

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

УПРАВЛЕНИЕ БОЛЬШИМИ СИСТЕМАМИ

Выпуск 45
Сентябрь 2013

**СБОРНИК
ТРУДОВ**

ISSN 1819-2467

Регистрационный номер Эл. №ФС77-44158 от 09 марта 2011 г.

Москва – 2013

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

**УПРАВЛЕНИЕ
БОЛЬШИМИ
СИСТЕМАМИ**

СБОРНИК ТРУДОВ

Выпуск 45

Москва – 2013

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. Алескеров 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), канд. техн. наук Куливец С.Г. (ИПУ 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. Гераскин 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): 0421200023.

© ИПУ РАН, 2013

СОДЕРЖАНИЕ

Системный анализ

Жилиякова Л. Ю.

Эргодические циклические ресурсные сети. II. Большие ресурсы..... 5

Щербатов И. А., Проталинский О. М.

Сложные слабоформализуемые многокомпонентные технические системы..... 30

Математическая теория управления

Кожевников А. С., Рыбаков К. А.

Спектральный метод анализа стохастических систем с разрывами траекторий, описываемыми случайной смесью эрланговских распределений..... 47

Шевляков А. А.

Квазиканонический вид уравнений, описывающих движение вертолета..... 72

Информационные технологии в управлении

Асратян Р. Э., Лебедев В. Н.

Защита электронных сообщений в мульти-сетевой среде..... 95

Душкин Д. Н.

Автоматизированное определение классов чувствительности веб-сервисов..... 112

Пырлина И. В.

Риски и выбор оптимальных проектов сервис-ориентированная архитектура информационных систем..... 132

Смирнов В. А., Гусев М. Н., Фархадов М. П.

Функция модуля акустического моделирования в системе автоматического анализа неструктурированной речевой информации..... 181

Управление в социально-экономических системах

- Гафарова Е. А.**
Имитационные модели комплексного регионального развития..... 206
- Губанов Д. А., Чхартишвили А. Г.**
Концептуальный подход к анализу онлайн-социальных сетей..... 222
- Ратнер С. В.**
Управление технологическим портфелем энергетической компании: сценарный подход..... 237
- Стежкин А. А.**
Управление устойчивостью банковской сети с учетом отраслевых рисков..... 264
- Трахтенгерц Э. А.**
Использование двух сетецентрических принципов модификации экономических целей и стратегий в кризисной ситуации..... 264

Управление в медико-биологических и экологических системах

- Мирзоян Г. Л.**
Модель поведения лечебно-профилактического учреждения на территориальном рынке медицинских услуг..... 330

Технические и программные средства управления

- Каравай М. Ф., Подлазов В. С.**
Расширенный обобщенный гиперкуб как отказоустойчивая системная сеть для многопроцессорных систем..... 344

УДК 519.177

ББК 22.18

ЭРГОДИЧЕСКИЕ ЦИКЛИЧЕСКИЕ РЕСУРСНЫЕ СЕТИ. II. БОЛЬШИЕ РЕСУРСЫ

Жилякова Л. Ю.¹

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

В работе исследуется функционирование эргодических нерегулярных ресурсных сетей при больших ресурсах. Найдена формула для порогового значения ресурса T . Показано, что при больших ресурсах циклические сети всегда устойчивы, колебаний в них не происходит. Найдена формула компонент вектора предельного состояния, соответствующих всем неаттрактивным вершинам. Сформулировано условие аттрактивности вершин в циклических сетях.

Ключевые слова: ресурсная сеть, стохастическая матрица, эргодическая цепь, аттракторы, предельное состояние.

1. Введение

Ресурсная сеть – динамическая пороговая потоковая модель, функционирующая в дискретном времени. Ресурс в ней – неотрицательная величина, хранящаяся в вершинах, и распределяемая ими по исходящим ребрам на каждом такте. Каждое ребро проводит ресурс, пропорциональный своей пропускной способности (и не превосходящий ее). Если на каждом такте все вершины сети отдают весь свой ресурс, это означает, что суммарный ресурс сети меньше порогового значения. Если хотя бы одна вершина отдает по полной пропускной способности в каждое исходящее ребро и при этом имеет отличный от нуля

¹ Людмила Юрьевна Жилякова, кандидат физико-математических наук (zhilyakova.ludmila@gmail.com)

остаток на сколь угодно больших тактах t , ресурс в сети превосходит пороговое значение. При ресурсах, меньших порога, сеть сходна с моделями рассеяния на графах [10, 11]; при ресурсах, больших порога, некоторые вершины изменяют правило функционирования, – для них имеется сходство с целочисленными пороговыми моделями, описанными в [8, 9, 12].

Настоящая работа исследует функционирование эргодических циклических сетей при больших ресурсах. Она является продолжением статьи «Эргодические циклические ресурсные сети. I. Колебания и равновесные состояния при малых ресурсах» [4] При малых ресурсах эргодическая нерегулярная сеть описывается циклической цепью Маркова [5, 7]. При произвольном начальном состоянии в сети возникают периодические колебания между d предельными векторами, названные *предельным циклом*, где d – количество циклических классов в сети. В том случае, когда все d векторов совпадают, в сети при малых ресурсах возникает равновесие.

В указанной статье были получены следующие результаты.

1. Найдены пределы d сходящихся последовательностей степеней стохастической матрицы сети. Эти предельные матрицы состоят из циклических перестановок d предельных векторов при единичном суммарном ресурсе, находящемся в начальном состоянии целиком в одном из циклических классов.

2. Доказано, что все предельные векторы являются собственными векторами одной и той же предельной матрицы: предела степеней стохастической матрицы, кратных d .

3. Получено необходимое и достаточное условие на начальное состояние, при котором все предельные векторы равны и в сети достигается равновесие.

Настоящая работа посвящена функционированию циклических сетей при ресурсе, большем порогового значения. Как будет показано, при большом суммарном ресурсе свойство цикличности перестает оказывать влияние на процессы перераспределения ресурса, происходящие в сети, и ее поведение становится сходным с поведением регулярных сетей. Однако в отличие от регулярных сетей, в некоторых циклических сетях

становится возможным по свойствам пропускных способностей определить, какие из вершин являются потенциальными аттракторами.

2. Основные определения

2.1. РЕСУРСНЫЕ СЕТИ

Ресурсная сеть (в дальнейшем просто «сеть») представляет собой ориентированный граф $G = (V, E)$, $|V| = n$, с матрицей пропускных способностей $R = (r_{ij})_{n \times n}$.

$Q(t) = (q_1(t), \dots, q_n(t))$ – состояние сети в момент t ; $q_i(t) \geq 0$ – количество ресурса в вершине v_i .

Состояние сети Q^* называется *предельным*, если оно асимптотически достижимо из некоторого начального состояния $Q(0)$.

r_i^{in} и r_i^{out} – входная и выходная пропускные способности вершины v_i соответственно.

Правила функционирования сети. В момент t вершина v_i отдает в смежную ей вершину v_m :

r_{im} единиц ресурса, если $q_i(t) > r_i^{out}$ (правило 1);

$\frac{r_{im}}{r_i^{out}} q_i(t)$ единиц ресурса, если $q_i(t) \leq r_i^{out}$ (правило 2).

Множество вершин, для которых $q_i(t) \leq r_i^{out}$, называется зоной $Z^-(t)$. Вершины из $Z^-(t)$ функционируют по правилу 2. $Z^+(t)$ – множество вершин, ресурс которых больше их выходной пропускной способности, они функционируют по правилу 1. (Эти и другие определения описаны более подробно в ряде работ. Среди них, например, [2, 3, 6]).

T – *пороговое значение ресурса*, такое, что при $W \leq T$ все вершины, начиная с некоторого t' , переходят в зону $Z^-(t)$; при $W > T$ зона $Z^+(t)$ не пуста, начиная с некоторого t'' .

Для предельного состояния Q^* обозначим эти зоны через Z^* и Z^{+*} .

Введем обозначение: $\Delta r_i = r_i^{in} - r_i^{out}$. По знаку Δr_i вершины делятся на три класса:

- 1) *вершины-приемники*, для которых $\Delta r_i > 0$;
- 2) *вершины-источники*, для которых $\Delta r_i < 0$;
- 3) *нейтральные вершины*, для которых $\Delta r_i = 0$.

Вершины, способные при $W > T$ из некоторого начального состояния перейти в Z^{+*} , называются *потенциальными аттракторами*. Потенциальными аттракторами могут быть некоторые приемники и некоторые нейтральные вершины, причем первые способны притягивать ресурс и поэтому названы *активными аттракторами*, вторые могут лишь сохранить ресурс, которым обладали в начальном состоянии. Такие аттракторы называются *пассивными*. Потенциальные аттракторы в регулярных сетях и предельные состояния при больших ресурсах были исследованы в [2, 3].

В d -циклической сети существует d подпоследовательностей векторов состояний, сходящихся к предельным векторам Q_1^*, \dots, Q_d^* , последовательно сменяющим друг друга при $t \rightarrow \infty$. Если при некотором начальном состоянии они равны, в сети достигается *равновесное состояние*. Если векторы Q_1^*, \dots, Q_d^* равны при любом начальном состоянии, будем говорить, что в сети существует предельное состояние Q^* :

$$Q_1^* = \dots = Q_d^* = Q^*.$$

Поток в ресурсной сети.

Ресурс, выходящий из вершины v_i по ребру (v_i, v_j) в момент t , приходит в вершину v_j в момент $t + 1$; между моментами t и $t + 1$ он находится в ребре (v_i, v_j) . Этот ресурс назовем потоком $f_{ij}(t)$. Общий поток сети описывается матрицей $F(t) = (f_{ij}(t))_{n \times n}$.

Величиной потока будем называть сумму:

$$f_{sum} = \sum_{i=1}^n \sum_{j=1}^n f_{ij}(t).$$

Введем следующие обозначения.

$\sum_{j=1}^n f_{ij}(t) = f_i^{out}(t)$; $\sum_{i=1}^n f_{ij}(t) = f_j^{in}(t+1)$. Кроме того, положим

по определению $f_j^{in}(0) = 0$.

Матрицу предельного потока, если он существует, обозначим через F^* . Суммарный предельный поток тогда будет

$$f_{sum}^* = \sum_{i=1}^n \sum_{j=1}^n f_{ij}^* .$$

$F^{out}(t) = (f_1^{out}(t), \dots, f_n^{out}(t))$ – вектор исходящего потока;

$F^{in}(t) = (f_1^{in}(t), \dots, f_n^{in}(t))$ – вектор входящего потока.

Пределы $\lim_{t \rightarrow \infty} F^{out}(t)$ и $\lim_{t \rightarrow \infty} F^{in}(t)$, если они существуют, обозначим через F^{in*} и F^{out*} соответственно.

2.2. ЦИКЛИЧЕСКИЕ СЕТИ И ИХ ТОПОЛОГИЯ

Эргодическая сеть – сеть, граф которой сильно связан. Это определение соответствует определению эргодической цепи Маркова, данному в [4]. Эргодическую сеть, наибольший общий делитель всех циклов которой больше единицы, будем называть *циклической*. Число d , равное НОД длин всех циклов, – период сети. Множество ее вершин делится на d циклических классов. Стохастическая матрица R' такой сети имеет ровно d собственных значений, равных по модулю единице: $\lambda_1 = |\lambda_2| = \dots = |\lambda_d| = 1$, т.е. является импримитивной с индексом импримитивности d .

По топологии циклические сети могут сильно различаться. Самым простым и наиболее легким для исследования является элементарный цикл из n вершин. Он содержит n циклических классов и при малых ресурсах на каждом такте ресурс переходит из некоторой вершины в следующую. Пороговое значение для таких сетей и потенциальные аттракторы могут быть найдены лишь исходя из пропускных способностей ребер.

Более сложный случай – *простые циклы*. Простой цикл состоит из нескольких циклов, пересекающихся по вершинам, но не пересекающихся по ребрам. Все эти циклы обязаны не быть

взаимно простыми. Если НОД длин этих циклов равен единице, сеть состоит из одного циклического класса; она регулярна и для нее выполняются все результаты, полученные для регулярных сетей.

Общий случай циклической сети – сеть, состоящая из циклов, пересекающихся либо только по вершинам, либо по ребрам.

3. Функционирование циклических сетей с различными топологиями при $W > T$

3.1. ЭЛЕМЕНТАРНЫЕ ЦИКЛЫ

Рассмотрим на примерах функционирование сети, состоящей из однородного и неоднородного элементарного цикла.

Пример 1. Пусть сеть представлена однородным циклом с пятью вершинами и ее матрица пропускных способностей имеет вид:

$$(1) \quad R = \begin{pmatrix} 0 & 5 & 0 & 0 & 0 \\ 0 & 0 & 5 & 0 & 0 \\ 0 & 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 0 & 5 \\ 5 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad r_{sum} = 25.$$

Она однородна и квазисимметрична, т.е. является эйлеровой нерегулярной сетью. Каждая ее вершина – потенциальный аттрактор. Нетрудно убедиться, что насыщение в этой сети наступает, когда все вершины имеют по $r_i^{out} = 5$ единиц ресурса, и, таким образом, $T = r_{sum} = 25$.

При начальном состоянии $Q(0) = (30, 0, 0, 0, 0)$ имеем рис. 1.

Из рис. 1 видно, что за первые пять тактов все вершины получили ресурс, равный своей пропускной способности (вершина 1 оставила себе излишки $W - T$), и на этом изменения компонент вектора состояния завершились. Сеть не прекращает функционировать, но ее состояние остается неизменным.

Рассмотрим сеть, представляющую собой неоднородный цикл, с матрицей пропускных способностей

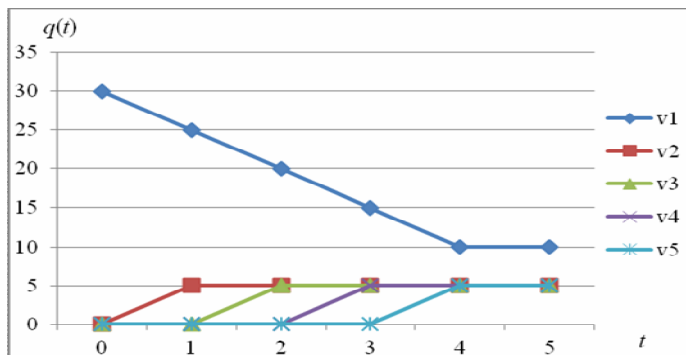


Рис. 1. Функционирование элементарного цикла при $W > T$

$$(2) R = \begin{pmatrix} 0 & 5 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \\ 5 & 0 & 0 & 0 & 0 \end{pmatrix}, r_{sum} = 21.$$

Пороговое значение ресурса T в такой сети достигается, когда поток в каждом ребре равен минимальному разрезу. Рассмотрим примеры, иллюстрирующие изменения поведения сети при увеличении ресурса от малых значений и переход через пороговое значение.

Пример 2. Пусть в сети с матрицей (2) начальное состояние $Q(0) = (7, 0, 0, 0, 0)$ (см. рис. 2).

Поскольку суммарный ресурс превосходит пропускные способности ребер, минимальная из которых $r_{33} = 3$, он делится на три неравные части, и дальше в сети наблюдается три цикла длины 5: с ресурсом 3, 3 и 1.

Таким образом, если элементарный цикл неоднороден и ресурс в начальном состоянии аккумулирован в одной из вершин,

он перемещается между вершинами порциями, – существует несколько последовательных циклов движения ресурса. В каждом из них (за исключением, возможно, одного) ресурс равен минимальной пропускной способности цикла. В одной (заключительной) порции перемещаются «остатки»: ресурс, равный

$$W - \left\lfloor \frac{W}{\min_{i,j} r_{ij}} \right\rfloor \min_{i,j} r_{ij}, \text{ где } \lfloor \cdot \rfloor - \text{целая часть числа.}$$

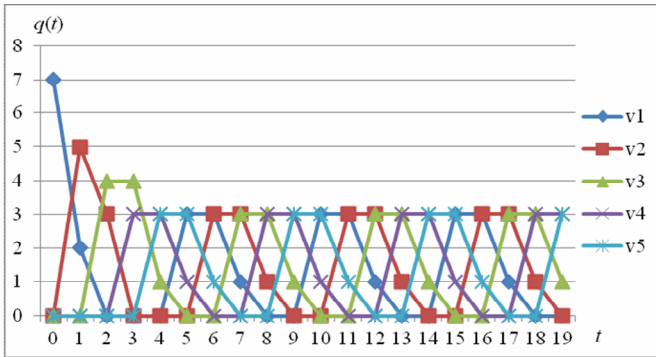


Рис. 2. Поток, не превосходящий минимальный разрез

Таблица 1. Протокол работы сети

t	v_1	v_2	v_3	v_4	v_5
0	7,000	0,000	0,000	0,000	0,000
1	2,000	5,000	0,000	0,000	0,000
2	0,000	3,000	4,000	0,000	0,000
3	0,000	0,000	4,000	3,000	0,000
4	0,000	0,000	1,000	3,000	3,000
5	3,000	0,000	0,000	1,000	3,000
6	3,000	3,000	0,000	0,000	1,000
7	1,000	3,000	3,000	0,000	0,000
8	0,000	1,000	3,000	3,000	0,000
9	0,000	0,000	1,000	3,000	3,000
10	3,000	0,000	0,000	1,000	3,000
...

Отсюда следует, что для любого неоднородного элементарного цикла существует значение ресурса, при котором поток стабилизируется за конечное число тактов и сеть приходит в равновесное состояние из любого начального. Поток в каждом ребре будет равен минимальному разрезу, который в данном случае равен минимальной пропускной способности ребер цикла. Суммарный равновесный поток равен пороговому значению ресурса T ; они вычисляются по формуле:

$$\tilde{F}^{in} = \tilde{F}^{out} = (\min r_{ij}, \dots, \min r_{ij}); T = n \cdot \min r_{ij},$$

где n – число вершин; $\min r_{ij}$ – минимальная пропускная способность цикла.

При $W > T$ предельный поток существует и равен T для любого значения суммарного ресурса, а излишки сверх T скапливаются лишь в вершинах v_k , для которых $k = \arg \min_i r_{ij}$.

Так, для сети, заданной матрицей (2), $T = 15$.

Пример 3. Пусть в сети с матрицей (2) начальное состояние $Q(0) = (15, 0, 0, 0, 0)$. В таком случае в ней за шесть тактов происходит выравнивание ресурса и стабилизация состояний и потока (рис. 3).

В неоднородном цикле при $W > T$, как уже было сказано, ресурс скапливается в вершинах с минимальной выходной пропускной способностью v_k . Такие вершины будут *потенциальными аттракторами цикла*. Если аттрактор один, предельное состояние единственно. Ресурс $W - T$ в предельном состоянии окажется в аттракторе (рис. 4). Отсюда следует, что в сети с матрицей (2) потенциальным аттрактором является вершина v_3 .

Излишек ресурса $W - T$ перешел в третью вершину. Предельное состояние описывается вектором $Q^* = (3, 3, 8, 3, 3)$.

Интересно, что в последних двух примерах вершина v_1 , обладая суммарным ресурсом сети в начальном состоянии, сначала отдает его весь и только затем получает часть обратно.

Если вершин с минимальной выходной пропускной способностью несколько, предельное состояние зависит от начального состояния и от расположения этих вершин в цикле. Если они идут друг за другом, активным аттрактором будет только

первая из них, остальные смогут лишь удержать излишек ресурса, находящийся в них в начальном состоянии, но новые излишки будут задерживаться только в первой вершине. Если между такими вершинами есть другие вершины, каждая из них будет активным потенциальным аттрактором, и количество ресурса в них в предельном состоянии будет зависеть от начального состояния.

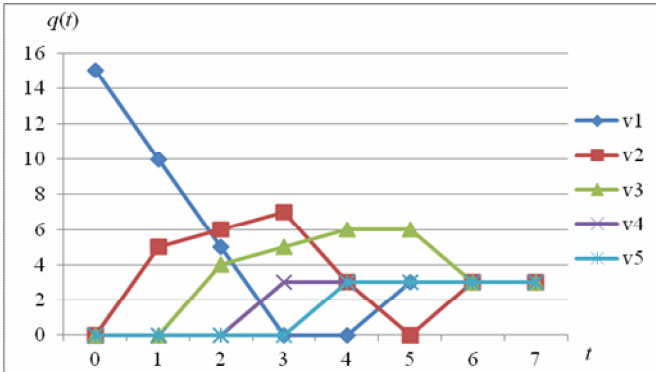


Рис. 3. Ресурс $W = T = 15$ выравнивается.
Вектор предельного состояния $Q^* = (3, 3, 3, 3, 3)$

Пример 4. Пусть в сети с матрицей (2) начальное состояние $Q(0) = (20, 0, 0, 0, 0)$.

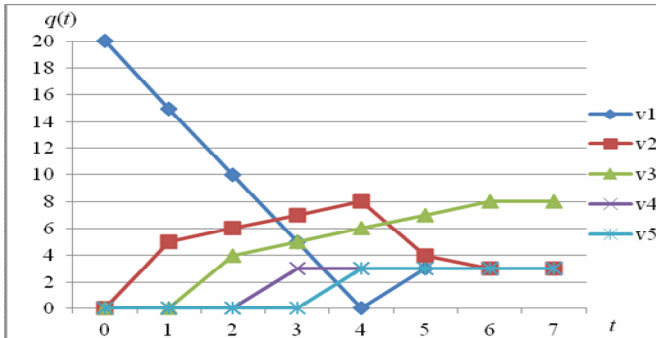


Рис. 4. Функционирование сети при $W > T$.

Пример 5. Рассмотрим неоднородный цикл из семи вершин с двумя потенциальными аттракторами v_2 и v_5 – они имеют минимальные выходные пропускные способности.

$$(3) \quad R = \begin{pmatrix} 0 & 5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 6 \\ 7 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}; \quad r_{sum} = 31; \quad T = 2 \cdot 7 = 14.$$

$W = 17 > T$. Начальное состояние $Q(0) = (17, 0, 0, 0, 0, 0, 0)$.

Из рис. 5 видно, что в зону Z^{+*} перешел тот аттрактор, который первым успел задержать ресурс у себя, т.е. вершина v_2 .

При начальном состоянии $Q(0) = (0, 0, 0, 17, 0, 0, 0)$ предельным состоянием будет $Q^* = (2, 2, 2, 2, 5, 2, 2)$, т.е. весь излишек останется во втором потенциальном аттракторе v_5 .

Если ресурс $W > T$ в начальном состоянии находится в нескольких вершинах, его излишек $W - T$ распределяется между аттракторами в пропорции, которая зависит от пропускных способностей всех ребер.

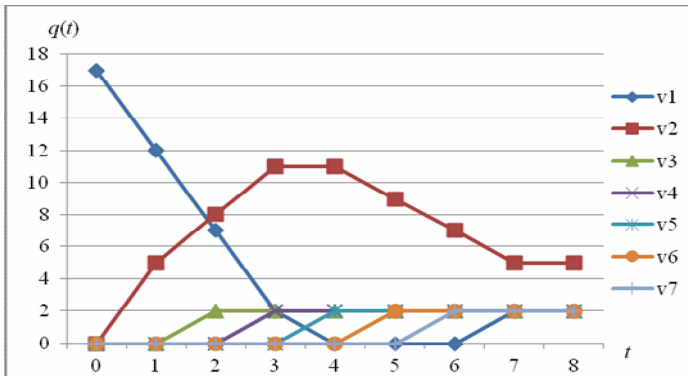


Рис. 5. Цикл с двумя аттракторами. Вектор предельного состояния $Q^* = (2, 5, 2, 2, 2, 2, 2)$

Так, например, для матрицы (3) при начальном состоянии $Q(0) = (8, 0, 0, 9, 0, 0, 0)$ предельное состояние будет $Q^* = (2, 2, 2, 2, 5, 2, 2)$, т.е. все излишки вновь окажутся в аттракторе v_5 . Чтобы вершина v_2 тоже получила излишек, ресурс в вершине v_1 должен быть больше 8. При начальном состоянии $Q(0) = (9, 0, 0, 8, 0, 0, 0)$ предельное состояние $Q^* = (2, 3, 2, 2, 4, 2, 2)$. Однако при перемещении ресурса, равного 8, в начальном состоянии из вершины v_4 в v_3 , т.е. при $Q(0) = (9, 0, 8, 0, 0, 0, 0)$, предельным состоянием вновь будет вектор $Q^* = (2, 2, 2, 2, 5, 2, 2)$ – все излишки забирает аттрактор v_5 . Это происходит оттого, что ресурс должен проделать более длинный путь по циклу до возвращения в его начало: вершину v_1 , служащую источником для аттрактора v_2 . При $Q(0) = (9, 0, 8, 0, 0, 0, 0)$ вершина v_1 пуста три такта (со второго по четвертый), и этого достаточно, чтобы ресурс из аттрактора v_2 переместился в аттрактор v_5 (рис. 6). Из рисунка видно, что к тому моменту (такт 5), как ресурс вершины v_1 , становится вновь ненулевым и она могла бы поддерживать баланс в вершине v_2 , вершина v_2 уже имеет ресурс, равный единице, т.е. не имеет излишков.

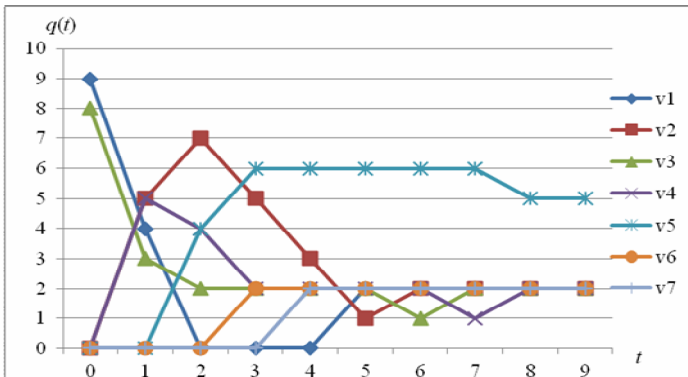


Рис. 6. Цикл с двумя аттракторами. Вектор предельного состояния $Q^* = (2, 2, 2, 2, 5, 2, 2)$

Обобщим все результаты, полученные при рассмотрении элементарных циклов при больших ресурсах.

Утверждение 1. Для циклической ресурсной сети с n вершинами, представленной элементарным циклом:

1. Пороговое значение ресурса T вычисляется по формуле: $T = n \cdot r_{min}$, где $r_{min} = \min r_{ij}$.

2. При $W = T$ предельный поток и предельное состояние существуют и единственны. Вектор предельного потока равен вектору предельного состояния: $F^{in*} = F^{out*} = Q^* = (r_{min}, \dots, r_{min})$.

3. При $W > T$ предельное состояние и поток существуют, причем предельный поток определяется единственным образом: $F^{in*} = F^{out*} = (r_{min}, \dots, r_{min})$. Предельное состояние единственно в том и только в том случае, когда сеть имеет единственный потенциальный аттрактор. В противном случае распределение ресурса сверх порогового значения $W - T$ зависит от начального состояния.

4. Вершина v_k является потенциальным аттрактором элементарного цикла в том и только в том случае, если для нее выполняется $k = \operatorname{argmin}_i r_{ij}$. Если два и более потенциальных аттрактора расположены в цикле непосредственно друг за другом, то активным аттрактором является только тот, для которого $r_i^{in} > r_i^{out}$; остальные аттракторы пассивны.

3.2. ПРОСТЫЕ ЦИКЛЫ И ЦИКЛЫ С ОБЩИМИ РЕБРАМИ

Рассмотрим функционирование неоднородной сети, состоящей из двух циклических классов. Циклы в ней пересекаются только по вершинам.

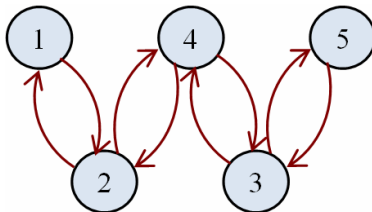


Рис. 7. Двудольный граф с пятью вершинами, $d = 2$

Пример 6. Пусть сеть задана матрицей (3)

$$(3) \quad R = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 2 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 & 3 \\ 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \end{pmatrix}, \quad r_{sum} = 15.$$

При заведомо большом значении ресурса (большем r_{sum}) $W = 20$ любое начальное состояние приводит к одному и тому же предельному состоянию $Q^* = (14,5; 2; 1,25; 1,5; 0,75)$.

Отсюда видно, что потенциальный аттрактор этой сети единственен, им является вершина v_1 . Пороговое значение ресурса для этой сети находится из того условия, что при $W \geq T$ сумма всех компонент вектора предельного состояния, кроме первой, постоянна и равна 5,5, а первая компонента при $W = T$ равна суммарной выходной пропускной способности вершины, т.е. единице: $T = 6,5$. При $W = T$ предельное состояние описывается вектором $\tilde{Q} = (1; 2; 1,25; 1,5; 0,75)$ для любого начального состояния. Этим же вектором описывается поток при любом $W > T$.

При суммарном ресурсе, меньшем 6,5, в сети существует два предельных вектора.

Векторы Q_1^{1*} и Q_2^{1*} , формирующие предельные матрицы $R_1^{1\infty}$ и $R_2^{1\infty}$, имеют вид:

$$Q_1^{1*} = (0, 0,616, 0,384, 0, 0), \quad Q_2^{1*} = (0,308, 0, 0, 0,462, 0,230).$$

Собственный вектор стохастической матрицы R' ищется по формуле

$$(4) \quad Q^{1*} = \frac{Q_1^{1*} + Q_2^{1*}}{2}.$$

$$Q^{1*} = (0,154, 0,313, 0,192, 0,231, 0,115).$$

Проверим, выполняется ли для вершины v_1 признак аттрактивности, введенный для вершин регулярных сетей.

Для аттрактора с номером j_k должно выполняться условие:

$$j_k = \arg \min_i \frac{r_i^{out}}{q_i^{1*}}.$$

Вычислим значение $\frac{r_i^{out}}{q_i^{1*}}$ для всех вершин сети. В качестве

вектора предельного состояния возьмем вектор (4).

$$\frac{r_1^{out}}{q_1^{1*}} \approx \frac{1}{0,154} \approx 6,5, \quad \frac{r_2^{out}}{q_2^{1*}} \approx \frac{4}{0,313} \approx 12,8, \quad \frac{r_3^{out}}{q_3^{1*}} \approx \frac{5}{0,192} \approx 26,$$

$$\frac{r_4^{out}}{q_4^{1*}} \approx \frac{3}{0,231} \approx 13, \quad \frac{r_5^{out}}{q_5^{1*}} \approx \frac{2}{0,115} \approx 17,4.$$

Как видно, минимум этого отношения действительно достигается на первой вершине, более того, его значение равно 6,5, что совпадает со значением T , как это и было доказано для регулярных сетей [3].

Из примера следует, что для циклической сети при определении порогового значения ресурса T и при ее функционировании с большим суммарным ресурсом наблюдаются те же закономерности, что и для регулярных сетей. Докажем, что такое совпадение имеет место для любой эргодической циклической сети.

4. Пороговое значение T , предельные состояния и потоки в циклических сетях при больших ресурсах

В регулярных сетях существование и единственность порогового значения T вытекали из единственности вектора предельного состояния при малых ресурсах. Все компоненты вектора Q^* при увеличении W вплоть до значения T увеличивались пропорционально, пока одна или несколько из них не принимали значения, равные своим суммарным выходным пропускным способностям. Это значение суммарного ресурса и было пороговым. Вершины с ресурсом, достигшим выходных пропускных способностей, – потенциальные аттракторы сети. При дальней-

шем увеличении ресурса пропорциональность компонент нарушается, и все излишки сверх T накапливаются в аттракторах.

В циклической сети существует d предельных векторов, в общем случае не равных друг другу. Кроме того, каждый из этих векторов зависит от начального состояния. Единственность порогового значения в такой сети уже неочевидна.

Будем говорить, что если в циклической сети d предельных векторов равны при любом начальном состоянии, в ней существует предельное состояние Q^* : $Q_1^* = \dots = Q_d^* = Q^*$.

Т е о р е м а 1. В эргодической d -циклической ресурсной сети существует единственное пороговое значение T , такое, что:

при $W < T$ все вершины за конечное число тактов переходят на правило 2, и в сети имеется d предельных векторов;

при $W \geq T$ потоки стабилизируются, и в сети достигается глобальное равновесие при любом начальном состоянии. Предельный поток существует и единственен. Предельное состояние существует; оно не зависит от начального состояния в том и только в том случае, когда сеть имеет один аттрактор.

Доказательство. В сети, все вершины которой функционируют по правилу 2, существует предельный цикл из d предельных векторов. Если в начальном состоянии некоторые вершины функционируют по правилу 1, но все они за конечное число тактов N переходят на правило 2, то, взяв $Q(N+1)$ за новое начальное состояние, вновь получим предельный цикл из d векторов. Верхняя граница суммарного ресурса, при котором все вершины функционируют по правилу 2, равна T . T существует для любой сети.

Докажем, что значение T единственно и не зависит от начального состояния. При $W = T$ по крайней мере одна вершина имеет ресурс, равный своей выходной пропускной способности. Пусть это вершина v_j . Для нее $q_j^* = r_j^{out}$ в каждом из d предельных векторов.

