ВИЧ-инфекции в популяции, могут быть оценены из экспертных данных, публикуемых в литературе. Проблемой является поиск конкретной информации по отдельным территориям России, так как данные различных исследований зачастую не поддаются сравнению.

Не поддаются прямой оценке из данных статистических наблюдений параметры процессов, связанные с дезадаптацией и адаптацией индивидов до развития наркологической патологии. Анализ данных по алкоголизации и наркотизации позволяет предположить наличие квазистационара в численностях групп различного уровня социальной дезадаптации и оценивать недостающие параметры путем решения оптимизационной задачи.

Предлагаемый подход позволяет идентифицировать параметры модели для семнадцати регионов Российской Федерации из сорока, для которых доступны все необходимые статистические данные. В таблице 3 приведены семь основных параметров, характеризующих моделируемую популяцию в данных территориях. Символ k_i обозначает долю i -го компартмента во всей популяции, τ_i – среднее время пребывания индивида в группе i в годах и λ – средняя по популяции сила ВИЧ-инфекции – базовый параметр оценки инфекционных свойств территории. Первые шесть величин соответствуют подмодели социальной дезадаптации (рис. 5).

Величины, приведенные в таблице 3, указывают на неоднородность протекания процессов социальной дезадаптации и, как возможное следствие, неоднородность распространения ВИЧ-инфекции в регионах России.

Таблица 3. Оценки некоторых параметров модели для субъектов Российской Федерации.

Субъект РФ	k_S	k_A	k_D	τ_S	τ_A	τ_D	$\lambda (x10^{-3})$
Орловская область	0,12	0,019	0,001	0,6	11,2	13,2	0,1
Рязанская область	0,01	0,019	0,001	1,4	8,5	6,8	0,12
г. Москва	0,31	0,009	0,003	2,4	12,5	21,5	0,17
Мурманская область	0,05	0,012	0,002	2,0	5,4	3,9	0,36
Новгородская область	0,04	0,031	0,003	0,1	8,7	6,0	0,17
г. Санкт-Петербург	0,30	0,008	0,003	3,6	8,7	11,0	0,52
Краснодарский край	0,42	0,017	0,005	3,9	7,8	9,2	0,11
Республика Башкортостан	0,43	0,017	0,002	2,3	10,7	12,9	0,15
Республика Татарстан	0,44	0,012	0,003	3,8	7,6	10,2	0,16
Нижегородская область	0,06	0,029	0,002	2,6	19,8	20,4	0,1
Пензенская область	0,09	0,020	0,001	0,5	6,9	4,9	0,07
Самарская область	0,22	0,019	0,008	0,6	9,6	18,2	0,72
Ульяновская область	0,14	0,020	0,004	6,7	6,7	12,3	0,41
Челябинская область	0,21	0,019	0,003	0,3	10,7	13,2	0,35
Омская область	0,27	0,014	0,004	5,4	9,5	16,1	0,04
Приморский край	0,21	0,020	0,005	0,3	12,9	8,2	0,27
Хабаровский край	0,28	0,019	0,003	11,6	14,8	13,7	0,06

Причинами наблюдаемой неоднородности могут являться такие социально-демографические факторы как плотность населения и уровень занятости, а также уровень доступности

психоактивных веществ. Так, например, при примерно одинаковой доле больных хроническим алкоголизмом (k_A) в регионах России в городах Москва и Санкт-Петербург данный параметр оказывается заметно ниже. В то же время уровень наркотизации популяции (k_D) заметно выше в Самарской области, Краснодарском и Приморском краях – пограничных регионах.

Более сложные связи лежат в основе формирования группы повышенного риска зависимости от алкоголя и наркотиков (k_S). Изучение причин повышения данного риска является предметом психологии и социологии. Полученные нами значения в результате процедуры оптимизации параметров подмодели социальной дезадаптации указывают на высокую чувствительность данной величины к региональным различиям населения России.

Величины τ_A и τ_D являются обратными к константам скорости социальной адаптации индивидов в состояниях с хроническим алкоголизмом и наркоманией. Сами константы скорости могут быть оценены напрямую из данных статистики, а в табл. 3 обратные значения приведены для удобства сопоставления с параметром τ_S . Полученные значения иллюстрируют, насколько замедляются процессы адаптации в более дезадаптированных состояниях.

Для простоты расчетов на начальном этапе мы предположили, что параметры подмодели социальной дезадаптации не зависят от возраста индивидов. Данное предположение может показаться не совсем верным. К сожалению, нами не было обнаружено достоверных данных о виде возрастной динамики данных величин. Выдвигая различные предположения о виде зависимости данных параметров от возраста потенциально можно повысить адекватность модели.

Заключение

Целью данной работы была разработка методов решения задач контроля ВИЧ-инфекции на территории России. Обзор литературы по данному вопросу, а также анализ данных, указывают на необходимость построения моделей, учитывающих как инфекционную динамику в группах риска, так и условия изме-

нения численности данных групп в результате внешних воздействий.

Предложена детерминированная популяционная модель, рассматривающая два взаимосвязанных процесса: распространение инфекции и социальную дезадаптацию. Предлагаемая модель принадлежит к более общему классу задач, нежели традиционные модели распространения инфекций, передаваемых половым путем, в неоднородной популяции.

Нам удалось сформулировать условия существования положений равновесия, соответствующих эндемическому типу эпидемического процесса, а также разработать подходы к оценке параметров модели задачи с использованием статистических данных и экспертных оценок.

Литература

1. АНДЕРСОН Р. М., МЭЙ Р. М. *Инфекционные болезни человека: динамика и контроль* – М.: Мир, "Научный мир", 2004.
2. *ВИЧ/СПИД в США: нерешенные вопросы противодействия эпидемии: Аналитический материал.* / Е.М.КОЖОКИН, А.И.МАЗУС, Ю.Н. ГЛУЩЕНКО, А.Я.ОЛЬШАНСКИЙ. — М., 2006.
3. ДЕНИСОВ Б. П., САКЕВИЧ В. И. *Динамика эпидемии ВИЧ/СПИД* // Социологические исследования – 2004 – № 1. – С. 75-85.
4. ДЕНИСОВ Б. П., САКЕВИЧ В. И. *Прогноз возможных демографических последствий эпидемии ВИЧ/СПИДа в РФ* // Демография – 2004 – №5 – С. 149-160.
5. КУПРЯШКИНА-МАКГИЛЛ С.В. *Влияние грантов глобального фонда на политику в области ВИЧ/СПИДа в Украине* // Русский журнал "СПИД, рак и общественное здоровье" – 2010 – Т. 14 № 1(23) – С. 27.
6. ПОКРОВСКИЙ В.В., ЛАДНАЯ Н.Н., СОКОЛОВА Е.В., БУРАВЦОВА Е. В. *Информационный бюллетень "ВИЧ-инфекция"*. Федеральный НМЦ ПБ СПИД Министерства Здравоохранения РФ N 33, 2009.

7. AALEN O. O., FAREWELL V. T., ANGELIS D. DE., DAY N. E., GILL O. N. *New therapy explains the fall in AIDS incidence with a substantial rise in number of persons on treatment expected* // AIDS – 1999 – № 13 – P. 103-108.
8. ALISTAR S., OWENS D., BRANDEAU M. *Effectiveness and cost-effectiveness of expanding drug treatment programs and HIV antiretroviral therapy in a mixed HIV epidemic: an analysis for Ukraine* // Русский журнал "СПИД, рак и общественное здоровье" – 2010 – Т. 14 № 1(23). С. 44.
9. ALMEDA J, CASABONA J, SIMON B, GERARD B, REY D, PURO V, *Proposed recommendations for the management of HIV post-exposure prophylaxis after sexual, injecting drug or other exposures in Europe.* // Eurosurveillance Monthly – 2004 – Vol. 9, Issue 6. – P. 5-6.
10. ANDERSON RM, GUPTA S, MAY RM. *Potential of community-wide chemotherapy or immunotherapy to control the spread of HIV-1.*// Nature. – 1991 – 350 – P. 356-359.
11. ARAZOZA H, LOUNES R, HOANG T, INTERLAN Y. *Modeling HIV epidemic under contact tracing--the Cuban case.* // J Theor Med – 2000 – 2 – P. 267-274.
12. BACAER N., PRETORIUS C., AUVERT B. *An age-structured model for the Potential Impact of Generalized Access to Antiretrovirals on the South African HIV Epidemic* // Bulletin of Mathematical Biology – 2010.
13. CASTILLO-CHAVEZ C. *Mathematical and statistical approaches to AIDS epidemiology* – Springer Verlag, 1989.
14. COOKE, K., L., YORKE, J., A. *Some equations modelling growth processes and gonorrhoea epidemics.* // Math. Biosci. – 1973 – No 16. – P. 75-101.
15. HABERMAN S., *Actuarial review of models for describing and predicting the spread of HIV-infection and AIDS* // JIA – 1990 – No 117 – P. 319-405.
16. HSIEH, Y.-H., COOKE K. *Behaviour change and treatment of core group and bridge population: its effect on the spread of HIV/AIDS.* // IMA J. Math. Appl. Med. Biol. – 2000. No 17 – P. 213-241.

17. HYMAN J. M., Li J., STANLEY E. A. *Sensitivity studies of the differential infectivity and stage progression models for the transmission of HIV* // LA-UR-99-2253 – 1999.
18. *Joint United Nations Programme on HIV/AIDS* [Электронный ресурс] URL: <http://unaids.org> (дата доступа – 30.09.2011).
19. LONGINI I. M., CLARK W. S., BYERS R. H., WARD J. W., DARROW W. W., LEMP G. F., HETHCOTE H. W. *Statistical analysis of the stages of HIV infection using a Markov model* // *Statistics in Medicine* – 1989 – 8 – P. 831-843.
20. NELSON G. W., PERELSON A. S. *A mechanism of immune escape by slow-replicating HIV strains* // *JAIDS* – 1992 – 5 – P. 82-93.
21. NOSOVA E.A., ROMANYUKHA A.A. *Regional index of HIV infection risk based on factors of social disadaptation* // *Russian Journal of Numerical Analysis and Mathematical Modelling* – 2009 – Volume 24, Issue 4 – P. 325-340.
22. NOWAK M. A., MAY R. M. *Mathematical biology of HIV infections: antigenic variation and diversity threshold* // *Math. Biosci.* – 1991 – 106 – P. 1-21.
23. SLOOT P. M. A., IVANOV S. V., BOUKHANOVSKY A. V., VAN DE VIJVER D., BOUCHER C. *HIV Population Dynamics on Complex Networks* // *European Conference on Complex Systems* – 2007 – P. 1-2.
24. SNEDECOR S. J. *Comparison of three kinetic models of HIV-1 implications for optimization of treatment* // *J. Theor. Biol.* - 2003 – 221 – P. 519-541.
25. WALKER D. *Cost and cost-effectiveness of HIV/AIDS prevention strategies in developing countries: is there an evidence base?* // *Health and Policy Planning* – 2000 – 18(1) – P. 4-17.
26. WILSON D. P., KAHN J., BLOWER S. M. *Predicting the epidemiological impact of antiretroviral allocation strategies in Kwasulu-Natal: the effect of urban-rural divide* // *PNAS* – 2006 – 103(38) – P. 14228-14233.

MODELING SPREAD OF HIV AS RESULT OF SOCIAL MALADJUSTMENT IN POPULATION

Aleksey Romanyukha, Institute of Numerical Mathematics of RAS, Moscow, Doctor of Science, professor (eburg@inm.ras.ru).

Ekaterina Nosova, Federal Public Health Institute, Moscow, research assistant (nosova@mednet.ru).

Abstract: The aim of the research is to construct methods for estimation of effectiveness of HIV-infection spread counteraction measures for Russian territories. It has been found that regional inequity in HIV prevalence is closely linked with prevalence of alcohol and drug abuse, which could be considered as a value of social maladjustment in population.

Accounting influence of these processes on HIV risk-group formation allows explaining regional epidemic scale inequity and evaluate effectiveness of counteraction measures.

We introduce the model, put the problem of model identification based on the data of Russian regions, and present results of data analysis.

Keywords: mathematical model, social maladjustment, HIV epidemiology.

*Статья представлена к публикации
членом редакционной коллегии В.Н. Новосельцевым*

МОДЕЛИ АВТОМАТИЗИРОВАННОГО ПРОГНОЗИРОВАНИЯ ЭЛЕКТРИЧЕСКИХ НАГРУЗОК ПРОМЫШЛЕННЫХ ПРЕДПРИЯТИЙ

Васильев Д. А.¹, Колоколов М. В.²

*(Саратовский государственный технический
университет, Саратов)*

Иващенко В. А.³

*(Учреждение Российской академии наук
Институт проблем точной механики и управления РАН,
Саратов)*

Предложены модели автоматизированного прогнозирования электрических нагрузок промышленных предприятий, обеспечивающие повышение надежности и экономичности электроснабжения потребителей электрической энергией.

Ключевые слова: автоматизация, прогнозирование, фильтр Брауна, искусственные нейронные сети, лингвистические переменные, гибридная модель.

1. Введение

Существенное место в управлении электроэнергетикой промышленных предприятий занимает прогнозирование электрических нагрузок, которое является важным как для производителя, так и для потребителя электрической энергии [4].

¹ Дмитрий Анатольевич Васильев, кандидат технических наук, доцент (vda230373@yandex.ru).

² Максим Владимирович Колоколов, студент (killgreen@rambler.ru).

³ Владимир Андреевич Иващенко, доктор технических наук, старший научный сотрудник (irtmuran@san.ru).

Для производителя электроэнергии прогноз нагрузок значим с точки зрения оптимизации поставки и резервирования электроэнергии, удобства проведения профилактических работ и обеспечения безопасности функционирования электроэнергетической системы (ЭЭС). Потребителю прогноз необходим для минимизации издержек, связанных с уплатой штрафов при превышении лимитов по мощности и с переплатой за заявленные, но не использованные мощности, а также с простым технологического оборудования в случае дефицита мощности в ЭЭС.

2. Анализ объекта прогнозирования

Рассматриваемые модели прогнозирования активных нагрузок ориентированы на промышленных потребителей электроэнергии (абонентов) с присоединенной мощностью (нагрузкой) 750 кВА и выше, которые оплачивают электропотребление по двухставочному тарифу, включающему плату:

- за заявленную мощность (основная ставка);
- за фактически потребленную электрическую энергию (дополнительная ставка).

Заявленная мощность представляет собой максимальную активную мощность, отпускаемую абонентам в часы максимума нагрузки ЭЭС.

Графики изменения активных нагрузок промышленных предприятий малой и средней энергоемкости (от 2 до 18 МВт) имеет явно выраженные максимумы, первый из которых совпадает с часами утреннего максимума (с 9 до 12 часов), а второй (послеобеденный) несколько смещен относительно вечернего максимума (с 15 ч. 30 мин. до 17 ч.) нагрузки ЭЭС. На рис. 1 представлен фрагмент графика активной нагрузки для типичного промышленного предприятия средней энергоемкости.

Оперативный прогноз нагрузки выполняется в часы максимума энергосистемы с периодичностью сбора информации в системе, равной $\Delta t = 5$ минутам. Интервал упреждения определяется компромиссом между точностью прогнозирования и возможностью реализации энергодиспетчером управляющих воз-

действий по снижению нагрузки с целью исключения превышения заявленной предприятием мощности (или установленного энергосистемой более жесткого ограничения) и составляет $t^* = 2\Delta t = 10$ минут.

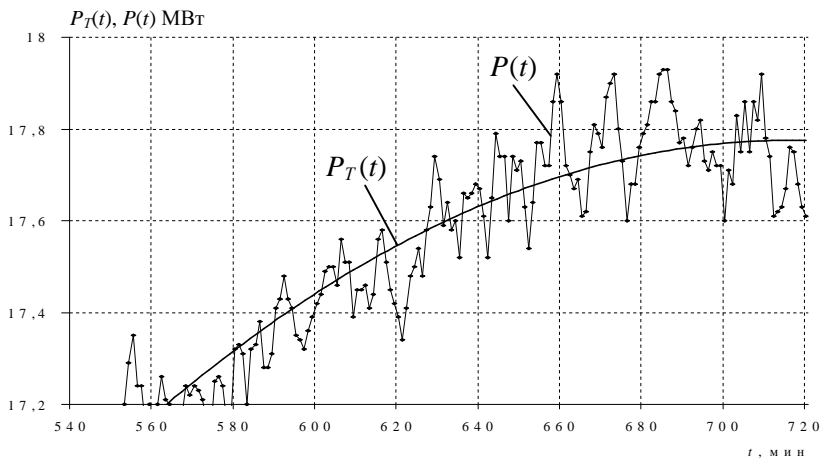


Рис. 1. Графики изменения активной нагрузки предприятия $P(t)$ и ее тренда $P_T(t)$

Качественный прогноз в часы максимума ЭЭС особенно важен для предприятий, поскольку за превышение заявленной ими мощности на этом отрезке времени они платят штраф (оплачивают превышение по повышенному тарифу) или подвергаются принудительному отключению (отключаются один или несколько вводов электроэнергии в предприятие).

Современные подходы к прогнозированию электропотребления предприятий представлены в работах в [1-6].

3. Фильтры Брауна

Для прогнозирования активной нагрузки предприятия используется полиномиальная модель вида

$$P(t) = a_0 + a_1 t + a_2 t^2,$$

где a_0, a_1, a_2 – параметры модели, определяемые с помощью метода экспоненциального сглаживания [8]; t – текущее время.

При этом прогнозное значение нагрузки $P(t)$ выражается рядом Тейлора

$$P(t + \Delta t) = P(t)_t + \frac{dP(t)}{dt} \Delta t + \frac{1}{2!} \frac{d^2 P(t)}{dt^2} \Delta t^2 + \dots + \frac{1}{n!} \frac{d^n P(t)}{dt^n} \Delta t^n,$$

члены которого связаны рекуррентной формулой

$$S_t^n(P(t)) = \alpha S_t^{n-1}(P(t)) + (1 - \alpha) S_{t-1}^n(P(t)),$$

определяющей экспоненциальную среднюю n -го порядка для ряда. Параметр сглаживания $\alpha \in [0, 1]$ учитывает влияние исходного ряда наблюдений на результаты прогнозирования.

В качестве начального условия используется соотношение

$$S_0^1(P(t)) = \dots = S_0^n(P(t)) = P(t_0),$$

где $P(t_0)$ – значение нагрузки, соответствующее моменту времени t_0 , с которого начинается прогноз.

Для фильтра Брауна второго порядка:

$$P(t) = a_0 = 3 S_t(P(t)) - 3 S_t^2(P(t)) + S_t^3(P(t));$$

$$\frac{dP(t)}{dt} = a_1 = \frac{\alpha^2}{2(1-\alpha)^2} ((6-5\alpha) S_t(P(t)) - 2(5-4\alpha) S_t^2(P(t)) + (4-3\alpha) S_t^3(P(t)));$$

$$\frac{d^2 P(t)}{dt^2} = a_2 = \frac{\alpha^2}{(1-\alpha)^2} (S_t(P(t)) - 2S_t^2(P(t)) + S_t^3(P(t)));$$

$$S_0^1(P(t)) = S_0^2(P(t)) = S_0^3(P(t)) = P(t_0).$$

Качество прогнозирования оценивается максимальной относительной ошибкой прогноза (абсолютной ошибкой прогноза, отнесенной к фактическому значению прогнозируемой нагрузки) δ_{\max} . Оно существенно зависит от выбора значений параметров модели – параметра сглаживания α и объема предыстории k , по которой выполняется прогноз.

Путем статистического анализа определены значения этих параметров, позволяющие выполнять прогнозирование с минимумом ошибки: $\alpha = 0,1-0,3$; $k = 4-6$ при прогнозе на $t^* = \Delta t = 5$ минут и $\alpha = 0,1-0,2$; $k = 4-9$ при прогнозе на $t^* = 2\Delta t = 10$ минут.

Реализация управляющих воздействий при глубоком регулировании активной нагрузки приводит к существенному изменению ее статистических характеристик. Поэтому при прогнозировании в данных условиях используется процедура коррекции.

При этом прогнозируемый процесс на отрезке коррекции $[t - k \Delta t; t]$ представляется в виде

$$P'(t') = P(t) + P_{\Sigma}(t),$$

где $P'(t')$ – откорректированный процесс, $t' \in [t - k \Delta t; t]$;

$$P_{\Sigma}(t) = \sum_{j=1}^J \sum_{i=1}^{I_j} P_{ij}(t) \beta_{ij} - \text{суммарная нагрузка потребителей элек-}$$

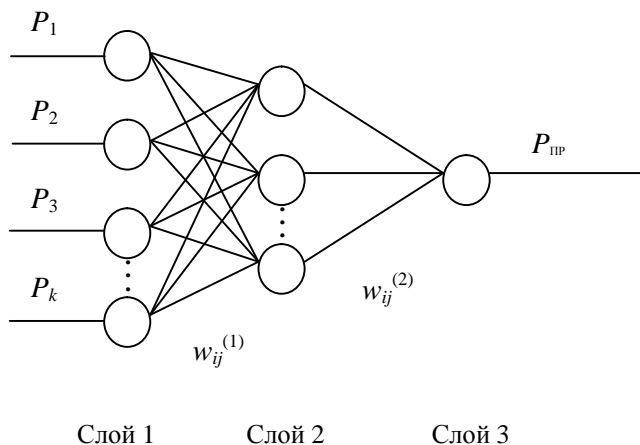
троэнергии, режим работы которых изменялся на отрезке коррекции; $P_{ij}(t)$ – нагрузка i -го потребителя, используемого для регулирования активной нагрузки (потребителя-регулятора) j -го уровня электрической сети, изменившего режим работы; $\beta_{ij} = \{-1, +1\}$ – признак, характеризующий изменения режима работы i -го потребителя-регулятора j -го уровня сети: -1 – означает, что потребитель-регулятор отключен; $+1$ – ПР подключен.

Исследования данной модели для графиков электрических нагрузок ОАО «Саратовское электроагрегатное производственное объединение», ОАО «Саратовстройстекло» и ОАО «Трансмаш» (г. Энгельс) показали, что для $\alpha = 0,15$ и $k = 5$ $\delta_{\max} \leq 1,18\%$ при прогнозировании на $t^* = \Delta t = 5$ минут и $\delta_{\max} \leq 1,5\%$ – на $t^* = 2\Delta t = 10$ минут.

4. Искусственные нейронные сети

Наряду со статистическими методами для прогнозирования нагрузок использовались искусственные нейронные сети (ИНС) типа «многослойный персептрон» с полным набором связей архитектуры $r_1 - r_2 - 1$, где r_1 , r_2 , и 1 – количество элементов в

первом (входном), втором (скрытом) и третьем (выходном) слоях сети (см. рис. 2).



P_1, \dots, P_k – входные и $P_{пр}$ – выходной сигналы сети;
 $w_{ij}^{(1)}$ и $w_{ij}^{(2)}$ – весовые коэффициенты связей между элементами
первого и второго, второго и третьего слоев сети;
 i, j – номера элементов в слоях сети

Рис. 2. Структура искусственной нейронной сети

При анализе ИНС данной архитектуры установлено, что для обеспечения требуемой точности прогноза достаточно одного скрытого слоя с числом нейронов, равным $r_2 = \lceil (r_1 + 1) / 2 \rceil$, где $\lceil \cdot \rceil$ – оператор округления до ближайшего целого числа, большего r_2 .

Анализ работы ИНС при различном характере изменений электрических нагрузок, типичных для предприятий малой и средней энергоемкости, показали, что для обеспечения требуемой точности и времени обучения сети приемлемыми являются персептроны архитектур: 5–3–1 для прогноза на $t^* = \Delta t = 5$ минут и 13–7–1 для прогноза на $t^* = 2\Delta t = 10$ минут.

В связи с изменением условий потребления предприятием электроэнергии параметры модели прогнозирования требуют

периодической адаптации, которая осуществляется по истечении каждых суток и сопровождается процедурой переобучения ИНС на новых данных.

Обучение сети осуществляется с помощью комбинированного метода, построенного на основе использования алгоритма обратного распространения ошибки и стохастического алгоритма Коши, позволяющего выходить из локальных минимумов функции ошибок.

Для сетей структуры 5–3–1 при 500 – 700 итерациях обучения прогнозирование нагрузки на $t^* = \Delta t = 5$ минут обеспечивается с $\delta_{\max} \leq 0,99\%$ (на 0,2% меньше, чем при использовании фильтра Брауна), а на $t^* = 2\Delta t = 10$ минут – с $\delta_{\max} \leq 1,19\%$ (на 0,49% меньше, чем при использовании фильтра Брауна).

Дальнейшее повышение точности оперативного прогнозирования активной нагрузки промышленных предприятий требует использования адаптивных моделей.

Трудно провести четкую грань между адаптивными и неадаптивными моделями. Так, рассмотренные в [6] регрессионные модели содержат некоторый элемент адаптации, заключающийся в пересчете параметров модели при получении нового значения прогнозируемого процесса. Такая степень адаптации весьма незначительна. К тому же с течением времени она уменьшается в связи с увеличением общего количества наблюдений [7].

Фильтры Брауна, адаптируемые за счет вариации параметра сглаживания и придания различной значимости наблюдениям ретроспективной выборки, обладают недостатком, связанным с запаздыванием прогнозных значений. Недостатком же ИНС является плохое реагирование на изменения характера прогнозируемых процессов на интервалах между обучениями, которые существенны при использовании моделей в режиме реального времени.

Исходя из этого, предлагается гибридная модель адаптации, исключаяющая вышеотмеченные недостатки.

5. Гибридная модель

Предлагается адаптируемая в режиме реального времени гибридная модель, составляющими которой являются рассмотренные выше модели – фильтр Брауна и ИНС (рис. 3).