Матрица R^l импримитивна с индексом импримитивности d . Это означает, что ее степень R^{nd} разлагается в прямую сумму d регулярных матриц, соответствующих циклическим классам $R_{ii}'^d$ [1]. Предельные векторы сети выражаются через собственные векторы этих матриц. Вершина v_j является потенциальным аттрактором в некотором циклическом классе с номером i . Если с первого такта все вершины функционировали по правилу 2, для него выполняется: $Q_i(0)R_{ii}'^{\infty} = Q_{di}^* = W_i\pi_i$, где W_i – суммарный ресурс в i -м циклическом классе; π_i – вектор предельных вероятностей соответствующей цепи Маркова. Для сети, описываемой регулярной матрицей $R_{ii}'^d$, пороговое значение существует и единственно. Обозначим его через T_i . Тогда при $W_i = T_i$ имеем $(q_{di}^*)_j = r_j^{out}$, где $(q_{di}^*)_j$ – j -я компонента вектора Q_{di}^* . Но чтобы ресурс вершины v_j был равен r_j^{out} на каждом такте, суммарный ресурс i -го циклического класса должен быть равен T_i также на каждом такте. Если сеть функционирует по правилу 2, ресурс циркулирует по циклическим классам, не перемешиваясь. Если раскрасить ресурс в каждом классе в свой цвет, ресурсы разных цветов на любом такте будут целиком в одном классе. Пороговое значение T определяется как максимальный ресурс, при котором все вершины функционируют по правилу 2 при $t \rightarrow \infty$. Тогда при $t \rightarrow \infty$ каждый циклический класс на каждом такте должен содержать T_i ресурса. А пороговое значение ресурса сети вычисляется по формуле $T = dT_i$.

Поскольку каждая компонента имеет одинаковое количество ресурса, поток в сети стабилизируется, и все предельные векторы одинаковы.

При $W > T$ как минимум один аттрактор (пусть это вновь будет вершина v_j) за конечное число тактов переходит на правило 1. То есть существует такой момент времени t' , что при $t \geq t'$ вершина v_j функционирует по правилу 1. На каждом такте она отдает ресурс, равный r_j^{out} , и получает ресурс, не меньший r_j^{out} . Тогда в нее можно добавить петлю с пропускной способностью

$q_j(t) - r_j^{out}$, по которой на каждом такте будет проходить неотрицательный поток, равный $q_j(t) - r_j^{out}$. Функционирование сети при этом не изменится. Но наличие петли влечет за собой потерю цикличности, и сеть превращается в регулярную. В регулярной сети поток существует и единственен. Предельное состояние единственно во всех вершинах, кроме потенциальных аттракторов. Если в сети единственный аттрактор, предельное состояние также единственно. \square

Докажем, что для вершин в произвольной d -циклической сети выполняется тот же критерий аттрактивности, что и для вершин в несимметричной регулярной сети.

Теорема 2. В эргодической d -циклической ресурсной сети вершина v_j является потенциальным аттрактором, если и только если $j = \arg \min_i \frac{r_i^{out}}{q_i^{1}}$, где вектор Q^{1*} определяется по формуле*

$$(5) \quad Q^{1*} = \frac{1}{d} \sum_{k=1}^d Q_k^{1*},$$

векторы Q_k^{1*} ($k = 0, \dots, d$) – предельные векторы при $W = 1$ и произвольном начальном состоянии.

Доказательство. Из теоремы 1 следует, что при $W = T$ предельное состояние сети при любом начальном состоянии единственно, и все ее d предельных векторов равны. Но для $W \leq T$ выполняется равенство: $Q_1^* + \dots + Q_d^* = dTQ^{1*}$, где вектор Q^{1*} , вычисляемый по формуле (5), – единственный положительный собственный вектор стохастической матрицы R' . Непосредственно отсюда следует, что предельное состояние сети $Q^* = Q_1^* = \dots = Q_d^* = TQ^{1*}$. Таким образом, функционирование сети начинается описываться регулярной матрицей A , состоящей из n строк Q^{1*} :

$$A = \mathbf{1} \cdot Q^{1*}.$$

Для сети, функционирование которой описывается матрицей A , справедливы все результаты, полученные для регулярных сетей. Таким образом, для того, чтобы вершина v_j была аттрак-

тором, необходимо и достаточно выполнение условия $j = \arg \min_i \frac{r_i^{out}}{q_i^{1*}}$ [3]. □

С л е д с т в и е 1. В эйлеровых d -циклических сетях при больших ресурсах предельное состояние полностью зависит от начального, так как каждая вершина в них – потенциальный аттрактор.

С л е д с т в и е 2. В эргодической d -циклической ресурсной сети пороговое значение T определяется по формуле

$$T = \min_i \frac{r_i^{out}}{q_i^{1*}}.$$

Таким образом, при $W > T$ для циклических сетей можно сформулировать теоремы о предельном состоянии и потоке, аналогичные теоремам о регулярных сетях. Их доказательства следуют из теорем 1–2 и следствия 2.

Т е о р е м а 3 (о предельном состоянии). В эргодической d -циклической ресурсной сети при $W \geq T$ значения неаттрактивных компонент вектора предельного состояния $Q^* = (q_1^*, \dots, q_n^*)$ вычисляются по формуле

$$q_i^* = q_i^{1*} \cdot T, \quad i \neq j_k, j_k - \text{номера аттракторов,}$$

где:

1) вектор Q^{1*} определяется по формуле (5),

$$2) T = \min_i \frac{r_i^{out}}{q_i^{1*}},$$

$$3) j_k \text{ определяются из условия: } j_k = \arg \min_i \frac{r_i^{out}}{q_i^{1*}}.$$

Оставшийся ресурс распределяется между потенциальными аттракторами.

Проиллюстрируем теорему следующим примером.

Пример 7. Рассмотрим функционирование неоднородной сети, содержащей циклы с общими ребрами.

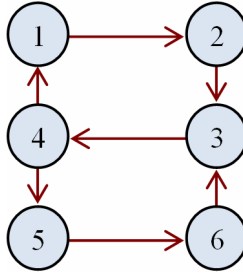


Рис. 8. Циклы с общими ребрами

В этой сети четыре циклических множества: $\{v_1, v_5\}$, $\{v_2, v_6\}$, $\{v_3\}$ и $\{v_4\}$.

Пусть матрица пропускных способностей сети имеет вид

$$R = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 \\ 2 & 0 & 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}, \quad r_{sum} = 18.$$

Тогда при помещении в начальном состоянии ресурса $W = 1$ в любую из шести вершин предельные векторы будут иметь вид

$$Q_1^{1*} = (0,286; 0; 0; 0; 0,714; 0),$$

$$Q_2^{1*} = (0; 0,286; 0; 0; 0; 0,714),$$

$$Q_3^{1*} = (0, 0, 1, 0, 0, 0), \quad Q_4^{1*} = (0, 0, 0, 1, 0, 0).$$

Собственный вектор стохастической матрицы находится как их среднее арифметическое (формула (5)):

$$Q^{1*} = (0,0715; 0,0715; 0,25; 0,25; 0,1785; 0,1785).$$

Вычислим значение отношения $\frac{r_i^{out}}{q_i^{1*}}$ для каждой вершины.

$$\frac{r_1^{out}}{q_1^{1*}} \approx \frac{1}{0,0715} \approx 14, \quad \frac{r_2^{out}}{q_2^{1*}} \approx \frac{3}{0,0715} \approx 42, \quad \frac{r_3^{out}}{q_3^{1*}} \approx \frac{2}{0,25} = 8,$$

$$\frac{r_4^{out}}{q_4^{1*}} \approx \frac{7}{0,25} = 28, \quad \frac{r_5^{out}}{q_5^{1*}} \approx \frac{4}{0,1785} \approx 22,4, \quad \frac{r_6^{out}}{q_6^{1*}} \approx \frac{1}{0,1785} \approx 5,6.$$

Минимум $\frac{r_i^{out}}{q_i^{1*}}$ достигается в вершине v_6 .

Таким образом, v_6 – единственный потенциальный аттрактор этой сети. По следствию 2 из теоремы 1 $T \approx 5,6$.

При $W = T$ предельное состояние при любом начальном:

$$\tilde{Q} = (0,4; 0,4; 1,4; 1,4; 1; 1).$$

Этому же вектору равны предельные входной и выходной потоки.

При $W > T$, например, $W = 20$, при любом начальном состоянии предельное состояние равно

$$Q^* = (0,4; 0,4; 1,4; 1,4; 1; 15,4).$$

Излишек ресурса аккумулировался в аттракторе. Остальные компоненты предельного вектора не изменились.

Теорема 4 (о предельном потоке). В эргодической d -циклической ресурсной сети при $W \geq T$ предельный поток существует и единственен и определяется по формуле:

$$f_i^{in*} = f_i^{out*} = q_i^{1*} T; \quad F^{in*} = F_i^{out*} = Q^{1*} T = \tilde{Q},$$

где

1) вектор Q^{1*} определяется по формуле (5),

$$2) T = \min_i \frac{r_i^{out}}{q_i^{1*}}.$$

З а м е ч а н и е. В регулярных сетях при разделении ресурса на «малый» и «большой», пороговое значение T , по которому проходит эта граница, относился к малым ресурсам. Это связано с тем, что при $W = T$ сеть описывается регулярной цепью Маркова, как и при меньших значениях суммарного ресурса, и предельное состояние в ней единственно. То есть в регулярных сетях суммарный ресурс разбивается на значения $W \leq T$ и $W > T$. В циклических сетях значение T входит в «большой» ресурс, так как при $W \geq T$ сеть имеет предельное состояние, в то время как при $W < T$ – лишь предельный цикл из d предельных векторов.

5. Заключение

В работе исследованы процессы перераспределения больших ресурсов в эргодических циклических сетях. Рассмотрены примеры функционирования сетей с большим ресурсом, имеющих различные топологии. Доказано, что циклические сети, как и регулярные, имеют единственное пороговое значение ресурса T , не зависящее от начального состояния, такое что при $W < T$ все вершины переходят на правило 2, в сети существует d предельных векторов, в общем случае различных. При $W \geq T$ предельные векторы равны для любого начального состояния. Предельный поток единственен, суммарный поток равен T . Найдены формулы для векторов предельного состояния в неаттрактивных вершинах и потока – во всех вершинах. Сформулирован и доказан критерий аттрактивности вершин.

Из полученных результатов следует, что при больших ресурсах функционирование сети становится регулярным и описывается регулярной матрицей A .

Литература

1. ГАНТМАХЕР Ф.Р. *Теория матриц*. – М.: Физматлит. 2004. – 560 с.
2. ЖИЛЯКОВА Л.Ю. *Несимметричные ресурсные сети. II. Потoki при больших ресурсах и их стабилизация* // Автоматика и телемеханика. – 2012. – №6. – С. 103–118.
3. ЖИЛЯКОВА Л.Ю. *Несимметричные ресурсные сети. III. Исследование предельных состояний* // Автоматика и телемеханика. – 2012. – №7. – С. 67–77.
4. ЖИЛЯКОВА Л.Ю. *Эргодические циклические ресурсные сети. I. Колебания и равновесные состояния при малых ресурсах* // Управление большими системами. – Вып. 43. – М.: ИПУ РАН, 2013. – С. 34 – 54.
5. КЕМЕНИ ДЖ., СНЕЛЛ ДЖ. *Конечные цепи Маркова*. – М.: Наука, 1970. – 271 с.

6. КУЗНЕЦОВ О.П. *Однородные ресурсные сети. I. Полные графы* // Автоматика и телемеханика. – 2009. – №11. – С. 136–147.
7. РОБЕРТС Ф.С. *Дискретные математические модели с приложениями к социальным, биологическим и экологическим задачам*. – М. Наука, 1986. – 496 с.
8. BJÖRNER A., LOVASZ L., SHOR P. *Chip-firing games on graphs* // Europ. J. Comb. – 1991. – №12. – P. 283–291.
9. BJÖRNER A., LOVASZ L. *Chip-firing game on directed graphs* // J. Algebraic Combinatorics. – 1992. – №1. – P. 305–328.
10. BLANCHARD PH., VOLCHENKOV D. *Random Walks and Diffusions on Graphs and Databases: An Introduction* (Springer Series in Synergetics). – Springer-Verlag – Berlin–Heidelberg, 2011. – 262 p.
11. LOVASZ L., WINKLER P. *Mixing of Random Walks and Other Diffusions on a Graph* // Surveys in Combinatorics / Ed. P. Rowlinson. – London Math. Soc. Lecture Notes Series 218. – Cambridge Univ. Press, 1995. – P. 119–154.
12. PRISNER E. *Parallel Chip Firing on Digraphs* // Complex Systems. – 1994. – №8. – P. 367–383.

ERGODIC CYCLIC RESOURCE NETWORKS. II. HIGH LEVELS OF RESOURCE

Ludmila Zhilyakova, Institute of Control Sciences of RAS, (Moscow, Profsoyuznaya st., 65), cand. sc., senior scientist, zhilyakova.ludmila@gmail.com.

Abstract: We study processes in ergodic nonregular resource networks at high levels of resource. The formula for the threshold value T is derived. It was shown that there are no oscillations if $W \geq T$; the network becomes stable from every initial state when resource level is high. The formulae for the vectors of limit state and the limit flow are derived. The criterion of attractivity of vertices is formulated.

Keywords: diffusion on graphs, resource network, attractors, limit state.

Статья представлена к публикации членом редакционной коллегии О.П. Кузнецовым

*Поступила в редакцию 06.02.2013.
Опубликована 30.09.2013.*

УДК 681.51
ББК 30.1

СЛОЖНЫЕ СЛАБОФОРМАЛИЗУЕМЫЕ МНОГОКОМПОНЕНТНЫЕ ТЕХНИЧЕСКИЕ СИСТЕМЫ

Щербатов И. А.¹, Проталинский О. М.²
(Астраханский государственный технический
университет, Астрахань)

В работе введено понятие сложной слабоформализуемой многокомпонентной технической системы. Показаны механизмы образования компонентных структур указанного класса систем, приведен способ описания компонент. Введено классификационное разделение компонент на активные и пассивные, сформулированы основные рациональные поведенческие признаки компонент. В качестве активной компоненты может выступать робот, а в качестве пассивной – слабоформализуемый технологический процесс.

Ключевые слова: компонента, сложная техническая система, неопределенность, организационная структура системы.

1. Введение

Вопросы анализа и синтеза сложных технических систем сопряжены с рядом трудностей. Эффективность применения универсальных или специальных методов анализа и синтеза указанного класса систем обусловлена возможностью отнесения их к тому или иному типу. Таким образом, выделение типа сложной системы на основе специфицированных уникальных признаков позволяет повысить эффективность моделирования и управления такими системами. Традиционные методы анализа

¹ Иван Анатольевич Щербатов, кандидат технических наук, доцент (sherbatov2004@mail.ru).

² Олег Мирославович Проталинский, доктор технических наук, профессор (prot@astu.org).

сложных технических систем подразумевают выделение подсистем внутри системы и описание связей между ними. Получение точного описания типа подсистемы и выделение признаков, позволяющих ее идентифицировать, является крайне актуальной задачей для получения точного описания сложной технической системы, поведение коалиции подсистем которой не будет существенным образом отличаться от оригинала [4].

Существует значительное число публикаций, в которых с различных позиций дается определение термину *сложная система*. Например, в [13] указывается на существенные трудности по описанию, пониманию, предсказанию, управлению, проектированию и изменению сложных систем. В [11] предлагается наряду с числом элементов, силой их взаимодействия, изменчивости и разнообразия учитывать также цели деятельности системы. Кроме того, существует значительное число классификационных делений сложных систем [2].

Сложная техническая система – сложная система, включающая в себя технический (технологический) процесс [12, 15]. Признаки *слабоформализуемых сложных технических систем* введены в [7].

В данной работе вводится новое свойство сложной слабоформализуемой технической системы (ССТС) – *многокомпонентность*, т.е. наличие некоторой специфической коалиции структурных элементов.

Термин *компонента* при анализе и синтезе сложных систем используется в различных контекстах. Например, в [2] компонента – совокупность однородных элементов системы, не обладающая свойствами системы.

Кроме того, существует определение компоненты как объекта классического анализа, используемого при имитационном моделировании, т.е. данный термин относится не к системам, а к их модельному представлению [1].

В этой связи крайне важно определить термин *компонента* применительно к классу сложных слабоформализуемых многокомпонентных технических систем.

Целью данной работы является выделение слабоформализуемых многокомпонентных технических систем как отдельного класса сложных систем.

2. Сложная слабоформализуемая техническая система

Существует ряд основных признаков сложных систем. Основным признаком является развитая уровневая структура взаимодействия подсистем и элементов внутри подсистем. Современные сложные системы являются многомерными, нелинейными и многосвязанными. Поэтому становится возможным выделить свойств, возникающих при системном анализе таких систем, их моделировании и синтезе систем управления ими:

- развитая уровневая псевдоиерархичность, когда взаимодействие происходит не только сверху вниз, но и наоборот, в связи с чем возникает потребность в реализации адаптивных, в том числе интеллектуальных систем управления с изменяемой структурой (свойство А1);

- множественность описания, фиксирующая зачастую огромное число вариантов математического описания сложных систем, а значит, требует применения высокоскоростных методов и алгоритмов ранжирования и выбора возможной реализации модели системы (системного описания) (свойство А2);

- различные виды неопределенности, которые зачастую выливаются в слабопрогнозируемые динамические процессы, возникающие в сложной системе в результате взаимодействия подсистем (свойство А3);

- присутствие синергетического эффекта [5], приводящее к многовариантности поведения подсистем и элементов подсистем (свойство А4).

Таким образом *ССТС* – сложная техническая система, обладающая признаками А1, А2, А3 и А4, функционирующая в условиях неопределенности, обусловленной действием двух факторов: невозможностью получения данных (свойство А3.1) и принципиальным незнанием закономерностей протекания процесса (свойство А3.2).

Незнание закономерностей протекания процесса определяется неопределенностью структуры (свойство А.3.2.1) и неопределенностью цели (свойство А.3.2.2). Неопределенность вида А.3.2.1 приводит к тому, что процесс построения математической модели крайне сложен (не возможен в ряде случаев). Не-

определенность вида А.3.2.2 характеризует отсутствие возможности сформулировать цель.

3. Понятие компоненты

В [1] введено понятие *компоненты* применительно к реализации замкнутой, изолированной от внешнего мира модели для целей имитационного моделирования. То есть *компонента*-подсистема, не допускающая (не требующая) дальнейшей декомпозиции для целей моделирования, а совокупность таких компонент образует модель рассматриваемой ССТС. Устройство компоненты похоже на объект классического анализа и обладает характеристиками, событиями и методами, однако несколько модифицированными.

Понятие компоненты в рамках данной работы используется для описания не моделей ССТС, а их организационных структур, т.е. осуществляется переход к рассмотрению слабоформализуемых многокомпонентных технических систем (МС).

Компонента – множество элементов ССТС, образующих единую элементарную структуру для достижения некоторой локальной цели в процессе ее функционирования [9].

Таким образом, возникает еще одно крайне важное интегративное вариативное свойство ССТС – *многокомпонентность* (А5). Наличие данного свойства в ССТС генерирует новый класс ССТС – МС. Наличие данного аспекта, являющегося крайне важным в системно-синергетической парадигме рассмотрения ССТС, не позволяет рассматривать ССТС с позиций классического системного анализа. Многокомпонентность приводит к необходимости разработки новых концептуальных принципов системного анализа таких систем [10].

Сложная слабоформализуемая многокомпонентная техническая система – это сложная система, обладающая кортежем свойств $\langle A_1, A_2, A_3, A_4, A_5 \rangle$ и образованная ограниченным множеством компонент $S = \{K_1, K_2, \dots, K_l\}$.

При системном анализе одной МС число образованных в ее составе компонент может различаться ($l = var$). А значит, различные варианты объединения элементов МС в компоненты

постулируют проблему множественности компонентного описания МС.

Компонента обладает рядом характерных признаков, которые должны быть описаны при ее выделении в МС:

$$(1) K = \{(X, Y, W), P_{ij}, M_k, I, L, T, N\}$$

где X – входные, Y – выходные и W внутрикомпонентные переменные; P_{ij} – связи между элементами в компоненте; M_k – механизмы образования связей; I – цель образования (функционирования) компоненты; L – уровень вложенности – элементарная или составная (образованная совокупностью элементарных компонент); T – тип, формализуемая (с помощью регулярных методов), слабоформализуемая (с помощью методов искусственного интеллекта), не формализуемая (в силу ряда причин не поддающаяся описанию с помощью известных методов); N – тип неопределенности, (статистические и не статистические), в которых функционирует компонента.

4. Компоненты и подсистемы

Обособление компоненты не только как важной составляющей понятийного аппарата выделяемого класса сложных слабоформализуемых систем, но и структурной единицы процесса функционирования МС требует существенно более пристального внимания уделить раскрытию коренных отличий термина «компонента» от термина «подсистема» применительно для целей системного анализа МС. Для удобства представим различия в табличной форме (таблица 1).

Таблица 1. Различия между «подсистемой» и «компонентой»

	Подсистема	Компонента
Образование на этапе жизненного цикла системы [6]	Проектирование	Функционирование
Наличие цели/подцели (наличие свойств системы) [2]	Да	Да

	Подсистема	Компонента
Ограниченная интеллектуальность	Выделение не подразумевается	Учитывается при формировании компоненты и принятии решения
Реорганизация и модификация структуры	Нет	Да, с учетом изменения/коррекции глобальной цели
Применимость к модельному представлению [1]	Да	Да
Учет механизмов образования структурной единицы [10]	Нет	Да
Образование на основе типа неопределенности	Нет	Да
Однородность информационных связей	Не всегда	Да

5. Карта компоненты

Для реализации модельного представления компоненты в форме табличного описания по результатам системного анализа реализована оригинальная табличная структура – карта компоненты. Структура карты компоненты МС представлена в таблице 2.

В карте компоненты фиксируются связи между элементами, формирующими компоненту, а также приводятся описания самих элементов. Кроме того, приводятся описания механизмов образования связей между элементами внутри компоненты. Наряду с целью образования компоненты, другие виды целей можно разделить на подцели и привести графическое изображение дерева целей с указанием вложенности и описания. В предложенной форме карты компоненты должен быть также указан уровень вложенности компоненты и тип, к которому она относится. Полученная таким образом карта может применяться для построения математической модели компоненты и ее реализации в виде имитационной модели. Переменные разделяются на два вида – количественные

(X, Y, W) и качественные (X^*, Y^*, W^*) , формализуемые с помощью методов искусственного интеллекта.

Таблица 2. Карта компоненты

Переменные		Уровень вложенности L	Элементарная/составная
Тип	Название переменной		
Входные X, X^*	$x_1 - \dots;$ $x_1^* - \dots$	Тип T	формализуемая/слабоформализуемая/ не формализуемая
Выходные Y, Y^*	$y_1 - \dots;$ $y_1^* - \dots$		
Внутрикомпонентные W, W^*	$w_1 - \dots;$ $w_1^* - \dots$	Иерархия целей I	
Связи между элементами внутри компоненты P_{ij}		Графическая структура	Описание целей
Графическая структура	Описание элементов		$I_{11} - \dots$ $I_{21} - \dots$
	$\mathcal{E}_{11} - \dots$ $\mathcal{E}_{21} - \dots$		
Механизмы образования связей M_k		Тип неопределенности	
Связь	Механизм образования	Статистическая	$N_{11} - \dots$
$\mathcal{E}_{11} - \mathcal{E}_{21}$...	Не статистическая	$N_{21} - \dots$

6. Понятие компонентной структуры

Введенное понятие компоненты и МС неразрывно связано с множественностью организационных структур, которые могут варьироваться в зависимости от целей системного анализа, моделирования или управления.

Взаимодействие внутри класса МС осуществляется на основе типовых организационных структур. С позиций синергетического подхода это означает, что существует некоторое ограниченное множество устойчивых состояний ССТС. Дадим определение организационной структуры внутри ССТС.

Организационная структура – совокупность взаимодействующих подсистем внутри ССТС и механизмов образования связей между ними. Организационная структура образованная более чем двумя обособленными подсистемами является сложной в смысле множественности связей между подсистемами:

$$(2) OS = \langle (S_1, \dots, S_j), P_{ij}(S_{ij}), M_k(P_{ij}) \rangle$$

где OS – сложная организационная структура ССТС (понятие введено по аналогии с [3]); s_j – подсистема внутри ССТС; P_{ij} – связь между i -й и j -й подсистемами; M_k – механизм образования связей P_{ij} .

С учетом (1) введем понятие *компонентной структуры* MC – совокупность взаимодействующих компонент внутри MC и механизмов образования связей между ними:

$$(3) KS = \left\{ (K_1, K_2, \dots, K_l), P_{ij}(K_{ij}), M_k(P_{ij}) \right\}$$

Для раскрытия понятия компоненты и формулирования доказательной базы целесообразности использования понятия компоненты в сложных системах приведем механизмы образования компонентных структур.

Необходимость выделения в составе MC структурных образований, отвечающих определению компоненты, обусловлена удобством представления для целей управления. На этапе проектирования выделяются и анализируются подсистемы MC , которые в процессе ее функционирования будут образовывать новые структурные формации (компоненты), отвечающие требованиям общности на основе единой цели или однотипной неопределенности. Таким образом, если в процессе функционирования удастся выделить компонентную структуру, состоящую из однотипных компонент, то управления MC становится значительно эффективнее. Покажем данный аспект на примере образования компонентных структур, состоящих из двух типов компонент.

7. Образование компонент и компонентных структур в MC

Компонента $K = \{e_{ij}\}$ есть множество элементов $e_{ij} \subset S$, входящих в различные подсистемы s_j и образующих единую

структуру. Компоненты – организационные структуры, которые формируются в МС двумя возможными способами:

- объединение части элементов одной подсистемы $K = \{e_{ij}\}$, где $j = const$ – номер подсистемы, а i – индекс элемента в подсистеме подсистемы s_j , $i \leq \gamma$ (γ – мощность множества элементов подсистемы s_j);

- объединение элементов подсистем $K = \bigcup_j \{e_{ij}\}$, где $j = var$ – номер подсистемы, а i – показывает номер элемента подсистемы s_j , $i \leq n$ (n – число подсистем в S).

Объединение элементов подсистем и образование компонент в процессе функционирования происходит двумя возможными способами: для достижения единой цели; на основании агрегирования элементов системы по типу неопределенности.

Образование компонент на основании единой цели происходит следующим образом. Пусть в МС S имеется j локальных целей (j – число компонент), достижение которых позволяет достигать глобальную цель I .

В случае изменения глобальной цели МС происходит коррекция локальных целей, которая приводит к изменению структурной организации системы и образованию компонент. При этом в общем случае $j \neq l$, где l – число вновь образованных компонент. Стоит отметить, что элементы подсистем образуют компоненты таким образом, чтобы они совместно позволяли достигать единую локальную цель.

Объединение компонент на основе типа неопределенности основано на предположении, что для перевода МС в целевое состояние и обеспечения устойчивости этого состояния ряд подсистем и их элементов могут образовывать структуры, на составные части которых действуют схожие по своей природе неопределенности. Поэтому устранение или компенсация некоторого вида неопределенности в двух компонентах значительно сложнее, чем внутри одной компоненты. Образование компонент на основе типа неопределенности происходит следующим образом.

Пусть в исходной МС существуют 2 компоненты – k_1 и k_2 . Допустим, что элементы исходных компонент $e_{1i} \rightarrow k_1$ и $e_{2i} \rightarrow k_2$

обладают свойством структурной неопределенности, а значит, этим свойством обладают и сами компоненты. Тогда для устранения данного вида неопределенности, который влияет на обе компоненты, возможно образование новой организационной структуры, состоящей из компонент \overline{k}_1 и \overline{k}_k , одна из которых (\overline{k}_1) свойством структурной неопределенности не обладает.

Компоненты находятся во взаимодействии между собой и внешней средой, реализуя это взаимодействие посредством организационных отношений. Одним из ключевых аспектов является устойчивость организационной структуры МС. При этом взаимное расположение компонент может изменяться, но эта модификация не должна нарушать связи, обеспечивающие функционирование МС для выполнения предусмотренных потребительских функций и достижения глобальной цели.

8. Устойчивость компонентной структуры МС

Под *устойчивостью* в общем случае понимается возможность компонентной структуры МС перейти в другую компонентную структуру, не нарушая связей и не порождая новых свойств и отношений, при любом изменении входных величин $x \in X$, поступлении возмущающих воздействий $f \in F$ и влиянии неопределенностей $\zeta \in Z$.

Таким образом,

$$\forall x \in X, f \in F, \zeta \in Z \exists u \in U : KS' \xrightarrow{u} \overline{KS} \rightarrow I,$$

где u – управляющее воздействие, переводящее систему в целевое устойчивое состояние и корректирующее ее компонентную структуру $KS' \xrightarrow{u} \overline{KS}$ для достижения глобальной цели I .

Поэтому обеспечение инвариантности МС по отношению к влиянию имеющихся неопределенностей различных типов обеспечит ее устойчивость при условии, что возмущающие воздействия компенсируются системой управления.

Коммуникационные возможности между компонентами МС обеспечиваются за счет связей. Сделаем два замечания.

Замечание 1. Компонентная структура типа «компонента – компонента» является простейшей.

Замечание 2. Особенности взаимодействия простейших компонентных структур могут быть распространены на любое число взаимодействующих компонент.

Особенности организационного взаимодействия компонент в составе компонентных структур заключаются в вариативности свойств элементарных актов воздействия компонент друг на друга, могут возникать ранее не рассматривавшиеся свойства, что приводит к появлению синергетического эффекта внутри МС.

Поэтому каждую компоненту можно отнести к типу активных или пассивных в поведенческом смысле. Это означает, что именно поведение технической системы внутри организации компонент, образующих единое целое, влияет на ее принадлежность к тому или иному типу.

9. Типы компонент и компонентных структур

Активная компонента – техническая система, обладающая признаками рационального поведения, осознанного планирования действий и реализации этого плана (таблица 3).

В сравнении с человеком-оператором, который может также присутствовать как отдельная составляющая (для упрощения может рассматриваться в качестве условной компоненты), активная компонента обладает ограниченной самостоятельностью. При этом переход к рассмотрению человека в составе компоненты потребует формализации компоненты как моноэнергетической системы, что может существенным образом затруднить вопросы анализа МС, поэтому при рассмотрении типовых компонентных структур оператор выделяется как самостоятельная единица организационного взаимодействия.

Таблица 3. Признаки рационального поведения компонент

Признак	Примечание
Наличие органов технического зрения и сенсорного очувствления	Позволяет строить сенсорные карты среды с выбранным временным интервалом

Признак	Примечание
Наличие механизма понимания сенсорной карты и распознавания сложившейся ситуации	Распознавание образа сложившейся ситуации с учетом имеющейся неопределенности
Умение формировать план поведения с учетом имеющейся сенсорной карты	Позволяет исключить влияние неопределенности на поведение подсистемы
Наличие механизма реализации плана поведения	Адаптация в меняющихся внешних условиях

Пассивная компонента – техническая компонента, не обладающая признаками рационального поведения.

Разбиение компонент внутри МС строго на два типа (активную и пассивную) обеспечивает рассмотрение разнородных подсистем не с классических позиций непрерывного и дискретного производства в статическом или динамическом режимах, а с позиций интеллектуальных самоорганизующихся систем с заданным поведением. Это обеспечивает переход к рассмотрению поведенческих аспектов технических систем, обладающих рациональными поведенческими признаками. Термин рациональный в данном случае характеризует способность компоненты варьировать стратегию поведения в составе компонентной структуры МС с учетом изменяющихся условий внешней среды.

Например, в качестве активной компоненты МС может выступать робототехническая система [8], а в качестве пассивной слабоформализуемый технологический процесс (СФТП) [7] (рис. 1).

В самом общем случае существует 5 вариантов простейшего взаимодействия компонент внутри компонентных структур МС: «пассивная компонента – пассивная компонента»; «оператор – пассивная компонента»; «пассивная компонента – активная компонента»; «активная компонента – активная компонента»; «оператор – активная компонента».

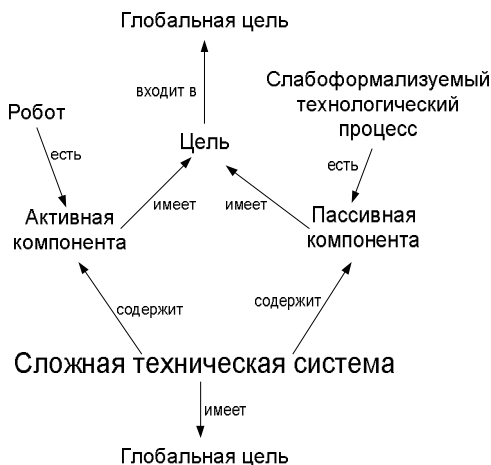


Рис. 1. Семантическая схема МС

10. Применение предложенного подхода

Рассмотрим проблему анализа сложной системы и ее отношение к классу МС. На принадлежность той или иной сложной технической системы к классу МС оказывает существенное влияние взаимное расположение компонент друг относительно друга внутри ограниченного пространства образующего МС с учетом направленности и силы взаимного влияния внутрисистемных связей, порождающих межкомпонентные и межэлементные отношения.

Покажем влияние наличия различных рациональных поведенческих признаков на степень интеллектуальности компонент выделенных видов (табл. 4).