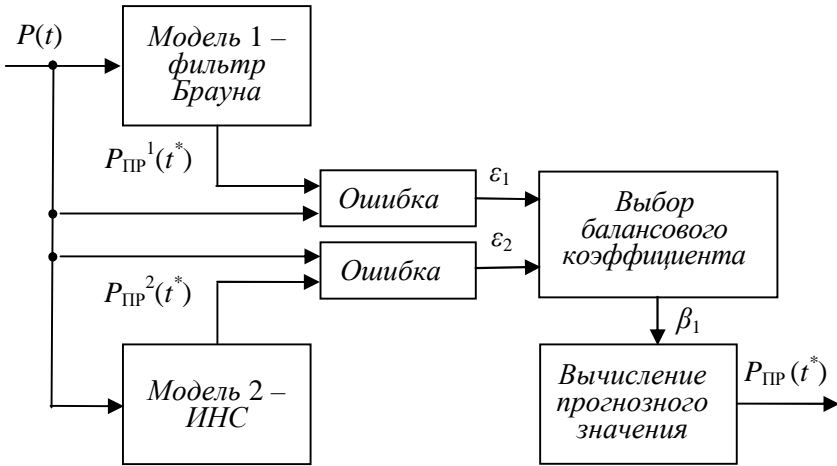


Рис. 3. Структура гибридной модели

Получаемое на ее основе прогнозное значение

$$P_{\text{ПР}}(t^*) = \sum_{i=1}^2 b_i P_{\text{ПР}}^i(t^*)$$

и ошибка прогноза

$$\varepsilon_{\text{ПР}}(t^*) = \sum_{i=1}^2 b_i \varepsilon_{\text{ПР}}^i(t^*),$$

где $P_{\text{ПР}}^i(t^*)$, $\varepsilon_{\text{ПР}}^i(t^*)$ – результаты и ошибки прогнозирования входящих в состав гибридной модели фильтра Брауна ($i = 1$) и ИНС ($i = 2$); β_1 , $\beta_2 = (1 - \beta_1)$ – коэффициенты, определяющие их долевое участие в результате прогноза.

Для вычисления значений коэффициентов β_1 , β_2 , обеспечивающих минимум ошибки прогноза, предлагается использовать систему нечеткого вывода. Для этого определяются две входные

лингвистические переменные (ЛП) «Ошибка прогноза ε_1 », «Ошибка прогноза ε_2 » и одна выходная – «Коэффициент β_1 ». Термы этих ЛП определены значениями «малое», «среднее», «большое» и представлены в виде треугольных и трапециевидных функций. Функции принадлежности для ЛП имеют вид, представленный на рис. 4, 5.

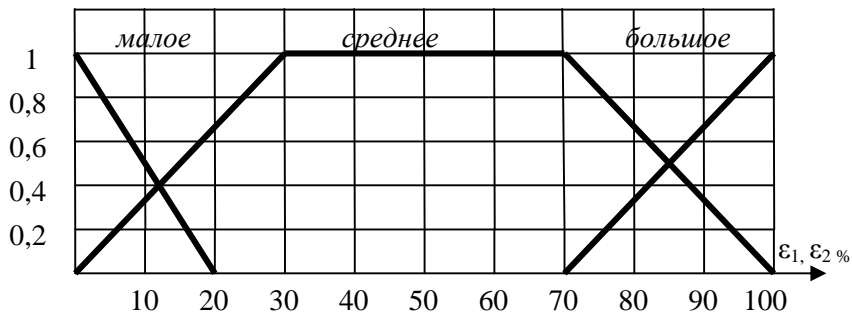


Рис. 4. Функции принадлежности для входных лингвистических переменных

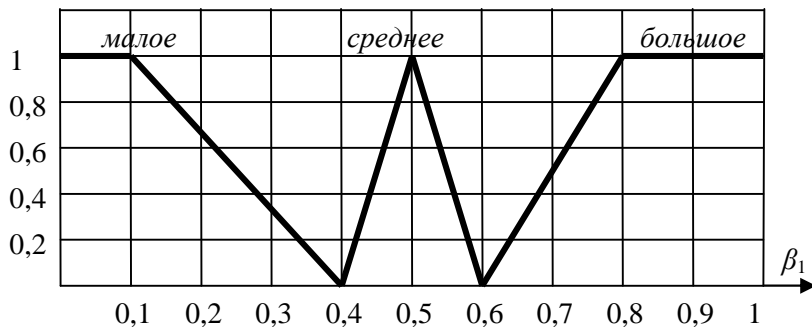


Рис. 5. Функции принадлежности для выходной лингвистической переменной

Определение текущих значений коэффициентов гибридной модели осуществляется на основе базы продукционных правил:

1. Если «Ошибка прогноза ε_1 » – «малое» И «Ошибка прогноза ε_2 » – «большое» ТО «Коэффициент β_1 » – «большое».

2. Если «Ошибка прогноза ε_1 » – «большое» И «Ошибка прогноза ε_2 » – «малое» ТО «Коэффициент β_1 » – «малое».

3. Если «Ошибка прогноза ε_1 » – «большое» И «Ошибка прогноза ε_2 » – «большое» ТО «Коэффициент β_1 » – «среднее».

4. Если «Ошибка прогноза ε_1 » – «малое» И «Ошибка прогноза ε_2 » – «малое» ТО «Коэффициент β_1 » – «среднее».

5. Если «Ошибка прогноза ε_1 » – «среднее» И «Ошибка прогноза ε_1 » – «среднее» ТО «Коэффициент β_1 » – «среднее».

6. Если «Ошибка прогноза ε_1 » – «среднее» И «Ошибка прогноза ε_2 » – «большое» ТО «Коэффициент β_1 » – «большое».

7. Если «Ошибка прогноза ε_1 » – «большое» И «Ошибка прогноза ε_2 » – «среднее» ТО «Коэффициент β_1 » – «малое».

8. Если «Ошибка прогноза ε_1 » – «среднее» И «Ошибка прогноза ε_1 » – «малое» ТО «Коэффициент β_1 » – «малое».

9. Если «Ошибка прогноза ε_1 » – «малое» И «Ошибка прогноза ε_2 » – «среднее» ТО «Коэффициент β_1 » – «большое».

В основу выбора коэффициента β_1 положен алгоритм Мамдани.

Апробирование гибридной модели на ряде промышленных предприятий г. Саратова показало, что она обеспечивает более точный прогноз, чем входящие в ее состав модели. При прогнозировании на один интервал упреждения, равный 5 минутам, максимальная относительная ошибка не превышает 0,89%, что на 0,29% меньше, чем для фильтра Брауна, и на 0,1%, чем для перцептрона, а при прогнозировании на два интервала (10 минут) – 0,91%, что на 0,59% меньше, чем для фильтра Брауна, и на 0,28%, чем для перцептрона (рис. 6).

На рис. 6 жирными линиями представлены прогнозные, а тонкими – фактические значения нагрузок.



a



б



в

*Рис. 6. Результаты прогнозирования активной нагрузки:
а – с помощью фильтра Брауна, б – с помощью ИНС,
в – с помощью гибридной модели*

6. Заключение

1. Исследование характера изменения электрических нагрузок промышленных предприятий показало возможность их прогнозирования с помощью фильтра Брауна второго порядка и трехслойного персептрона. Определены параметры фильтра и персептрона, обеспечивающие минимум ошибки прогноза.

2. Предложена гибридная модель прогнозирования, обладающая более высокой точностью прогноза за счет адаптации к характеру изменения электрических нагрузок в режиме реального времени.

Литература

1. БОЛЬШОВ Л.А., КАНЕВСКИЙ М.Ф., САВЕЛЬЕВА Е.А., ТИМОНИН В.А., ЧЕРНОВ С.Ю. *Прогнозирование электропотребления: современные подходы и пример исследования* // Известия академии наук. Энергетика. 2004. № 6. С. 74 – 93.
2. ВАСИЛЬЕВ Д.А., ИВАЩЕНКО В.А., ЛУКЬЯНОВ Д.В. *Прогнозирование активных нагрузок промышленных предприятий на основе распознавания образов* // Известия высших учебных заведений. Проблемы энергетики. 2011. № 3-4. С. 122 – 126.
3. ИВАЩЕНКО В.А., КОЛОКОЛОВ М.В., ВАСИЛЬЕВ Д.А. *Прогнозирование электропотребления промышленных предприятий на основе статистических методов и искусственных нейронных сетей* // Вестник Саратовского государственного технического университета. 2010. № 2 (45). С. 110 – 115.
4. ИВАЩЕНКО В.А., РЕЗЧИКОВ А.Ф. *Автоматизированное управление электропотреблением промышленных предприятий: концепция и основные задачи* // Мехатроника, автоматизация, управление. 2006. № 3. С. 52 – 56.
5. ИВАЩЕНКО В.А., ВАСИЛЬЕВ Д.А., РЕЗЧИКОВ А.Ф. *Методы прогнозирования электрических нагрузок в условиях АСУ электропотреблением промышленных предприятий* // Мехатроника, автоматизация, управление. 2006. № 7. С. 52 – 55.

6. КАМАЕВ В.А., ЩЕРБАКОВ М.В., ПАНЧЕНКО Д.П., ЩЕРБАКОВА Н.Л. *Применение коннективистских систем для прогнозирования потребления электроэнергии в торговых центрах // Управление большими системами. 2010. Вып. 31. С. 92 – 107.*
7. ЛУКАШИН Ю.П. *Адаптивные методы краткосрочного прогнозирования временных рядов: учеб. пособ. М.: Финансы и статистика. 2003.*
8. *Теория прогнозирования / под. ред. С.А. Саркисяна. М.: Высш. шк., 1977.*

MODELS FOR AUTOMATED ELECTRIC LOADS FORECASTING ON INDUSTRIAL ENTERPRISES

Dmitry Vasiliev, Saratov State Technical University, Saratov, Cand. Sc., assistant professor (vda230373@yandex.ru).

Maxim Kolokolov, Saratov State Technical University, Saratov, student (killgreen@rambler.ru).

Vladimir Ivaschenko, Establishment of the Russian Academy of Sciences Institute of problems of Precision Mechanics and Control of RAS, Saratov, Doctor of Science, Senior Researcher (iptmuran@san.ru).

Abstract: We suggest models of automated electric loads forecast for industrial enterprises. These models help to save energy and to increase reliability of consumers' power supply.

Keywords: automation, forecasting, Brown's filter, artificial neural networks, linguistic variables, hybrid model.

*Статья представлена к публикации
членом редакционной коллегии Д.А. Новиковым*

УДК 65.014.12
ББК 34.42-05

СИСТЕМА КАЛЕНДАРНОГО ПЛАНИРОВАНИЯ ПРОЦЕССОВ КОНСТРУКТОРСКО- ТЕХНОЛОГИЧЕСКОЙ ПОДГОТОВКИ МЕЛКОСЕРИЙНОГО МАШИНОСТРОИТЕЛЬНОГО ПРОИЗВОДСТВА

Денисов А. Р.¹, Белянкин М. В.²

*(Костромской государственной университет имени Н.А.
Некрасова, Кострома)*

Рассмотрены вопросы создания системы календарного планирования конструкторско-технологической подготовки мелкосерийного машиностроительного производства (КТТП), использующей технологию управления потоками работ (Workflow). Обосновывается, что в условиях мелкосерийного машиностроения необходим автоматизированный синтез Workflow-диаграмм, для чего предложена формальная модель процессов КТТП, базирующаяся на принципах управления конфигурацией, а также информационное и алгоритмическое обеспечение предлагаемой системы.

Ключевые слова: конструкторско-технологическая подготовка, календарное планирование, управление конфигурацией, мелкосерийное производство.

1. Введение

Одной из основных заявленных целей внешней политики РФ является вступление во Всемирную торговую организацию. Россия будет вынуждена открыть свои собственные рынки для иностранных компаний, что неизбежно приведет к усилению

¹ Артем Руфимович Денисов, кандидат технических наук, доцент (inf_service@ksu.edu.ru).

² Михаил Вячеславович Белянкин, аспирант (inf_service@ksu.edu.ru).

внутренней конкуренции [4]. При этом в ряде областей промышленности, например, в машиностроении, для отечественных предприятий внешние рынки останутся закрытыми. Одной из причин этого является низкая степень автоматизации систем управления предприятием и его производственными процессами, без чего невозможно решать задачи управления качеством продукции, повышения эффективности и прозрачности документооборота на всех этапах жизненного цикла выпускаемой продукции, внедрения интегрированной логистической поддержки потребителей.

Повысить конкурентоспособность отечественных предприятий можно через внедрение современных производственных и управленческих технологий, основанных на процессном подходе. Данный подход является необходимым условием с точки зрения большинства международных стандартов, в том числе, стандартов управления качеством ISO 9000, и должен быть внедрен на всех этапах жизненного цикла продукции, включая и конструкторско-технологическую подготовку производства (КТПП). Важной задачей при реализации процессного подхода является календарное планирование производственных процессов. В частности, необходимость планирования на этапе КТПП регламентируется ГОСТом [2 п. 5.2.5, 6.1.2, 7.2.2, 7.3.4, 7.4.4.].

Для решения задачи календарного планирования процессов КТПП принято использовать т.н. PDM-системы (Product Data Management). Для этого в данные системы включены специальные модули, которые позволяют представить процессы КТПП в виде совокупности Workflow-диаграмм [3, 7]. Подобная функция есть в большинстве представленных на рынке PDM-систем, включая SmarTeam, WindChill, T-Flex DOCs, Аскон Лоцман:PLM и др.

При этом следует заметить, что Workflow-моделирование является очень трудоемким процессом и, соответственно, его применение при планировании процессов КТПП целесообразно только в случае устоявшейся номенклатуры выпускаемых изделий, то есть на предприятиях с крупно- и среднесерийным производством. В мелкосерийном же машиностроении, которое характеризуется постоянными изменениями в номенклатуре изго-

тавливаемых изделий, использование данного механизма либо будет высокочрезмерно затратным, либо будет носить общий характер (работы будут соответствовать этапам КТПП [1]). В любом случае, это сделает календарное планирование процессов КТПП неэффективным.