В [14] введен показатель, характеризующий степень интеллектуальности компонент $\gamma \in [0, 1]$. Чем выше значение γ , тем большее количество признаков интеллектуальности присутствует у компоненты, что позволяет отнести ее к типу активных компонент.

Таблица 4. Признаки интеллектуальности компонент

Активная компонента	Признаки рационального поведения				
	№ 1	№ 2	№ 3	№ 4	№ 5
Слабоформализуемый технологический процесс					
Гибкая автоматизированная линия					
Сварочный робот					
Мобильный робот					
Группа роботов					

Степень интеллектуальности γ :

$$(4) \quad \gamma = \frac{\sum_{j=1}^l \sum_{i=1}^m w_i \rho_i^j}{l}$$

где $\rho_i^j = \begin{cases} 1, \text{признак присутствует,} \\ 0, \text{признак отсутствует;} \end{cases}$ w_i – степень важности (вес)

i -го признака, $\sum_i w_i = 1$.

Степень интеллектуальности позволяет выделять компоненты и элементы, которые могут функционировать самостоятельно (на данном технологическом этапе развития эта самостоятельность не может быть полной), сообразно с необходимостью достижения общей цели МС.

Пусть имеется 5 рациональных поведенческих признаков, обладающих одинаковым весом $w_i = 0, 2$, $i = 1, \dots, 5$ (таблица 3). Тогда для компонент, указанных в таблице 3 величина γ будет равна: $\gamma_1 = 0$; $\gamma_2 = 0,2$; $\gamma_3 = 0,4$; $\gamma_4 = 0,8$; $\gamma_5 = 1$.

11. Заключение

Противоречие, заключающееся в том, что МС являясь сложной системой, не может быть формализована с помощью методов, разработанных для сложных систем в классическом понимании, должно быть разрешено на основании разработки и

применения новых (модифицированных) методов. Данные методы должны быть реализованы для учета свойства многокомпонентности.

На основании введенного понятия компонентной структуры МС возможно описать типовые компонентные структуры, а также их взаимодействие для целей математического моделирования и управления МС. Выделение обособленных компонент в компонентной структуре МС позволяет формализовать подходы для системного анализа рассматриваемого класса сложных систем [10], что значительно повышает эффективность построения и реализации систем управления ими.

Подходы, описанные в статье, позволяют перейти к разработке компонентного подхода для целей математического моделирования и синтеза методов управления, в том числе группового, компонентами в рамках компонентных структур МС.

Литература

1. БРОДСКИЙ Ю.И., ПАВЛОВСКИЙ Ю.Н. *Разработка инструментальной системы распределенного имитационного моделирования* // Информационные технологии и вычислительные системы. – 2009. – №4. – С. 9–21.
2. ВОЛКОВА В.Н., ДЕНИСОВ А.А. *Теория систем и системный анализ*. – М.: Изд. Юрайт, 2010. – 680 с.
3. ГУБКО М.В., КОРГИН Н.А., НОВИКОВ Д.А. *Классификация моделей анализа и синтеза организационных структур* // Управление большими системами. – 2004. – №6. – С. 5–21.
4. ДУБОВ В.М., КАПУСТЯНСКАЯ Т.И., ПОПОВ С.А., ШАРОВ А.А. *Проблематика сложных систем (концептуальные основы модельных представлений)* / Под общ. ред. С.А. Попова. – СПб.: «Элмор», 2006. – 184 с.
5. КОЛЕСНИКОВ А.А. *Синергетические методы управления сложными системами: Теория системного синтеза*. – М.: КомКнига, 2006. – 240 с.
6. ОСТРОВСКИЙ Г.М., ВОЛИН Ю.М. *Технические системы в условиях неопределенности: анализ гибкости и*

- оптимизация: учебное пособие.* – М.: БИНОМ. Лаборатория знаний, 2008. – 319 с.
7. ПРОТАЛИНСКИЙ О.М. *Применение методов искусственного интеллекта при автоматизации технологических процессов*: Моногр. / Астрахан. гос. техн. ун-т. – Астрахань: Изд-во АГТУ, 2004. – 184 с.
 8. ЩЕРБАТОВ И.А., ПРОТАЛИНСКИЙ И.О. *Исследование эффективности группового управления роботами методом имитационного моделирования* // Вестник Саратовского государственного технического университета. – 2010. – №4(50). – С. 34–37.
 9. ЩЕРБАТОВ И.А. *Понятие компоненты слабоформализуемой многокомпонентной технической системы* // Сб. трудов XXV Междунар. науч. конф. «Математические методы в технике и технологиях ММТТ-25». Т.8. – Волгоград: Волгогр. гос. техн. ун-т, 2012. – С. 207–209.
 10. ЩЕРБАТОВ И.А. *Концепция системного анализа сложных слабоформализуемых многокомпонентных систем в условиях неопределенности* // Современные технологии. Системный анализ. Моделирование. – 2013. – №2. – С. 28–35.
 11. BAR-YAM YANEER. *About engineering complex systems: Multiscale analysis and evolutionary engineering* // Engineering self-organizing systems: methodologies and applications. – 2005. – Vol. 3464. – P. 16-31.
 12. HUBKA V., EDER W.E., *Theory of Technical Systems.* – Berlin: Springer-Verlag, 1988. – 132 p.
 13. MAGEE C., DE WECK O.L., *Complex System Classification* // Fourteenth Annual International Symposium of the International Council on Systems Engineering (INCOSE), Toulouse, France, June 20-24, 2004. – P. 24–42.
 14. SHCHERBATOV I.A. *Classification of pure formalized complex multicomponent technical systems under conditions of uncertainty* // Вестн. Астрахан. гос. техн. ун-та. Серия: «Управление, вычислительная техника и информатика». – 2012. – №2. – С. 9–13.

15. SIMON H. *The Sciences of the Artificial*. – Cambridge, MA: The MIT Press, 1982. – 217 p.

COMPLEX ILL-FORMALIZABLE MULTICOMPONENT TECHNICAL SYSTEMS

Ivan Shcherbatov, Astrakhan State Technical University, Astrakhan, Cand.Sc., assistant professor (sherbatov2004@mail.ru).

Oleg Protalinskiy, Astrakhan State Technical University, Astrakhan, Doctor of Science, professor (prot@astu.org).

Abstract: We suggest a concept of a complex ill-formalizable multi-component technical system. For such systems we reveal mechanisms of organization structure formation and introduce the method for components' description. We suggest classifying components into active and passive ones and explain main rational behavioral features of components. Examples of an active component (a robot) and a passive component (an ill-formalizable technological process) are given.

Keywords: component, complex technical system, uncertainty, organization structure of the system.

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

*Поступила в редакцию 14.12.2012.
Опубликована 30.09.2013.*

УДК 519.63 + 519.246

ББК 22.161.6 + 22.193 + 22.171

СПЕКТРАЛЬНЫЙ МЕТОД АНАЛИЗА СТОХАСТИЧЕСКИХ СИСТЕМ С РАЗРЫВАМИ ТРАЕКТОРИЙ, ОПИСЫВАЕМЫМИ СЛУЧАЙНОЙ СМЕСЬЮ ЭРЛАНГОВСКИХ РАСПРЕДЕЛЕНИЙ¹

Кожевников А. С.², Рыбаков К. А.³

(Московский авиационный институт, Москва)

В статье рассматриваются стохастические системы управления с импульсными воздействиями, которые образуют гиперэрланговские потоки событий и приводят к разрывам траекторий системы. Решается задача нахождения плотности вероятности вектора состояния. В основе решения лежит использование спектральной формы математического описания систем управления.

Ключевые слова: гиперэрланговское распределение, задача анализа, скачкообразный процесс, спектральный метод, стохастическая система.

1. Введение

Многие окружающие нас явления и закономерности (природные, технические, экономические) имеют случайный характер, что позволяет их описывать с помощью случайных процессов. Для описания моделей явлений, учитывающих влияние случайных факторов, как правило, применяют стохастические дифференциальные уравнения (при условии, конечно, что время можно считать непрерывным).

¹ Работа выполнена при поддержке РФФИ (проект №12-08-00892-а).

² Александр Сергеевич Кожевников, аспирант (echequit@yandex.ru).

³ Константин Александрович Рыбаков, кандидат физико-математических наук, доцент (rkoffice@mail.ru).

Используя уравнения с диффузионной и скачкообразной компонентами, можно моделировать поведение довольно сложных систем, учитывая как непрерывные случайные воздействия, так и импульсные, приводящие к разрывам траекторий. Однако часто для описания скачкообразной компоненты ограничиваются общим пуассоновским процессом (или пуассоновской случайной мерой) [19, 23, 24]. Такой выбор обусловлен развитым математическим аппаратом (стохастические дифференциальные уравнения с пуассоновской составляющей, уравнения Колмогорова–Феллера), но следствием выбора общего пуассоновского процесса является использование только показательного закона распределения для промежутков времени между последовательными импульсными воздействиями (разрывами траекторий). В частности, для случайных величин, имеющих показательное распределение, среднее значение и стандартное отклонение совпадают, что вносит некоторые ограничения при решении прикладных задач [6].

Для того чтобы снять такие ограничения, можно использовать модели систем со случайным периодом квантования – частный случай систем со случайной структурой [4, 7, 12, 13, 20]. В них появлением разрывов траекторий управляет вспомогательный марковский процесс с конечным множеством состояний, задаваемый начальным распределением и интенсивностями переходов. Использование таких моделей приводит к ряду трудностей, в основном связанных со сложностью решения уравнений для нахождения плотности вероятности вектора состояния – системы обобщенных уравнений Фоккера–Планка–Колмогорова (интегро-дифференциальных уравнений в частных производных), решение уравнений для моментов тоже сложнее, чем для моделей, использующих только общий пуассоновский процесс. Вторая сложность связана с тем, что большое количество параметров модели (интенсивности, параметры распределений приращений, которые получает вектор состояния при разрывах траекторий для различных переходов между состояниями вспомогательного марковского процесса) затрудняет их идентификацию. Поэтому целесообразно, взяв за основу модель

системы со случайным периодом квантования, рассмотреть некоторые частные случаи с обозримым числом параметров. Один из таких частных случаев сводится к описанию промежутков времени между последовательными импульсными воздействиями эрланговским законом распределения, т.е. когда состояния вспомогательного марковского процесса последовательно сменяются «по кругу», а переход из некоторого, заранее выбранного, состояния в следующее сопровождается разрывом траектории: моменты появления разрывов траекторий образуют эрланговский поток событий. Такие модели изучены, например, в [4, 10, 22]. Модель, рассматриваемая в этой статье, допускает, что моменты появления разрывов траекторий образуют гиперэрланговский поток событий и формируется два типа приращений при разрыве траектории, имеющих различные законы распределения. Распределение промежутков времени между последовательными разрывами можно задать смесью эрланговских распределений. Использование такого гиперэрланговского распределения позволяет включить показательное и эрланговское распределения как частные случаи и расширить множество решаемых прикладных задач. Аналогичный подход применен в [22, 25].

В последнее время системы, описываемые стохастическими дифференциальными уравнениями с диффузионной и скачкообразной компонентами, все чаще используются для моделирования цен акций [9, 21, 24]. Разумеется, задачи финансовой математики составляют только часть прикладных задач, при решении которых целесообразно использовать математический аппарат теории систем со случайным периодом квантования. Например, можно рассматривать асинхронные системы управления сложными техническими объектами [4], механические системы с импульсными воздействиями [22], изучать влияние случайных импульсных воздействий и помех на электрические цепи [2, 19] и др.

Все существующие методы анализа систем со случайной структурой, допускающих при смене структуры разрывы траекторий, можно применять и при анализе систем со случайным

периодом квантования. Один из них – метод Монте-Карло [1, 8]. Он сводится к многократному непосредственному моделированию реализаций случайного процесса и их статистической обработке. Существенное преимущество этого метода состоит в простоте реализации и универсальности. К его недостаткам можно отнести необходимость проведения большого количества вычислений для получения достаточной точности, особенно при оценке плотности вероятности вектора состояния.

Другие методы, так или иначе, связаны с уравнениями для нахождения плотности вероятности вектора состояния – обобщенными уравнениями Фоккера–Планка–Колмогорова: методы, основанные на задании структуры плотности вероятности вектора состояния, например, полигауссовская аппроксимация; переход к моментам; методы, связанные с представлением плотности вероятности или связанной с ней характеристики в виде функционального ряда: степенного, ортогонального; численные методы). Подробное описание таких методов содержится в монографиях [7, 20].

Предлагаемый метод решения основан на ортогональных разложениях функций с использованием спектральной формы математического описания систем управления [12, 18], он является развитием подхода, применяемого для более простых стохастических систем [2, 10, 11].

Отметим, что представление решения задачи в виде ортогонального ряда довольно часто используется при построении приближенно-аналитических методов. Но, как правило, для этого выбирается конкретная система ортогональных или биортогональных функций и выводятся соотношения для нахождения коэффициентов ряда. При таком подходе, конечно же, используются свойства выбранных функций (например, рекуррентные соотношения, связывающие эти функции; формулы, устанавливающие связи функций и их производных и т.п.). Основное отличие предлагаемого метода состоит в использовании произвольной ортонормированной системы, при этом соотношения для определения коэффициентов ряда не зависят от выбранных функций. Они представляются матричными уравне-

ниями, для которых можно получить формулы точного решения или сформировать методiku приближенного решения. Свойства выбранных функций также используются, но несколько иначе, а именно от этих свойств зависит структура матриц, которые ставятся в соответствие линейным операторам (операторам умножения, дифференцирования, интегрирования и др.), определяющим исходную задачу. Выбор той или иной системы ортонормированных функций может привести к тому, что соответствующая матрица будет треугольной, ленточной, симметрической или кососимметрической. Для многих базисных систем, например, для полиномов Лежандра, тригонометрических функций, функций Уолша и Хаара, обобщенных функций Лагерра (включающих полиномы и функции Лагерра) и обобщенных функций Эрмита (включающих полиномы и функции Эрмита), такие матрицы известны [12, 14, 16–18].

Для рассмотренной модели стохастической системы с разрывами траекторий, описываемыми случайной смесью эрланговских распределений, получено аналитическое решение матричных уравнений для коэффициентов разложения функций, по которым определяется плотность вероятности вектора состояния. Это же решение можно использовать и как приближенное, получая конечное число коэффициентов разложения и представляя плотность вероятности в виде частичной суммы ряда. Для апробации предложенной методики решен пример задачи анализа одномерной нелинейной стохастической системы с двумя типами приращений при разрыве траектории: случайным гауссовским и детерминированным; промежутки времени между последовательными разрывами описываются эрланговским и показательным распределениями.

2. Постановка задачи

Будем предполагать, что поведение модели системы управления описывается случайным процессом в непрерывном времени, являющимся аддитивной смесью диффузионного и специ-

ального скачкообразного процессов и удовлетворяющим стохастическому дифференциальному уравнению Ито [13]:

$$(1) \quad dX(t) = f(t, X(t))dt + \sigma(t, X(t))dW(t) + dQ(t), \quad X(t_0) = X_0,$$

в котором $X \in R^n$ – вектор состояния; $t \in T$, $T = [t_0, t_1]$ – заданный отрезок времени функционирования системы;

$f(t, x): T \times R^n \rightarrow R^n$ – вектор-функция размеров $n \times 1$;

$\sigma(t, x): T \times R^n \rightarrow R^{n \times s}$ – матричная функция размеров $n \times s$;

$W(t)$ – s -мерный стандартный винеровский процесс, который не зависит от начального состояния X_0 , определяемого заданной плотностью вероятности $\varphi_0(x)$.

Случайный процесс $Q(t)$ представляется в виде

$$Q(t) = \sum_{i=1}^{J(t)} (\xi_i Y_1(\tau_i) + (1 - \xi_i) Y_2(\tau_i)),$$

где $J(t)$ – гиперэрланговский процесс, ассоциированный со случайным потоком событий [4], состоящих в том, что вектор состояния X получает приращения $Y_1(\tau_i) \in R^n$ или $Y_2(\tau_i) \in R^n$ в случайные моменты времени τ_1, τ_2, \dots . Случайный вектор $Y_1(\tau_i)$ характеризуется плотностью вероятности $q_1(t, y)$, а случайный вектор $Y_2(\tau_i)$ – плотностью вероятности $q_2(t, y)$; $t = \tau_i$. Выбор приращения $Y_1(\tau_i)$ или $Y_2(\tau_i)$ зависит от случайной величины ξ_i , принимающей значения 1 с вероятностью $\lambda_1/(\lambda_1 + \lambda_2)$ и 0 с вероятностью $\lambda_2/(\lambda_1 + \lambda_2)$ (случайные величины ξ_1, ξ_2, \dots имеют распределение Бернулли):

$$X(\tau_i) = X(\tau_i - 0) + \left\{ \begin{array}{l} Y_1(\tau_i), \quad \xi_i = 1 \\ Y_2(\tau_i), \quad \xi_i = 0 \end{array} \right\}.$$

Заданные положительные числа λ_1 и λ_2 , а также натуральные числа N_1 и N_2 определяют гиперэрланговский закон распределения промежутков времени $\Delta\tau_i = \tau_i - \tau_{i-1}$, $i = 1, 2, \dots$ ($\tau_0 = t_0$), который является эрланговским с параметрами λ_1 и N_1 , если $\xi_i = 1$, или эрланговским с параметрами λ_2 и N_2 , если $\xi_i = 0$. Случайные величины ξ_i независимы, поэтому выбор закона распределения для случайного приращения – $q_1(t, y)$ или $q_2(t, y)$ – в момент времени τ_i не зависит от предыстории.

Возможно и другое описание схемы появления событий (разрывов в траекториях вектора состояния). Для этого рассмот-

рим случайный процесс $K(t)$ с конечным множеством состояний $\{1, 2, \dots, N\}$, где $N = N_1 + N_2 - 1$. Интенсивности переходов задаются следующим образом (см. рис. 1): смена состояний $1 \rightarrow 2, 2 \rightarrow 3, \dots, N_1 - 1 \rightarrow N_1$ и $N_1 \rightarrow 1$ происходит с интенсивностью λ_1 , а смена состояний $1 \rightarrow N_1 + 1, N_1 + 1 \rightarrow N_1 + 2, \dots, N - 1 \rightarrow N$ и $N \rightarrow 1$ – с интенсивностью λ_2 ; другие переходы невозможны. При переходе из состояния с номером N_1 в состояние с номером 1 вектор X получает случайное приращение $Y_1(\tau_i)$, а при переходе из состояния с номером N в состояние с номером 1 – случайное приращение $Y_2(\tau_i)$, что соответствует разрыву (скачку) траектории процесса $X(t)$ (см. рис. 2).

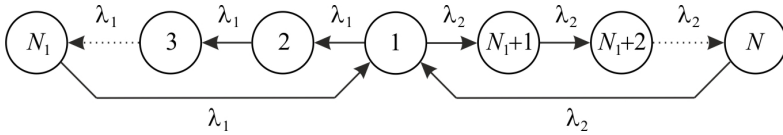


Рис. 1. Граф состояний случайного процесса $K(t)$

Если предположить, что $\lambda_1 > 0, \lambda_2 = 0$, то $J(t)$ будет эрланговским процессом порядка N_1 , а процессу $K(t)$ будет соответствовать кольцевой граф состояний (аналогично при $\lambda_1 = 0, \lambda_2 > 0$). Этот случай рассмотрен в работе [10].

Заметим, что постановка задачи может быть изменена с предположением, что интенсивности λ_1 и λ_2 зависят от времени, т.е. $\lambda_1 = \lambda_1(t), \lambda_2 = \lambda_2(t)$. Далее будем рассматривать именно этот вариант.

Предполагается [24], что функции $f(t, x)$ и $g(t, x) = \sigma(t, x) \times \times \sigma^T(t, x)$ кусочно-непрерывны по t для всех $X \in R^n$; при фиксированном $t \in T$ функция $f(t, x)$ имеет непрерывные и ограниченные частные производные первого порядка, а функция $g_{ij}(t, x)$ – непрерывные и ограниченные частные производные второго порядка по координатам вектора состояния, $i, j = 1, 2, \dots, n$. Кроме того, $\lambda_1(t)$ и $\lambda_2(t)$ непрерывны на множестве T , а случайные величины, задаваемые плотностями вероятности $q_1(t, y)$ и $q_2(t, y)$, имеют конечные моменты второго порядка.

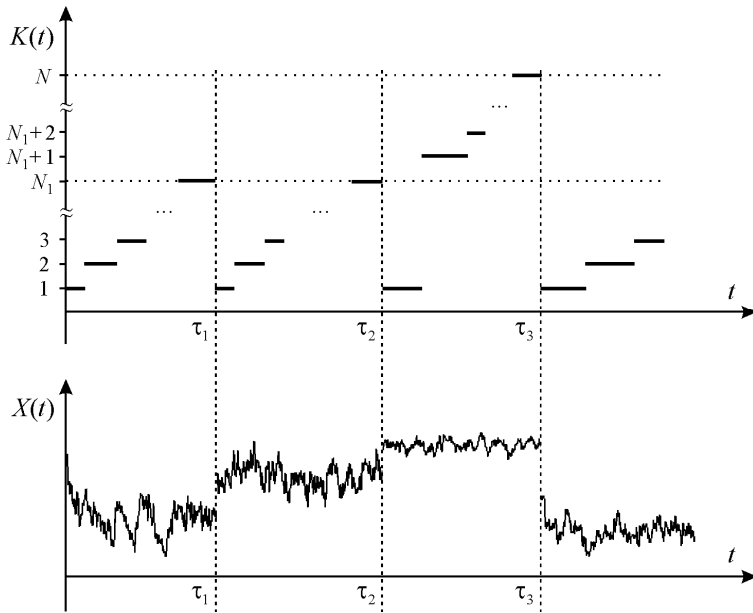


Рис. 2. Выборочные траектории процессов $K(t)$ и $X(t)$

Эти условия (иногда к ним добавляется дополнительное условие на матрицу диффузии $g(t, x)$, требуемое для разрешимости приведенных ниже параболических уравнений) обеспечивают существование и единственность сильного решения уравнения (1), однако они вносят слишком много ограничений при решении прикладных задач. Вопросы, связанные с ослаблением приведенных условий, рассмотрены в [3, 5]. В частности, можно рассматривать уравнения с разрывным коэффициентом сноса $f(t, x)$ или вырожденной матрицей диффузии $g(t, x)$, достаточно часто встречающиеся в задачах управления. Отметим, что в [3, 5] изучаются стохастические дифференциальные уравнения без скачкообразной компоненты или с пуассоновской составляющей, тем не менее эти результаты могут быть обобщены и на рассматриваемый случай. Кроме того, можно понимать решение уравнения (1) в слабом смысле, тем более что далее ставится задача нахождения плотности вероятности вектора состояния.

Таким образом, рассматривается стохастическая система с расширенным вектором состояния, непрерывная часть которого – X , а дискретная – $K \in \{1, 2, \dots, N\}$. Тогда плотность вероятности $\varphi(t, x)$ вектора X может быть представлена в виде суммы:

$$\varphi(t, x) = \sum_{k=1}^N \varphi^{(k)}(t, x),$$

где функции $\varphi^{(k)}(t, x)$ удовлетворяют системе обобщенных уравнений Фоккера–Планка–Колмогорова [4, 7, 12, 20]:

$$(2) \quad \frac{\partial \varphi^{(1)}(t, x)}{\partial t} = A\varphi^{(1)}(t, x) - (\lambda_1(t) + \lambda_2(t))\varphi^{(1)}(t, x) + \lambda_1(t) \int_{R^n} q_1(t, x - z)\varphi^{(N_1)}(t, z)dz + \lambda_2(t) \int_{R^n} q_2(t, x - z)\varphi^{(N)}(t, z)dz,$$

$$(3) \quad \frac{\partial \varphi^{(k)}(t, x)}{\partial t} = A\varphi^{(k)}(t, x) - \lambda_1(t)\varphi^{(k)}(t, x) + \lambda_1(t)\varphi^{(k-1)}(t, x),$$

$k = 2, \dots, N_1,$

$$(4) \quad \frac{\partial \varphi^{(N_1+1)}(t, x)}{\partial t} = A\varphi^{(N_1+1)}(t, x) - \lambda_2(t)\varphi^{(N_1+1)}(t, x) + \lambda_2(t)\varphi^{(1)}(t, x),$$

$$(5) \quad \frac{\partial \varphi^{(k)}(t, x)}{\partial t} = A\varphi^{(k)}(t, x) - \lambda_2(t)\varphi^{(k)}(t, x) + \lambda_2(t)\varphi^{(k-1)}(t, x),$$

$k = N_1 + 2, \dots, N,$

в которой

$$(6) \quad A\varphi^{(k)}(t, x) = -\sum_{i=1}^n \frac{\partial}{\partial x_i} \left[f_i(t, x)\varphi^{(k)}(t, x) \right] + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2}{\partial x_i \partial x_j} \left[g_{ij}(t, x)\varphi^{(k)}(t, x) \right], \quad k = 1, 2, \dots, N,$$

$$g_{ij}(t, x) = \sum_{r=1}^s \sigma_{ir}(t, x)\sigma_{jr}(t, x), \quad i, j = 1, 2, \dots, n.$$

Для случайного процесса $K(t)$ начальное состояние фиксировано: $K(t_0) = 1$, поэтому с учетом заданной плотности вероятности $\varphi_0(x)$ начального состояния X_0 имеем

$$(7) \quad \varphi^{(1)}(t_0, x) = \varphi_0(x), \quad \varphi^{(k)}(t_0, x) = 0, \quad k = 2, \dots, N.$$

Последние два слагаемых в правой части уравнения (2) для дальнейших преобразований целесообразно представить в операторной форме, определив

$$(8) \quad H_r \varphi(t, x) = \lambda_r(t) \int_{R^n} q_r(t, x - z) \varphi(t, z) dz, \quad r = 1, 2,$$

для всех допустимых функций $\varphi(t, x)$; H_r – линейный оператор, а именно композиция оператора умножения на функцию $\lambda_r(t)$ и оператора Фредгольма с ядром $q_r(t, x - z)$.

Задача анализа стохастической системы управления, описываемой уравнением (1), заключается в нахождении плотности вероятности $\varphi(t, x)$ вектора состояния X при заданных функциях $f(t, x)$, $\sigma(t, x)$, $\lambda_1(t)$, $\lambda_2(t)$, $q_1(t, y)$, $q_2(t, y)$ и $\varphi_0(x)$. Далее предполагается, что это решение существует в классе функций $W_2^{1,1}(T \times R^n)$, т.е. плотность вероятности $\varphi(t, x)$ и ее производные первого порядка принадлежат пространству $L_2(T \times R^n)$, при этом решение $\{\varphi^{(1)}(t, x), \varphi^{(2)}(t, x), \dots, \varphi^{(N)}(t, x)\}$ системы (2)–(5) понимается в обобщенном смысле, т.е. эта совокупность удовлетворяет соотношениям

$$\begin{aligned} & \int_{T \times R^n} \psi^{(1)}(t, x) \frac{\partial \varphi^{(1)}(t, x)}{\partial t} dt dx = \int_{T \times R^n} \varphi^{(1)}(t, x) A^* \psi^{(1)}(t, x) dt dx - \\ & - \int_{T \times R^n} \psi^{(1)}(t, x) (\lambda_1(t) + \lambda_2(t)) \varphi^{(1)}(t, x) dt dx + \\ & + \int_{T \times R^n} \psi^{(1)}(t, x) \left(\lambda_1(t) \int_{R^n} q_1(t, x - z) \varphi^{(N_1)}(t, z) dz + \right. \\ & \left. + \lambda_2(t) \int_{R^n} q_2(t, x - z) \varphi^{(N)}(t, z) dz \right) dt dx, \\ & \int_{T \times R^n} \psi^{(k)}(t, x) \frac{\partial \varphi^{(k)}(t, x)}{\partial t} dt dx = \int_{T \times R^n} \varphi^{(k)}(t, x) A^* \psi^{(k)}(t, x) dt dx - \\ & - \int_{T \times R^n} \psi^{(k)}(t, x) (\lambda_1(t) \varphi^{(k)}(t, x) - \lambda_1(t) \varphi^{(k-1)}(t, x)) dt dx, \\ & k = 2, \dots, N_1, \end{aligned}$$

$$\begin{aligned}
 & \int_{T \times R^n} \psi^{(N_1+1)}(t, x) \frac{\partial \varphi^{(N_1+1)}(t, x)}{\partial t} dt dx = \\
 & = \int_{T \times R^n} \varphi^{(N_1+1)}(t, x) A^* \psi^{(N_1+1)}(t, x) dt dx - \\
 & - \int_{T \times R^n} \psi^{(N_1+1)}(t, x) \left(\lambda_2(t) \varphi^{(N_1+1)}(t, x) - \lambda_2(t) \varphi^{(1)}(t, x) \right) dt dx, \\
 & \int_{T \times R^n} \psi^{(k)}(t, x) \frac{\partial \varphi^{(k)}(t, x)}{\partial t} dt dx = \int_{T \times R^n} \varphi^{(k)}(t, x) A^* \psi^{(k)}(t, x) dt dx - \\
 & - \int_{T \times R^n} \psi^{(k)}(t, x) \left(\lambda_2(t) \varphi^{(k)}(t, x) - \lambda_2(t) \varphi^{(k-1)}(t, x) \right) dt dx, \\
 & k = N_1 + 2, \dots, N,
 \end{aligned}$$

для любых бесконечно дифференцируемых функций $\psi^{(1)}(t, x)$, $\psi^{(2)}(t, x)$, ..., $\psi^{(N)}(t, x)$ с компактными носителями, где A^* – сопряженный оператор:

$$\begin{aligned}
 A^* \psi^{(k)}(t, x) &= \sum_{i=1}^n f_i(t, x) \frac{\partial \psi^{(k)}(t, x)}{\partial x_i} + \\
 &+ \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n g_{ij}(t, x) \frac{\partial^2 \psi^{(k)}(t, x)}{\partial x_i \partial x_j}, \quad k = 1, 2, \dots, N.
 \end{aligned}$$

3. Применение спектральной формы математического описания

Сведем задачу анализа к поиску коэффициентов разложения $\varphi_{i_0 i_1 \dots i_n}$ функции $\varphi(t, x)$ по ортонормированным функциям в пространстве $L_2(T \times R^n)$. Обозначим полную систему ортонормированных функций $\{e(i_0, i_1, \dots, i_n, t, x)\}_{i_0, i_1, \dots, i_n=0}^{\infty}$, предполагая, что они представляются в виде произведения

$$e(i_0, i_1, \dots, i_n, t, x) = q(i_0, t) \cdot p(i_1, \dots, i_n, x), \quad i_0, i_1, \dots, i_n = 0, 1, 2, \dots,$$

причем функции $\{q(i_0, t)\}_{i_0=0}^{\infty}$ и $\{p(i_1, \dots, i_n, x)\}_{i_1, \dots, i_n=0}^{\infty}$ образуют полные ортонормированные системы в пространствах $L_2(T)$ и $L_2(R^n)$ соответственно.

Применим спектральное преобразование [11, 12] к системе уравнений (2)–(5) с учетом условий (7), тогда

$$\begin{aligned}
 & P(n+1, n+1) \cdot \Phi^{(1)}(n+1, 0) - q(1, 0; t_0) \otimes \Phi_0(n, 0) = \\
 (9) \quad & = A(n+1, n+1) \cdot \Phi^{(1)}(n+1, 0) - (\Lambda_1(n+1, n+1) + \\
 & + \Lambda_2(n+1, n+1)) \cdot \Phi^{(1)}(n+1, 0) + H_1(n+1, n+1) \cdot \\
 & \cdot \Phi^{(N_1)}(n+1, 0) + H_2(n+1, n+1) \cdot \Phi^{(N)}(n+1, 0),
 \end{aligned}$$

$$\begin{aligned}
 (10) \quad & P(n+1, n+1) \cdot \Phi^{(k)}(n+1, 0) = A(n+1, n+1) \cdot \Phi^{(k)}(n+1, 0) - \\
 & - \Lambda_1(n+1, n+1) \cdot \Phi^{(k)}(n+1, 0) + \Lambda_1(n+1, n+1) \cdot \Phi^{(k-1)}(n+1, 0), \\
 & k = 2, \dots, N_1,
 \end{aligned}$$

$$\begin{aligned}
 (11) \quad & P(n+1, n+1) \cdot \Phi^{(N_1+1)}(n+1, 0) = A(n+1, n+1) \cdot \Phi^{(N_1+1)}(n+1, 0) - \\
 & - \Lambda_2(n+1, n+1) \cdot \Phi^{(N_1+1)}(n+1, 0) + \Lambda_2(n+1, n+1) \cdot \Phi^{(1)}(n+1, 0),
 \end{aligned}$$

$$\begin{aligned}
 (12) \quad & P(n+1, n+1) \cdot \Phi^{(k)}(n+1, 0) = A(n+1, n+1) \cdot \Phi^{(k)}(n+1, 0) - \\
 & - \Lambda_2(n+1, n+1) \cdot \Phi^{(k)}(n+1, 0) + \Lambda_2(n+1, n+1) \cdot \Phi^{(k-1)}(n+1, 0), \\
 & k = N_1 + 2, \dots, N.
 \end{aligned}$$

В этих соотношениях $P(n+1, n+1)$ – спектральная характеристика оператора дифференцирования по времени с учетом значения функции в начальный момент; $A(n+1, n+1)$, $H_1(n+1, n+1)$ и $H_2(n+1, n+1)$ – спектральные характеристики операторов A , H_1 и H_2 , определенных выражениями (6) и (8); $\Lambda_1(n+1, n+1)$ и $\Lambda_2(n+1, n+1)$ – спектральные характеристики операторов умножения на функции $\lambda_1(t)$ и $\lambda_2(t)$ соответственно; $\Phi^{(k)}(n+1, 0)$ – спектральные характеристики функций $\varphi^{(k)}(t, x)$, $k = 1, 2, \dots, N$. Все перечисленные спектральные характеристики определены относительно системы $\{e(i_0, i_1, \dots, i_n, t, x)\}_{i_0, i_1, \dots, i_n=0}^{\infty}$. Далее, $q(1, 0; t_0)$ – матрица-столбец значений функций базисной системы $\{q(i_0, t)\}_{i_0=0}^{\infty}$ в точке t_0 ; $\Phi_0(n, 0)$ – спектральная характеристика плотности вероятности $\varphi_0(x)$ начального состояния X_0 , определенная относительно системы $\{p(i_1, \dots, i_n, x)\}_{i_1, \dots, i_n=0}^{\infty}$.