2. Формальная модель подсистемы календарного планирования процессов КТПП

Повысить эффективность календарного планирования можно через автоматизацию синтеза Workflow-диаграмм. Для этого целесообразно использовать принципы управления конфигурацией [5,6,8]. В проектируемом изделии выделяются объекты конфигурации, которые обычно являются типовыми для предприятия, что позволяет осуществлять планирование процессов КТПП с достаточной степенью точности. Тогда проектируемое изделие можно представить как:

$$(1) \quad Pu_i = \langle Ok_i, Cв_i, Doc_i, T_i, R_i, Tв_i \rangle, \quad i = 1, \dots, |Pu|,$$

где Pu – множество всех проектируемых на предприятии изделий; Ok_i – множество объектов конфигурации i -го изделия; $Cв_i$ – множество всех связей между объектами конфигурации (ОК) в архитектуре i -го изделия; Doc_i – множество всех документов, создаваемых в процессе проектирования i -го изделия; T_i – трудоемкость проектирования i -го изделия; R_i – другие виды расходов на проектирование i -го изделия; $Tв_i$ – время проектирования i -го изделия.

При этом каждый объект конфигурации имеет вид:

$$(2) \quad Ok_{ij} = \langle B_{ij}, ПГ_{ij}, DOк_{ij}, Tok_{ij}, Rok_{ij}, Tн_{ij}, Tk_{ij} \rangle, \quad i = 1, \dots, |Pu|, \\ j = 1, \dots, |Ok_i|,$$

где B_{ij} – важность Ok_{ij} в изделии; $ПГ_{ij}$ – проектная группа, ответственная за проектирование Ok_{ij} ; $DOк_{ij}$ – множество всех документов, создаваемых в процессе проектирования Ok_{ij} :

$$(3) \quad \forall i : Doc_i = \left(\bigcup_{j=1}^{|Ok_i|} DOк_{ij} \right) \cup Du_i ;$$

Tok_{ij} – трудоемкость изготовления объекта конфигурации Ok_{ij} :

$$(4) \quad \forall i: T_i = Tu_i + \sum_{j=1}^{|Ok_{ij}|} Tok_{ij};$$

Rok_{ij} – множество других расходов на проектирование Ok_{ij} ;

$$(5) \quad \forall i: R_i = Ru_i + \sum_{j=1}^{|Ok_{ij}|} Rok_{ij};$$

Tn_{ij} , Tk_{ij} – время начала и окончания проектирования Ok_{ij} ; Du_i – множество всех документов, непосредственно связанных с проектированием i -го изделия; Tu_i и Ru_i – трудоемкость и другие виды расходов, непосредственно связанные с проектированием i -го изделия.

Пример конфигурации станка класса ШЛПС, выпускаемого на одном из предприятий Костромской области, приведен на рис. 1.

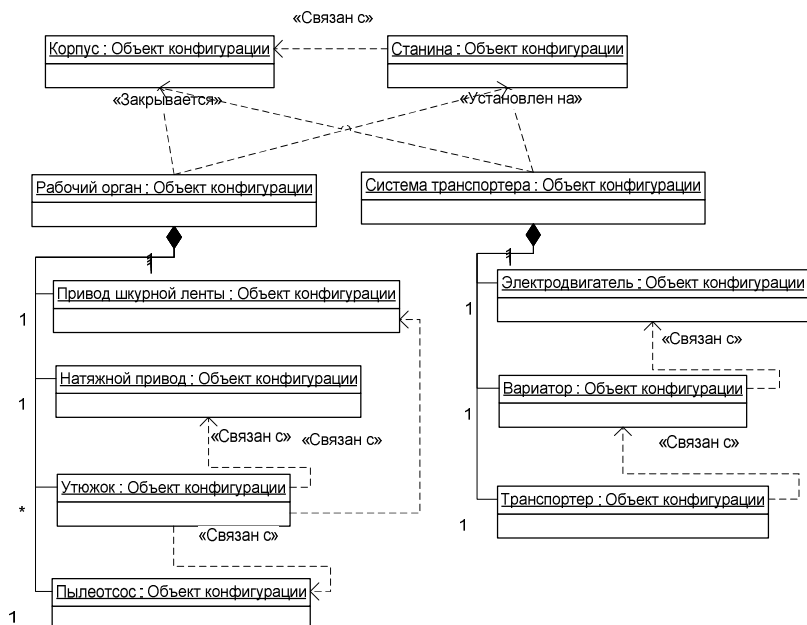


Рис. 1. Конфигурация станка класса ШЛПС

Таким образом, становится возможным определение времени проектирования i -го изделия:

$$(6) \quad \forall i: T_{\mathcal{B}_i} = \sup_{\substack{x=1, \dots, |O_{K_i}| \\ y=1, \dots, |O_{K_i}|}} (T_{K_{ix}} - T_{H_{iy}}).$$

Для этого необходимо определить количество специалистов различного профиля, входящих в проектные группы:

$$(7) \quad ПГ_{ij} = \langle PГ_{ij}, M_{C_{ij}} \rangle, \quad i = 1, \dots, |Pu|, \quad j = 1, \dots, |O_{K_i}|,$$

где $PГ_{ij}$ – руководитель проектной группы $ПГ_{ij}$; $M_{C_{ij}}$ – множество специалистов различного профиля, входящих в проектную группу $ПГ_{ij}$:

$$(8) \quad M_{C_{ij}} = \{ \langle П_{C_{ijk}}, K_{C_{ijk}}, T_{C_{ijk}} \rangle, \quad k = 1, \dots, |M_{C_{ij}}| \}, \quad i = 1, \dots, |Pu|, \\ j = 1, \dots, |O_{K_i}|,$$

где $П_{C_{ijk}}$ – профиль соответствующих специалистов; $K_{C_{ijk}}$ – количество специалистов соответствующего профиля, входящих в проектную группу; $T_{C_{ijk}}$ – время выполнения k -го вида проектных работ для $O_{K_{ij}}$:

$$(9) \quad \forall i, \forall j: T_{O_{K_{ij}}} \approx \sum_{k=1, \dots, |M_{C_{ij}}|} (T_{C_{ijk}} \cdot K_{C_{ijk}}); \\ \forall i, \forall j: T_{K_{ij}} \approx T_{H_{ij}} + \sum_{k=1, \dots, |M_{C_{ij}}|} T_{C_{ijk}}.$$

Таким образом, целью календарного планирования является определение для каждого ОК числа проектировщиков, необходимого для достижения максимального сокращения сроков проектирования всего i -го изделия, а именно:

$$(10) \quad \forall i=1, \dots, |Pu|: T_{\mathcal{B}_i} \rightarrow \min.$$

При решении данной задачи необходимо учитывать:

– для проектирования любого ОК может быть задействовано только ограниченное число проектировщиков:

$$(11) \quad \forall i, \forall j, \forall k: M_n(O_{K_{ij}}, П_{C_{ijk}}) \leq K_{C_{ijk}} \leq M_x(O_{K_{ij}}, П_{C_{ijk}}),$$

где $M_n(O_{K_{ij}}, П_{C_{ijk}})$, $M_x(O_{K_{ij}}, П_{C_{ijk}})$ – минимальное и максимальное количество специалистов профиля $П_{C_{ijk}}$, необходимое для проектирования $O_{K_{ij}}$;

– штатное расписание предприятия также ограничивает число одновременно задействованных специалистов:

$$(12) \quad \forall i, \forall j, \forall k : \sum_{\substack{x=1, \dots, |OK_i|, \\ y=1, \dots, |MC_{ij}|, \\ T_{ij} = T_{ix} \vee T_{ij} = T_{ix}, \\ Pc_{ijk} = Pc_{ixy}}} Kc_{ixy} \leq Кол(Pc_{ijk}),$$

где $Кол(Pc_{ijk})$ – количество специалистов профиля Pc_{ijk} на предприятии.

– при планировании процессов проектирования необходимо учитывать связи, ранее определенные при построении архитектуры изделия:

$$Cv_{ij} = \langle OK\delta_{ij}, OKo_{ij}, Tcv_{ij} \rangle, \quad i = \overline{|Pu|}, \quad j = \overline{|Cv_i|}$$

$$(13) \quad \forall i, \forall j, \exists x, y = 1, \dots, |OK_i| : \left(\begin{array}{l} OK\delta_{ij} = OK_{ix} \wedge \\ \wedge OKo_{ij} \subseteq OK_{iy} \wedge T_{ix} \leq T_{iy} \end{array} \right),$$

где $OK\delta_{ij}$ – образующий элемент связи Cv_{ij} ; OKo_{ij} – второй элемент связи Cv_{ij} ; Tcv_{ij} – тип образуемой связи Cv_{ij} .

3. Информационное и алгоритмическое обеспечение подсистемы календарного планирования процессов КТПП

В соответствии с представленной формальной моделью были разработаны информационное (рис. 2) и алгоритмическое обеспечение процессов КТПП.

Алгоритм планирования процессов КТПП имеет вид:

1. Оценка важности B_{ij} каждого OK_{ij} в изделии, выбор формы контроля качества проектных работ при его разработке, определение максимального количества итераций цикла его проектирования $ИТ(B_{ij})$.

2. Определение «внешних» (по отношению к ОК) контрольных точек и формирование «внешних» итерационных циклов проектирования через:

– Формирование квазиоптимального плана выполнения проектных работ при заданном количестве специалистов, для чего был

использован «жадный» (GRASP) алгоритм с применением FFD¹ (первоначально располагают в убывающем порядке) эвристики:

$$(14) \forall i, \forall x, y=1, \dots, |O_{k_i}|: ((T_{ok_{ix}} > T_{ok_{iy}}) \rightarrow (T_{n_{ix}} \leq T_{n_{iy}})).$$

Тогда с учетом выявленных правил и ограничений (12)–(14) время начала проектирования любого ОК можно определить как:

$$(15) \forall i, \forall j: \left(T_{n_{ij}} = \min_{t \in [0; T_{\theta_i}]} (t) \right).$$

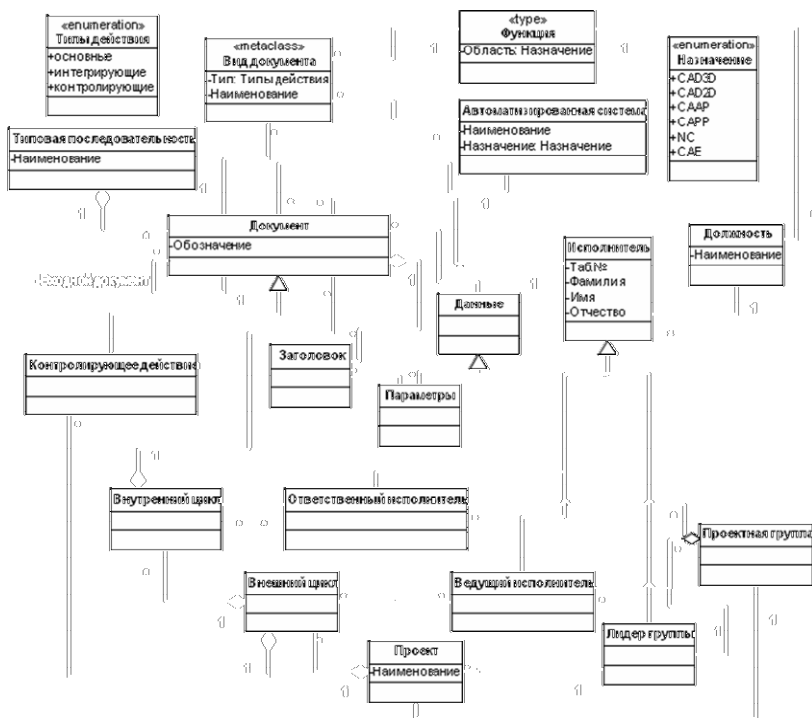


Рис. 2. Информационное обеспечение подсистемы календарного планирования процессов КТПП

– формирование квазиоптимального по критерию (10) распределения проектировщиков для разработки ОК изделия, где

¹ Выполняется только в случае соблюдения ограничений (12) и (13).

время проектирования ОК определяется по формуле (9), в которой время выполнения каждого вида работ оценивается через:

$$(16) \forall i, \forall j, \forall k : Tc_{ijk} \approx \frac{Tum_{ijk} \cdot ИТ(B_{ij})}{Kc_{ijk}},$$

где Tum_{ijk} – время выполнения k -го вида проектных работ для Ok_{ij} за одну итерацию, определяемое опытным путем.

Данная задача оптимизации решается методом пошагового улучшения (рис. 3) первоначального плана, в котором во всех ОК каждый вид работ выполняется минимально-возможным числом один проектировщиков Mn (Ok_{ij} , Pc_{ijk}).

3. Определение количества задействованных проектировщиков KCu_i в каждом виде работ ПС:

$$(17) \forall i, \forall PC : KCu_i(PC) = \max_{t \in [0; T\theta_i]} \left(\sum_{\substack{x=1, \dots, |Ok_i|, \\ k=1, \dots, |Mc_{ij}|, \\ Tn_{ix} \leq t, Tk_{ix} \geq t, \\ PC_{ixk} = PC}} Kc_{ixk} \right).$$

4. Назначение непосредственных исполнителей проектирования ОК изделия, формирование проектных групп.

4. Апробация подсистемы календарного планирования процессов КТПП

Для оценки эффективности предлагаемой подсистемы были проведены вычислительный эксперимент и апробация в реальных производственных условиях.

Вычислительный эксперимент¹ проводился с целью оценки временных затрат на календарное планирование процессов КТПП, в результате которого было выявлено, что эти затраты напрямую зависят от количества выполненных «жадных» процедур формирования календарных планов.

¹ Эксперимент проводился на ЭВМ Intel Pentium D 3000.