Спектральная характеристика $\Phi(n+1, 0)$ плотности вероятности $\varphi(t, x)$, называемая также обобщенной характеристической

функцией [11, 12], выражается следующим образом ($\Phi(n+1, 0)$ – многомерная гиперстолбцовая матрица, образованная искомыми коэффициентами разложения $\varphi_{i_0 i_1 \dots i_n}$):

$$(13) \quad \Phi(n+1, 0) = \sum_{k=1}^N \Phi^{(k)}(n+1, 0).$$

В основе соотношений (9)–(12) лежат определения, форма представления и свойства спектральных характеристик функций и линейных операторов, подробно изложенные в [12, 17, 18]. Определение спектральной характеристики, аналогичной $H_1(n+1, n+1)$ и $H_2(n+1, n+1)$, дано в [10, 12]. В этих же работах приведено представление спектральной характеристики $A(n+1, n+1)$ с помощью спектральных характеристик операторов дифференцирования и умножения.

Система уравнений (9)–(12) – это система линейных матричных уравнений относительно неизвестных спектральных характеристик $\Phi^{(k)}(n+1, 0)$, или система линейных неоднородных алгебраических уравнений относительно элементов матриц $\Phi^{(k)}(n+1, 0)$, т.е. коэффициентов разложения функций $\varphi^{(k)}(t, x)$ по функциям $\{e(i_0, i_1, \dots, i_n, t, x)\}_{i_0, i_1, \dots, i_n=0}^{\infty}$. Перейдем к решению этой системы.

Из уравнений (10)–(12) следует, что

$$\begin{aligned} & (P(n+1, n+1) - A(n+1, n+1) + \Lambda_1(n+1, n+1)) \cdot \Phi^{(k)}(n+1, 0) = \\ & = \Lambda_1(n+1, n+1) \cdot \Phi^{(k-1)}(n+1, 0), \quad k = 2, \dots, N_1, \\ & (P(n+1, n+1) - A(n+1, n+1) + \Lambda_2(n+1, n+1)) \cdot \\ & \cdot \Phi^{(N_1-1)}(n+1, 0) = \Lambda_2(n+1, n+1) \cdot \Phi^{(1)}(n+1, 0), \\ & (P(n+1, n+1) - A(n+1, n+1) + \Lambda_2(n+1, n+1)) \cdot \Phi^{(k)}(n+1, 0) = \\ & = \Lambda_2(n+1, n+1) \cdot \Phi^{(k-1)}(n+1, 0), \quad k = N_1 + 2, \dots, N, \end{aligned}$$

т.е.

$$\begin{aligned} & \Phi^{(k-1)}(n+1, 0) = \Lambda_1^{-1}(n+1, n+1) \cdot (P(n+1, n+1) - \\ & - A(n+1, n+1) + \Lambda_1(n+1, n+1)) \cdot \Phi^{(k)}(n+1, 0), \quad k = 2, \dots, N_1, \end{aligned}$$

$$\begin{aligned}\Phi^{(1)}(n+1, 0) &= \Lambda_2^{-1}(n+1, n+1) \cdot (P(n+1, n+1) - A(n+1, n+1) + \\ &+ \Lambda_2(n+1, n+1)) \cdot \Phi^{(N_1-1)}(n+1, 0), \\ \Phi^{(k-1)}(n+1, 0) &= \Lambda_2^{-1}(n+1, n+1) \cdot (P(n+1, n+1) - \\ &- A(n+1, n+1) + \Lambda_2(n+1, n+1)) \cdot \Phi^{(k)}(n+1, 0), \quad k = N_1 + 2, \dots, N,\end{aligned}$$

или кратко

$$\begin{aligned}\Phi^{(k-1)}(n+1, 0) &= W_1(n+1, n+1) \cdot \Phi^{(k)}(n+1, 0), \quad k = 2, \dots, N_1, \\ \Phi^{(1)}(n+1, 0) &= W_2(n+1, n+1) \cdot \Phi^{(N_1+1)}(n+1, 0), \\ \Phi^{(k-1)}(n+1, 0) &= W_2(n+1, n+1) \cdot \Phi^{(k)}(n+1, 0), \quad k = N_1 + 2, \dots, N,\end{aligned}$$

где

$$\begin{aligned}W_1(n+1, n+1) &= \Lambda_1^{-1}(n+1, n+1) \cdot (P(n+1, n+1) - \\ &- A(n+1, n+1) + \Lambda_1(n+1, n+1)), \\ W_2(n+1, n+1) &= \Lambda_2^{-1}(n+1, n+1) \cdot (P(n+1, n+1) - \\ &- A(n+1, n+1) + \Lambda_2(n+1, n+1)).\end{aligned}$$

Таким образом,

$$\begin{aligned}\Phi^{(1)}(n+1, 0) &= W_1^{k-1}(n+1, n+1) \cdot \Phi^{(k)}(n+1, 0), \\ \Phi^{(k)}(n+1, 0) &= W_1^{N_1-k}(n+1, n+1) \cdot \Phi^{(N_1)}(n+1, 0), \quad k = 1, \dots, N_1, \\ \Phi^{(1)}(n+1, 0) &= W_2^{k-N_1}(n+1, n+1) \cdot \Phi^{(k)}(n+1, 0), \\ \Phi^{(k)}(n+1, 0) &= W_2^{N-k}(n+1, n+1) \cdot \Phi^{(N)}(n+1, 0), \quad k = N_1 + 1, \dots, N,\end{aligned}$$

в частности

$$(14) \quad \Phi^{(1)}(n+1, 0) = W_1^{N_1-1}(n+1, n+1) \cdot \Phi^{(N_1)}(n+1, 0),$$

$$(15) \quad \Phi^{(1)}(n+1, 0) = W_2^{N_2-1}(n+1, n+1) \cdot \Phi^{(N)}(n+1, 0).$$

Значит,

$$(16) \quad \begin{aligned}W_1^{N_1-1}(n+1, n+1) \cdot \Phi^{(N_1)}(n+1, 0) &= \\ = W_2^{N_2-1}(n+1, n+1) \cdot \Phi^{(N)}(n+1, 0),\end{aligned}$$

т.е.

$$(17) \quad \begin{aligned}\Phi^{(N_1)}(n+1, 0) &= [W_1^{N_1-1}(n+1, n+1)]^{-1} \cdot \\ \cdot W_2^{N_2-1}(n+1, n+1) \cdot \Phi^{(N)}(n+1, 0),\end{aligned}$$

или

$$(18) \quad \Phi^{(N)}(n+1, 0) = [W_2^{N_2-1}(n+1, n+1)]^{-1} \cdot \\ \cdot W_1^{N_1-1}(n+1, n+1) \cdot \Phi^{(N_1)}(n+1, 0).$$

Перепишем уравнение (10) с учетом (14)–(16):

$$\begin{aligned} & (\Lambda_1(n+1, n+1) \cdot W_1^{N_1}(n+1, n+1) + \Lambda_2(n+1, n+1) \cdot \\ & \cdot W_1^{N_1-1}(n+1, n+1) - H_1(n+1, n+1) - H_2(n+1, n+1) \cdot \\ & \cdot [W_2^{N_2-1}(n+1, n+1)]^{-1} \cdot W_1^{N_1-1}(n+1, n+1)) \cdot \Phi^{(N_1)}(n+1, 0) = \\ & = q(1, 0; t_0) \otimes \Phi_0(n, 0), \end{aligned}$$

или

$$\begin{aligned} & (\Lambda_2(n+1, n+1) \cdot W_2^{N_2}(n+1, n+1) + \Lambda_1(n+1, n+1) \cdot \\ & \cdot W_2^{N_2-1}(n+1, n+1) - H_2(n+1, n+1) - H_1(n+1, n+1) \cdot \\ & \cdot [W_1^{N_1-1}(n+1, n+1)]^{-1} \cdot W_2^{N_2-1}(n+1, n+1)) \cdot \Phi^{(N)}(n+1, 0) = \\ & = q(1, 0; t_0) \otimes \Phi_0(n, 0), \end{aligned}$$

следовательно,

$$\begin{aligned} \Phi^{(N_1)}(n+1, 0) &= (\Lambda_1(n+1, n+1) \cdot W_1^{N_1}(n+1, n+1) - H_1(n+1, n+1) + \\ & + \{\Lambda_2(n+1, n+1) - H_2(n+1, n+1) \cdot [W_2^{N_2-1}(n+1, n+1)]^{-1}\} \cdot \\ & \cdot W_1^{N_1-1}(n+1, n+1))^{-1} \cdot (q(1, 0; t_0) \otimes \Phi_0(n, 0)), \end{aligned}$$

или

$$\begin{aligned} \Phi^{(N)}(n+1, 0) &= (\Lambda_2(n+1, n+1) \cdot W_2^{N_2}(n+1, n+1) - H_2(n+1, n+1) + \\ & + \{\Lambda_1(n+1, n+1) - H_1(n+1, n+1) \cdot [W_1^{N_1-1}(n+1, n+1)]^{-1}\} \cdot \\ & \cdot W_2^{N_2-1}(n+1, n+1))^{-1} \cdot (q(1, 0; t_0) \otimes \Phi_0(n, 0)). \end{aligned}$$

Выразим спектральную характеристику $\Phi(n+1, 0)$, принимая во внимание (13):

$$\begin{aligned}
 \Phi(n+1, 0) &= \sum_{k=1}^{N_1} \Phi^{(k)}(n+1, 0) + \sum_{k=N_1+1}^N \Phi^{(k)}(n+1, 0) = \\
 (19) \quad &= \sum_{k=1}^{N_1} W_1^{N_1-k}(n+1, n+1) \cdot \Phi^{(N_1)}(n+1, 0) + \\
 &+ \sum_{k=1}^{N_2-1} W_2^{N_2-k-1}(n+1, n+1) \cdot \Phi^{(N)}(n+1, 0).
 \end{aligned}$$

В [10] было показано, что

$$\begin{aligned}
 (20) \quad \sum_{k=1}^N W^{N-k}(n+1, n+1) &= \left(E(n+1, n+1) - W^N(n+1, n+1) \right) \cdot \\
 &\cdot \left(E(n+1, n+1) - W(n+1, n+1) \right)^{-1},
 \end{aligned}$$

поэтому с учетом (17)

$$\begin{aligned}
 \Phi(n+1, 0) &= \left[\left(E(n+1, n+1) - W_1^{N_1}(n+1, n+1) \right) \cdot \right. \\
 &\cdot \left(E(n+1, n+1) - W_1(n+1, n+1) \right)^{-1} + \\
 &+ \left(E(n+1, n+1) - W_2^{N_2-1}(n+1, n+1) \right) \cdot \\
 &\cdot \left(E(n+1, n+1) - W_2(n+1, n+1) \right)^{-1} \cdot \\
 &\left. \cdot \left[W_2^{N_2-1}(n+1, n+1) \right]^{-1} \cdot W_1^{N_1-1}(n+1, n+1) \right] \cdot \Phi^{(N_1)}(n+1, 0),
 \end{aligned}$$

или с учетом (18)

$$\begin{aligned}
 \Phi(n+1, 0) &= \left[\left(E(n+1, n+1) - W_1^{N_1}(n+1, n+1) \right) \cdot \right. \\
 &\cdot \left(E(n+1, n+1) - W_1(n+1, n+1) \right)^{-1} \cdot \left[W_1^{N_1-1}(n+1, n+1) \right]^{-1} \cdot \\
 &\cdot W_2^{N_2-1}(n+1, n+1) + \left(E(n+1, n+1) - W_2^{N_2-1}(n+1, n+1) \right) \cdot \\
 &\left. \cdot \left(E(n+1, n+1) - W_2(n+1, n+1) \right)^{-1} \right] \cdot \Phi^{(N)}(n+1, 0).
 \end{aligned}$$

Таким образом,

$$\begin{aligned}
 \Phi(n+1, 0) = & \left[\left(E(n+1, n+1) - W_1^{N_1}(n+1, n+1) \right) \cdot \right. \\
 & \cdot \left(E(n+1, n+1) - W_1(n+1, n+1) \right)^{-1} + \left(E(n+1, n+1) - \right. \\
 & \left. - W_2^{N_2-1}(n+1, n+1) \right) \cdot \left(E(n+1, n+1) - W_2(n+1, n+1) \right)^{-1} \cdot \\
 (21) \cdot & \left. \left[W_2^{N_2-1}(n+1, n+1) \right]^{-1} \cdot W_1^{N_1-1}(n+1, n+1) \right] \cdot \left(\Lambda_1(n+1, n+1) \cdot \right. \\
 & \cdot W_1^{N_1}(n+1, n+1) - H_1(n+1, n+1) + \{ \Lambda_2(n+1, n+1) - \\
 & \left. - H_2(n+1, n+1) \cdot \left[W_2^{N_2-1}(n+1, n+1) \right]^{-1} \} \cdot \right. \\
 & \left. \cdot W_1^{N_1-1}(n+1, n+1) \right)^{-1} \cdot (q(1, 0; t_0) \otimes \Phi_0(n, 0)),
 \end{aligned}$$

или

$$\begin{aligned}
 \Phi(n+1, 0) = & \left[\left(E(n+1, n+1) - W_1^{N_1}(n+1, n+1) \right) \cdot \right. \\
 & \cdot \left(E(n+1, n+1) - W_1(n+1, n+1) \right)^{-1} \cdot \left[W_1^{N_1-1}(n+1, n+1) \right]^{-1} \cdot \\
 & \cdot W_2^{N_2-1}(n+1, n+1) + \left(E(n+1, n+1) - W_2^{N_2-1}(n+1, n+1) \right) \cdot \\
 (22) \cdot & \left. \left(E(n+1, n+1) - W_2(n+1, n+1) \right)^{-1} \right] \cdot \left(\Lambda_2(n+1, n+1) \cdot \right. \\
 & \cdot W_2^{N_2}(n+1, n+1) - H_2(n+1, n+1) + \{ \Lambda_1(n+1, n+1) - \\
 & \left. - H_1(n+1, n+1) \cdot \left[W_1^{N_1-1}(n+1, n+1) \right]^{-1} \} \cdot \right. \\
 & \left. \cdot W_2^{N_2-1}(n+1, n+1) \right)^{-1} \cdot (q(1, 0; t_0) \otimes \Phi_0(n, 0)),
 \end{aligned}$$

– решение рассматриваемой задачи в спектральной форме математического описания.

Для нахождения решения задачи анализа в пространстве функций времени и вектора состояния требуется применить формулу обращения [12]:

$$\varphi(t, x) = \sum_{i_0=0}^{\infty} \sum_{i_1=0}^{\infty} \dots \sum_{i_n=0}^{\infty} \varphi_{i_0 i_1 \dots i_n} \cdot e(i_0, i_1, \dots, i_n, t, x), \quad (t, x) \in T \times R^n.$$

Обычно приближенно определяется конечное число коэффициентов $\varphi_{i_0 i_1 \dots i_n}$, поскольку задача нахождения всех коэффициентов разложения не является тривиальной. В этом случае бесконечные матрицы в (9)–(13) заменяются конечными матрицами, тогда

$$\varphi(t, x) \approx \sum_{i_0=0}^{L_0-1} \sum_{i_1=0}^{L_1-1} \dots \sum_{i_n=0}^{L_n-1} \varphi_{i_0 i_1 \dots i_n} \cdot e(i_0, i_1, \dots, i_n, t, x),$$

где натуральные числа L_0, L_1, \dots, L_n – выбранные порядки усечения спектральных характеристик, влияющие на точность решения. Проверка корректности расчетов обычно осуществляется путем сравнения результатов при использовании различных наборов базисных систем или усечений спектральных характеристик [18].

Для апробации разработанного метода рассмотрим стохастическую систему, заданную уравнением Ито

$$dX(t) = a \operatorname{sign} X(t) dt + \sigma dW(t) + dQ(t), \quad X(0) = X_0,$$

где $X \in R$ – состояние системы, $t \in [0, 1]$; a, σ – числовые параметры; $W(t)$ – скалярный стандартный винеровский процесс; $Q(t)$ – скачкообразный процесс, определяемый интенсивностями λ_1, λ_2 , порядками N_1, N_2 и плотностями вероятности $q_1(t, y), q_2(t, y)$.

Для численных расчетов положим $a = -0,5$; $\sigma = 0,75$; начальное состояние X_0 имеет нормальное распределение с математическим ожиданием $m_0 = 1$ и дисперсией $D_0 = 0,5$; интенсивности и порядки заданы следующим образом: $\lambda_1 = 6$; $\lambda_2 = 1,5$, $N_1 = N_2 = 2$; плотность распределения $q_1(t, y)$ приращения Y_1 – гауссовская с параметрами $m_1 = 0,5$ и $D_1 = 0,1$, $q_2(t, y) = \delta(y + 0,2)$, т.е. величина приращения Y_2 постоянна и равна $-0,2$.

При решении задачи спектральным методом в качестве базисной системы $\{q(i_0, t)\}_{i_0=0}^{\infty}$ были выбраны косинусоиды [12, 17, 18], а в качестве базисной системы $\{p(i_1, x)\}_{i_1=0}^{\infty}$ – функции Эрмита [14] с параметрами $m = 1$ и $D = 1$, порядки усечения $L_0 = L_1 = 16$.

Результаты вычислений для плотности вероятности состояния системы приведены на рис. 3, а на рис. 4 показаны ее сечения в моменты времени 0,0; 0,5; 1,0.

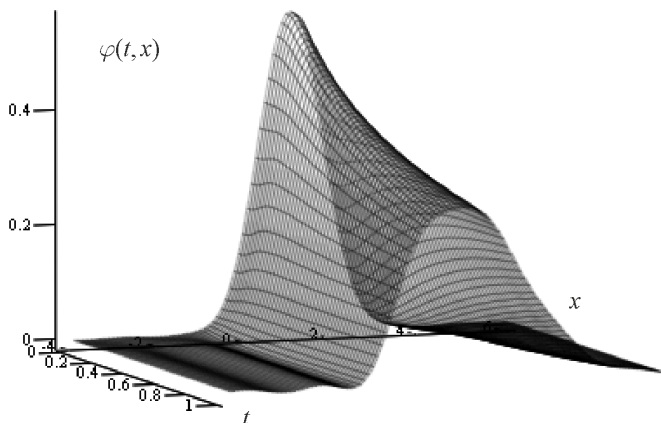


Рис. 3. Плотность вероятности состояния системы

Далее, на рис. 5 показано, как меняется плотность вероятности в конечном моменте времени в зависимости от порядка N_1 (дополнительные расчеты проводились для случаев $N_1 = 1$ и $N_1 = 3$, при этом интенсивность λ_1 определялась из условия $\lambda_1/N_1 = \text{const}$, которое обеспечивает постоянство среднего числа скачков в единицу времени для эрланговского закона с параметрами λ_1 и N_1 , т.е. $N_1 = 1, \lambda_1 = 3$ и $N_1 = 3, \lambda_1 = 9$). Видно, что максимальное значение для плотности вероятности и математическое ожидание состояния смещаются вправо. Это объясняется ростом вероятности выбора приращения Y_1 , а не Y_2 : при $\lambda_1 = 3$ эта вероятность равна $3/(3 + 1,5) = 2/3$, при $\lambda_1 = 6 - 6/(6 + 1,5) = 4/5$ и при $\lambda_1 = 9 - 9/(9 + 1,5) = 6/7$.

Асимметрию можно объяснить положительной средней величиной приращения Y_1 , которая по модулю превосходит детерминированное приращение Y_2 , и, конечно, соотношением интенсивностей λ_1, λ_2 .

Наконец, на рис. 6 показано, как зависит плотность вероятности в конечном моменте времени от различных усечений спектральных характеристик (дополнительные расчеты проводились для случаев $L_0 = L_1 = 8$ и $L_0 = L_1 = 24$ при номинальных параметрах $N_1 = 2, \lambda_1 = 6$).

Как следует из графиков, при выбранных порядках усечения (8; 16; 24) разница в приближенных решениях не столь заметна, однако при другом масштабе видно (см. рис. 7), что с ростом усечения растет и точность. В первую очередь это проявляется в минимизации интервалов, где оценка плотности вероятности отрицательна (появление таких интервалов при приближенном решении неизбежно и вызвано свойствами выбранных базисных функций).

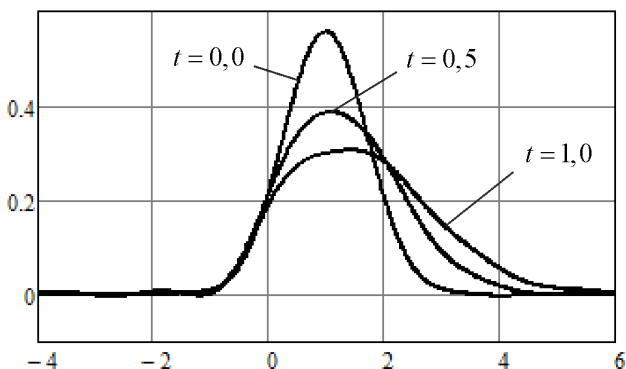


Рис. 4. Сечения плотности вероятности состояния системы в различные моменты времени

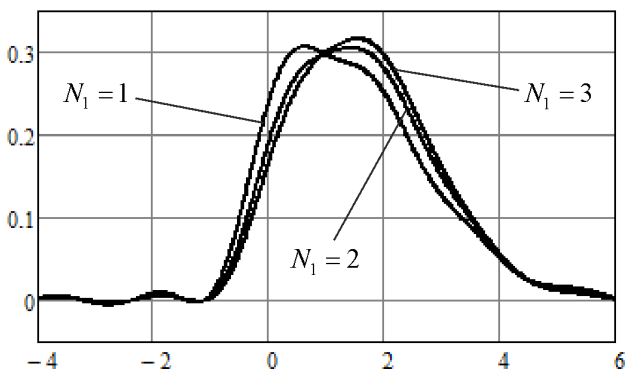


Рис. 5. Сечения плотности вероятности состояния системы в конечном моменте времени при различных значениях параметров λ_1 и N_1

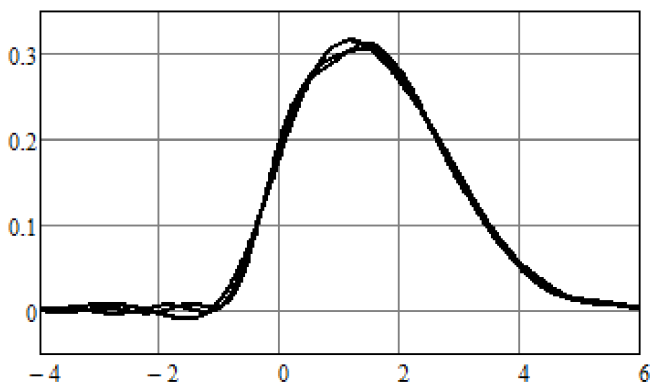


Рис. 6. Сечения плотности вероятности состояния системы в конечном моменте времени при различных порядках усечения

Отметим, что все численные расчеты были выполнены с помощью специализированного программного обеспечения Spectrum [15].

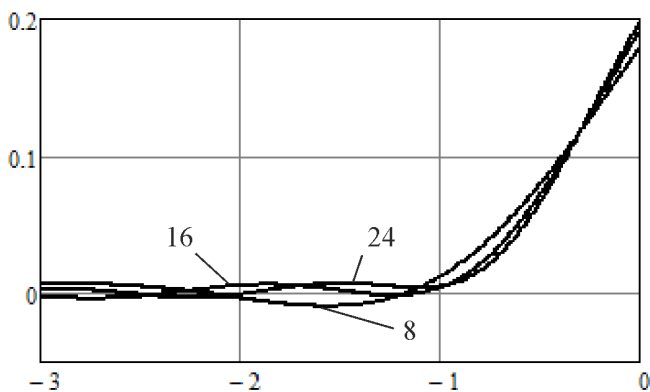


Рис. 7. Сечения плотности вероятности состояния системы в конечном моменте времени при различных порядках усечения

4. Заключение

В работе рассмотрено применение спектральной формы математического описания к задаче вероятностного анализа стохастических систем, которые характеризуются наличием разрывов (скачков) траекторий, образующих гиперэрланговский поток событий. Получены соотношения для нахождения плотности вероятности вектора состояния в спектральной форме математического описания систем управления. Использование гиперэрланговских потоков дает возможность учитывать более сложный характер поведения траекторий вектора состояния, а применение спектральной формы математического описания позволяет свести систему интегро-дифференциальных уравнений к системе линейных алгебраических уравнений относительно коэффициентов разложения искомой плотности вероятности по некоторой ортонормированной системе функций.

Рассмотренный подход упрощает процесс решения задачи анализа, делая его удобным для применения современных высокопроизводительных вычислительных систем.

Литература

1. АВЕРИНА Т.А. *Модифицированный алгоритм статистического моделирования систем со случайным периодом квантования* // Вестник СГТУ. – 2011. – №4(62), Вып. 4. – С. 212–218.
2. АВЕРИНА Т.А., РЫБАКОВ К.А. *Новые методы анализа воздействия пуассоновских дельта-импульсов в задачах радиотехники* // Журнал радиоэлектроники. – 2013. – №1. – [Электронный ресурс]. – URL: <http://jre.cplire.ru/jre/content.html> (дата обращения: 09.09.2013).
3. АНУЛОВА С.В., ВЕРЕТЕННИКОВ А.Ю., КРЫЛОВ Н.В. И ДР. *Стохастическое исчисление* // Итоги науки и техники. Современные проблемы математики. Фундаментальные направления. Т. 45. – М.: ВИНТИ, 1989. – С. 42–79.

4. АРТЕМЬЕВ В.М., ИВАНОВСКИЙ А.В. *Дискретные системы управления со случайным периодом квантования*. – М.: Энергоатомиздат, 1986. – 96 с.
5. ВЕРЕТЕННИКОВ А.Ю. *О сильных решениях стохастических уравнений Ито со скачками* // Теория вероятностей и ее применения. – 1987. – Т. 32., Вып. 1. – С. 159–163.
6. ГОРИЦКИЙ Ю.А., КАЗАКОВ В.А. *Дискретизация случайных процессов с конечным множеством состояний и эрланговским временем пребывания* // Известия РАН. Теория и системы управления. – 2011. – №6. – С. 14–27.
7. КАЗАКОВ И.Е., АРТЕМЬЕВ В.М., БУХАЛЕВ В.А. *Анализ систем случайной структуры*. – М.: Физматлит, 1993. – 272 с.
8. КОЖЕВНИКОВ А.С. *Программное обеспечение для статистического моделирования и анализа случайных процессов со скачками, описывающих динамику цен акций предприятий авиационной отрасли* // Труды МАИ. – 2012. – № 59. – [Электронный ресурс]. – URL: <http://www.mai.ru/science/trudy> (дата обращения: 09.09.2013).
9. КОЖЕВНИКОВ А.С. *Математические модели динамики цены акций с гиперэрланговскими скачками* // Научный альманах. Вып. 17: Материалы IX научно-практической конференции молодых ученых и студентов «Инновационный менеджмент в аэрокосмической промышленности». – М.: Изд-во «Доброе слово», 2013. – С. 180–186.
10. КОЖЕВНИКОВ А.С., РЫБАКОВ К.А. *Анализ нелинейных стохастических систем управления с импульсными воздействиями, образующими эрланговские потоки событий* // Научный вестник МГТУ ГА. – 2012. – №184(10). – С. 37–45.
11. ПАНТЕЛЕЕВ А.В., РЫБАКОВ К.А. *Анализ нелинейных стохастических систем управления в классе обобщенных характеристических функций* // Автоматика и телемеханика. – 2011. – №2. – С. 183–194.

12. ПАНТЕЛЕЕВ А.В., РЫБАКОВ К.А., СОТСКОВА И.Л. *Спектральный метод анализа нелинейных стохастических систем управления*. – М.: Вузовская книга, 2006. – 392 с.
13. ПУГАЧЕВ В.С., СИНИЦЫН И.Н. *Стохастические дифференциальные системы. Анализ и фильтрация*. – М.: Наука, 1990. – 632 с.
14. РОМАНОВ В.А., РЫБАКОВ К.А. *Спектральные характеристики операторов умножения, дифференцирования и интегрирования в базисе обобщенных функций Эрмита // Труды МАИ*. – 2010. – №39. – [Электронный ресурс]. – URL: <http://www.mai.ru/science/trudy> (дата обращения: 09.09.2013)
15. РЫБАКОВ К.А. *Программное обеспечение спектрального метода Spectrum // Труды МАИ*. – 2003. – №14. – [Электронный ресурс]. – URL: <http://www.mai.ru/science/trudy> (дата обращения: 09.09.2013).
16. РЫБАКОВ К.А. *Спектральные характеристики операторов умножения, дифференцирования и интегрирования в базисе обобщенных функций Лагерра // Дифференциальные уравнения и процессы управления*. – 2012. – №1. – С. 114–141. – [Электронный ресурс]. – URL: <http://www.math.spbu.ru/diffjournal> (дата обращения: 09.09.2013).
17. РЫБИН В.В. *Моделирование нестационарных непрерывно-дискретных систем управления спектральным методом в системах компьютерной математики*. – М.: Изд-во МАИ, 2011. – 220 с.
18. СОЛОДОВНИКОВ В.В., СЕМЕНОВ В.В. *Спектральная теория нестационарных систем управления*. – М.: Наука, 1974. – 336 с.
19. ТИХОНОВ В.И., МИРОНОВ М.А. *Марковские процессы*. – М.: Советское радио, 1977. – 488 с.
20. ФЕДОСОВ Е.А., ИНСАРОВ В.В., СЕЛИВОХИН О.С. *Системы управления конечным положением в условиях противодействия среды*. – М.: Наука, 1989. – 272 с.

21. CONT R., TANKOV P. *Financial Modelling with Jump Processes*. – Chapman and Hall, 2004. – 560 p.
22. IWANKIEWICZ R., NIELSEN S.R.K. *Advanced Methods in Stochastic Dynamics of Non-Linear Systems*. – Aalborg tekniske Universitetsforlag, 1999. – 276 p.
23. HANSON F.B. *Applied Stochastic Processes and Control for Jump-Diffusions: Modeling, Analysis, and Computation*. – SIAM, 2007. – 472 p.
24. ØKSENDAL B., SULEM A. *Applied Stochastic Control of Jump Diffusions*. – Springer, 2007. – 266 p.
25. TELLIER M., IWANKIEWICZ R. *Response of linear dynamic systems to non-Erlang renewal impulses: stochastic equation approach* // Probabilistic Engineering Mechanics. – 2005. – Vol. 20, №4. – P. 281–295.

**SPECTRAL METHOD FOR STOCHASTIC SYSTEMS
WITH DISCONTINUOUS TRAJECTORIES DESCRIBED
BY RANDOM MIXTURE OF ERLANG DISTRIBUTIONS**

Alexander Kozhevnikov, Moscow aviation institute, graduate student (exequit@yandex.ru).

Konstantin Rybakov, Moscow aviation institute, Cand.Sc., assistant professor (rkoffice@mail.ru).

Abstract: We consider stochastic control systems with impulses generated by hyper-Erlang flows of events leading to jumps in system trajectories. We employ the spectral form of mathematical description of the system to find the probability density function of its states.

Keywords: analysis problem, hyper-Erlang distribution, jump process, spectral method, stochastic system.

*Статья представлена к публикации
членом редакционной коллегии А.П. Курдюковым*

Поступила в редакцию 17.01.2013.

Опубликована 30.09.2013.

УДК 517.977.1
ББК 3 965-01я7

КВАЗИКАНОНИЧЕСКИЙ ВИД УРАВНЕНИЙ, ОПИСЫВАЮЩИХ ДВИЖЕНИЕ ВЕРТОЛЕТА¹

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

Для уравнений, описывающих движение вертолета, находятся индексы приводимости к регулярному квазиканоническому виду. Выполняется преобразование к квазиканоническому виду, анализируется подсистема, определяющая нулевую динамику. Находятся выходы, относительно которых система является минимально-фазовой.

Ключевые слова: квазиканонический вид, нелинейные динамические системы, нулевая динамика.

Введение

Беспилотные летательные аппараты (БПЛА) являются важным объектом приложения результатов теории управления. Это связано, в первую очередь, с их возрастающим количеством и все более важными ролями, которые приходится выполнять беспилотным аппаратам. В настоящий момент применяются БПЛА различных форм, размеров и компоновок. Выбор конструкции обусловлен решаемыми задачами и условиями применения. Наряду с другими, используется традиционная вертолетная схема с двумя винтами: главным и хвостовым.

При внедрении БПЛА возникает необходимость их оснащения автоматическими системами управления для обеспечения

¹ Работа выполнена при финансовой поддержке РФФИ, грант №12-07-00329-а, и Министерства образования и науки Российской Федерации (соглашение 14.В37.21.0370).

² Андрей Анатольевич Шевляков (aash29@gmail.com).

действий в автономном и полуавтономном режиме. Таким образом, актуальной является задача синтеза алгоритмов управления для динамической системы.

Один из подходов к решению задачи управления нелинейной динамической системой основывается на преобразовании системы к специальному виду, для которого метод решения соответствующей задачи управления известен. Преимущество данного класса методов в том, что они позволяют избежать линеаризации рассматриваемой системы уравнений. Благодаря этому используемая математическая модель может быть более адекватна исследуемому явлению.

В работе [12] система уравнений, описывающих движение вертолета, приводится к нормальной форме [11] относительно некоторых заданных выходов. За счет этого траектория системы стабилизируется по части переменных.

Другим примером специального вида являются системы канонического вида [1], которые с помощью линеаризации обратной связью можно преобразовать в линейную систему, записанную в канонической форме Бруновского [9].

Условия приводимости к каноническому виду хорошо известны [1], однако не всякую аффинную систему можно преобразовать к этому виду. Поэтому среди аффинных систем выделяют системы, которые преобразуются к квазиканоническому виду [3]. Такие системы содержат подсистему, которая линеаризацией обратной связью преобразуется в каноническую форму Бруновского, и подсистему общего вида. Основные теоретические положения о преобразовании в некоторой открытой области аффинных систем со скалярным управлением к квазиканоническому виду приведены в [3], а с векторным управлением — в [2]. Вычисление максимального индекса приводимости к регулярному квазиканоническому виду для случая скалярного управления описано в [7].

При росте размерности исследуемой системы анализ рангов соответствующих распределений может представлять существенные технические трудности. Для решения возникающих задач

использована среда аналитических вычислений Maple и разработанная автором библиотека quasiPack.

Структура статьи следующая: в первом разделе приводятся известные по работам [1] и [13] результаты о преобразовании к квазиканоническому виду и вычислении индексов приводимости. В том же разделе приводится описанная в [10] математическая модель вертолета. В разделе 2 известные методики, описанные в вышеназванных работах, применяются для преобразования уравнений движения вертолета к квазиканоническому виду. Если при преобразовании системы к нормальной форме выходы задаются заранее, то проведенный анализ позволяет сказать, какие выходы и, следовательно, нормальные формы возможны для данной системы. Также можно утверждать, что максимальным индексом приводимости для рассматриваемой системы является $r = (3, 3, 3, 3)$, и преобразование к квазиканоническому виду позволяет стабилизировать положение равновесия по части переменных. В разделе 3 для решения задачи стабилизации положения равновесия применяется метод виртуальных выходов. Подробный анализ полученного управления и области стабилизации не проводится, цель данной работы состоит в том, чтобы показать применимость данного метода к рассматриваемой системе.

1. Квазиканонический вид системы

Рассмотрим аффинную систему с векторным управлением

$$(1) \quad \dot{x} = A(x) + \sum_{i=1}^m B_i(x)u_i,$$

где $x \in R^n$, $u = (u_1, \dots, u_m)^T \in R^m$, $A(0) = 0$, $A(x) = (a_1(x), \dots, a_n(x))^T$, $B(x) = (B_1(x), \dots, B_m(x))$, $B_j(x) = (b_j^1(x), \dots, b_j^n(x))^T$, $j = \overline{1, m}$, $i = \overline{1, n}$, $a_i(x), b_i^j(x) \in C^\infty(\Omega)$, Ω — открытое множество, содержащее положение равновесия $x = 0$.

Среди таких систем выделяют [3] системы специального вида (2)

$$\begin{aligned}
 \dot{z}_1^1 &= z_2^1, \\
 &\dots, \\
 \dot{z}_{r_1-1}^i &= z_{r_1}^1, \\
 \dot{z}_{r_1}^1 &= f_1(z, \eta) + g_{11}(z, \eta)u_1 + \dots + g_{1m}(z, \eta)u_m, \\
 &\dots, \\
 \dot{z}_1^m &= z_2^m, \\
 &\dots, \\
 \dot{z}_{r_m-1}^m &= z_{r_m}^m, \\
 \dot{z}_{r_m}^m &= f_m(z, \eta) + g_{m1}(z, \eta)u_1 + \dots + g_{mm}(z, \eta)u_m, \\
 \dot{\eta}_1 &= q_1(z, \eta) + p_{11}(z, \eta)u_1 + \dots + p_{1m}(z, \eta)u_m, \\
 &\dots, \\
 \dot{\eta}_{n-r} &= q_{n-r}(z, \eta) + p_{n-r, 1}(z, \eta)u_1 + \dots + p_{n-r, m}(z, \eta)u_m,
 \end{aligned}$$

где $z = (z_1, \dots, z_r)^T, z \in R^r, r = r_1 + \dots + r_m, \eta = (\eta_1, \dots, \eta_{n-r})^T, \eta \in R^{n-r}$. Такой вид называют [2, 3] квазиканоническим видом системы (1), а упорядоченный набор (r_1, \dots, r_m) — мультииндексом приводимости.

Теорема 1. [2]. Для того чтобы для аффинной стационарной системы (1) в области Ω существовали переменные, в которых она имеет квазиканонический вид, необходимо и достаточно существование таких функций $\phi_i(x) \in C^\infty(\Omega), i = \overline{1, m}$, которые в Ω являются решением системы уравнений в частных производных первого порядка

$$(3) \quad \text{ad}_A^k B_j \phi_i = 0, \quad k = \overline{0, r_i - 2}, \quad i, j = \overline{1, m},$$

и для которых соотношения

$$\begin{aligned}
 z_k^i &= A^{k-1} \phi_i(x), \quad k = \overline{1, r_i}, \quad i = \overline{1, m}, \\
 \eta_j &= \psi_j(x), \quad j = \overline{1, n-r}
 \end{aligned}$$

задают в Ω гладкую невырожденную замену переменных.

Хотя теорема (1) имеет характер теоремы существования, она является конструктивной в том смысле, что при известном мультииндексе приводимости (r_1, \dots, r_m) решение системы (3), проверка возможности получить с использованием найденных функций $\phi_i, i = \overline{1, m}$, замены по части переменных z , а также

выбор координатных функций $\eta_j = \Psi_i(x)$, $j = \overline{1, n-r}$, представляют только технические трудности. Основной проблемой является определение мультииндекса приводимости.

1.1. ВЫЧИСЛЕНИЕ ИНДЕКСОВ ПРИВОДИМОСТИ

Семейство гладких векторных полей X_1, \dots, X_k , определенных в области Ω , порождает в Ω гладкое распределение $F = span(X_1, \dots, X_k)$, где линейная оболочка векторных полей рассматривается над кольцом гладких функций [1]. В каждой точке $x \in \Omega$ $F(x) = span(X_1(x), \dots, X_k(x))$ есть линейное подпространство касательного пространства, порожденное векторами $X_1(x), \dots, X_k(x)$. Под размерностью распределения F в точке x понимают размерность соответствующего линейного подпространства $F(x)$. Распределение называют регулярным в Ω , если его размерность постоянна во всех точках $x \in \Omega$. Для проверки регулярности распределения составляют функциональную матрицу W , по столбцам которой записывают координатные функции векторных полей X_i , $i = 1, \dots, k$, и исследуют в Ω ее ранг. Предположим, что распределение

$$\mathfrak{B} = span(B_1, \dots, B_m)$$

в окрестности точки x^0 имеет размерность m . В противном случае для анализа необходимо выбрать другую точку, в которой указанное условие выполнено. Пусть

$$\Delta_1 = span(X_1, \dots, X_k), \Delta_2 = span(Y_1, \dots, Y_m).$$

Распределение

$$\Delta_3 = span(X_1, \dots, X_k, Y_1, \dots, Y_m)$$

будем записывать в виде

$$\Delta_3 = span(\Delta_1, \Delta_2).$$

Для распределения

$$\Delta_4 = span([X_i, Y_j], i = \overline{1, k}, j = \overline{1, m})$$

будем использовать обозначение

$$\Delta_4 = [\Delta_1, \Delta_2].$$

Введем следующие обозначения:

$$\begin{aligned} \mathfrak{B}_A &= A + \mathfrak{B} = (A + B : B \in \mathfrak{B}), \\ [\Delta, \Gamma] &= \text{span}([X, Y] : X \in \Delta, Y \in \Gamma), \\ \text{ad}_A^l \mathfrak{C} &= \text{span}(\text{ad}_A^l X : X \in \mathfrak{C}). \end{aligned}$$

Рассмотрим систему (1). Предположим, что распределение \mathfrak{B} имеет в Ω постоянную размерность m . В соответствие системе (1) можно поставить набор распределений

$$(4) \quad \begin{aligned} \mathfrak{B}^0 &= \mathfrak{B}, \\ \mathfrak{B}^1 &= \text{span}(\mathfrak{B}, [\mathfrak{B}_A, \mathfrak{B}]), \\ &\dots \\ \mathfrak{B}^j &= \text{span}(\mathfrak{B}^{j-1}, [\mathfrak{B}_A, \mathfrak{B}^{j-1}]), \\ \mathfrak{L}^0 &= \mathfrak{B}, \\ \mathfrak{L}^1 &= \text{span}(\text{ad}_A^1 \mathfrak{B}, \overline{\mathfrak{B}}), \\ &\dots \\ \mathfrak{L}^k &= \text{span}(\text{ad}_A^k \mathfrak{B}, \overline{\mathfrak{B}^{k-1}}). \end{aligned}$$

Далее всюду будем предполагать, что распределения (4) имеют постоянную размерность в окрестности рассматриваемой точки x^0 . Рассмотрим следующий набор чисел n_i :

$$\begin{aligned} n_0 &= \dim \mathfrak{B}^0 = m \\ n_1 &= \dim \mathfrak{L}^1 - \dim \overline{\mathfrak{B}^0}, \\ &\vdots \\ n_i &= \dim \mathfrak{L}^i - \dim \overline{\mathfrak{B}^{i-1}}, \\ &\vdots \end{aligned}$$

Данные числа показывают, насколько увеличивается размерность распределения $\overline{\mathfrak{B}^{i-1}}$ при пополнении его векторными полями, принадлежащими распределению $\text{ad}_A^i \mathfrak{B}$. В [13] показано, что последовательность n_0, \dots, n_i, \dots является невозрастающей. Ей можно поставить в соответствие набор чисел

$$(5) \quad \hat{r}_i = |n_j \geq i, j \geq 1|.$$

Полученные в [13] результаты сформулируем в виде следующей теоремы.

Теорема 2. Пусть векторное поле A является гладким в некоторой окрестности точки x^0 , распределение $\text{span}(B_1, \dots, B_m)$ в этой окрестности гладкое и имеет размерность m . Тогда

– существует окрестность точки x^0 , в которой система (1) эквивалентна регулярной системе квазиканонического вида с индексами приводимости \hat{r}_i , задаваемые соотношением (5);

– если (1) приводится к регулярному квазиканоническому виду с индексами r_1, \dots, r_m , то для каждого $i = 1, 2, \dots, m$ $r_i \leq \hat{r}_i$.

Для упрощения использования теоремы 2 нами был разработан следующий алгоритм. Пусть среди чисел r_i есть лишь q попарно различных. Обозначим их через $l_i, i = \overline{1, q}$, а через k_i – кратности соответствующих чисел. Тогда для нахождения индексов приводимости l_i и определения их кратности k_i удобно использовать специальную таблицу. Для заполнения второго столбца вычисляются числа

$$n_{i-1} = \dim \mathcal{L}^{i-1} - \dim \overline{\mathcal{B}^{i-2}}.$$

Вычисления производятся до тех пор, пока очередное число не будет равным нулю. Номер, соответствующий последнему ненулевому элементу второго столбца, равен значению l_1 , а его кратность k_1 равна последнему ненулевому значению n_i . Затем вычисляется третий столбец по формуле

$$d_{i2} = n_{i-1} - k_1.$$

Индекс последнего ненулевого значения во втором столбце равен l_2 , а само последнее ненулевое значение равно k_2 . Аналогично заполняются остальные столбцы таблицы до тех пор, пока очередной столбец не окажется состоящим целиком из нулевых значений. Тогда последнее ненулевое значение в каждом столбце равно k_i , а номер строки, в которой оно стоит равен l_i . Таким образом можно вычислить все различные индексы приводимости l_i и их кратности k_i . Данная таблица является обобщением на квазиканонический вид результатов, полученных в [3, 4].

Таблица 1. Вычисление индексов приводимости.

i	$n_{i-1} = \dim \mathfrak{L}^{i-1} - \dim \mathfrak{B}^{i-2}$	$d_{i2} = n_{i-1} - k_1$	$d_{i3} = d_{i2} - k_2$
1	m	$n_0 - k_1$	$n_0 - k_1 - k_2$
...	*	$n_1 - k_1$	$n_1 - k_1 - k_2$
...	*	k_2	*
...	*	0	0
...	*	0	0
s	k_1	0	0
$s + 1$	0	0	0

1.2. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ДВИЖЕНИЯ ВЕРТОЛЕТА

Рассмотрим модель движения вертолета, заданную уравнениями

$$\begin{aligned} \dot{r} &= V, \\ m\dot{V} &= F, \\ \dot{\Omega} &= D\omega_b, \\ J_b\dot{\omega}_b &= M_b - \omega_b \times J\omega_b, \end{aligned}$$

где $\bar{r} = (x, y, z)$ – радиус-вектор центра масс в земной системе координат, $\Omega = (\psi, \theta, \gamma)$ – углы Крылова ориентации вертолета,

$$\omega_b = \begin{pmatrix} \sin \theta & 0 & 1 \\ \cos \gamma \cos \theta & \sin \gamma & 0 \\ -\sin \gamma \cos \theta & \cos \gamma & 0 \end{pmatrix} \begin{pmatrix} \dot{\psi} \\ \dot{\theta} \\ \dot{\gamma} \end{pmatrix},$$

$$D = \begin{pmatrix} 0 & \frac{\sin \gamma}{\cos \theta} & \frac{\cos \gamma}{\cos \theta} \\ 0 & \cos \gamma & -\sin \gamma \\ 1 & -\frac{\sin \gamma \sin \theta}{\cos \theta} & -\frac{\cos \gamma \sin \theta}{\cos \theta} \end{pmatrix}.$$

Следуя [10], будем использовать следующие обозначения для сил и моментов, действующих на вертолет:

- F^{vf} – сила, создаваемая вертикальным стабилизатором;
- F^{ht} – сила, создаваемая горизонтальным стабилизатором;
- F^{fus} – сила, сопротивления фюзеляжа;
- M^m – реактивный момент, создаваемый несущим винтом;
- M^t – реактивный момент, создаваемый рулевым винтом;

F^{mr} – сила, создаваемая несущим винтом;
 F^{tr} – сила, создаваемая рулевым винтом;
 G – сила тяжести.

Пренебрежем ввиду малости силами F^{vf} и F^{ht} аэродинамического сопротивления, действующими на стабилизаторы. Матрица перехода из связанной системы координат в земную неподвижную систему координат имеет вид

$$R = \begin{pmatrix} \cos \theta \cos \psi & -\sin \theta \cos \psi \cos \gamma + \sin \gamma \sin \psi & \sin \theta \cos \psi \sin \gamma + \sin \psi \cos \gamma \\ \sin \theta & \cos \theta \cos \gamma & -\cos \theta \sin \gamma \\ -\cos \theta \sin \psi & \sin \theta \sin \psi \cos \gamma + \cos \psi \sin \gamma & -\sin \theta \sin \psi \sin \gamma + \cos \psi \cos \gamma \end{pmatrix}.$$

Результирующая сила F , действующая на центр масс вертолета, находится по формуле

$$F = R(F_b^{mr} + F_b^{tr}) + F^{fus} + G.$$

Вектор силы F^{mr} , создаваемой несущим винтом, перпендикулярен плоскости несущего винта. Значит, в системе координат, связанной с главным винтом вертолета,

$$F^{mr} = (0 \quad T^{mr} \quad 0)^T.$$

В связанной с корпусом системе координат F^{mr} имеет вид

$$(6) \quad F_b^{mr} = \begin{pmatrix} -\sin \alpha \cos \beta \\ \cos \alpha \cos \beta \\ \sin \beta \end{pmatrix} T^{mr}.$$

Компонентами F_x^{tr} и F_y^{tr} силы F^{tr} пренебрегают вследствие их малости [10], тогда сила F^{tr} , создаваемая рулевым винтом, в связанной системе координат будет иметь вид

$$F_b^{tr} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} T^{tr}.$$

Сила тяжести G , записанная в земной системе координат, имеет вид

$$G = - \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} mg.$$

Сила F^{fus} сопротивления фюзеляжа находится по формуле

$$F^{fus} = -\frac{1}{2}\rho\hat{V} \begin{pmatrix} c_x S_x^{fus} V_x \\ -c_y S_y^{fus} (V_y + V_{imr}) \\ c_z S_b^{fus} V_z \end{pmatrix},$$

где $\hat{V} = \sqrt{V_x^2 + V_z^2 + (V_y + V_{imr})^2}$. Главный момент M , создаваемый внешними силами, действующими на вертолет, записанный в связанной системе координат, рассчитывается по формуле

$$M = M_b^m + M_b^t + M_b^{mr} + M_b^{tr} + M_b^{fus}.$$

Согласно [12] реактивные моменты несущего и рулевого винтов можно найти по формулам

$$M_b^m = \begin{pmatrix} M_x^m \\ M_y^m \\ M_z^m \end{pmatrix} = - \begin{pmatrix} a_x^m \beta \\ 0 \\ a_z^m \alpha \end{pmatrix} - \begin{pmatrix} -\sin \alpha \cos \beta \\ \cos \alpha \cos \beta \\ \sin \beta \end{pmatrix} Q^m,$$

$$M_b^t = - \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} Q^t,$$

где величины Q^m и Q^t задаются соотношениями

$$Q^m = c^m (T^{mr})^{1,5} + d^m, Q^t = c^t (T^{tr})^{1,5} + d^t,$$

а $c^m, c^t, d^m, d^t, a_x^m, a_z^m$ являются постоянными и определяются для конкретной модели вертолета. Моменты $M_b^{mr}, M_b^{tr}, M_b^{fus}$ находятся по формулам

$$M_b^{mr} = r^{mr} \times F_b^{mr},$$

$$M_b^{tr} = r^{tr} \times F_b^{tr},$$

$$M_b^{fus} = r^{fus} \times F_b^{fus},$$

где r^{mr}, r^{tr}, r^{fus} – точки приложения соответствующих сил. Для того чтобы найти координаты сил $F_b^{vf}, F_b^{ht}, F_b^{fus}$, воспользуемся

матрицей R перехода из связанной с корпусом в земную систему координат. Получим

$$\begin{aligned} F_b^{ht} &= R^T F^{ht}, \\ F_b^{vf} &= R^T F^{vf}, \\ F_b^{fus} &= R^T F^{fus}. \end{aligned}$$

Чтобы рассматриваемая система стала аффинной, выберем в качестве управлений скорости изменения величин сил тяги, создаваемых рулевым \dot{T}^{tr} и несущим \dot{T}^{mr} винтами, а также скорости изменения углов наклона несущего винта $\dot{\alpha}$ и $\dot{\beta}$. Аффинная система, описывающая динамику полета вертолета, имеет вид

$$(7) \quad \begin{aligned} m\dot{V} &= F, \\ \dot{\Omega} &= D\omega_b, \\ J_b\dot{\omega}_b &= M_b - \omega_b \times J\omega_b, \\ \dot{T}^{mr} &= w_1, \\ \dot{T}^{tr} &= w_2, \\ \dot{\alpha} &= w_3, \\ \dot{\beta} &= w_4. \end{aligned}$$

Для того чтобы найти положение равновесия системы (7), а также силы и моменты, при которых вертолет неподвижно висит в воздухе, необходимо решить следующую систему

$$(8) \quad \begin{aligned} F &= 0, \\ D\omega_b &= 0, \\ M_b - \omega_b \times J\omega_b &= 0. \end{aligned}$$

Для упрощения зададим численные значения параметров вертолета. Эти параметры взяты из [10] и не приводятся здесь ввиду громоздкости. Одно из положений равновесия, найденное с использованием системы Matlab, задается следующими соотношениями:

$$\begin{aligned} r_x^0 &= 0, & V_x^0 &= 0, & \psi^0 &= 0, & \omega_{bx}^0 &= 0, \\ r_y^0 &= 0, & V_y^0 &= 0, & \theta^0 &\approx 0.2068, & \omega_{by}^0 &= 0, \\ r_z^0 &= 0, & V_z^0 &= 0, & \gamma^0 &\approx 0.0642, & \omega_{bz}^0 &= 0, \\ T^{mr0} &\approx 82,4436, & \alpha^0 &\approx -0,2073, \\ T^{tr0} &\approx 4,4828, & \beta^0 &\approx -0,1171. \end{aligned}$$

2. Преобразование уравнений движения вертолета к квазиканоническому виду

Сделаем следующую замену переменных

$$\begin{aligned} x_1 &= r_x, & x_4 &= V_x, & x_7 &= \psi, \\ x_2 &= r_y, & x_5 &= V_y, & x_8 &= \theta - \theta^0, \\ x_3 &= r_z, & x_6 &= V_z, & x_9 &= \gamma - \gamma^0, \\ x_{15} &= \alpha - \alpha^0, & x_{16} &= \beta - \beta^0, \\ x_{10} &= \omega_{bx}, & x_{13} &= T^{mr} - T^{mr0}, \\ x_{11} &= \omega_{by}, & x_{14} &= T^{tr} - T^{tr0}, \\ x_{12} &= \omega_{bz}. \end{aligned}$$

Данная замена является линейной и невырожденной в R^{16} . Матрицу R , функции $F_b^r = F_b^{mr} + F_b^{tr}$, F_b^{fus} , D , M_b , записанные в новых переменных x_i , обозначим \hat{R} , $\hat{F}_b^r = \hat{F}_b^{mr} + \hat{F}_b^{tr}$, \hat{F}_b^{fus} , \hat{D} , \hat{M}_b соответственно. Исследуемая система примет вид

$$\begin{aligned} \dot{x}_1 &= x_4, \\ \dot{x}_2 &= x_5, \\ \dot{x}_3 &= x_6, \\ \dot{x}_{\overline{4,6}} &= \frac{1}{m} \left[\hat{R}(x_7, x_8, x_9) \hat{F}_b^r(x_{13}, x_{14}, x_{15}, x_{16}) + \right. \\ &\quad \left. + \hat{F}_b^{fus}(x_4, x_5, x_6) + G \right], \\ (9) \quad \dot{x}_{\overline{7,9}} &= \hat{D}(x_7, x_8, x_9) x_{\overline{10,12}}, \\ \dot{x}_{\overline{10,12}} &= J_b^{-1} [\hat{M}_b(x_4, x_5, x_6, x_{13}, x_{14}, x_{15}, x_{16}) - \\ &\quad - x_{\overline{10,12}} \times J_b x_{\overline{10,12}}], \\ \dot{x}_{13} &= u_1, \\ \dot{x}_{14} &= u_2, \\ \dot{x}_{15} &= u_3, \\ \dot{x}_{16} &= u_4. \end{aligned}$$

Здесь через $x_{\overline{i,j}}$ обозначен вектор-столбец переменных $(x_i, x_{i+1}, \dots, x_j)^T$. Исследовав размерности распределений в окрестности точки $x = 0$, получим, что эти распределения регулярны. Далее вычислим $n_0 = \dim \mathfrak{G} = 4$, $n_1 = \dim \mathfrak{L}^1 - \dim \mathfrak{G}^0 = 4$, $n_2 = \dim \mathfrak{L}^2 - \dim \mathfrak{G}^1 = 4$, $n_3 = \dim \mathfrak{L}^3 - \dim \mathfrak{G}^2 = 0$.

Таблица 2. Определение индексов приводимости системы (9)

i	$n_i = \dim \mathfrak{L}^i - \dim \overline{\mathfrak{B}^{i-1}}$	$d_{i2} = n_{i-1} - k_1$
1	4	0
2	4	0
3	4	0
4	0	0

Таким образом, согласно таблице 2, $k_1 = 4, l_1 = 3$. Отсюда можно сделать вывод о том, что максимальные индексы приводимости к регулярному квазиканоническому виду $\hat{r}_1 = 3, \hat{r}_2 = 3, \hat{r}_3 = 3, \hat{r}_4 = 3$.

Рассмотрим случай, когда мультииндекс приводимости максимальный и равен $r = (3, 3, 3, 3)$. Решив систему уравнений (3), получим, что в качестве ϕ_i можно выбрать гладкие функции $\phi_i(x_1, x_2, x_3, x_7, x_8, x_9), i = \overline{1, 4}$.

Положим

$$z_1^1 = x_1, z_1^2 = x_2, z_1^3 = x_3, z_1^4 = x_7.$$

Дифференцируя функции $z_1^i, i = \overline{1, 4}$, в силу системы (9), найдем

$$(10) \quad \begin{aligned} z_2^1 &= \dot{z}_1^1 = x_4, \\ z_2^2 &= \dot{z}_1^2 = x_5, \\ z_2^3 &= \dot{z}_1^3 = x_6, \\ z_2^4 &= \dot{z}_1^4 = \hat{D}_1 x_{\overline{10,12}}, \\ z_3^{1-3} &= \dot{z}_2^{1-3} = \frac{1}{m} \hat{R} \hat{F}_b^r + \frac{1}{m} \hat{F}^{fus} + g, \\ z_3^4 &= \dot{z}_2^4 = \hat{D}_1 x_{\overline{10,12}} + \hat{D}_1 J_b^{-1} [\hat{M}_b - x_{\overline{10,12}} \times J_b x_{\overline{10,12}}], \end{aligned}$$

где \hat{D}_1 – первая строка матрицы \hat{D} .

Тогда для преобразования системы к квазиканоническому виду можно сделать следующую замену

$$\begin{aligned}
 (11) \quad & z_1^1 = x_1 \\
 & z_2^1 = x_4, \\
 & z_1^2 = x_2, \\
 & z_2^2 = x_5, \\
 & z_1^3 = x_3, \\
 & z_2^3 = x_6, \\
 & z_1^4 = x_7, \\
 & z_2^4 = \hat{D}_1 x_{\overline{10,12}}, \\
 & z_3^{1-3} = \frac{1}{m} \hat{R} \hat{F}_b^r + \frac{1}{m} \hat{F}^{fus} + g, \\
 & z_3^4 = \dot{\hat{D}}_1 x_{\overline{10,12}} + \hat{D}_1 J_b^{-1} [\hat{M}_b - x_{\overline{10,12}} \times J_b x_{\overline{10,12}}], \\
 & \eta_1 = x_8, \\
 & \eta_2 = x_9, \\
 & \eta_3 = \hat{D}_2 x_{\overline{10,12}}, \\
 & \eta_4 = \hat{D}_3 x_{\overline{10,12}}.
 \end{aligned}$$

Операции по преобразованию к квазиканоническому виду были выполнены с помощью библиотеки quasiPack [5] для системы компьютерной алгебры Maple.

Запишем систему (9) в виде заготовки, т.е. с правой частью, зависящей от переменных z и x . Это удобно сделать, так как при таком подходе результирующее управление, записанное с учетом (11), получится сразу в исходных переменных.

$$\begin{aligned}
 & z_2^1 = \dot{z}_1^1 \\
 & z_2^2 = \dot{z}_1^2 \\
 & z_2^3 = \dot{z}_1^3 \\
 & z_2^4 = \dot{z}_1^4 = \hat{D}_1 x_{\overline{10,12}} \\
 & z_3^{1-3} = \dot{z}_2^{1-3} = \frac{1}{m} \hat{R} \hat{F}_b^r + \frac{1}{m} \hat{F}^{fus} + g \\
 & z_3^4 = \dot{z}_2^4 = \dot{\hat{D}}_1 x_{\overline{10,12}} + \hat{D}_1 J_b^{-1} [\hat{M}_b - x_{\overline{10,12}} \times J_b x_{\overline{10,12}}] \\
 & \dot{z}_3^{1-3} = \frac{1}{m} \hat{R} \hat{F}_b^r + \frac{1}{m} \hat{R} (\hat{F}_b^r)'_{x_{13,16}} u + \frac{1}{m} \dot{\hat{F}}^{fus} \\
 & \dot{z}_3^4 = \ddot{\hat{D}}_1 x_{\overline{10,12}} + 2\dot{\hat{D}}_1 J_b^{-1} [\hat{M}_b - x_{\overline{10,12}} \times J_b x_{\overline{10,12}}] + \\
 & + \hat{D}_1 J_b^{-1} [(\hat{M}_b)'_{x_{4,6}} \dot{x}_{4,6} + (\hat{M}_b)'_{x_{13,16}} u - \\
 & - \dot{x}_{\overline{10,12}} \times J_b x_{\overline{10,12}} - x_{\overline{10,12}} \times J_b \dot{x}_{\overline{10,12}}],
 \end{aligned}$$

$$\begin{aligned} \dot{\eta}_{1-2} &= \hat{D}_{23}x_{\overline{10,12}}, \\ \dot{\eta}_{3-4} &= \hat{D}_{23}x_{\overline{10,12}} + \\ &+ \hat{D}_{23}J_b^{-1}[\hat{M}_b - x_{\overline{10,12}} \times J_b x_{\overline{10,12}}]. \end{aligned}$$

Расположим новые переменные в порядке $z_1^{\overline{1,3}}, z_2^{\overline{1,3}}, z_1^4, \eta_1, \eta_2, z_2^4, \eta_3, \eta_4, z_3^{\overline{1,3}}, z_3^4$, а старые – $x_{\overline{1,16}}$. Матрица Якоби замены переменных в блочной форме имеет вид

$$\frac{\partial(z, \eta)}{\partial(x)} = \begin{pmatrix} E & 0 & 0 & 0 & 0 \\ 0 & E & 0 & 0 & 0 \\ 0 & 0 & E & 0 & 0 \\ 0 & 0 & \hat{D}'_{\Omega}x_{\overline{10,12}} & \hat{D} & 0 \\ 0 & 0 & \frac{1}{m}R'_{\Omega}F & 0 & \frac{1}{m}RF'_{\nu} \\ 0 & 0 & (z_3^4)'_{\Omega} & (z_3^4)'_{\omega} & \hat{D}_1I^{-1}M'_{\nu} \end{pmatrix}.$$

Из этой записи видно, что невырожденность матрицы Якоби равносильна невырожденности блочной матрицы

$$\begin{pmatrix} \hat{D} & 0 \\ 0 & \frac{1}{m}RF'_{\nu} \\ (\ddot{\psi})'_{\omega} & \hat{D}_1I^{-1}M'_{\nu} \end{pmatrix},$$

а так как \hat{D} невырождена, то – невырожденности матрицы

$$\begin{pmatrix} \frac{1}{m}RF'_{\nu} \\ \hat{D}_1I^{-1}M'_{\nu} \end{pmatrix}.$$

Матрица R невырождена, следовательно невырожденность этой матрицы равносильна невырожденности следующей матрицы:

$$\begin{pmatrix} F'_{\nu} \\ \hat{D}_1I^{-1}M'_{\nu} \end{pmatrix}.$$

Заметим, что для невырожденности последней матрицы достаточно, чтобы матрица F'_{ν} или $\hat{D}_1I^{-1}M'_{\nu}$ имела максимальный

ранг. Проверка показывает, что матрица Якоби замены координат невырождена в точке

$$x_i = 0, i = \overline{1, 16},$$

и, следовательно, в некоторой ее окрестности. Обратную замену в явном виде выписать не удастся, однако это не мешает записать систему в квазиканоническом виде в виде заготовки, что позволяет построить стабилизирующее управление для задачи стабилизации заданного положения или задачи реализации заданного программного движения по части переменных z . При этом отсутствие в явном виде обратной замены переменных затрудняет исследование поведения переменных η . Уравнения на переменные $\dot{\eta} = P(z, \eta)$ называют также уравнениями, определяющими нулевую динамику. Их правая часть известна нам как выражение, зависящее от переменных x . Можно провести исследование устойчивости нулевого положения равновесия по первому приближению, воспользовавшись формулой производной от сложной функции:

$$\frac{\partial P}{\partial(z, \eta)} = \frac{\partial P}{\partial x} \frac{\partial x}{\partial(z, \eta)}.$$

Чтобы вычислить последний множитель, применим теорему об обратной функции.

$$\left. \frac{\partial x}{\partial(z, \eta)} \right|_{(z=0, \eta=0)} = \left(\left. \frac{\partial(z, \eta)}{\partial x} \right|_{x=0} \right)^{-1}.$$

Вычислив таким образом линеаризацию подсистемы, определяющей нулевую динамику, и положив равными нулю переменные z , получим первое приближение подсистемы нулевой динамики:

$$\begin{pmatrix} 0, 0 & 0, 0 & 1, 000135 & 0, 0 \\ 0, 0 & 0, 0 & 2, 033510 \cdot 10^{-13} & 1, 0 \\ -137, 623846 & -2, 981054 & 0, 0 & 0, 0 \\ 2, 918847 & -254, 846348 & 0, 0 & 0, 0 \end{pmatrix}$$

Анализ показывает, что собственные значения этой матрицы чисто мнимые, поэтому по первому приближению нельзя судить

об устойчивости нулевой динамики. В этом случае можно воспользоваться, например, численной оценкой устойчивости нулевой динамики [6].