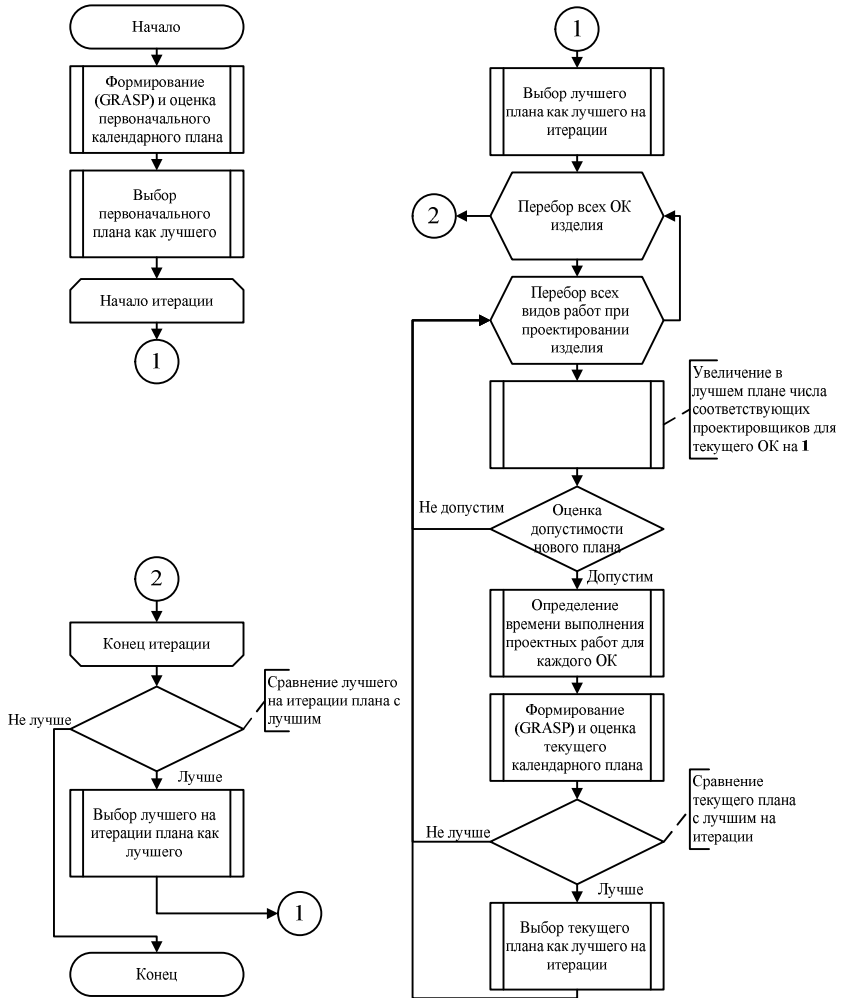


Рис. 3. Схема алгоритма формирования сетевого плана проектных работ

Из приведенного алгоритма видно, что максимально возможное количество таких процедур КП_i для *i*-го изделия составит:

$$(18) \quad КП_i = 1 + N_i \cdot M \cdot \prod_{j=1, \dots, |M|} (KC_j - 1), \quad \forall i = 1, \dots, |Pu|,$$

где N_i – количество ОК в i -м изделии; M – количество видов проектных работ, выполняемых в процессе КТПП на предприятии; KC_j – количество специалистов, выполняющих соответствующих вид проектных работ.

В результате эксперимента было выявлено, что на формирование (крайний случай) календарного плана для 1000 ОК на 300 специалистов (4 отдела по 75 человек) потребовалось примерно 324000 «жадных» процедур, что заняло около 4,5 часа. При этом критический путь в полученном календарном плане (время разработки ОК каждой службой выбиралось от 100 до 300 часов) составил – 10886 часов или 6 лет.

Время же формирования календарного плана для 100 объектов конфигурации на 300 специалистов (4 отдела по 75 человек) составило 63201 цикл или примерно 4 минуты.

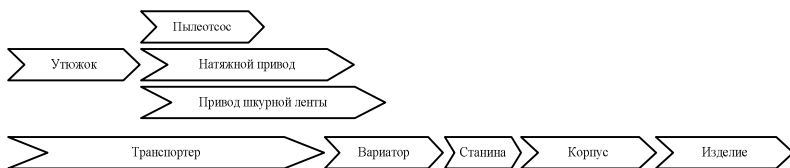


Рис. 4. Схема расчета критического пути проектирования станка ШЛПС-9

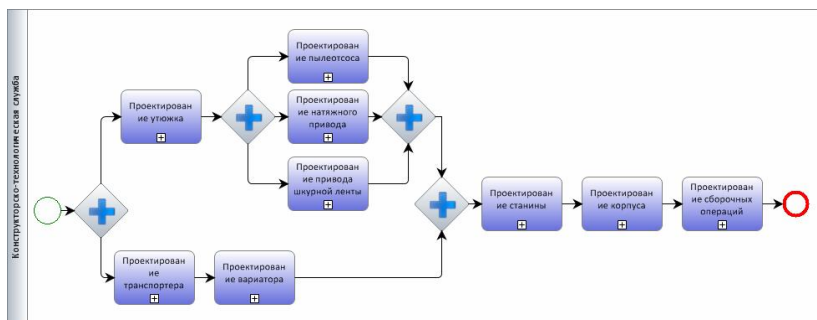


Рис. 5. Workflow диаграмма проектирования станка ШЛПС-9

Предлагаемый алгоритм был также опробован при планировании проектных работ для станка ШЛПС-9 (рис. 1), в результате чего был сформирован сетевой план, представленный на рисунке 4. Workflow-диаграмма проектирования данного станка приведена на рисунке 5.

Проведенные вычислительный эксперимент и экспертный анализ полученных результатов апробации показали высокую эффективность предлагаемой системы, что позволяет рекомендовать ее для внедрения в условия мелкосерийных машиностроительных предприятий.

Литература

1. ГОСТ 2.103–68. ЕСКД. *Стадии разработки*. – М., 1971.
2. ГОСТ Р ИСО 10006–2005. *Системы менеджмента качества. Руководство по менеджменту качества при проектировании*. М.: 2005.
3. КОСТРОВ А.В. *Основы информационного менеджмента*: учеб. для вузов.– 2-е изд., перераб. и доп. М.: Финансы и статистика, 2008.
4. *Народнохозяйственные последствия присоединения России к ВТО*: доклад национального инвестиционного совета и Российской академии наук. 2002. URL: <http://www.wto.ru/ru/content/documents/docs/NIC1.doc> (дата обращения: 08.05.11).
5. СТАРОДУБОВ В. *Управление конфигурацией: задачи, стандарты и реализация* // CAD/CAM/CAE Observer. 2006. №4 (28). С. 30 – 33.
6. *Теория и методы управления конфигурацией*. – URL: <http://www.cals.ru/material/mater/UK.pdf> (дата обращения: 08.05.11).
7. ЯБЛОЧНИКОВ Е.И., МОЛОЧНИК В.И., ФОМИНА Ю.Н. *Реинжиниринг бизнес-процессов проектирования и производства*: учеб. пособие. СПб.: СПбГУИТМО, 2008.
8. *MIL-HDBK-61. Military handbook. Configuration management guidance*. Department of Defense, USA, 1997.

SYSTEM OF DESIGN PROCESSES SCHEDULING FOR SMALL-SCALE MACHINE-BUILDING MANUFACTURE

Artem Denisov, Kostroma State University named after N.A. Nekrasov, Cand.Sc., assistant professor (inf_service@ksu.edu.ru).

Mikhail Belyankin, Kostroma State University named after N.A. Nekrasov, post graduate student (inf_service@ksu.edu.ru).

Abstract: Problems are considered of workflow-based scheduling system development for design processes of small-scale machine-building manufacture. Automated construction of design process workflow diagrams is proved to be required for small-scale machine-manufacturing. Algorithms for workflow diagrams automated construction are implemented in an information system. The algorithms use a formal model of design processes based on configuration management principles.

Keywords: production design process, scheduling, configuration management, small-scale engineering manufacture.

*Статья представлена к публикации
членом редакционной коллегии Г.Н. Каляновым*

УДК 004.942

ББК 3 + 31

СРАВНЕНИЕ ЭФФЕКТИВНОСТИ АЛГОРИТМОВ УПРАВЛЕНИЯ С УЧЁТОМ ТОЧНОСТИ ДАННЫХ И РЕАЛИЗАЦИИ РЕШЕНИЙ

Куров Б. Н.¹

*(Учреждение Российской академии наук
Институт проблем информатики РАН, Москва)*

Рассмотрена усовершенствованная версия метода сравнения эффективности алгоритмов управления с учётом точности данных и реализации решений. Метод выбора эффективного алгоритма, соответствующего условиям применения, инвариантен к типу задач управления, сложности сравниваемых алгоритмов и условий применения. Приведён пример сравнения алгоритмов распределения нагрузок в энергосистеме.

Ключевые слова: сравнение эффективности алгоритмов управления, точность данных и реализации решений, s-моделирование.

1. Введение

При разработке и выборе алгоритмов решения оптимизационных задач управления, рассчитанных на программную реализацию, необходимо оценивать их сравнительную эффективность.

В 1972 г. В.Д. Ильин предложил идею метода сравнения эффективности алгоритмов управления, учитывающего точность данных и реализации решений. Первая версия метода, конкретизированная применительно к алгоритмам распределения активных нагрузок в энергосистеме, была опубликована в [14]. С тех пор метод совершенствуется и применяется для оценки эффек-

¹ Борис Николаевич Куров, кандидат технических наук, доцент
(bnkurov@ipiran.ru).

тивности алгоритмов решения различных задач управления: оперативного распределения нагрузок в энергосистеме [11], расчета потоков мощности по линиям электропередач [15], идентификации установившихся режимов электрической сети при управлении в условиях неполной информации [16], идентификации и оперативных расчётов потоков мощности по межсистемным линиям энергообъединений [17], распределения ресурсов [5, 7, 8].

Совершенствованию обсуждаемого метода способствовало развитие методологии *s*-моделирования [9, 10, 18–20] программируемых задач, изучаемых как конструктивные объекты [6, 9, 10, 12].

Особое значение для развития метода имело применение языка *TSM* спецификации объектов *s*-моделирования [6, 9, 10], так как записи постановок задач, методов и алгоритмов на языке *TSM* имеют вид, привычный для разработчиков программно реализуемых алгоритмов и программистов².

Актуальность метода сравнения эффективности алгоритмов решения различных задач управления будет возрастать по мере нормализации экономических отношений [13].

2. Постановка задачи

Задача оценки сравнительной эффективности алгоритмов заключается в том, чтобы из заданной совокупности программно реализуемых алгоритмов решения некоторой задачи управления выбрать наиболее эффективный в заданных условиях применения. Рассматриваемый метод сравнения эффективности алгоритмов инвариантен к типу задач управления, сложности сравниваемых алгоритмов и условий применения.

Чтобы конкретизировать изложение метода, рассмотрим упрощённый пример сравнения эффективности алгоритмов распределения активных нагрузок между тепловыми станциями энергосистемы (ЭС). Пусть для каждой станции известны экономические характеристики, которые в сравниваемых алгорит-

² Запись алгоритма на языке *TSM* может применяться как задание на программирование.

мах представлены разными математическими моделями. Прогнозируемую суммарную нагрузку энергосистемы требуется распределить между станциями так, чтобы на заданном отрезке времени минимизировать суммарные затраты на топливо в энергосистеме при выполнении условий баланса мощностей и ограничений на изменения нагрузок станций. Вычисленные по сравниваемым алгоритмам вектора распределения нагрузок являются конкретными представителями сравниваемых алгоритмов.

Предположим, что ЭС включает n электростанций, экономичность работы которых определяется расходными характеристиками – функциями затрат на топливо $y|j|(x|j|)$, где $x|j|$ – нагрузка j -й станции. Известна суммарная нагрузка энергосистемы $x|c|$ (c – помета), включающая потери в электрической сети. Задача оптимального распределения нагрузок между электростанциями заключается в минимизации затрат

$$(1) \quad y = \sum|j|, j = 1...n| y|j|(x|j|) \rightarrow \min$$

при ограничениях, заданных условием баланса в ЭС:

$$(2) \quad \sum|j| x|j| - x|c| = 0,$$

и допустимыми изменениями нагрузок станций:

$$(3) \quad x|j|_{\min} \leq x|j| \leq x|j|_{\max}.$$

Чтобы облегчить дальнейшее восприятие записи формул на языке *TSM*³, приведём выражения (1)–(3) в традиционной записи:

$$y = \sum_{j=1}^n y_j(x_j) \rightarrow \min,$$

$$\sum_{j=1}^n x_j - x_c = 0,$$

$$x_j^{\min} \leq x_j \leq x_j^{\max}.$$

Решение сформулированной задачи сводятся к решению системы нелинейных уравнений [2]:

$$dy|j|/dx|j| + \lambda = 0,$$

$$\sum|j| x|j| - x|c| = 0.$$

На практике значения функций $y|j|(x|j|)$ получают путём натуральных испытаний для заданного набора значений $x|j|$. То есть $y|j|(x|j|)$ — это табличные функции. В алгоритмах оптимального распределения предусматривается их интерполирование.

³ Подробное описание языка *TSM* приведено в [10].