Результаты численного моделирования позволяют сформулировать гипотезу о том, что нулевая динамика является устойчивой, но не асимптотически устойчивой.

Для стабилизации нулевого положения равновесия системы (9) по всем переменным можно использовать метод виртуальных выходов [8], позволяющий в некоторых случаях найти новые выходы, относительно которых нулевая динамика системы является асимптотически устойчивой.

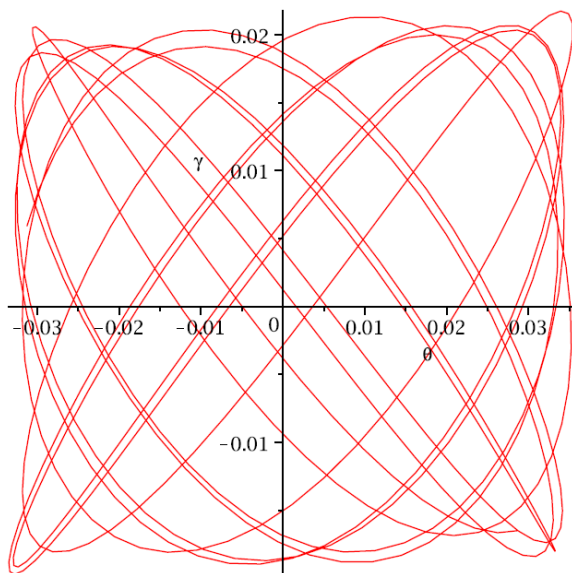


Рис. 1. Сечение фазового пространства по переменным θ и γ

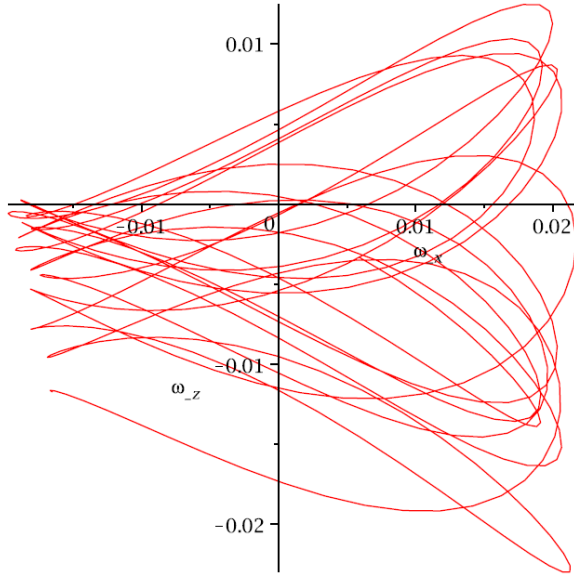


Рис. 2. Сечение фазового пространства по переменным ω_x и ω_z

3. Метод виртуальных выходов

Рассмотрим систему (1) в нормальной форме по Исидори относительно выходов y :

$$\begin{aligned}
 \dot{z}_1^1 &= z_2^1, \dots, \dot{z}_{r_1-1}^1 = z_{r_1}^1, \\
 \dot{z}_{r_1}^1 &= f_1(z, \eta) + g_{11}(z, \eta)u_1 + \dots + g_{1m}(z, \eta)u_m, \\
 &\dots, \\
 \dot{z}_1^m &= z_2^m, \dots, \dot{z}_{r_m-1}^m = z_{r_m}^m, \\
 \dot{z}_{r_m}^m &= f_m(z, \eta) + g_{m1}(z, \eta)u_1 + \dots + g_{mm}(z, \eta)u_m, \\
 \dot{\eta} &= q(z, \eta), \\
 y &= (z_1^1, z_1^2, \dots, z_1^m)^T,
 \end{aligned}
 \tag{12}$$

матрица $(g_{ij}(0, 0))_{i,j=1,\overline{m}}$ невырождена.

Систему уравнений

$$\dot{\eta} = q(0, \eta)
 \tag{13}$$

называют уравнениями нулевой динамики (нулевой динамикой). Если ее положение равновесия $\eta = 0$ асимптотически устойчи-

во, то аффинную систему (1) с виртуальным выходом $y = \phi(x)$ называют минимально фазовой (в точке $x = 0$) [11].

Метод виртуальных выходов применим к системам, в которых $q(z, \eta) = q(z^1, z^2, \eta)$. Чтобы преобразовать систему (9) к такому виду, выберем в качестве выходов

(14)

$$y_1 = z_1^1 = x_4, y_2 = z_1^2 = x_5, y_3 = z_1^3 = x_6, y_4 = z_1^4 = \hat{D}_1 x_{10,12}.$$

По этим выходам можно построить замену, преобразующую систему к квазиканоническому виду с индексом приводимости $r = (2, 2, 2, 2)$.

Задача поиска нового выхода сводится к поиску матрицы K , которая удовлетворяет неравенству

$$R = (A - B^1 K + B^2 (E + K B^2)^{-1} (-K A + K B^1 K)) < 0,$$

где $z^1, z^2 \in R^m, \eta \in R^l, l = n - 2m$,

$$A = \left. \frac{\partial q}{\partial \eta} \right|_{z=0, \eta=0}, B^1 = \left. \frac{\partial q}{\partial z^1} \right|_{z=0, \eta=0}, B^2 = \left. \frac{\partial q}{\partial z^2} \right|_{z=0, \eta=0}.$$

Одним из способов найти K является численная минимизация функции $\max_K (Re(\lambda_i))$, λ_i — собственные значения матрицы R , по элементам матрицы K . С помощью среды Maple было найдено решение

$$K = \begin{pmatrix} 1, 29 & 0, 621 & 0, 716 & 0, 701 & -3, 48 & -1, 88 & -0, 778 & -0, 0319 \\ -1, 61 & -0, 140 & -0, 685 & -1, 94 & -2, 41 & -1, 28 & -0, 335 & -1, 18 \\ -0, 948 & -1, 38 & -1, 21 & -1, 02 & -0, 156 & -2, 57 & -0, 284 & -0, 759 \\ -1, 29 & -0, 842 & -1, 06 & -1, 0 & -0, 776 & -1, 19 & 0, 425 & -7, 98 \end{pmatrix},$$

дающее для матрицы R собственные числа

$$\begin{pmatrix} -10, 8 \\ -0, 757 + 2, 029 i \\ -0, 757 - 2, 029 i \\ -0, 675 + 0, 947 i \\ -0, 675 - 0, 947 i \\ -0, 687 + 0, 356 i \\ -0, 687 - 0, 356 i \\ -0, 661 \end{pmatrix}.$$

Тогда новые выходы могут быть вычислены по формуле

$$(15) \quad \bar{y} = y + K\eta.$$

Построим новую замену переменных

$$\begin{aligned} \bar{z}_1^{1-4} &= \bar{y}_{1-4}, \\ \bar{z}_2^{1-4} &= \dot{\bar{z}}_1^{1-4}, \\ \bar{\eta} &= \eta. \end{aligned}$$

Нормальная форма системы будет иметь вид

$$(16) \quad \begin{aligned} \dot{\bar{z}}_1^{1-3} &= z_2^{1-3}, \\ \dot{\bar{z}}_1^4 &= z_2^4, \\ \dot{\bar{z}}_2^{1-3} &= \bar{f}_{1-3}(x) + \bar{g}_{123}(x)u, \\ \dot{\bar{z}}_2^4 &= \bar{f}_4(x) + \bar{g}_4(x)u, \\ \dot{\bar{\eta}}_{1-3} &= x_{4-6}, \\ \dot{\bar{\eta}}_4 &= \hat{D}_{10,12}x_{10,12}, \\ \dot{\bar{\eta}}_{5-6} &= \hat{D}_{23}x_{10,12}, \\ \dot{\bar{\eta}}_{7-8} &= \hat{D}_{23}x_{10,12} + \hat{D}_{23}J_b^{-1}[\hat{M}_b - x_{10,12} \times J_b x_{10,12}]. \end{aligned}$$

Функции $\bar{f}(x)$ и $\bar{g}(x)$ вычислены с помощью системы Maple и не приводятся здесь ввиду громоздкости. Система (9),(15) является минимально-фазовой, т.е. ее нулевая динамика является локально асимптотически устойчивой. После построения замены переменных, соответствующей выходам (15), переменные z могут быть стабилизированы линеаризацией обратной связью. Таким образом достигается стабилизация нулевого положения по всем переменным.

4. Результаты

Проведено исследование возможности преобразования системы уравнений, описывающей движение вертолета, к квазилинейному виду. Проанализирована устойчивость подсистемы нулевой динамики по первому приближению и с помощью численного моделирования. Осуществлена стабилизация положения равновесия методом виртуальных выходов.

5. Выводы и перспективы

Приведенные результаты позволяют сделать вывод о том, что управление с помощью преобразования к квазиканоническому виду и метод виртуальных выходов могут применяться к системе уравнений, описывающих движение вертолета. Рассмотренная система имеет сравнительно высокую размерность ($n = 16$), и ее преобразование к квазиканоническому виду стало возможно благодаря современным компьютерным средствам аналитических вычислений. В отличие от работы [12] учитываются аэродинамические силы, действующие на вертолет, рассматривается движение в трехмерном пространстве, достигается стабилизация по всем переменным. Использование метода виртуальных выходов позволило обойтись без применения линеаризации в окрестности положения равновесия. Возможен выбор функций ϕ_i , удовлетворяющих системе уравнений (3) и определяющих преобразование к квазиканоническому виду из каких-либо других соображений.

Литература

1. КРАСНОЩЕЧЕНКО В.И., КРИЩЕНКО А.П. *Нелинейные системы: геометрические методы анализа и синтеза*. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2005. – 520 с.
2. КРИЩЕНКО А.П., КЛИНКОВСКИЙ М.Г. *Преобразование аффинных систем с управлением и задача стабилизации* // Дифференциальные уравнения. – 1992. – Т. 28. – С. 1945–1952.
3. КРИЩЕНКО А.П. *Преобразование нелинейных систем и стабилизация программных движений* // Труды МВТУ им. Н.Э. Баумана. – 1988. – № 512. – С. 69 – 87.
4. КРИЩЕНКО А.П. *Преобразование многомерных аффинных управляемых систем* // Управляемые нелинейные системы. – 1991. – № 2. – С. 5–14.
5. *Свидетельство о государственной регистрации программы №2012617543. quasiPack* / С.Б. Ткачев, А.А. Шевля-

- ков. Зарегистрировано в Реестре программ для ЭВМ 21 августа 2012 г.
6. ТКАЧЕВ С.Б., ШЕВЛЯКОВ А.А. *Оценка нулевой динамики неминимально-фазовых аффинных систем* // Управление в технических системах: Труды конференции с международным участием. СПб., 2010. – С. 48 – 51.
 7. ТКАЧЕВ С.Б., ШЕВЛЯКОВ А.А. *Преобразование аффинных систем со скалярным управлением к квазиканоническому виду* // Вестник МГТУ им. Н.Э. Баумана. Естественные науки. – 2013. – №1. – С. 3 – 16.
 8. ТКАЧЕВ С.Б. *Стабилизация неминимально фазовых аффинных систем с векторным управлением* // Наука и образование. МГТУ им. Н.Э. Баумана. Электрон. журн. – 2012. – №10. [Электронный ресурс]. – URL: <http://technomag.edu.ru/doc/450613.html> (дата обращения: 02.02.13).
 9. BRUNOVSKY P.A. *A classification of linear controllable systems* // Kybernetika. – 1970. – Vol. 6, №3. – P. 176–188.
 10. GAVRILETS V., METTLER B., FERON E. *Dynamic Model for a Miniature Aerobatic Helicopter* // MIT-LIDS report – 2003. – No. LIDS-P-2580. – 22 p.
 11. ISIDORI A. *Nonlinear control systems*. – London: Springer-Verlag, 1995. – 587 p.
 12. KOO T.J, SASTRY S. *Output Tracking Control Design of a Helicopter Model Based on Approximate Linearization* // Proc. 37th IEEE Conference on Decision & Control. Tampa (Florida, USA), 1998. – Vol. 4. – P. 3635–3640.
 13. MARINO R. *On the largest feedback-linearizable subsystem* // Systems & Control Letters. – 1986. – №6. – P. 345–351.

HELICOPTER MODEL AND ITS QUASICANONICAL FORM

Andrey Shevlyakov, Institute of Control Sciences of RAS Moscow (aash29@gmail.com).

Abstract: We consider the mathematical model of a helicopter described by a system of ordinary differential equations. These equations are transformed to the quasicanonical form. Then we analyse the subsystem governing the zero dynamics of the system and find the outputs with respect to which the system is a minimum-phase one.

Keywords: quasicanonical form, nonlinear systems, zero dynamics.

*Статья представлена к публикации
членом редакционной коллегии А. П. Курдюковым*

Поступила в редакцию 03.02.2013.

Опубликована 30.09.2013.

УДК 004.738
ББК 30

ЗАЩИТА ЭЛЕКТРОННЫХ СООБЩЕНИЙ В МУЛЬТИ-СЕТЕВОЙ СРЕДЕ

Асратян Р. Э.¹, Лебедев В. Н.²

(ФГБУН Институт проблем управления РАН, Москва)

Рассматривается новое архитектурное решение задачи защиты информационных запросов и обеспечения безопасных взаимодействий на базе технологии .NET в гетерогенной мульти-сетевой среде, основанное на применении технологии прокси-серверов. Суть решения заключается в организации специальных серверов-посредников, исполняющих роль шлюзов между информационно-управляющими системами и глобальной сетью и выполняющих функции проверки и формирования электронных подписей во входящих и исходящих SOAP-сообщениях.

Ключевые слова: распределенные системы, интернет-технологии, сетевые протоколы, веб-сервисы, электронная подпись.

1. Введение

Методам и средствам управления информационной безопасностью в территориально-распределенных информационных системах уже давно уделяется большое внимание. Интерес к ним еще более усилился в последние годы в связи с появлением систем, обеспечивающих межведомственное или транснациональное информационное взаимодействие в сложной мульти-сетевой среде, включающей множество сетей разного размера и

¹ Рубен Эзрасович Асратян, кандидат технических наук, доцент (rea@ipu.ru).

² Виталий Николаевич Лебедев, кандидат технических наук, доцент (lebvini@ipu.ru).

административной подчиненности. Вопросы, связанные с защитой и аутентификацией электронных сообщений (информационных запросов и ответов), а также с защитой информационных ресурсов от несанкционированного доступа из сети, в этих условиях стали особенно актуальны.

Однако многие разработчики подобных систем уже успели почувствовать, что созданные еще в 90-х годах универсальные интернет-технологии защиты информации в сети имеют с этой точки зрения ряд важных недостатков.

- Аутентификация и проверка целостности данных в универсальных технологиях сетевой защиты осуществляется или на уровне IP-датаграмм (IPsec), или же на уровне блоков данных, на которые разбивается информационный поток в TCP-канале (SSL и TLS [1, 4, 8]) с помощью встроенного механизма электронной подписи (ЭП). Однако в этих условиях не существует никакой связи между электронным сообщением и подобной ЭП. Зарегистрировать и сохранить документ электронного сообщения вместе с удостоверяющей его подписью попросту невозможно.

- В системах электронного документооборота уже давно сложилась практика использования сразу нескольких ЭП для подтверждения корректности одного и того же документа (например, подписи исполнителя и одной или нескольких утверждающих подписей руководителей или организаций). В последние годы возникло понимание необходимости применения подобной схемы аутентификации и для межведомственных или транснациональных электронных сообщений. Между тем, универсальные методы сетевой защиты данных не предоставляют возможности применения нескольких ЭП для удостоверения одной и той же информации.

Именно по этой причине в последние годы был разработан ряд международных стандартов на внедрение ЭП в электронные документы в формате XML, которые нашли применение в сетевой архитектуре .NET для защиты информационных запросов и ответов (так называемых SOAP-сообщений) [5, 7]. Эти стандарты вместе с соответствующим программным обеспечением,

разработанным рядом криптопровайдеров, создали основу для решения задачи защиты данных в сети на уровне электронных сообщений. Тем не менее, разработчики распределенных систем все еще сталкиваются в этой области с рядом проблем. Эти проблемы связаны с отсутствием универсальных архитектурных решений задачи встраивания упомянутых средств криптозащиты электронных сообщений в уже существующие или вновь разрабатываемые информационные системы. В этих условиях разработчики вынуждены тратить время и силы для поиска собственных ответов на следующие вопросы:

- Должны ли средства информационной защиты встраиваться в клиентские и сервисные компоненты систем или же занимать в ней какое-то особое выделенное положение?
- Каким образом осуществлять «перехват» электронных сообщений для их криптообработки (формирования и проверки ЭП)?
- Как совместить аутентификацию электронных сообщений с защитой информационных ресурсов в мульти-сетевой среде?

В данной работе описывается новый подход к решению данной задачи, основанный на технологии прокси-серверов, т.е. серверов-посредников, размещаемых на границах сетей и обеспечивающих как аутентификацию электронных сообщений на основе сетевого протокола HTTP/SOAP [5, 7], так и защиту информационных ресурсов внутри сетей от несанкционированного доступа. Главное преимущество этого подхода заключается в его высокой универсальности, выражающейся в независимости от архитектур взаимодействующих информационных систем, их операционных и языковых платформ и средств их разработки.

2. Аутентификации SOAP-сообщений и защита информационных ресурсов

Основу мульти-сетевой среды распределенных информационных систем обычно составляют частные локальные сети организаций, соединенные одной или несколькими глобальными

ми сетями. Статус частной сети предполагает высокую степень изолированности от остальной среды: прямое сетевое соединение между программами, работающими в разных частных сетях невозможно без применения специальных средств (например, трансляторов IP-адресов в датаграммах или серверов-посредников). Желание администраторов «спрятать» информационные ресурсы в частных сетях и возрастающее недоверие к низкоуровневым (т.е. основанным на IP-адресах и портах) средствам ограничения доступа - межсетевым экранам - можно признать важной тенденцией последних лет. Появление технически изощренных форм «взлома» межсетевых экранов, основанных на фальсификации IP-адресов в IP-датаграммах, значительно поколебало веру в их надежность. Сегодня многие разработчики склоняются к мнению, что высоконадежные средства информационной защиты должны быть основаны на ЭП в составе электронных сообщений (если, разумеется, удаленное взаимодействие целиком построено на технологии электронных сервисов в архитектуре .NET).

Структура SOAP-сообщения с электронной подписью, внедренной в его заголовок в соответствии со стандартами консорциумов W3C и OASIS, проиллюстрирован на рис. 1 (предполагается, что soap и wsse – префиксы пространств имен <http://schemas.xmlsoap.org/soap/envelope/> и <http://docs.oasis-open.org/wss/2004/01/oasis-200401-wss-wssecurity-secext-1.0.xsd> соответственно). Как видно из рисунка, внедрение ЭП осуществляется с помощью специальных тегов (например, тегов security) с сохранением корректной XML-структуры документа. Важно отметить, что стандарт допускает размещение любого количества ЭП в заголовке сообщения. Каждая ЭП может относиться (т.е. удостоверять) или ко всему телу сообщения или же к отдельному фрагменту тела, ограниченного каким-либо тегом.

Применение технологии прокси-серверов для аутентификации электронных сообщений и защиты информационных ресурсов в частных сетях основано на следующих принципах.

- Функции «перехвата», защиты и аутентификации SOAP-сообщений реализуются в выделенных прокси-серверах, размещаемых на границах сетей.

- Каждый прокси-сервер уполномочивается формировать ЭП организации в SOAP-сообщениях, исходящих из частной сети организации, и проверять ЭП во входящих SOAP-сообщениях, т.е. поступающих в нее. Это означает, что ему должны быть доступны открытый и закрытый криптографические ключи организации, необходимые для формирования ЭП.

- Авторизация и проверка прав доступа к данным осуществляются в адресуемых электронных сервисах на основе имен подписантов, содержащихся в ЭП. Чтобы упростить им эту задачу, прокси-сервер извлекает имена подписантов (подтвержденные доверенным удостоверяющим центром) из всех проверенных ЭП и помещает их в заголовок SOAP-сообщения.

- Все входящие информационные запросы (ответы), не прошедшие аутентификацию, не только не обрабатываются соответствующими электронными сервисами (клиентами), но вообще не допускаются в частную сеть.

Другими словами, при данном подходе прокси-сервер выполняет функции своего рода высокоуровневого межсетевого экрана, работающего на уровне SOAP-сообщений и ЭП (а не на общепринятом уровне IP-датаграмм). Важно подчеркнуть, что в сочетании с жесткими мерами безопасности (отключение всех сетевых служб, кроме самого прокси-сервера, запрет на прямую маршрутизацию IP-датаграмм с одного сетевого интерфейса на другой на «пограничных» серверах) межсетевой экран подобного типа способен обеспечить наиболее высокий уровень информационной безопасности, практически исключая возможность «взлома» с помощью технических средств. Далее в статье мы будем называть прокси-серверы, оснащенные средствами формирования и проверки ЭП в SOAP-сообщениях, «прокси-серверами с криптозащитой» (ПСК).

На рис. 2 проиллюстрирована защита SOAP-сообщений в сложной мульти-сетевой среде, включающей глобальную сеть, несколько частных сетей организаций и ведомственную сеть

(например, организованную с помощью технологии виртуальных частных сетей [1]). Как видно из рисунка, информационный запрос от рабочей станции PC , направленный к серверу WS_3 , проходит в форме SOAP-сообщения три ПСК, размещенных на границах частной сети LAN_1 и ведомственной сети (P_1), ведомственной и глобальной сетей (P_0) и, наконец, глобальной сети и LAN_3 (P_3). Первый из серверов формирует и помещает в SOAP-сообщение подпись организации. Второй проверяет подпись организации и добавляет «утверждающую» подпись ведомства. Наконец, сервер P_3 , защищающий LAN_3 , проверяет обе подписи у входящего сообщения и, после успешной аутентификации, пропускает его в частную сеть к адресуемому информационному ресурсу.

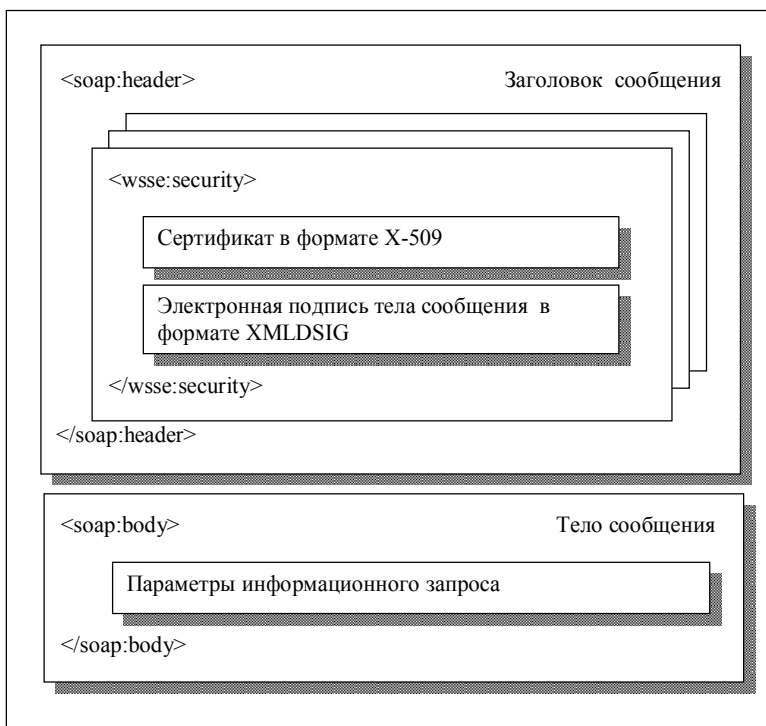


Рис. 1. Электронные подписи в составе SOAP-сообщения

Важно отметить, что документ входящего сообщения может быть не только обработан, но и зарегистрирован в WS_3 вместе со всеми содержащимися в нем ЭП.

Разумеется, приведенная в данном примере обработка электронного сообщения была бы невозможна, если бы на прокси-серверы не обеспечивали выполнение еще одной важной функции в дополнение к функциям криптозащиты: маршрутизации информационных запросов в мульти-сетевой среде. В данной статье мы не будем подробно останавливаться на данной функции; отметим лишь, что маршрутизация в данном случае осуществляется не по IP-адресам, а по символическим интернет-именам информационных ресурсов. Это позволяет обеспечить корректный доступ к электронным сервисам в удаленных частных сетях, даже если несколько сервисов имеют одинаковый IP-адрес [2, 3].

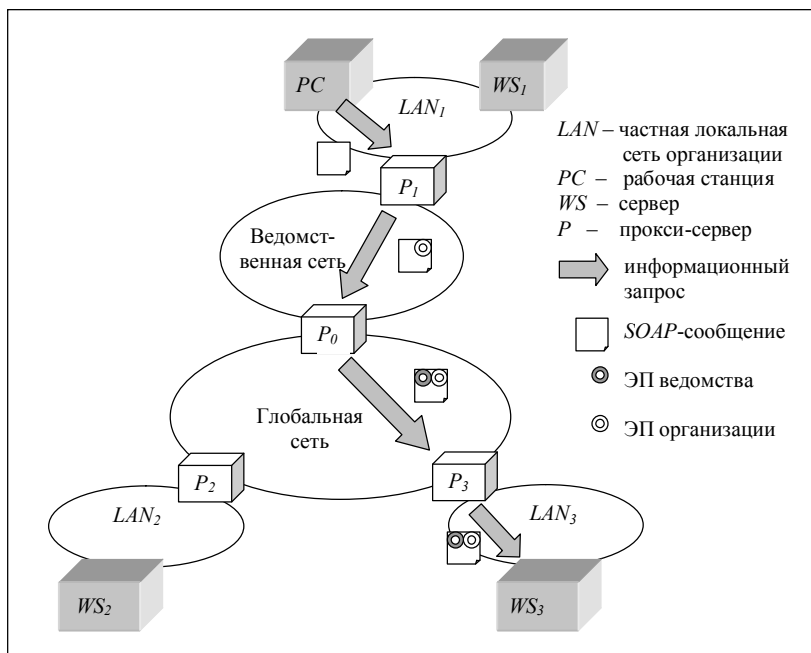


Рис. 2. Защита SOAP-сообщений в мульти-сетевой среде

К недостаткам описанного подхода следует отнести дополнительные временные задержки, вносимые ПСК и функциями формирования и проверки ЭП в электронных сообщениях. С целью оценки этих задержек была проведена серия экспериментов с опытной реализацией ПСК. Цель экспериментов заключалась в сравнении времен выполнения запросов к модельным электронным сервисам в условиях прямого обращения (без ПСК) и при обращении через ПСК, выполняющий проверку ЭП в каждом запросе и формирование ЭП в ответе на запрос. Важно подчеркнуть, что измерялось не только время выполнения одиночных запросов, но и скорость обработки (число обработанных запросов в секунду), достигаемая при одновременной обработке пакета информационных запросов. Другими словами, мы попытались провести сравнение с учетом распараллеливания обработки в множестве программных потоков (и в ПСК, и службе электронных сервисов для обработки каждого запроса создается отдельный программный поток). Скорость обработки вычислялась как частное от деления числа запросов в пакете на полное время его выполнения.

Эксперименты проводились в условиях скоростного (100 Мбит/сек.) Ethernet в среде Windows Server 2003 на достаточно скромном сервере, оснащенный одноядерным процессором Intel с частотой 2,8 ГГц и 2 Гб оперативной памяти. Размер SOAP-сообщения колебался в пределах от 2 до 4 Кб, число элементов (тегов) в сообщении – от 50 до 100.

На рис. 3 показан характерный график, который дает представление о дополнительных задержках, вносимых ПСК, при обращении к двум модельным электронным сервисам с фиксированным временем выполнения (задержкой), равным 0,5 с и 1 с соответственно. В эксперименте измерялась скорость обработки (число обработанных запросов в секунду), достигаемая при обработке группы запросов, т.е. при одновременном обращении сразу нескольких клиентов. По горизонтальной оси графика отложено количество запросов в группе, а по вертикальной – скорость обработки. Первая пара кривых (прорисованная сплошными линиями) относится к более быстрому сервису

(0,5 с), а вторая (прорисованная прерывистыми линиями) – к более медленному (1 с). При этом более тонкая кривая соответствует ситуации с прямым обращением к модельному сервису (без криптозащиты), а более толстая – с обращением через ПСК с формированием и проверкой ЭП.

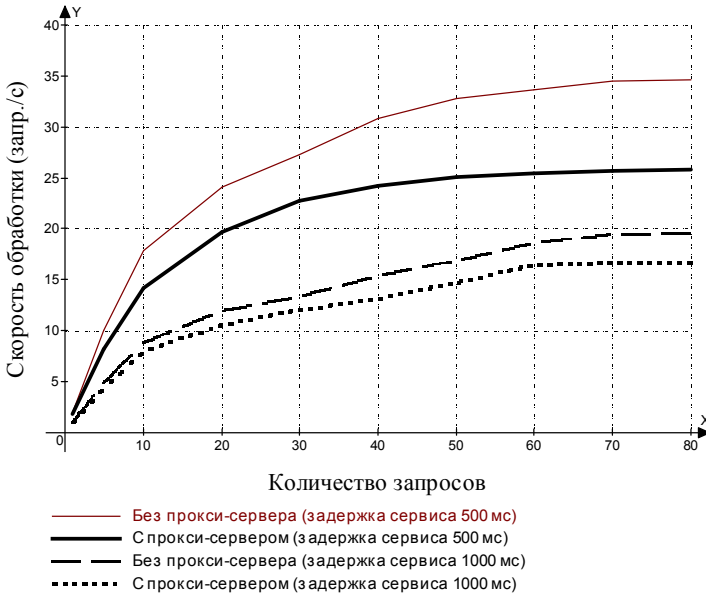


Рис. 3. Пример графика зависимости скорости обработки от числа одновременных запросов

При подаче одиночного запроса на вход сервиса со временем выполнения 0,5 с полное время обработки составило 500 мс без ПСК и 546 мс с ПСК (2 запроса в секунду и 1,88 соответственно). Как видно из графика, при росте числа одновременных запросов растет и скорость обработки, что объясняется «многоканальной» природой и ПС и электронного сервиса, т.е. положительным эффектом от распараллеливания обработки в нескольких программных потоках. Например, при одновременной подаче 10 запросов время их выполнения составило 560 мс без ПСК и 703 мс с ПСК (17,8 и 14,2 запросов в секунду соответ-

венно). Как видно из графика, при превышении числа запросов в группе значения 20-30 рост кривых сильно замедляется, а при превышении значения 70-80 – прекращается вовсе, достигая предельной производительности.

Как видно из первой пары кривых, при обращениях к быстрому (0,5 с) сервису применение ПСК существенно замедляет обработку практически при любом числе запросов в группе. Но для более медленного сервиса (1 с) относительная разница между кривыми значительно уменьшается. Другими словами, для электронных сервисов с собственным временем выполнения более 1 с влияние ПСК становится менее существенным.

На рис. 4 приведен график зависимости среднего времени выполнения информационного запроса от числа запросов в пакете при обращении к тем же двум модельным сервисам и в тех же обозначениях. В отличие от рис. 3, здесь все кривые демонстрируют устойчивый рост. Легко видеть, что и по этому важному показателю влияние ПСК является довольно ощутимым при обращении к более быстрому сервису (0,5 с) и менее ощутимым при обращении к более медленному сервису (1 с). Например, при поступлении 10 одновременных запросов соотношение значений среднего времени обработки запроса с ПСК и без ПСК равно 1,28 в первом случае и 1,12 во втором.

В целом результаты экспериментов позволяют сформулировать следующие выводы.

- Применение ПСК в целом сохраняет положительный эффект от многопоточного распараллеливания обработки.
- ПСК вполне позволяет поддерживать скорость обработки до нескольких и даже нескольких десятков запросов в секунду, что обычно бывает достаточным для большинства информационно-управляющих систем.
- Чем больше время выполнения электронного сервиса, тем менее заметна задержка, вносимая ПСК. В частности, в проведенных экспериментах наблюдалось заметное снижение скорости обработки при одновременном обращении сразу нескольких клиентов к быстрым (со временем выполнения менее 1 с) электронным сервисам через ПСК и не столь существенное при

обращения к более медленным (со временем выполнения более 1 с).

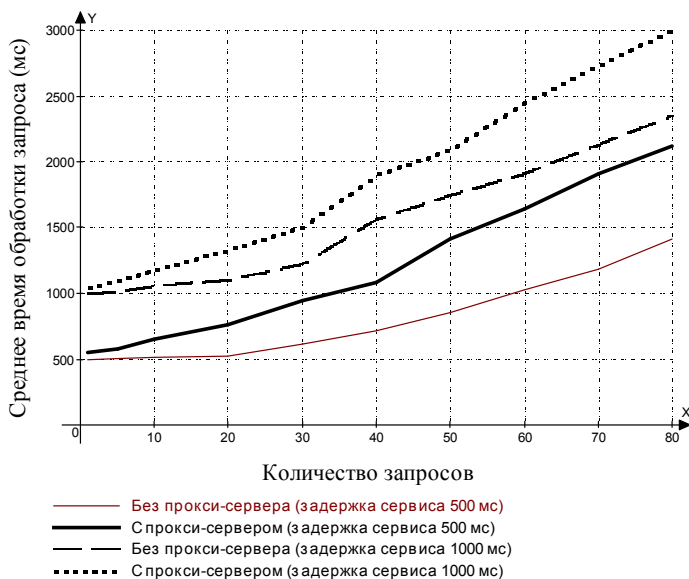


Рис. 4. Пример графика зависимости среднего времени обработки одного запроса от количества одновременных запросов

3. Проблема сохранения подписи исполнителя в частной сети

Важный недостаток описанной выше организации информационной защиты заключается в отсутствии в ней электронной подписи исполнителя – лица, непосредственно инициировавшего информационный запрос с помощью клиентского приложения. Без ЭП исполнителя, формируемой клиентским приложением одновременно с электронным документом запроса на основе личных ключей защиты, возникает угроза искажения или фальсификации запроса еще в пределах исходной частной сети. Очевидно, что безопасное поведение ПСК должно вклю-

чать проверку подписи исполнителя перед формированием подписей организации и/или ведомства в SOAP-документе. Сетевая архитектура .NET содержит специальные средства, позволяющие сформировать ЭП исполнителя в SOAP-документе запроса еще в пределах создавшего его приложения.

Однако здесь возникает важная проблема – проблема «посредников». Если между клиентским приложением и ПСК имеются один или несколько дополнительных обработчиков информационного запроса, то ЭП исполнителя может оказаться потерянной.

Рассмотрим следующий пример. Предположим, что в частной сети организации находятся одно или несколько клиентских рабочих мест для подготовки информационных запросов и сервер приложений с локальным электронным сервисом для централизованной регистрации этих запросов в БД (см. рис. 5). Логика обработки предполагает, что клиенты не обращаются непосредственно к удаленному сервису, но обращаются к локальному сервису-регистратору, который регистрирует информационный запрос, дополняет его добавочными параметрами (например, регистрационным номером, датой и временем регистрации) и уже после этого осуществляет обращение к удаленному сервису через ПСК. Результаты выполнения запроса также регистрируются в БД, после чего передаются клиенту. Очевидно, что на выходе сервиса-регистратора появляется совершенно новый SOAP-документ уже без ЭП исполнителя (сервис попросту не может сформировать ее повторно, так как не располагает необходимыми для этого личными ключами защиты).

Эффективный способ решения этой проблемы дает новый механизм – копирование ЭП исполнителя из входного SOAP-сообщения в выходное без повторного формирования в локальных электронных сервисах. Этот механизм «наследования» ЭП работает по следующим правилам:

- Клиентское приложение формирует в исходящем SOAP-сообщении одну или несколько ЭП, каждая из которых относится к какому-то фрагменту тела сообщения ограниченному опре-

деленным тегом (причем разным ЭП соответствуют разные теги).

- Получив клиентское SOAP-сообщение, локальный электронный сервис проверяет все содержащиеся в нем ЭП и сохраняет его «как есть» перед началом обработки.

- Сформировав собственное исходящее SOAP-сообщение для обращения к удаленному сервису, локальный сервис начинает сопоставление входящего (клиентского) и исходящего сообщений.

- Для каждой ЭП, обнаруженной во входящем сообщении, локальный сервис определяет имя связанного с ним тега и отыскивает тег с таким же именем в выходном сообщении. Если выходное сообщение содержит тег с таким же именем, то соответствующий ему фрагмент тела сообщения полностью замещается аналогичным фрагментом входного сообщения (см. рис. 5), а соответствующее ЭП копируется из входного сообщения в выходное.

Как видно из приведенного описания, механизм «наследования» ЭП опирается на два предположения:

- фрагменты входного и выходного SOAP-сообщений, соответствующие «подписанным» тегам с одинаковыми именами, действительно содержат одну и ту же информацию,

- каждый «подписанный» тег не повторяется в теле сообщения несколько раз.

Вернемся к рассмотрению примера с локальным сервисом-регистратором. Предположим, что метод сервиса-регистратора, к которому обращается клиентское приложение, имеет следующую спецификацию:

int LocalQuery(ParamList QueryParams) ;

Здесь QueryParams – формальный параметр класса ParamList, содержащий набор параметров информационного запроса к удаленному сервису (структура этого класса для нас безразлична).

Предположим, также, что вызываемый метод удаленного сервиса имеет следующую спецификацию:

int RemoteQuery(string RegNum, string DateTime, ParamList QueryParams) ;

Здесь RegNum и DateTime – формальные параметры, предназначенные для передачи регистрационного номера и времени запроса, а QueryParams – формальный параметр класса ParamList, содержащий набор параметров информационного запроса. Будем считать, что описания классов ParamList на сервисе регистраторе и на удаленном сервисе совершенно одинаковы, а сервис-регистратор просто переносит полученные параметры информационного запроса в удаленный вызов. В этом случае и входящее и исходящее SOAP-сообщения будут содержать теги <QueryParams>, содержащие одинаковые данные. Если клиентское приложение сформирует ЭП для фрагмента SOAP-сообщения, ограниченного тегом <QueryParams>, то эта ЭП может быть «унаследована» выходным сообщением сервиса-регистратора.

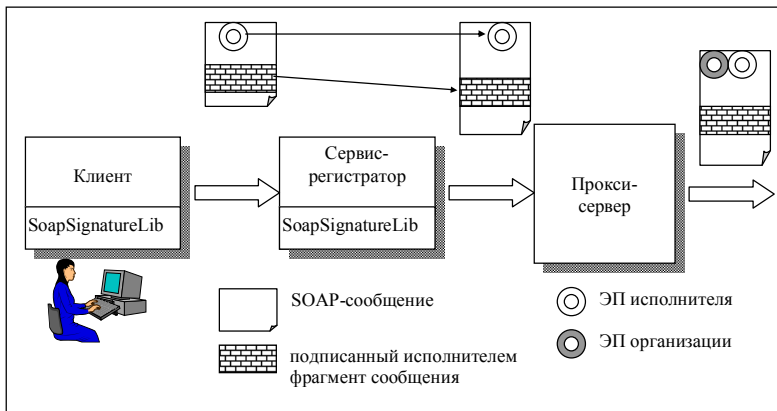


Рис. 5. Механизм наследования ЭП исполнителя в сервисе-посреднике

Описанные принципы «наследования» ЭП исполнителя были реализованы в библиотеке функций SoapSignatureLib, предназначенной для формирования и проверки ЭП непосредственно в клиентских и сервисных приложениях для организации безо-

пасного взаимодействия с ПСК (см. рис. 5). Работа библиотеки основана на механизмах Client Message Inspector и Dispatch Message Inspector в сетевой архитектуре .NET. Будучи подключенной к приложению, библиотека берет на себя такие функции, как «перехват» входящих и исходящих SOAP-сообщений, формирование и проверка ЭП, причем работа библиотеки совершенно не зависит от архитектуры приложения или структуры электронных сервисов. Следует оговориться, что библиотеку SoapSignatureLib можно использовать только в системах, ориентированных на работу в среде Microsoft Framework 4.0 или старше. Другими словами, в отличие от ПСК, она не обеспечивает платформенной независимости.

4. Заключение

В условиях отсутствия общего архитектурного решения для задачи криптозащиты электронных сообщений в мульти-сетевой среде разработчики вынуждены расходовать время и силы для создания собственных технических решений, привязанных к особенностям конкретных систем, платформам и средствам программирования. Авторы попытались показать, что технология прокси-серверов может послужить основой общего решения этой задачи.

Описанный подход был реализован в форме дополняющих друг друга программных продуктов для среды Win32: ПСК (в форме Windows-службы [1, 6]) и библиотеки функций SoapSignatureLib для обработки ЭП в клиентских приложениях и электронных сервисах. Оба продукта реализованы на языке C# в среде Microsoft Visual Studio с использованием сертифицированных криптосредств компании «КриптоПро».

В данной работе не рассматривались вопросы шифрования электронных сообщений в сети. Однако следует отметить, что предложенный подход вполне совместим со средствами шифрования, реализованными в IPsec или TLS/SSL.

Разработанные методы и средства управления информационной безопасностью прошли испытания в составе Системе

межведомственного электронного взаимодействия (СМЭВ), представляющей собой федеральную государственную информационную систему, предназначенную для интеграции информационных ресурсов федеральных органов государственной власти на базе единых стандартов защиты информации и сетевых технологий. Испытания показали, что описанный подход может быть использован в качестве типового решения для задачи организации безопасного взаимодействия в СМЭВ и других территориально-распределенных информационных системах, базирующихся на сетевой архитектуре .NET.

Литература

1. АНДРЕЕВ А.Г., БЕЗЗУБОВ Е.Ю., ЕМЕЛЬЯНОВ М.М. И ДР. *Windows 2000: Server и Professional*. – СПб.: «БХВ-Санкт-Петербург», 2001. – 1055 с.
2. АСРАТЯН Р.Э., ЛЕБЕДЕВ В.Н. *Организация защищенного http-взаимодействия в мульти-сетевой среде* // Управление большими системами. – 2012 - №36. - С. 285-300.
3. АСРАТЯН Р.Э., ЛЕБЕДЕВ В.Н. *Прокси-серверы в распределенных гетерогенных мульти-сетевых средах* // Проблемы управления. – 2013 - №2.- С. 45-50.
4. ДЖАМСА К., КОУП К. *Программирование для Internet в среде Windows*. – СПб.: Питер, 1996. – 659 с.
5. МАК-ДОНАЛЬД М., ШПУШТА М. *Microsoft ASP.NET 3.5 с примерами на C# 2008 и Silverlight 2 для профессионалов*. - М.: Вильямс, 2009. – 1408 с.
6. СНЕЙДЕР Й. *Эффективное программирование TCP/IP. Библиотека программиста*. – СПб.: Символ-Плюс, 2002. – 320 с
7. ШАПОШНИКОВ И.В. *Web-сервисы Microsoft .NET*. – СПб.: БХВ – Петербург, 2002. – 336 с.
8. ХАНТ К. *TCP/IP. Сетевое администрирование*. – СПб.: Питер, 2007. – 816 с.

ORGANIZATION OF PROTECTED HTTP-INTERACTION IN MULTI-NETWORK ENVIRONMENT

Ruben Asratian, Institute of Control Sciences of RAS, Moscow, Cand.Sc., assistant professor (rea@ipu.ru).

Vitali Lebedev, Institute of Control Sciences of RAS, Moscow, Cand.Sc., assistant professor (lebvini@ipu.ru).

Abstract: We suggest an architectural solution based on proxy servers technology for the problem of query protection and the.NET technology-based interactions security in the heterogeneous multi-network environment. The idea is to use special intermediary servers playing the role of gateways between information systems and a global network and carrying out functions of digital signature generation and verification for input and output SOAP messages.

Keywords: distributed system, Internet technology, network protocol, Web-service, digital signature.

*Статья представлена к публикации
членом редакционной коллегии М.В. Губко*

*Поступила в редакцию 08.05.2013.
Опубликована 30.09.2013.*

АВТОМАТИЗИРОВАННОЕ ОПРЕДЕЛЕНИЕ КЛАССОВ ЧУВСТВИТЕЛЬНОСТИ ВЕБ-СЕРВИСОВ¹

Душкин Д. Н.²,
(ФГБУН Институт проблем управления
им. В.А. Трапезникова РАН, Москва)

*В работе представлена методика вычисления чувствительности — критерия сравнения веб-сервисов по производительности. Вводятся показатели, характеризующие чувствительность, описаны классы чувствительности, способ их получения и метод определения класса для произвольного веб-сервиса. В работе используются такие алгоритмы машинного обучения, как кластеризация с помощью алгоритма *k*-средних и классификация методом опорных векторов. Приводятся результаты вычислительного эксперимента, иллюстрирующего представленную методику.*

Ключевые слова: веб-сервисы, сервисно-ориентированная архитектура, машинное обучение, анализ чувствительности.

Введение

В последние несколько лет развития области информационных технологий (ИТ) наметился переход от всеобъемлющих информационных систем (ИС), включающих все необходимые для работы данные и функции, к распределенным системам, использующим внешние ресурсы [12]. Такой переход оправдан тем, что одна система зачастую уже не способна хранить весь объём необходимых данных и обеспечивать приемлемый уровень произво-

¹ Автор признателен Смирновой Н.В. и Фархадову М.П. за ценное обсуждение содержания статьи.

² Дмитрий Николаевич Душкин, аспирант, младший научный сотрудник (ddushkin@asmon.ru, тел. 8(916)978-58-96).

длительности вычислений. Одним из современных решений, позволяющих построить распределенную ИС, является использование сервисно-ориентированной архитектуры (СОА).

В основе сервисно-ориентированной архитектуры лежит принцип использования *веб-сервисов* — автономных вычислительных ресурсов, предоставляющих свои функции через сеть Интернет (или Интранет) посредством открытых протоколов обмена данными, не зависящих от платформ как самих ресурсов, так и связываемых с ними программных систем. Системы, использующие в своей работе веб-сервисы, называют системами с СОА.

В предыдущей авторской работе [2] проведен обзор и анализ существующих подходов выбора критериев сравнения веб-сервисов, подробно рассмотрен вопрос оценки чувствительности, дается формальное описание понятия чувствительности веб-сервисов, введены классы чувствительности и описана методика их выделения.

Основной мотивацией ввода нового критерия является отсутствие описания в современных научных публикациях таких критериев, которые описывали бы веб-сервис в долгосрочной перспективе, что важно в ситуациях, когда необходимо быть уверенным в обеспечиваемом уровне производительности при увеличении нагрузки. Определение такого критерия может быть полезно как при решении задачи рационального выбора архитектуры систем с СОА, так и при составлении документов, регламентирующих соглашение об уровне предоставления услуги.

Ввод разделения веб-сервисов на классы мотивирован большим числом критериев, характеризующих чувствительность, и вариациями их возможных значений. В таком случае формирование экспертом наборов продукционных правил, помогающих выделять отдельные классы, является трудоёмкой задачей. Поэтому целесообразно использовать методы машинного обучения, которые хорошо подходят для решения задач в случае неприменимости точных алгоритмов.

В настоящей работе описывается усовершенствованный ме-

тод выделения классов чувствительности и метод автоматизированного определения класса чувствительности произвольного веб-сервиса с помощью алгоритма машинного обучения «с учителем» — метода опорных векторов (англ. *Support vector machine, SVM*).

Работа состоит из трех основных частей. В первой части описан усовершенствованный метод выделения классов чувствительности. Во второй части рассмотрен вопрос автоматизированного определения класса чувствительности произвольного веб-сервиса и приведены результаты вычислительного эксперимента. В заключительной части сделаны выводы по проведенной работе и обозначены перспективные направления исследований.

Исследования проведены при финансовой поддержке РФФИ в рамках научного проекта № 12-07-31214 мол_а.

1. Выделение классов чувствительности

1.1. ОСНОВНЫЕ ОПРЕДЕЛЕНИЯ

Чувствительность веб-сервиса — критерий, на основе которого может быть оценена возможность обеспечения определенного уровня производительности веб-сервиса при возрастающей нагрузке. Таким образом, для определения чувствительности необходимо произвести тестирование веб-сервиса по *определенному плану* и вычислить значения ряда признаков, характеризующих чувствительность, используя результаты теста. Поскольку значения признаков различных веб-сервисов имеют большой разброс, целесообразно ввести *классы чувствительности*, характеризующие диапазоны значений признаков.

1.2. ФОРМАЛЬНАЯ ПОСТАНОВКА ЗАДАЧИ

Дано X — множество описаний веб-сервисов, Y — конечное множество классов чувствительности. Существует неизвестная целевая зависимость — отображение $y^*: X \rightarrow Y$, значения которой известны только на объектах конечной обучающей выборки X^m и соответствующих им классам y^m . Требуется построить алгоритм $a: X \rightarrow Y$, способный классифицировать произвольный

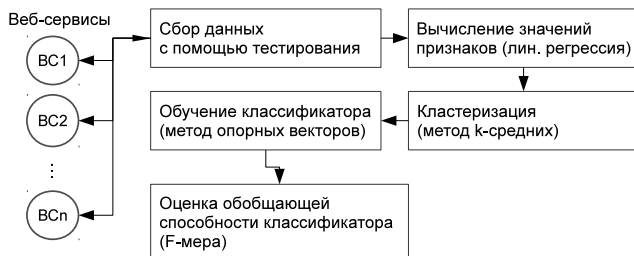


Рис. 1. Схематическая диаграмма предлагаемого комбинированного метода определения чувствительности

объект $x \in X$. Говоря менее формально, необходимо, используя предварительно обученный алгоритм классификации, присвоить класс чувствительности новому веб-сервису.

На рис. 1 представлено схематическое описание предлагаемого метода решения поставленной задачи. Конечная обучающая выборка X^m и соответствующие классы y^m вычисляются следующим образом:

- 1) Составляется случайная выборка веб-сервисов объемом m (раздел 1.3).
- 2) Вводится описание ряда признаков, характеризующих чувствительность веб-сервисов, и способов их вычисления (раздел 1.5).
- 3) Проводится тестирование выборки по определенному плану (раздел 1.4).
- 4) На основе результатов тестов вычисляется матрица «объект–признак» X^m .
- 5) При помощи визуализации матрицы X^m экспертом проводится эвристическое выделение схожих по значению показателей подмножеств веб-сервисов (раздел 1.6).
- 6) Формируются классы чувствительности на основе упорядоченных по предпочтению подмножеств веб-сервисов, при

этом во вводимой градации классов число ошибок является более значимым критерием (раздел 1.6).

- 7) Составляется множество меток y^m принадлежности веб-сервисов X^m к определенным классам чувствительности.

Пункт 7 при достаточно большом объеме выборки является крайне трудозатратным, поэтому целесообразно автоматизировать этот процесс, проведя кластеризацию данных с числом кластеров, равным числу классов чувствительности. Также результаты кластеризации могут служить индикатором правильности эвристического разделения на классы: если полученные кластеры достаточно обособлены друг от друга, то классы выбраны верно. Кластеризация проводится методом k -средних [6].

Получив матрицу «объект–признак» X^m и множество соответствующих классов y^m , возможно обучить и оценить алгоритм классификации «с учителем». В качестве такого алгоритма используется метод опорных векторов [11].

Далее поэтапно описывается весь процесс выделения классов чувствительности и организации автоматизированной классификации.

1.3. ФОРМИРОВАНИЕ ВЫБОРКИ

Для вычислительного эксперимента выборка была сформирована на основе данных из каталога API Directory [8], содержащего информацию о более чем 5000 различных веб-сервисах. Выбираются разнородные сервисы, реализующие функции картографии и геокодинга, предоставляющие информацию о различных показателях торговых бирж, о погоде, новостях и т.д. Все сервисы реализуют архитектуру REST [5], предоставляя свои функции по протоколу HTTP.

В работе используется выборка объемом $n = 100$ веб-сервисов.

1.4. ПЛАН ТЕСТА

Для вычисления признаков необходимо провести сбор необходимых данных посредством тестирования выборки веб-

сервисов. В процессе тестирования осуществляется последовательное выполнение итераций. Одна итерация длится 1 секунду. Количество запросов, отправляемых в течение одной итерации, зависит от её номера (см. далее). Запросы равномерно распределены в рамках секунды. В среднем один тест длится 15–25 минут в зависимости от скорости ответа на запросы веб-сервисом. Каждый веб-сервис тестируется 3 раза с разницей между тестами в 9 часов, результаты проведенных тестов усредняются. Такой подход к тестированию позволяет снизить разброс значений, возникающий из-за различной загруженности веб-сервисов в течение дня.

Пусть λ_{max} — число запросов в секунду, отправляемое в последней итерации, S — шаг теста (число, на которое увеличивается количество запросов в секунду в последующей итерации). Тогда можно вычислить общее число итераций в тесте N_{iter} :

$$(1) \quad N_{iter} = \left\lceil \frac{\lambda_{max}}{S} \right\rceil.$$

Пусть $\bar{r}_{j,k}^{(i)}$ — усреднённое по трем тестам значение времени обработки запроса, где $i = 1, \dots, n$ — номер веб-сервиса, n — объем выборки, $j = 1, \dots, N_{iter}$ — номер итерации, $k = 1, \dots, j \cdot S$ — номер запроса в рамках j -й итерации.

В работе используется следующий план теста: максимальное число запросов в секунду $\lambda_{max} = 300$, шаг теста $S = 10$, объем выборки $n = 100$. Такие значения параметров выбраны экспериментально, так как было установлено, что начиная с 280–290 запросов в секунду большинство веб-сервисов демонстрируют устойчивое поведение. Выборка состоит из 100 веб-сервисов: Google Maps, Яндекс Карты, Bing Maps, Nokia Maps, Twitter, Factolex, Quora и др.

1.5. ВЫЧИСЛЕНИЕ МАТРИЦЫ «ОБЪЕКТ-ПРИЗНАК»

Матрица «объект–признак» $\widetilde{X}^m \in \mathbb{R}^{m \times n}$, где m — число объектов, n — количество признаков, формируется на основе результатов тестов веб-сервисов. Ряды матрицы соответствуют объектам (веб-сервисам), столбцы — признакам. Обозначим $x^{(i)}$ i -й ряд

матрицы X . Опишем каждый веб-сервис следующим образом:

$$(2) \quad \mathbf{x}^{(i)} = \left[\mathbf{t}^{(i)}, \mathbf{d}^{(i)}, \mathbf{e}^{(i)} \right],$$

где $\mathbf{t}^{(i)}$ — вектор, описывающий изменение среднего времени обработки запросов i -м веб-сервисом при изменении нагрузки, $\mathbf{d}^{(i)}$ — вектор, описывающий изменение стандартного отклонения, $\mathbf{e}^{(i)}$ — вектор, описывающий изменение числа ошибок.

По каждому веб-сервису сначала вычислим среднее время обработки запросов, стандартное отклонение и число необработанных запросов по итерациям. Далее с целью устранения ограничения на однозначный план теста (шаг теста и максимальное число запросов в секунду) и более экономного описания характеристики чувствительности веб-сервиса произведем аппроксимацию значений признаков по итерациям полиномом первой степени с помощью алгоритма линейной регрессии [3]. Значения коэффициентов полученной регрессионной модели, а также суммарную квадратичную ошибку будем использовать в качестве признаков, характеризующих чувствительность веб-сервиса.

В дальнейших вычислениях среднего времени обработки и стандартного отклонения исключаются ошибочные запросы, т.е. такие запросы, которые не были обработаны в течение 30 секунд (время timeout) или веб-сервис вернул в ответе на запрос HTTP-код, отличный от «200 — ОК».

Примем за $t_j^{(i)}$ среднее время обработки запросов i -м веб-сервисом в рамках j -й итерации:

$$(3) \quad t_j^{(i)} = \frac{1}{j \cdot S} \sum_{k=0}^{j \cdot S} \bar{r}_{j,k}^{(i)}.$$

Воспользуемся регрессионной моделью вида

$$h_t(j \cdot S)^{(i)} = \theta_{t,0}^{(i)} + \theta_{t,1}^{(i)} \cdot (j \cdot S).$$

Параметр $\theta_{t,0}^{(i)}$ является показателем ординаты точки пересечения прямой с осью ординат, $\theta_{t,1}^{(i)}$ — тангенс угла наклона прямой. Необходимо подобрать такие параметры $\theta_t^{(i)}$, чтобы регрессионная модель наиболее точно описывала исходную зависи-

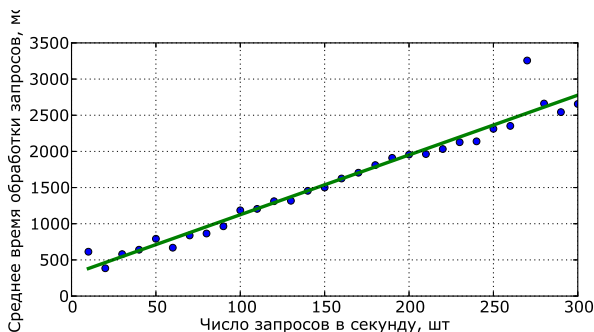


Рис. 2. Аппроксимация среднего времени обслуживания запросов от числа запросов в секунду. Точки — исходные данные, линия — аппроксимация

мость среднего времени обслуживания запросов от числа запросов в секунду. Подбор параметров производится с помощью метода наименьших квадратов. Суть метода в минимизации значения стоимостной функции, равной разности между фактическим значением $t_j^{(i)}$ и вычисленным регрессионной моделью значением $h_t(x)$:

$$(4) \quad \min_{\theta_0, \theta_1} J_t(\theta_t^{(i)}) = \sum_{j=1}^{N_{iter}} (t_j^{(i)} - h_t(j \cdot S))^{(i)2}.$$

С целью уменьшения диапазона возможных значений этого признака преобразуем тангенс угла наклона аппроксимирующей прямой, $\theta_{t,1}^{(i)}$, в градусы:

$$(5) \quad \alpha_t^{(i)} = \arctan(\theta_{t,1}^{(i)}) \frac{180}{\pi}.$$

С целью вычисления относительных величин дополним стоимостную функцию постоянным делителем, равным максимальному числу запросов. В итоге в качестве признаков, характеризующих изменение среднего времени обслуживания запросов при увеличении нагрузки, примем

$$(6) \quad \mathbf{t}^{(i)} = \left[\alpha_t^{(i)}, \theta_{t,0}^{(i)}, \frac{J_t(\theta_t^{(i)})^{(i)}}{\lambda_{max}} \right].$$

На рис. 2 синими кругами показаны усредненные по трем тестам данные веб-сервиса картографии «Яндекс.Карты», зеленой линией — аппроксимирующий полином. Видно, что в данном случае полином хорошо описывает тенденцию роста среднего времени обслуживания запросов при росте числа запросов в секунду, поэтому значение стоимостной функции $J(\theta_t^{(i)})$ будет мало. Значения признаков для примера на рис. 2: $\alpha_t^{(i)} = 83, 1$; $\theta_{t,1} = 298, 1$; $J_t(\theta_t)/\lambda_{max} = 80, 9$.

Аналогично проведем вычисления для стандартного отклонения времени обработки запросов:

$$(7) \quad d_j^{(i)} = \sqrt{\frac{1}{j \cdot S - 1} \sum_{k=0}^{j \cdot S} (\bar{r}_{j,k}^{(i)} - t_j^{(i)})^2},$$

$$\mathbf{d}^{(i)} = \left[\alpha_d^{(i)}, \theta_{d,0}^{(i)}, \frac{J_d(\theta_d^{(i)})^{(i)}}{\lambda_{max}} \right].$$

И для числа ошибок:

$$(8) \quad e_j^{(i)} = \sum_{k=0}^{j \cdot S} s_k, s_k = \begin{cases} 1 & \text{если запрос } \bar{r}_{j,k}^{(i)} \text{ не обработан,} \\ 0 & \text{если запрос } \bar{r}_{j,k}^{(i)} \text{ обработан;} \end{cases}$$

$$\mathbf{e}^{(i)} = \left[\alpha_e^{(i)}, \theta_{e,0}^{(i)}, \frac{J_e(\theta_e^{(i)})^{(i)}}{\lambda_{max}} \right].$$

В итоге каждый веб-сервис описывается следующим набором признаков:

$$\mathbf{x}^{(i)} = \left[\alpha_t^{(i)}, \theta_{t,0}^{(i)}, \frac{J_t(\theta_t^{(i)})^{(i)}}{\lambda_{max}}, \alpha_d^{(i)}, \theta_{d,0}^{(i)}, \frac{J_d(\theta_d^{(i)})^{(i)}}{\lambda_{max}}, \alpha_e^{(i)}, \theta_{e,0}^{(i)}, \frac{J_e(\theta_e^{(i)})^{(i)}}{\lambda_{max}} \right].$$

Нормализуем полученные данные. Для этого зададим функцию нормализации z_j :

$$(9) \quad z_j : x_j^{(i)} \mapsto \frac{x_j^{(i)} - E_j[x_j^{(i)}]}{\mu_j(x_j^{(i)})},$$

где E_j и μ_j — среднее и стандартное отклонение по j -му признаку соответственно. Примем за $X^m = z(\widetilde{X^m})$ нормализованную по признакам матрицу «объект-значение».

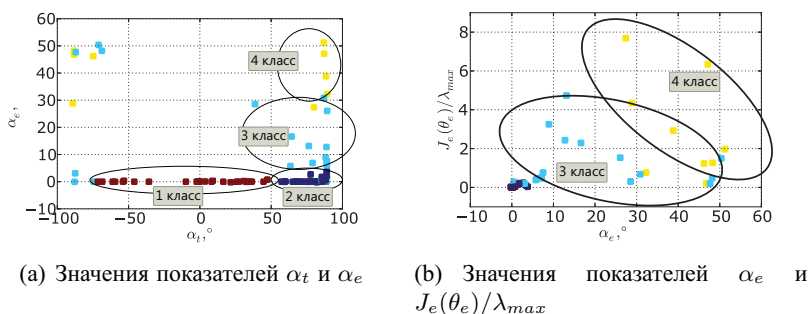


Рис. 3. Значения показателей выборки веб-сервисов, сгруппированных по кластерам

1.6. ОПИСАНИЕ КЛАССОВ ЧУВСТВИТЕЛЬНОСТИ

После проведения теста по заданному плану и вычисления матрицы «объект–признак» X^m можно визуализировать полученные данные. На рис. 3 различными цветами обозначены различные классы чувствительности, полученные в результате этапа кластеризации. Овалами обведены *примерные* границы классов веб-сервисов. Проиллюстрировать точные границы на данном типе графиков невозможно ввиду изображения только 2 признаков на каждом графике, в то время как каждый веб-сервис характеризуется 9 показателями.

Введем эвристическое разделение веб-сервисов по классам чувствительности: от предпочтительной низкой чувствительности к высокой.

Первый класс — низкая чувствительность — характеризуется *небольшим изменением среднего времени обслуживания запросов и стандартного отклонения, а также отсутствием необработанных запросов*. На рис. 3(а) можно увидеть, что ряд веб-сервисов имеют *отрицательное* значение признака α_t , угла аппроксимирующей прямой графика среднего времени обработки запроса, — это означает, что начальное среднее время обработки запросов при малой нагрузке *больше* конечного. Такое поведение характерно для ряда «облачных» веб-сервисов (при увели-

чении нагрузки динамически увеличивается мощность обслуживающего узла) или веб-сервисов с адаптивным распределителем нагрузки (англ. *Load balancer*) (при высокой утилизации ресурсов одного обслуживающего узла часть запросов передается на обслуживание узлам с меньшей утилизацией).

Второй класс характеризуется *более быстрым повышением среднего времени обслуживания запросов, стандартного отклонения и наличием небольшого числа необработанных запросов.*

Третий класс характеризуется *быстрым повышением среднего времени обслуживания запросов и стандартного отклонения, а также большим количеством необработанных запросов.*

Четвертый класс — высокая чувствительность — характеризуется в первую очередь *наибольшим числом необработанных запросов.* При небольшом повышении нагрузки быстро растет число необработанных запросов, среднее время обработки запросов и стандартное отклонение.

1.7. ФОРМИРОВАНИЕ МНОЖЕСТВА КЛАССОВ С ПОМОЩЬЮ КЛАСТЕРИЗАЦИИ

С целью проверки обоснованности эвристического разделения веб-сервисов, а также для последующей автоматизации процесса определения класса чувствительности используется кластеризация данных с помощью алгоритма k -средних [6]. В качестве входных данных используются нормализованная матрица «объект–признак» X^m , 4 центроида (по количеству эвристически определенных классов), первоначальное положение центроидов выбирается случайно, всего проводится 200 итераций, в качестве меры расстояния используется манхэттенское расстояние:

$$(10) \quad d_{ij} = |x_i - x_j|.$$

На рис. 3(а) и рис. 3(б) обозначены эвристически выделенные классы чувствительности. Принимая во внимание эвристическое описание классов чувствительности, можно графически обозначить классы на рисунке.

2. Определение класса чувствительности

2.1. КРАТКОЕ ОПИСАНИЕ МЕТОДА ОПОРНЫХ ВЕКТОРОВ

Решение задачи бинарной классификации при помощи метода опорных векторов заключается в поиске параметров функции, описывающей гиперплоскость, которая правильно разделяет набор данных на два класса. Гиперплоскость выбирается таким образом, чтобы расстояние от объектов разных классов до гиперплоскости было максимальным. Формально эту задачу оптимизации с ограничениями можно записать следующим образом:

$$\begin{aligned} \min_{\gamma, \mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^m \xi_i \\ y^{(i)}(\mathbf{w}^T \mathbf{x}^{(i)} + b) \geq 1 - \xi_i, i = 1, \dots, m, \\ \xi_i \geq 0, i = 1, \dots, m, \end{aligned}$$

где $\mathbf{x}^{(i)}$ — вектор признаков i -го объекта; $y^{(i)}$ — метка принадлежности i -го объекта к определенному классу; \mathbf{w} — направляющий гиперплоскость вектор; b (от англ. «bias» — смещение) — кратчайшее расстояние между разделяющей гиперплоскостью и началом координат; ξ — дополнительные переменные, характеризующие величину ошибки. Если $\xi_i > 0$, то предполагается, что объект x_i может лежать внутри разделяющего зазора или ошибочно отнесен не к тому классу. Коэффициент C — коэффициент регуляризации, который позволяет изменять отношение между максимизацией ширины разделяющей полосы и минимизацией суммарной ошибки классификации.