В сравниваемых далее алгоритмах используются различные способы построения интерполяционных функций. В первом из них (алгоритме A|1|) строится кубический сплайн [1] по точкам $(x|0|, y|0|)$, $(x|1|, y|1|)$, ..., $(x|N|, y|N|)$ произвольной расходной характеристики. Сплайн-функция $S(f, x)$ является многочленом третьей степени на каждом из отрезков $[x|i-1|, x|i|]$:

$$S(f, x) = f|i|(x) = a|i0| + a|i1|*x + a|i2|*x**2 + a|i3|*x**3$$

при $x|i-1| \leq x \leq x|i|$, $i = 1...N$, и удовлетворяет следующим условиям:

$$(4) f|i|(x|i|) = y|i|, f|1|(x|0|) = y|0|,$$

$$(5) f|k|; j|(x|j|) = f|k|; j+1|(x|j|), k = 1, 2; j = 1...(N-1),$$

где k – порядок производной.

Для отыскания коэффициентов $a|i0|$, $a|i1|$, $a|i2|$ и $a|i3|$ функции $S(f, x)$ методом прогонки [4] решается система линейных (относительно вторых производных сплайна в заданных точках z) уравнений. Разрешимость этой системы (при заданных нулевых краевых условиях) гарантирована [3], поскольку главная диагональ матрицы коэффициентов при неизвестных является доминирующей.

Сплайн $S(f, x)$ не всегда является наилучшим. Дело в том, что при интерполяции таким сплайном возникают посторонние точки перегиба, что требует алгоритмического контроля монотонности функции. Такого недостатка лишён кубический сплайн «с растяжением» [21], который применяется в конкурирующем алгоритме A|2|. Для кубического сплайна «с растяжением» вместо линейности вторых производных требуется линейность разности $f''; i|(x) - \alpha*f|i|(x)$ на $[x|i-1|, x|i|]$ (" – помета, обозначающая вторую производную от f):

$$(6) f''; i|(x) - \alpha*f|i|(x) = (z|i-1| - \alpha*y|i-1|)*(x|i-1| - x)/h|i| + \\ + (z|i| - \alpha*y|i|)*(x - x|i-1|)/h|i|, \alpha > 0$$

при сохранении условий (4) и (5).

Дифференцирование решения уравнения (6) приводит к системе $(N-1)$ -го линейного уравнения. Её решением (при заданных краевых условиях $f''; 1|(x|0|) = y''; 0|$, $f''; N|(x|N|) = y''; N|$) являются значения $z|0|$, $z|1|$, ..., $z|N|$.

Для кубического сплайна «с растяжением» решение получается аналогично. При $\alpha < 0,001$ кубический сплайн «с растяжени-

ем» близок к кусочно-линейной функции, а при $a > 50$ – к кубическому сплайну.

Третьим конкурирующим алгоритмом (A|3|) является алгоритм, для которого исходными данными служат характеристики относительных приростов электростанций [14], при построении которых для вычисления производных от таблично заданных расходных характеристик применяется метод численного дифференцирования. Используется трёхточечная формула для равномерной сетки:

$y|i + 1| == (y|i + 2| - y|i|)/(2*\delta x)$, где $y|i|$, $y|i + 1|$, $y|i + 2|$ – три последовательных значения функции, а δx – шаг дискретизации. Полученной табличной функции ставим в соответствие интерполирующую кусочно-линейную функцию. Вектор оптимальных нагрузок станций соответствует равенству относительных приростов.

Итак, в нашем примере сравниваем эффективность алгоритмов A|1|, A|2| и A|3|.

3. Сравнение решений без учета изменений состояния объекта

Решения, получаемые по конкурирующим алгоритмам, должны сравниваться посредством единого критерия оптимальности, соответствующего исходным расходным характеристикам. Для этих характеристик выполнена процедура интерполирования кусочно-линейными функциями $y|0; j|(x|j|)$. Сравним решения, полученные по двум конкурирующим алгоритмам, например, A|1| и A|2|. Схема сравнения без учёта изменения состояния объекта (детерминированное сравнение) такова.

Вначале определяются оптимальные векторы нагрузок $x(A|1|) = x|1j, j = 1...n|$ и $x(A|2|) = x|2j, j = 1...n|$ для алгоритмов A|1| и A|2|. Затем вычисляется разность расходов топлива, соответствующая векторам $x(A|1|)$ и $x(A|2|)$:

$$\begin{aligned} y(A|1|, A|2|) &= \\ &= \sum|j, j = 1...n| (y|0; 1j|(x|1j|) - y|0; 2j|(x|2j|)). \end{aligned}$$

Полученное значение $y(A|1|, A|2|)$ определяет сравнительную эффективность решений алгоритмов A|1| и A|2|. Если $y(A|1|, A|2|) > 0$, то следует предпочесть решение, полученное по

алгоритму $A|2|$, если же $y(A|1|, A|2|) < 0$ — по алгоритму $A|1|$. Аналогично проводится сравнение и для других пар алгоритмов.

4. Сравнение решений с учетом изменений состояния объекта

Случайным изменениям подвергаются исходные данные алгоритмов распределения нагрузок: характеристики электростанций и суммарная нагрузка системы. Ошибки вносятся и в вычисленные оптимальные значения нагрузок.

При оценке сравнительной эффективности алгоритмов (с учетом изменения состояния объекта при реализации решений) схема сравнения должна включать блоки имитации «истинного» значения суммарной нагрузки, случайных изменений компонент векторов $x(A|1|)$ и $x(A|2|)$, имитации поведения объекта при достижении баланса генерации и потребления, а также имитации «истинных» характеристик электростанций.

4.1. УЧЕТ СЛУЧАЙНОГО ХАРАКТЕРА НАГРУЗКИ СИСТЕМЫ

При работе сравниваемых алгоритмов векторы распределения нагрузок вычислены для прогнозируемого значения суммарной нагрузки системы $x|c|$. При реализации их в системе истинное значение $x|c|$ обычно отличается от расчётного. В предлагаемой схеме сравнения ошибка прогноза суммарной нагрузки является случайной величиной. В каждом испытании истинное значение $x|c|$ вычисляется следующим образом:

$$(7) \quad x|c| = (1 + d * u) * x|c|,$$

где d — коэффициент, задающий дисперсию ошибки прогноза; u — значение случайной величины с фиксированным законом распределения (закон распределения формируется путем обработки статистических данных об ошибках прогноза суммарной нагрузки).

Найденное значение $x|c|$ используется в блоке имитации поведения системы управления в процессе достижения баланса генерации и потребления путем изменения нагрузки выбранной регулирующей станции на величину разности $x|c| - x|c|$. В общем случае из-за ограничений нагрузки регулирующей станции предусматривается возможность изменения нагрузок нескольких станций.

4.2. УЧЕТ СЛУЧАЙНОГО ИЗМЕНЕНИЯ ВЕКТОРА РАСПРЕДЕЛЕНИЯ

Вычисленные по разным алгоритмам значения векторов нагрузок, как правило, не являются целыми числами и на практике не задаются в качестве нагрузок станций, поэтому они округляются. Округлённые и сбалансированные значения нагрузок подвергаются случайным изменениям, отражающим неточность их реализации станциями:

$$(8) \quad x^j; j = x^j; j^*(1 + r*u|j|),$$

где $x^j; j$ – j -я компонента вектора с округленным значением (например, кратным 5 или 10 МВт), причем $\sum |j| x^j; j = x|c|$ (предполагается, что $x|c|$ тоже кратно 5 или 10 МВт); r – коэффициент, задающий дисперсию ошибок реализации; $u|j|$ – значение случайной величины с заданным законом распределения.

Полученные компоненты $x^j; j$ используются в блоке имитации поведения системы управления, обеспечивающем выполнение условия $\sum |j| x^j; j = x^j; c|$.

4.3. УЧЕТ НЕОПРЕДЕЛЕННОСТИ ХАРАКТЕРИСТИК СТАНЦИЙ

Истинные характеристики станций неизвестны в силу ограниченной точности натурных испытаний и неконтролируемого изменения параметров оборудования с течением времени. Предположим, что известен закон распределения истинных характеристик, который используется при моделировании.

В рассматриваемом примере расчетными являются расходные характеристики, полученные путём кубической сплайн-интерполяции (в алгоритме $A|1|$), кубической сплайн-интерполяции «с растяжением» (в $A|2|$), характеристики относительных приростов (в $A|3|$). По расчетным характеристикам получаем оптимальные векторы распределения $x|1j, j = 1...n|$, $x|2j, j = 1...n|$ и $x|3j, j = 1...n|$, а по истинным определяем расход топлива ЭС. Сравнительная эффективность алгоритмов определяется в результате их попарного сравнения. Например, для алгоритмов $A|1|$ и $A|2|$ она вычисляется по формуле:

$$(9) \quad y(A|1|, A|2|) = \\ = \sum |s, s = 1...S| q|s|^* \sum |j| (y|s; j|(x|1j|) - y|s; j|(x|2j|)),$$

где $q|s|$ – вероятность того, что истинной будет характеристика $y|s; j|(x|j|)$; $x|1|j|$ и $x|2|j|$ – оптимальные нагрузки, полученные по алгоритмам $A|1|$ и $A|2|$

При положительном значении $y(A|1|, A|2|)$ лучшим является алгоритм $A|2|$, при отрицательном — алгоритм $A|1|$.

Функции $y|s; j|(x|j|)$ могут быть получены различными способами. Укажем два из них, которые соответствуют кусочно-линейному представлению характеристик:

$$(10) y|'; j| = y|j| + d^*u|j|*y|max; j|, x|'; j| = x|j|;$$

$$(11) y|'; j| = y|j|*(1 + d^*u|j|), x|'; j| = x|j|*(1 + r^*v|j|),$$

где d и r – коэффициенты, задающие разброс характеристик; $u|j|$ и $v|j|$ – значения случайных величин, распределенных по выбранным законам, причем значения $y|'; j|$ и $x|'; j|$ определяются в заданных точках характеристик, $y|max; j|$ – максимальное значение расхода топлива, полученное при натурных испытаниях.

В алгоритме получения характеристик предусмотрена проверка их неубывания. Характеристики, имеющие убывающие участки, бракуются.

5. Общий алгоритм оценки сравнительной эффективности решений

Взаимодействие основных блоков алгоритма сравнения решений покажем на примере сравнения алгоритмов $A|1|$ и $A|2|$:

1. Вычисляются расчётные характеристики путем их интерполяции кубическими сплайнами и кубическими сплайнами «с растяжением».

2. Производится оптимальное распределение по конкурирующим алгоритмам $A|1|$ и $A|2|$ для заданного значения суммарной нагрузки системы $x|c|$.

3. Округляются компоненты векторов $x(A|1|)$ и $x(A|2|)$.

4. Полученные значения приводятся в соответствии с условием $\sum|j| x|''; j| = x|c|$ путем изменения нагрузки регулирующих станций (где $x|''; j|$ – округленные значения компонент векторов $x(A|1|)$ и $x(A|2|)$).

5. Имитируется неточность поддержания значений $x|''; |(A|1|)$ и $x|''; |(A|2|)$ на станциях (в соответствии с (8)).

6. Имитируется случайное значение нагрузки $x|'$; $c|$ в соответствии с выражением (7).

7. Балансируются полученные в соответствии с п.5 векторы из условия равенства суммарной генерации значению $x|'$; $c|$.

8. Имитируются расходные характеристики в соответствии с (10) или (11).

9. По полученным характеристикам в соответствии с (9) вычисляется сравнительная эффективность решений алгоритмов $A|1|$ и $A|2|$.

Усреднение оценок сравнительной эффективности решений проводится по множеству истинных характеристик относительных приростов, а также по множеству случайных векторов распределения и случайных значений суммарной нагрузки системы. По этой же схеме сравниваются и другие алгоритмы.

6. Результаты сравнения

Сравнение алгоритмов выполнено применительно к ЭС из десяти станций с использованием реальных данных о составах работающего оборудования, о расходных характеристиках электростанций и об изменении суммарной нагрузки ЭС.

Суммарная нагрузка ЭС изменялась от 4200 до 5200 МВт. Сравнивались алгоритмы с расходными характеристиками, полученными путём кубической сплайн-интерполяции ($A|1|$), кубической сплайн-интерполяции с «растяжением» ($A|2|$) и характеристиками относительных приростов, построенными с применением формулы численного дифференцирования ($A|3|$).

В первой схеме сравнение проводилось без учета изменения состояния ЭС при реализации решений (детерминированное сравнение).

Во второй схеме осуществлялся учет округления компонент расчетных векторов распределения, изменения векторов при ведении режима станциями, случайного отклонения суммарной нагрузки от расчетного значения и учет неопределенности истинных характеристик станций.

Округление проводилось из условия кратности нагрузок 5 МВт, что составляло величину порядка 0,1%–0,5% генерации станций.

При получении истинных значений суммарной нагрузки и реализуемых векторов распределения использовался генератор случайных чисел с усеченным нормальным законом распределения и параметрами $(0, 1)$.

Для моделирования характеристик применялся генератор случайных чисел с равномерным законом распределения в интервале $(-1, 1)$.

При попарном сравнении всех алгоритмов использовалась функция оценки, построенная путём интерполяции кусочно-линейными функциями табличных расходных характеристик, полученных при натурных испытаниях.

Результаты моделирования показывают, что учёт изменения состояния объекта влияет на оценку сравнительной эффективности алгоритмов A|1|, A|2| и A|3|. Например, при значениях суммарной нагрузки 4200 и 4400 МВт при детерминированном сравнении эффективнее работал алгоритм A|1|, а при тех же значениях нагрузки с учётом изменения состояния объекта (ошибка прогноза $x|c|$ и реализации составляла 0,5%, а зона неопределённости характеристик – 5%) эффективнее уже был A|2|. В обоих случаях изменение оценок не превосходило 0,06%, что в абсолютном выражении составляло 0,9 т условного топлива в час. Сравнительная эффективность существенно зависит от суммарной нагрузки энергосистемы. Например, в диапазоне нагрузок 4200–4300 МВт эффективнее работал алгоритм A|3|, а в диапазоне 4430–4560 МВт – алгоритм A|1|. При детерминированной схеме сравнения наилучшие результаты показывали алгоритмы A|1| и A|2|, тогда как при учёте изменения состояния объекта перевес уже был на стороне алгоритма A|3|.

В каждой серии испытаний усреднение проводилось до получения устойчивых статистических оценок (50–200 испытаний) путем построения доверительных интервалов.

Оценка сравнительной эффективности алгоритмов для интервала значений суммарной нагрузки производилась путем усреднения результатов, полученных для отдельных значений этого интервала.

Основной вывод заключается в том, что усложнение s-моделей объекта моделирования и, как следствие, алгоритмов оптимизации не даёт гарантии улучшения результатов решений.

Приведённый пример сравнения эффективности алгоритмов показывает, что наиболее целесообразные стратегии управления должны основываться на применении совокупности программно реализованных алгоритмов, из которых осуществляется выбор, соответствующий заданным условиям функционирования.

7. Заключение

1. Рассмотрена усовершенствованная версия метода оценки сравнительной эффективности алгоритмов управления, в которой для описания постановок задач и алгоритмов их решения используется язык *TSM* спецификации объектов s-моделирования.

2. Метод инвариантен к типу задач управления, сложности сравниваемых алгоритмов и условий применения.

3. Область применения метода расширяется с ростом числа систем управления различного назначения.

Литература

1. АЛБЕРГ ДЖ., НИЛЬСОН Э., УОЛШ ДЖ. *Теория сплайнов и ее приложения*. – М.: Мир, 1972. – 316 с.
2. БЕРТСЕКАС Д. *Условная оптимизация и методы множителей Лагранжа*. – М.: Радио и связь, 1987. – 400 с.
3. ГАНТМАХЕР Ф.О. *Теория матриц*. – М.: Наука, 1966. – 576 с.
4. ГОДУНОВ С.К., РЯБЕНЬКИЙ В.С. *Введение в теорию разностных схем*. – М.: Физматгиз, 1962. – 340 с.
5. ИЛЬИН А.В. *Математическое обеспечение процессов преобразования ресурсов* // Системы и средства информатики. – 1999. – Вып. 9. – С. 159–177.
6. ИЛЬИН А.В. *Конструирование разрешающих структур на задачных графах системы знаний о программируемых задачах* // Информационные технологии и вычислительные системы. – 2007. – №3. – С. 30–36.
7. ИЛЬИН А.В., ИЛЬИН В.Д. *Интерактивный преобразователь ресурсов с изменяемыми правилами поведения* // Информационные технологии и вычислительные системы. –

2004. – №2. – С. 67–77.
8. ИЛЬИН А.В., ИЛЬИН В.Д. *Распределение ресурсов по обязательным и ориентирующим правилам: сравнительная эффективность алгоритмов* // Системы и средства информатики. – 2005. – Вып. 15. – С. 123–159.
 9. ИЛЬИН А.В., ИЛЬИН В.Д. *Основы теории s-моделирования*. – М.: ИПИ РАН, 2009. – 143 с.
 10. ИЛЬИН А.В., ИЛЬИН В.Д. *S-моделирование объектов информатизации*. – М.: ИПИ РАН, 2010. – 412 с. – URL: <http://smodeling.wordpress.com/2010/12/29/245/>.
 11. ИЛЬИН В.Д. *Оперативное распределение нагрузок в энергосистеме* // Сб. Автоматизированные системы контроля и управления. Киев, Ин-т кибернетики АН УССР – 1973. – С. 70–76.
 12. ИЛЬИН В.Д. *Система порождения программ*. – М.: Наука, 1989. – 264 с.
 13. ИЛЬИН В.Д. *Модель нормализованной экономики (НЭк-модель): основы концепции* // Управление большими системами: сборник трудов. – 2009. – №25. – С. 116–138.
 14. ИЛЬИН В.Д., КУРОВ Б.Н. *Сравнение алгоритмов распределения нагрузок с учётом изменения состояния энергосистемы при реализации решений* // Электричество. – 1972. – №9. – С. 7–10.
 15. ИЛЬИН В.Д., КУРОВ Б.Н. *Математическая модель для оперативного определения потоков мощности по ВЛ в автоматизированной системе диспетчерского управления* // Электричество. – 1973. – №11. – С. 23–28.
 16. ИЛЬИН В.Д., КУРОВ Б.Н. *Идентификация установившихся режимов электрической сети при управлении в условиях неполной информации* // Электричество. – 1976. – №1. – С. 18–22.
 17. ИЛЬИН В.Д., КУРОВ Б.Н., БАТАЛОВ А.Г. *Комплекс программ для идентификации и оперативных расчётов потоков мощности по межсистемным линиям энергообъединений* // Электричество. – 1980. – №8. – С. 53–54.
 18. ИЛЬИН В.Д., СОКОЛОВ И.А. *Символьное моделирование в человеко-машинной среде: основы концепции* // Информационные технологии и вычислительные системы. – 2008. –

№1. – С. 51–60.

19. ИЛЬИН В.Д., СОКОЛОВ И.А. *Символьная модель системы знаний информатики в человеко-автоматной среде* // Информатика и её применения. – 2007. – Т. 1. – С. 66–78.
20. ИЛЬИН В.Д., СОКОЛОВ И.А. *Информация как результат интерпретации сообщений на символьных моделях систем понятий* // Информационные технологии и вычислительные системы. – 2006. – №4. – С. 74–82.
21. SCHWEIKERT D.G. *An interpolation curve using a spline in tension* // Math. And Physics. – 1966. – Vol. 45. – P. 312–317.

COMPARISON OF CONTROL ALGORITHMS EFFICIENCY TAKING INTO ACCOUNT DATA ACCURACY AND SOLU- TIONS IMPLEMENTATION

Boris Kurov, Institute of Informatics Problems of RAS, Moscow, Cand. Sc., assistant professor (bnkurov@ipiran.ru).

Abstract: The article addresses to an improved version of the method of control algorithms efficiency comparison taking into account data accuracy solutions implementation. The method to choose an efficient algorithm that matches conditions of use is invariant to control problems type, complexity of algorithms compared and conditions of use. An example is given of algorithms comparison for division of load in a power system.

Keywords: control algorithms efficiency comparison, data accuracy and solutions implementation, s-modeling.

Статья представлена к публикации членом редакционной коллегии Д.А. Новиковым

УДК 681.518.5
ББК 32.817

ОПРЕДЕЛЕНИЕ ИНФОРМАЦИОННОЙ МОДЕЛИ ОБЪЕКТА ДИАГНОСТИКИ, ЗАДАННОГО СИГНАЛЬНЫМИ ГРАФАМИ ФУНКЦИОНАЛЬНЫХ ЭЛЕМЕНТОВ

Дубов А. В.¹

(Государственное образовательное учреждение высшего профессионального образования «Рязанский государственный радиотехнический университет», Рязань)

В работе описаны способы отыскания информационных моделей объектов диагностики с обратными связями, представленных сигнальными графами составляющих их функциональных элементов, позволяющие использовать их в качестве составных частей более сложных устройств.

Ключевые слова: сигнальный граф, функциональный элемент, регулярное выражение.

1. Введение

Для возможности диагностирования объект диагностики (ОД) должен быть разбиваем на связанные между собой функциональные элементы (ФЭ), каждый из которых может находиться в работоспособном или неработоспособном состояниях [4]. Причем, как отмечается в [2], для возможности сужения области рассмотрения выходов ФЭ и определения порядка проведения проверок выходов ФЭ из числа предположительно отказавших как при одиночных, так и при одновременных отказах, они должны быть заданы в виде сигнальных подграфов графа причинно-следственных связей ОД (рис. 1).

В данном рассмотрении сигнальным графом будем называть ориентированный граф, соответствующий линейным или

¹ Антон Владимирович Дубов, аспирант, (dubovanton@mail.ru).

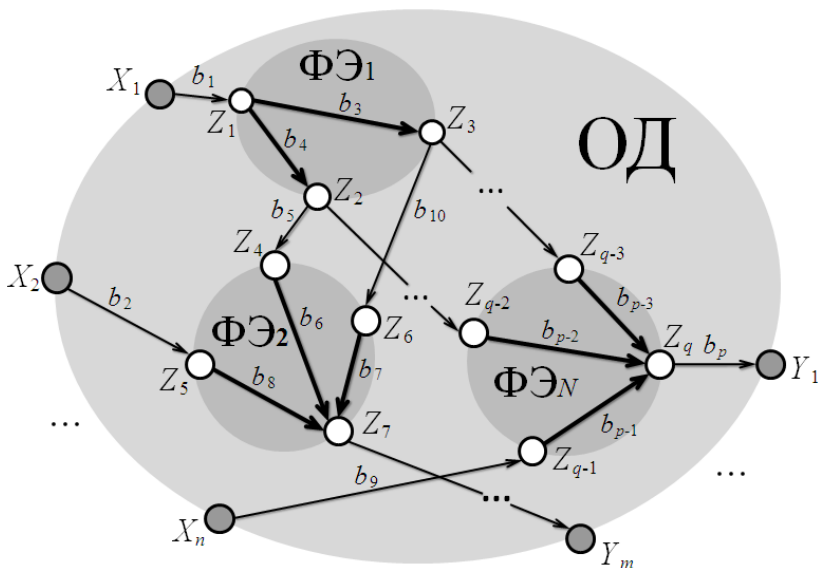


Рис. 1. Структура сигнального графа ОД

линеаризованным системам уравнений математической модели ОД. Вершины сигнального графа ОД соответствуют сигналам на входах $X_{\text{ОД}} = \{X_1, X_2, \dots, X_n\}$ и выходах $Y_{\text{ОД}} = \{Y_1, Y_2, \dots, Y_m\}$ ОД или на входах и выходах $Z_{\text{ОД}} = \{Z_1, Z_2, \dots, Z_q\}$ ФЭ, а ориентированные ребра – передаточным функциям (b_1, b_2, \dots, b_p), характеризующим связь между этими сигналами и образующим квадратную матрицу соединений B объекта диагностики. Входы ОД являются вершинами-источниками сигнального графа и отображают независимые сигналы. Выходы ОД являются вершинами-стоками и отображают зависимые сигналы. Вершины ФЭ, соответствующие входящим и исходящим ребрам, называют смешанными.

Построение сигнальных графов выполняют на основании следующих правил:

1. Сигналы передаются вдоль ребер только в направлении их ориентации.
2. Сигнал, проходящий вдоль какого-либо ребра, умножается на передачу этого ребра.

3. Сигнал, изображаемый какой-либо вершиной, является суммой всех сигналов, только приходящих в эту вершину.

4. Значение сигнала, изображаемого какой-либо вершиной, передается по всем ребрам, выходящим из нее.

Таким образом, граф (подграф) причинно-следственных связей ОД (ФЭ) устанавливает влияние какого-либо сигнала на входе ОД (ФЭ) на формирование того или иного сигнала на том или ином выходе ОД (ФЭ).

Так как ОД в свою очередь может являться составной частью более крупного устройства, то актуальной задачей является выявление информационных взаимосвязей между входами и выходами ОД (представление его в виде переходного графа причинно-следственных связей), по которым в дальнейшем можно определить аналитические функциональные модели ОД. Задача определения информационных моделей ОД прежде всего тесно связана с отысканием регулярных выражений (событий), описывающих функционирование ОД. Причем решение этой задачи для реальных систем часто усложняется наличием обратных связей внутри структуры ОД. В статье приводится алгоритм выявления регулярных выражений между входами X и выходами Y ОД, основанный на теории сигнальных графов.

2. Представление событий ОД

При функционировании ОД в каждый момент времени всегда удовлетворяются требования подачи входного сигнала $x(t)$ и выдачи выходного сигнала $y(t)$. В связи с этим алфавитное отображение, индуцируемое ОД, можно отнести к автоматным отображениям, для которых удобной формой является задание с помощью событий [1].

Рассмотрим эти события. Пусть $X = \{x_1, x_2, \dots, x_n\}$ – произвольный входной алфавит; $\sigma(X)$ – множество всех слов в этом алфавите. Любое подмножество множества $\sigma(X)$ называется событием в алфавите X .

Пусть DU – ОД с входным алфавитом $X = \{x_1, x_2, \dots, x_n\}$ и выходным алфавитом $Y = \{y_1, y_2, \dots, y_m\}$, который индуцирует частичное отображение f множества $\sigma(X)$ в $\sigma(Y)$. Событием R_j , представленным в объекте DU выходным сигналом y_j , называет-

ся множество всех слов $P \in \sigma(X)$ этого алфавита, для которых слово $f(P)$ определено и оканчивается буквой y_j . Если $W \subseteq Y$ – некоторое подмножество выходных сигналов, то событием, представленным в объекте DU множеством W , называется объединение событий, представленных всеми элементами этого множества. В том случае, когда W совпадает с алфавитом Y , соответствующее ему событие называется каноническим множеством событий R_1, R_2, \dots, R_m объекта DU .

Задание отображения f множества $\sigma(X)$ и $\sigma(Y)$ произвольного объекта DU с входным алфавитом $X = \{x_1, x_2, \dots, x_n\}$ и выходным алфавитом $Y = \{y_1, y_2, \dots, y_m\}$ эквивалентно заданию канонического множества событий R_1, R_2, \dots, R_m данного объекта. Отсюда вытекает, что произвольные отображения ОД можно задавать с помощью разбиений множества $\sigma(X)$ всех слов во входном алфавите на конечное число попарно непересекающихся событий. Эффективное описание конечных событий и некоторых классов бесконечных событий осуществляется при помощи алгебры событий.

Алгеброй событий в алфавите X называется множество всех событий в этом алфавите, на котором задана система трех операций: двух бинарных, называемых дизъюнкцией и конъюнкцией, и одной унарной, называемой итерацией.

Дизъюнкцией событий R и S называется событие P , обозначаемое $P = R + S$, которое образуется теоретико-множественным объединением событий R и S .

Произведением событий R и S называется событие U , обозначаемое $U = R \cdot S$ или $U = RS$, состоящее из всех слов вида $u = r \cdot s$, где $u \in U$, $r \in R$, $s \in S$. Слова события U образуются приписыванием справа любого слова события S к любому слову события R , и наоборот.

Итерацией события R называется событие, обозначаемое $\{R\}^*$, которое определяется как дизъюнкция пустого слова e , события R , события RR , события RRR и т.д. до бесконечности, т.е. $\{R\}^* = e + R + RR + RRR + \dots$.

При анализе ОД исследователь имеет дело с конечным числом событий, поэтому рассматриваемые события можно отнести к регулярным.

Регулярное выражение – формула в алгебре событий, причем одно и то же событие может быть по-разному выражено через одноэлементные события и операции дизъюнкции, конъюнкции и итерации. Основные свойства этих операций рассматриваются в [3].

3. Определение информационных моделей ОД

Тождественные преобразования, применяемые к сигнальным графам, можно распространить на графы ОД. Примеры сигнальных графов, соответствующих элементарным выражениям дизъюнкции и умножения, а также эквивалентные преобразования этих графов приведены на рис. 2.

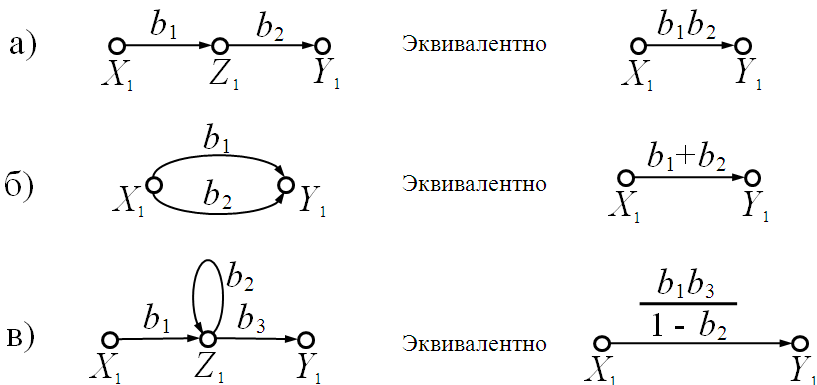


Рис. 2. Эквивалентные преобразования сигнальных графов

При устранении петель (рис. 2в) имеем $Z_1 = b_1 X_1 + b_2 Z_1$; $Y_1 = b_3 Z_1$. Из первого уравнения находим $Z_1 = \frac{b_1}{1 - b_2} X_1$, откуда $Y_1 = \frac{b_1 b_3}{1 - b_2} X_1$.

Задачу отыскания регулярных событий на выходах абстрактного ОД можно решить графическим и аналитическим методами.

3.1. ГРАФИЧЕСКИЙ МЕТОД

Отыскание регулярных событий, переводящих граф ОД из входных сигналов множества X в каждый выходной сигнал множества Y , сводится к нахождению ветвевых переходов из вершин X в Y . В этом случае граф ОД приводится к переходному графу, имеющему только вершины из множеств X и Y , а веса этих переходов как раз и являются искомыми регулярными событиями.

При сведении графа ОД к переходному графу с вершинами из множеств X и Y остальные вершины, принадлежащие множеству Z , должны быть удалены. Любой граф ОД можно привести к переходному графу с помощью элементарных преобразований, показанных на рис. 2. Сформулируем алгоритм отыскания регулярных событий, основанный на теории сигнальных графов.

1. По графу ОД определяем источники (вершины множества X) и стоки (вершины множества Y). Переходим к п.2.

2. Все параллельные дуги приводим к дугам с коэффициентом передачи, равным сумме коэффициентов исходных дуг (рис. 2а); последовательные – к дугам с коэффициентом передачи, равным произведению коэффициентов исходных дуг (рис. 2б); петли – по приведенной формуле (рис. 2в). Переходим к п.3.

3. Устраняем последовательно вершины графа ОД, входящие во множество Z .

Покажем работу алгоритма на примере.

Пусть дан ОД, граф которого показан на рис. 3а. Требуется найти регулярные события по выходам Y .

Источниками графа являются вершины X_1 , X_2 и X_3 , стоками – Y_1 и Y_2 . Исключая последовательно вершины графа Z_3 , Z_4 , получаем промежуточный граф (рис. 3б); Z_1 – промежуточный граф (рис. 3в), и Z_2 – получаем ветвевые переходы (рис. 3г), которые определяют искомые регулярные события Y_1 и Y_2 (информационную модель ОД):

$$Y_1 = \frac{b_1 b_2 b_4 b_6}{1 - b_2 b_3 - b_2 b_4 b_5} X_1 + \frac{b_4 b_6 b_7}{1 - b_2 b_3 - b_2 b_4 b_5} X_2,$$

$$Y_2 = \frac{b_1 b_2 b_8 b_{10}}{1 - b_2 b_3 - b_2 b_4 b_5} X_1 + \frac{b_7 b_8 b_{10}}{1 - b_2 b_3 - b_2 b_4 b_5} X_2 + b_9 b_{10} X_3.$$

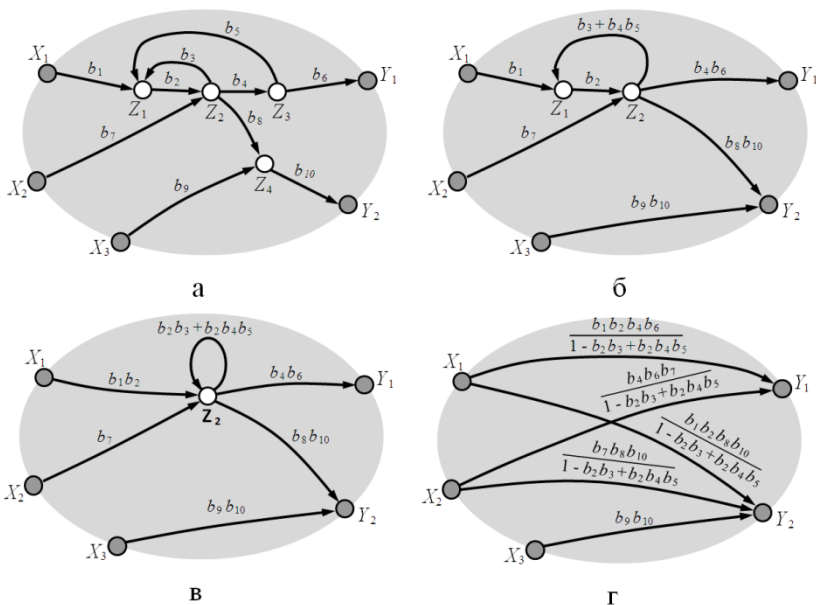


Рис. 3. Сигнальный граф ОД и его преобразования

3.2. АНАЛИТИЧЕСКИЙ МЕТОД

Аналитический метод состоит в нахождении регулярных событий для множества Y на выходах ОД путем решения системы линейных уравнений в алгебре событий. Запишем систему в виде уравнения

$$(1) R = R^T B + C,$$

где B – матрица соединений ОД; $R = \{X, Z, Y\}$ – вектор сигналов (событий) в вершинах ОД; $C = \begin{cases} e, & \text{если } R = \{X\}; \\ 0, & \text{если } R = \{Z, Y\}; \end{cases}$ где e характеризует пустую букву.

Пример. Пусть дан ОД, граф которого показан на рис. 3а. Найти регулярные события по выходам Y_1 и Y_2 . Уравнение (1) для данного случая будет иметь вид

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \\ Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \\ Y_1 \\ Y_2 \end{pmatrix}^T \begin{pmatrix} 0 & 0 & 0 & b_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & b_7 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & b_9 & 0 & 0 \\ 0 & 0 & 0 & 0 & b_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & b_3 & 0 & b_4 & b_8 & 0 & 0 \\ 0 & 0 & 0 & b_5 & 0 & 0 & 0 & b_6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & b_{10} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} e \\ e \\ e \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

По этому уравнению запишем следующую систему:

$$\begin{cases} X_1 = e, \\ X_2 = e, \\ X_3 = e, \\ Z_1 = b_1 X_1 + b_3 Z_2 + b_5 Z_3, \\ Z_2 = b_7 X_2 + b_2 Z_1, \\ Z_3 = b_4 Z_2, \\ Z_4 = b_9 X_3 + b_8 Z_2, \\ Y_1 = b_6 Z_3, \\ Y_2 = b_{10} Z_4. \end{cases}$$

Систему уравнений решаем методом последовательного исключения неизвестных. Получим

$$\begin{aligned} Z_2 &= b_7 X_2 + b_2 Z_1 = b_7 X_2 + b_2 (b_1 X_1 + b_3 Z_2 + b_5 Z_3) = b_7 X_2 + \\ &+ b_2 (b_1 X_1 + b_3 Z_2 + b_5 b_4 Z_2) \Rightarrow Z_2 = \frac{b_1 b_2 X_1 + b_7 X_2}{1 - b_2 b_3 - b_2 b_4 b_5}, \\ Y_1 &= b_6 Z_3 = b_6 b_4 Z_2 = \frac{b_1 b_2 b_4 b_6}{1 - b_2 b_3 - b_2 b_4 b_5} X_1 + \frac{b_4 b_6 b_7}{1 - b_2 b_3 - b_2 b_4 b_5} X_2, \\ Y_2 &= b_{10} Z_4 = b_{10} (b_9 X_3 + b_8 Z_2) = \\ &= b_{10} \left(b_9 X_3 + b_8 \frac{b_1 b_2 X_1 + b_7 X_2}{1 - b_2 b_3 - b_2 b_4 b_5} \right) = \\ &= \frac{b_1 b_2 b_8 b_{10}}{1 - b_2 b_3 - b_2 b_4 b_5} X_1 + \frac{b_7 b_8 b_{10}}{1 - b_2 b_3 - b_2 b_4 b_5} X_2 + b_9 b_{10} X_3, \end{aligned}$$

которые являются регулярными событиями на выходе ОД.

Нетрудно убедиться, что полученные регулярные события устанавливают алфавит и вес параметров ОД и характеризуют взаимосвязь входных и выходных характеристик ОД. С учетом этого регулярные события ОД можно представить в виде анали-

тических формул, которые являются функциональными моделями ОД. То есть рассматриваемое диагностируемое устройство может являться функциональным элементом (составной частью) для более крупного ОД.

4. Заключение

Рассмотренные графический и аналитический методы нахождения информационных моделей объектов диагностики дают возможность устанавливать влияние входов ОД на его выходы даже в том случае, когда исходные модели содержат обратные связи. Это позволяет использовать получаемые модели ОД в качестве функциональных элементов при построении более сложных устройств.

Литература

1. ГЛУШКОВ В.М. *Введение в кибернетику*. – Киев: Изд-во АН УССР, 1964. – 324 с.
2. ДУБОВ А.В., КАПРАНОВ А.П., СУСКИН В.В., ШЕВЧЕНКО В.Ф. *Об одном варианте решения технического диагностирования радиоэлектронных средств* / Управление большими системами. Выпуск 31. – М.: ИПУ РАН, 2010. – С. 363–377.
3. КУРОШ А.Г. *Лекции по общей алгебре*. – М.: Физматгиз, 1962. – 396 с.
4. СЕРДАКОВ А.С. *Автоматический контроль и техническая диагностика*. – «Техніка», 1971. – 244 с.

BUILDING INFORMATION MODEL OF DIAGNOSTICS OBJECT DEFINED BY SIGNAL GRAPHS OF FUNCTIONAL ELEMENTS

Anton Dubov, post-graduate student, Ryazan State Radio-engineering University, Ryazan (dubovanton@mail.ru).

Abstract: In this article we describe methods to build an information model for an object of diagnostics with feedback loops. An object of diagnostics is defined by signal graphs of its functional elements. The suggested methods can be used as components of more complicated devices.

Keywords: signal graph, functional element, regular expression.

*Статья представлена к публикации
членом редакционной коллегии М.Ф. Караваем*

***3-я Российская конференция с международным участием
«Технические и программные средства систем управления,
контроля и измерения» (УКИ-12)***

Приглашаем принять участие в конференции. Она состоится с 16 по 19 апреля 2012 года в Институте проблем управления имени В. А. Трапезникова РАН (г. Москва, Россия).

28 ноября 2011 – завершение приема заявок на участие и кратких текстов докладов

20 января 2012 – результаты рассмотрения докладов; последний срок подачи пленарных докладов.

10 марта 2012 – окончание приема финальных версий принятых докладов и внесения оргвзноса.

16-19 апреля 2012 – проведение конференции.

Почтовый адрес конференции: cmm-conf@mail.ru

Официальный сайт конференции: <http://cmm.ipu.ru>