Описанный метод подходит для случаев, когда возможно линейно разделить классы. Однако на практике такие случаи встречаются нечасто. Для обеспечения возможности нелинейного разделения классов метод опорных векторов был дополнен ядрами (kernel trick). Суть дополнения заключается в переводе исходных объектов в расширенное пространство, где возможно линейное разделение на классы. В настоящей работе используется ради-

альная базисная функция Гаусса (далее РБФ):

$$K(x, x') = \exp(-\gamma \|x - x'\|^2).$$

Если необходимо разделить данные на большее число классов, применяют процедуру обучения нескольких классификаторов по принципу «один против всех». В этом случае каждый классификатор определяет гиперплоскость, разделяющую данные одного класса от всех остальных.

2.2. ВЫБОР И ОЦЕНКА АЛГОРИТМА КЛАССИФИКАЦИИ

В качестве алгоритма классификации используется метод опорных векторов. Необходимо:

- 1) Выбрать подходящую оценку обобщающей способности алгоритма (раздел 2.3).
- 2) Выбрать наилучшую комбинацию ядра алгоритма классификации и его параметров по показателям полноты и точности с помощью процедур скользящего контроля (кросс-валидации) и сеточного поиска (англ. *grid search*) (раздел 2.4).

2.3. ОЦЕНКА ОБОБЩАЮЩЕЙ СПОСОБНОСТИ АЛГОРИТМА

На практике многие классы чувствительности являются смещенными (англ. *scewed*), т.е. включают себя малое число объектов по сравнению с другими классами. Например, в выборке, используемой в настоящей работе, объемом 100 веб-сервисов к 3-му классу относится только 4 объекта, в то время как к 1-му — 40. Поэтому целесообразно в качестве оценки обобщающей способности алгоритма использовать не просто отношение верно классифицированных веб-сервисов к их общему числу, но оценки обобщающей способности по каждому классу. Такой оценкой может служить F -мера [10].

Введем основные определения. Пусть C_i — множество объектов, принадлежащих i -му классу, u_i — множество объектов, которые алгоритм отнес к i -му классу. Тогда полнотой (англ. *recall*)

классификации объектов по i -му классу является отношение количества объектов, правильно отнесенных к i -му классу, к общему количеству объектов, относящихся к этому классу:

$$(11) \quad r_i(u_i) = \frac{|u_i \cap C_i|}{|C_i|}.$$

Точностью (англ. *precision*) классификации объектов по i -му классу является отношение количества объектов, правильно отнесенных к i -му классу, к общему количеству объектов, отнесенных к i -му классу:

$$(12) \quad p_i(u_i) = \frac{|u_i \cap C_i|}{|u_i|}.$$

Идеальный алгоритм обеспечивает 100% полноту и точность.

Для удобства полноту и точность сводят к одной оценке, называемой F -мерой (англ. *F-score*, F_1 -score):

$$(13) \quad F_i(u_i) = 2 \times \frac{p_i(u_i) \times r_i(u_i)}{p_i(u_i) + r_i(u_i)}.$$

Для идеального классификатора F -мера равна 1, для худшего — 0.

F -меру по всем классам можно записать как

$$(14) \quad F = \sigma \pm \mu,$$

где σ — среднее арифметическое, μ — среднеквадратическое отклонение по всем F_i .

2.4. СКОЛЬЗЯЩИЙ КОНТРОЛЬ И СЕТОЧНЫЙ ПОИСК

Скользящий контроль — процедура эмпирического оценивания обобщающей способности алгоритмов, обучаемых по прецедентам («с учителем») [1]. В общем случае исходная выборка разбивается на множество сочетаний двух подвыборок: обучающей и контрольной. Для каждого разбиения выполняется настройка алгоритма на обучающей подвыборке, затем оценивается ошибка на контрольной подвыборке. Оценкой скользящего контроля называется средняя по всем разбиениям величина ошибки на контрольных подвыборках.

В настоящей работе используется k -кратный скользящий контроль (англ. *k-fold cross-validation*), который заключается в

следующем: выборка случайным образом разбивается на k непесекающихся подмножеств одинаковой (или почти одинаковой) длины. Процедура кросс-валидации начинается с выбором первого подмножества как контрольного, оставшиеся $k - 1$ подмножеств используются как обучающая выборка. Процедура повторяется k раз, при этом каждое подмножество по одному разу используется как контрольное; k результатов оценки точности и полноты алгоритма усредняются.

Дополним k -кратный скользящий контроль техникой сеточного поиска (англ. *grid search*) [4]. Сеточный поиск — достаточно простой, но эффективный метод оптимизации модели алгоритма и значений параметров. Задаются множества возможных моделей и значений параметров и с помощью полного перебора возможных сочетаний выбирается такое сочетание, для которого оценка обобщающей способности (в нашем случае это F -мера) по итогам скользящего контроля будет наивысшей.

2.5. ВЫЧИСЛИТЕЛЬНЫЙ ЭКСПЕРИМЕНТ

Проведем процедуры скользящего контроля и сеточного поиска с целями выбора наилучшей комбинации параметров и оценки обобщающей способности алгоритма.

Разделим случайным образом исходную выборку на два подмножества: 80% — подмножество, используемое в процедурах сеточного поиска и скользящего контроля, 20% — контрольное подмножество.

Зададим $k = 3$ для процедуры k -кратного скользящего контроля.

Зададим множество параметров, среди которых необходимо найти наилучшую комбинацию с помощью метода сеточного поиска и кросс-валидации.

1) Для линейного ядра: $C = [1, 10, 100, 1000]$.

2) Для радиальной базисной функции: $\gamma = [0, 001; 0, 0001]$, $C = [1, 10, 100, 1000]$.

Используя библиотеку с открытым исходным кодом, содержащую множество алгоритмов машинного обучения, Scikit Learn 126

[9], осуществим вычислительный эксперимент с заданными параметрами. Результаты сеточного поиска и кросс-валидации (в качестве оценки используется средняя F -мера по всем классам) показаны в таблице 1.

Таблица 1. Результаты сеточного поиска и кросс-валидации

F -мера	Ядро	Параметры ядра
$0,303 \pm 0,010$	РБФ	$C = 1, \gamma = 0,001$
$0,303 \pm 0,010$	РБФ	$C = 1, \gamma = 0,0001$
$0,470 \pm 0,010$	РБФ	$C = 10, \gamma = 0,001$
$0,303 \pm 0,010$	РБФ	$C = 10, \gamma = 0,0001$
$0,799 \pm 0,019$	РБФ	$C = 100, \gamma = 0,001$
$0,470 \pm 0,010$	РБФ	$C = 100, \gamma = 0,0001$
$0,862 \pm 0,013$	РБФ	$C = 1000, \gamma = 0,001$
$0,799 \pm 0,019$	РБФ	$C = 1000, \gamma = 0,0001$
$0,862 \pm 0,022$	Линейное	$C = 1$
$0,848 \pm 0,013$	Линейное	$C = 10$
$0,787 \pm 0,013$	Линейное	$C = 100$
$0,787 \pm 0,013$	Линейное	$C = 1000$

Лучший набор параметров наблюдается для ядра РБФ с параметрами $C = 1000, \gamma = 0,001$. Проведем оценку классификатора с заданными параметрами на контрольном подмножестве. Стоит отметить, что результаты сеточного поиска и скользящего контроля, а также оценка F -меры на контрольном подмножестве зависят от первоначального случайного разбиения всей выборки на подмножества, поэтому конечное значение общей F -меры может варьироваться. После нескольких повторений всей процедуры обучения классификатора оценка на контрольном подмножестве F -мера = $0,955 \pm 0,025$.

Полученные результаты можно представить визуально. Для этого уменьшим размерность пространства признаков с 9 до 2 с помощью метода главных компонент [7], настроим параметры классификатора, используя полученное пространство признаков,

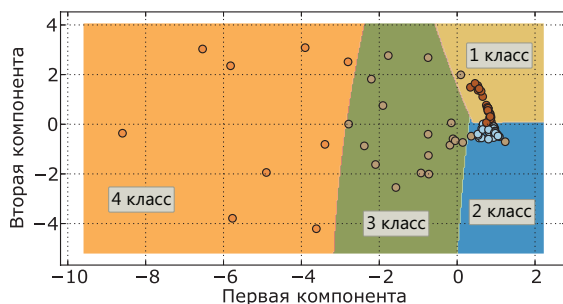


Рис. 4. Границы принятия решения классификатора по классам

и построим границы принятия решений по классам (рис. 4).

Методика определения класса чувствительности веб-сервиса демонстрирует определенную устойчивость. В результате вычислительного эксперимента каждый веб-сервис проходил тестирования 3 раза с разницей в 9 часов между тестами. В дальнейшем для проверки гипотезы устойчивости метода каждый результат теста использовался как входные данные для определения чувствительности. 83% всех веб-сервисов показывают одинаковое значение чувствительности за все три измерения, остальные 17% веб-сервисов только единожды изменяют класс, т.е. присутствовали последовательности классов вида 1 2 1, 3 2 2 и др.

3. Выводы

В работе проведен анализ и предложено решение проблемы получения численной характеристики чувствительности веб-сервиса. Показана практическая возможность эффективного выделения классов чувствительности, приведено эвристическое описание классов. Сформирована теоретическая и алгоритмическая основа программного комплекса для определения класса чувствительности произвольного веб-сервиса.

Преимуществами предложенного подхода являются независимость показателей от выбранного шага тестирования и невысокое начальное значение максимального числа запросов в се-

кунду. Таким образом методика может быть эффективна в отсутствии возможности полноценного тестирования веб-сервиса при высокой нагрузке.

Видится перспективным продолжение работ в данном направлении. На данный момент отсутствуют методики и программное обеспечение, которого могло бы провести полноценную оценку веб-сервиса и дать рекомендации по набору используемых веб-сервисов в системе с сервисно-ориентированной архитектурой при наличии функциональных и нефункциональных требований и предпочтений. Поэтому, как правило, при решении таких вопросов руководствуются экспертными рекомендациями веб-сервисов в виду отсутствия каких-либо строгих методик оценки.

Литература

1. ВОРОНЦОВ К.В. *Комбинаторный подход к оценке качества обучаемых алгоритмов* // Математические вопросы кибернетики / Под ред. О.Б. Лупанова. — Москва : Физматлит, 2004. — Т. 13. — С. 5–36. — [Электронный ресурс]. — URL: www.ccas.ru/frc/papers/voron04mpc.pdf (дата обращения: 30.09.2013).
2. ДУШКИН Д.Н. *Анализ чувствительности веб-сервисов в задаче выбора оптимальной конфигурации систем с сервисно-ориентированной архитектурой* // Управление большими системами. — 2012. — No. 40. — С. 164–182.
3. СТРИЖОВ В.В. *Методы выбора регрессионных моделей* — Москва: ВЦ РАН, 2010. — 60 с. — [Электронный ресурс]. — URL: <http://www.machinelearning.ru/wiki/images/5/52/Strijov-Krymova10Model-Selection.pdf> (дата обращения: 30.09.2013).
4. BERGSTRA J., BENGIO Y. *Random Search for Hyper-Parameter Optimization* // J. Mach. Learn. Res. — 2012. — Vol. 13. — P. 281–305. — [Электронный ресурс]. — URL:

- <http://dl.acm.org/citation.cfm?id=2188385.2188395> (дата обращения: 30.09.2013).
5. FIELDING R.T. *Architectural styles and the design of network-based software architectures*: Ph. D. thesis / Roy Thomas Fielding. — University of California, Irvine, 2000. — 162 p.
 6. MACQUEEN J.B. *Some Methods for Classification and Analysis of MultiVariate Observations* // Proc. 5th Berkeley Symposium on Mathematical Statistics and Probability / Ed. by L.M. Le Cam, J. Neyman. — Vol. 1. — University of California Press, 1967. — P. 281–297.
 7. PEARSON K. *On lines and planes of closest fit to systems of points in space* // Philosophical Magazine. — 1901. — Vol. 2, No. 6. — P. 559–572.
 8. *Programmable Web. API Directory*. — [Электронный ресурс]. — URL: <http://www.programmableweb.com/apis/directory> (дата обращения: 30.09.2013).
 9. *Scikit-learn: Machine Learning in Python* / F. Pedregosa, G. Varoquaux, A. Gramfort et al. // Journal of Machine Learning Research. — 2011. — Vol. 12. — P. 2825–2830.
 10. VAN RIJSBERGEN C.J. *Information Retrieval*. — 2nd edition. — Newton, MA, USA : Butterworth Heinemann, 1979. — ISBN: 0408709294.
 11. VAPNIK V.N. *The nature of statistical learning theory*. — New York, NY, USA : Springer-Verlag New York, Inc., 1995. — ISBN: 0-387-94559-8.
 12. WIRSING M., HOLZL M., KOCH N. AND OTHER *SENSORIA — Software Engineering for Service-Oriented Overlay Computers* // Rigorous Software Engineering for Service-Oriented Systems Lecture Notes in Computer Science. — 2011. — Vol. 6582. — P. 1–14.

AUTOMATED DETERMINATION OF SENSITIVITY CLASSES OF WEB SERVICES

Dmitry Dushkin, Institute of Control Sciences of RAS, Moscow,
Postgraduate (ddushkin@asmon.ru).

Abstract: We presents a method to calculate the sensitivity, which is a benchmark for web services performance. We introduce metrics which characterize sensitivity, describe the routine for their evaluation, and suggest the method of class definition for an arbitrary web service. The method uses several machine learning algorithms such as clustering via k-means and classification via support vector machines. At the end we present the results of a numerical experiment, which illustrates our method.

Keywords: web-services, service-oriented arcitecture, machine learning, sensitivity analysis.

*Статья представлена к публикации
членом редакционной коллегии М. В. Губко*

Поступила в редакцию 03.02.2013.

Опубликована 30.09.2013.

УДК 021.8 + 025.1
ББК 78.34

РИСКИ И ВЫБОР ОПТИМАЛЬНЫХ ПРОЕКТОВ: СЕРВИС-ОРИЕНТИРОВАННАЯ АРХИТЕКТУРА ИНФОРМАЦИОННЫХ СИСТЕМ

Пырлина И. В.¹

(ГУ Высшая школа экономики, Москва)

В статье предложен обзор известных моделей выбора проектов реализации информационных систем с сервис-ориентированной архитектурой (СОА), методов оценки рисков и многокритериального анализа информационных систем с различными типами архитектур, включая СОА. Статья также описывает предложенные методы анализа рисков от СОА и метод многокритериального анализа проектов информационных систем с СОА. Применение данных методов представлено с использованием прототипа инструмента, который состоит из двух подсистем – подсистема имитационного моделирования и подсистема многокритериального анализа.

Ключевые слова: сервис-ориентированная архитектура, операционные риски при сервис-ориентированном подходе, критерии эффективности архитектурного проекта, метод порогового агрегирования, ранжирование проектов.

1. Введение

В современных предприятиях сервис-ориентированная архитектура информационных систем (ИС) используется для создания более открытой и гибкой информационной среды предприятия, более сбалансированного ландшафта приложений

¹ Ирина Владимировна Пырлина, специалист, эксперт ИТ-архитектуры (irina.pyrlina@gmail.com).

без дублирования функциональности. Под функциональностью в данном случае понимается совокупность решаемых задач в рамках информационной системы, предоставленных пользователю в виде вызываемых функций. Она позволяет адаптировать инновации, среди которых мобильные технологии и «облачные» системы. Проблема выбора проекта внедрения информационной системы с сервис-ориентированной архитектурой становится более актуальной с появлением различных вариантов ее реализации от поставщиков решений.

Под проектом [6] понимается ограниченное во времени целенаправленное изменение отдельного приложения или информационной системы, а также реализация новой информационной системы с сервис-ориентированной архитектурой с определенными целями, достижение которых означает завершение проекта, а также с установленными требованиями к результату, срокам, рискам, бюджету и ресурсам. Фокусируясь больше на технической стороне реализации SOA, предприятия пренебрегают другими, не менее важными аспектами этой концепции. Однако успешность проекта реализации SOA зависит от учета всех аспектов готовности организационной архитектуры, специалистов и механизмов управления. Поэтому выбор проекта реализации ИС с сервис-ориентированной архитектурой на предприятии должен учитывать эти немаловажные аспекты.

В данной работе предлагается метод многокритериального анализа информационных систем с сервис-ориентированной архитектурой, который должен использоваться любым предприятием, принимающим решения по корпоративной архитектуре и планирующим портфель ИТ-проектов. Несмотря на то, что мы находимся в начале периода повсеместного использования новой архитектуры, если верить Gartner [29], в ландшафте большинства компаний все еще есть разнообразие систем и архитектур. Причинами отказа от перехода на архитектуру SOA, как показано в [29], являются в первую очередь недостаток экспертизы SOA, отсутствие или непонимание выгод для бизнеса, стоимость и недостаток необходимых для внедрения ресурсов, рискованность перехода

на относительно новую концепцию, сложность архитектуры, и отсутствие необходимости. Поэтому важным вопросом сегодня является разработка методики выбора проектов реализации информационных систем с сервис-ориентированной архитектурой, также как и обоснование эффективности и рисков решения. Рассмотрим предлагаемые подходы оценки рисков и выбора оптимального проекта.

2. Выбор оптимального проекта информационной системы с сервис-ориентированной архитектурой

Прежде чем перейти к определению методов выбора проектов, введем несколько ключевых понятий. Под информационной системой [7] полагаются «аппаратно-программные системы, которые поддерживают приложения с интенсивной обработкой данных». При этом приложением информационной системы (или прикладной программой) в данном случае считается «настройка над информационной системой, обеспечивающая решение некоторого комплекса задач».

Информационная система также может являться компонентой (или подсистемой) более сложной информационной системы [7]. Например, информационная система развития и управления персоналом и информационная система бухгалтерского учета являются компонентами информационной системы предприятия.

Информационные системы используют несколько ресурсов [7], к ним относятся средства вычислительной техники, системное и прикладное программное обеспечение, информационные, лингвистические и человеческие ресурсы.

Прикладные ресурсы (или прикладное программное обеспечение) – прикладные программы, которые имеют возможность публикации/поддержки технических сервисов; к их числу относятся не только программное обеспечение (ПО), но и системы аппаратного обеспечения. Прикладное ПО [7] используется для решения конкретного класса задач и обычно в

качестве таких средств используются коммерческие программные продукты: СУБД общего назначения, Web-серверы, системы управления документами, информационные системы, имеющие узкое назначение (используемые для поддержки конкретных процессов предприятия).

Под системой управления базой данных (СУБД) понимается база данных со специализированным программным обеспечением, которое позволяет структурированно хранить и обрабатывать данные [6].

Подсистема доступа – различные технологии пользовательского интерфейса для работы с системой («тонкий» клиент на базе Web-технологий и «толстый» клиент в виде прикладных Windows-приложений). Пользовательские ресурсы [7] в интернете – это, например, страницы Web-сайтов, разные доступные пользователям Web-документы, страницы HTML и другие форматы интерфейса.

Инфраструктурное программное обеспечение (или системное) включает [7] различные операционные системы, системные оболочки, служебные программы для поддержки серверного оборудования и работы администратора, сетевое программное обеспечение и телекоммуникационное оборудование.

Кроме того, для успешного оперирования информационной системы необходимы и технические ресурсы, к которым относятся сервера, помещение, электричество и другие.

2.1. МЕТОД КЛАССИФИКАЦИИ И ОЦЕНКИ РИСКОВ СЕРВИС-ОРИЕНТИРОВАННОЙ АРХИТЕКТУРЫ

Операционные риски¹ в рамках информационных систем с сервис-ориентированной архитектурой тесно связаны с ошибками работы ресурсов, потребляемых этими приложениями, в частности с ресурсами бизнес-сервисов.

¹ *Общее определение термина «операционный риск» приводится в [13, стр.478].*

Оценка рисков ИС с сервис-ориентированной архитектурой производится на базе статистики ошибок, собранной по анализируемым проектам. Качество информационной системы напрямую выражается через количество ошибок в различных компонентах системы, соответственно и риски ИС с сервис-ориентированной архитектурой представляются в виде рисков возникновения ошибок компонентов информационной системы. Используем следующую классификацию рисков ИС с сервис-ориентированной архитектурой:

1. Риски ввода-вывода результатов.
2. Функциональные риски.
3. Риски связующего программного обеспечения (ПО).
4. Риски обработки данных.
5. Системные риски.

При классификации ошибок мы будем также указывать приоритет ошибки (1 – очень высокий, 2 – высокий, 3 – средний, 4 – низкий) и сложность устранения ошибки или длительность рассмотрения, так как то, как быстро была решена проблема, характеризует сложность проблемы. Длительность рассчитывается от даты подачи заявления до даты, когда проблема считается решенной, т.е. выставлен статус «completed». Обоснование такой классификации рисков приводится в [13].

Предлагается использовать риски как отдельную группу критериев, влияющих на выбор проекта реализации ИС с сервис-ориентированной архитектурой. Для проверки предложенной классификации было проанализировано 4435 ошибок по семи проектам за период с 2006-2010 гг. Все ошибки разделены на четыре приоритета по степени убытка и важности их устранения («очень высокий» – «высокий» – «средний» – «низкий»). Собранная статистика удовлетворяла нескольким типам функции распределения в зависимости от типа ошибок. В случае типов ошибок ввода-вывода результатов, ошибок данных и ошибок связующего звена четко определяется экспоненциальное распределение, что видно и по графикам функций вероятностей (рис 1). В то время как системные

ошибки имеют распределение Пуассона. А функциональные ошибки имеют нормальное распределение.

Под приоритетом понимается серьезность наносимого убытка работе информационной системы и предприятию в целом при возникновении ошибки. Предлагается различать 4 типа приоритетов:

1. Очень высокий – ошибки, которые имеют или оказывают критическое влияние на продуктивную эксплуатацию системы, на процессы и операции предприятия, а также могут привести к краху или полной остановке работы системы.

2. Высокий приоритет имеют ошибки, которые оказывают существенное влияние на процессы и операции предприятия, а также ведут к заметному снижению производительности системы.

3. Средний приоритет имеют ошибки, которые оказывают влияние на работу системы, и соответственно могут повлиять на процессы и операции, однако не являются критическими.

4. Низкий приоритет получают ошибки, прозрачные для пользователя. Например, ошибки пользовательского интерфейса не нарушают возможность использования системы, однако делают использование некомфортным и требуют устранения.

В рамках данного анализа применение метода порогового агрегирования для системы с большим числом приоритетов (типов ошибок по рискам) нецелесообразно. А преимущества максиминного метода (или метода Гермейера) по сравнению с методом линейной свертки приведены в [9]. Поэтому для ранжирования данных по рискам была применена максиминная свертка (1). Определим $N = \{n_1, n_2, \dots, n_k\}$ – множество сообщений об ошибках, тогда агрегирование рисков происходит по следующему правилу:

Определим матрицу S^+ такую что,

$$(1) \quad \forall x \in T, \forall y \in P, S^+ = \{n(x, y)\},$$

где T – множество типов рисков; P – множество приоритетов; R – пространство рангов; $x_i, i \in [1, |T|]$, – количество ошибок по типам рисков.

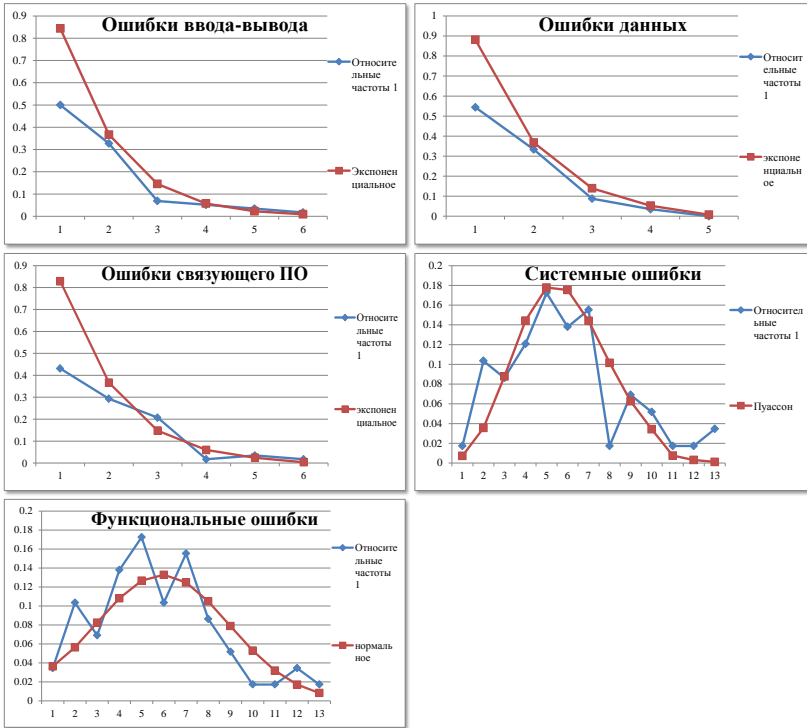


Рис. 1. Плотности вероятностей возникновения по типу рисков

В матрице S^+ на пересечении строки x и столбца y стоит число $n(x, y)$, которое равно количеству ошибок типа x по приоритету y . Выберем минимальное значение в каждой строке матрицы S^+ . А затем выберем наибольшее значение из выбранных. Результирующее «максиминное» значение (т.е. группа рисков) получает максимальный ранг. Далее повторим процедуру для оставшихся групп рисков. Таким образом, правило выглядит так: x признается лучшим и выбирается тогда и только тогда, когда

$$a(x) = \min_{b \in P} \{n(x, b)\},$$

где $a(x)$ – оценка, соответствующая риску x , которая позволяет определить упорядоченное отображение из множества типов рисков во множество рангов

$$(2) \quad T \rightarrow R: r[a(x)] \in \{\overline{1,4}\},$$

$$\text{такое, что } r[a(x)] = \begin{cases} 1, \max_{x \in T} \{a(x)\}; \\ 2, \max_{x \in T \setminus \{x_1\}} \{a(x)\}; \\ \dots \end{cases}$$

где r из R – ранг соответствующий некоторому типу рисков x .

Таким образом, в результате применения максиминного метода получим вектор значений с рангами. В случае максиминного критерия получается гарантированная нижняя оценка для всех $n(x, y)$ [9].

Таким образом, в данном случае $r(x_i)$ представляет собой вектор, значения которого приобретают ранг 1, 2, 3 или 4 в зависимости от того, какое из значений матрицы S^+ становится минимальным по приоритетам ошибок и одновременно максимальным по типам ошибок.

Переход от числовых оценок к рангам необходим для приведения результатов анализа рисков к виду, позволяющему сопоставление их с общими требованиями многокритериального анализа вариантов проектов по методу порогового агрегирования.

На основе метода перехода от числовых оценок к рангам самый высокий ранг получают функциональные риски для обоих проектов, описанный в [12]. При ранжировании были выбраны самые высокие из минимальных оценок, что говорит о наибольшем количестве ошибок и соответственно свидетельствует о наибольшем риске в соответствующем типе рисков. В данном случае первый (самый высокий) ранг присваивается наибольшей оценке риска, так как чем больше риск, тем хуже предлагаемый проект. Поэтому, в отличие от ранжирования по всем остальным критериям, которые будут рассматриваться в дальнейшем, в случае с рисками используется противоположный метод ранжирования: ранг 1

присваивается самому высокому риску, ранг 4 – самому низкому риску.

2.2. АНАЛИЗ ВАРИАНТОВ СЕРВИС-ОРИЕНТИРОВАННОЙ АРХИТЕКТУРЫ МЕТОДОМ ПОРОГОВОГО АГРЕГИРОВАНИЯ

Множество проектов информационных систем с сервис-ориентированной архитектурой оценивается по группам критериев. Подобные задачи принятия решения, содержащие множество критериев и возможных альтернатив решения, называют многокритериальными задачами [11]. В подобных задачах наиболее часто применяется метод линейной свертки, а также методы принятия решения с учетом относительной важности критериев [10]. Однако в ряде задач использование этих методов представляется неэффективным в силу «компенсаторного» характера критериев, т.е. возможности критериев при агрегировании компенсировать низкие оценки по одним критериям высокими оценками по другим. В отличие от упомянутых методов пороговый метод агрегирования [3, 16] основан на построении результирующего ранжирования по совокупности критериев, представленных четырьмя рангами. Агрегирование осуществляется с помощью порогового правила, которое позволяет построить бинарное отношение, определяющее предпочтения на множестве проектов.

В [12] описывается набор критериев для сервис-ориентированной архитектуры. Определенные группы критериев проекта, кроме рисков, ранжируются по четырех- или пятибалльной шкале. Группа рисков ранжируется сначала методом, описанным выше. А уже затем методом порогового агрегирования – по четырехбалльной шкале [16].

Формула преобразования ϕ из [16] для случая четырех рангов функции агрегирования имеет вид:

$$\phi(x) = \sum_{j=1}^2 C_{n-V_j(x)+3-j}^{4-j} + V_4(x) + 1 = \sum_{j=1}^2 \frac{(n - V_j(x) + 3 - j)!}{(4 - j)!(n - V_j(x))!} + V_4(x) + 1,$$

$$\phi(x) = \frac{(n - V_1(x) + 1)(n - V_1(x) + 2)}{6} + \frac{n - V_2(x) + 1}{2} + V_4(x) + 1,$$

где n – число критериев для сравнения альтернатив; $V_g(x)$ – число g -х оценок в записях векторов x или y ; $C_k^{k+1} = 0$ для всех $k \in [0, m-1]$, и $C_{-1}^0 = 1$.

Максимально возможное число классов эквивалентности этого слабого порядка при длине векторов, равной n , и при значении компонент вектора из $\{1, 2, 3, 4\}$ сводится к расчету числа неупорядоченных n выборок из 4 элементов с повторением:

$$\max \phi(x) = C_{n+3}^n = \frac{(n+3)!}{n!3!} = \frac{(n+1)(n+2)(n+3)}{6}.$$

А в случае пяти рангов функция агрегирования приобретает следующий вид:

$$\phi(x) = \sum_{j=1}^3 C_{n-V_j(x)+4-j}^{5-j} + V_5(x) + 1 = \sum_{j=1}^3 \frac{(n-V_j(x)+4-j)!}{(5-j)!(n-V_j(x))!} + V_5(x) + 1,$$

где n – число критериев для сравнения альтернатив; $V_g(x)$ – число g -х оценок в записях векторов x или y ; $C_k^{k+1} = 0$ для всех $k \in [0, m-1]$, и $C_{-1}^0 = 1$.

Максимальное число классов эквивалентности для пяти рангов приобретает вид

$$\max \phi(x) = C_{n+4}^n = \frac{(n+4)!}{n!4!} = \frac{(n+1)(n+2)(n+3)(n+4)}{24}.$$

Более подробно метод порогового агрегирования из [16] рассматривается в пункте 3.3.

В результате подход ранжирования проектов построения ИС с сервис-ориентированной архитектурой состоит из четырех этапов, приведенных на рис. 2.

Обследование	Первоначальное ранжирование	Повторное ранжирование	Результаты анализа
<ol style="list-style-type: none"> 1. Определение альтернатив 2. Анкетирование архитекторов решений <ul style="list-style-type: none"> • Сбор данных по архитектуре решений • Сбор статистики ошибок текущих систем 3. Оценка альтернатив по вектору критериев 	<ol style="list-style-type: none"> 1. Ранжирование рисков <ul style="list-style-type: none"> • Категоризация рисков по 5 группам • Ранжирование групп рисков максимальным методом 2. Ранжирование проектов по группам критериев <ul style="list-style-type: none"> • Внутренняя организация ИС • Организация изменений и политик • Готовность процессов и бизнес-сервисов • Минимизация операционных рисков 	<ol style="list-style-type: none"> 1. Перевод в ранги результатов первоначального ранжирования 2. Повторное ранжирование альтернатив методом порогового агрегирования 	<p>Выбор проекта с наибольшим рангом как наиболее эффективного проекта</p>

Рис. 2. Описание общего подхода

2.3. ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ РИСКОВ СЕРВИС-ОРИЕНТИРОВАННОЙ АРХИТЕКТУРЫ И ВЫБОР ПРОЕКТОВ В СРЕДЕ АНАЛИЗА

В данной работе построена модель для имитации работы информационной системы с сервис-ориентированной архитектурой в части возникновения ошибок в ее отдельных компонентах. При этом выполняется моделирование всего процесса обработки ошибок от подачи заявки на решение проблемы до ее разрешения (процедура моделирования рисков). На основе этих данных выполняется сбор статистики по тем группам рисков информационной системы с сервис-ориентированной архитектурой, по которым данная статистика не была получена в рамках практического применения предлагаемых в работе методов. С помощью имитационной модели можно будет сделать прогноз частоты возникновения рисков и, соответственно, оценок, которые используются в модели выбора наиболее эффективного проекта реализации сервис-ориентированной архитектуры информационной системы. Также выполняется генерация таблицы рангов на основе свободного определения числа приоритетов и критериев. Программная реализация предложенных методик выбора

эффективного проекта реализации информационных систем с СОА и анализа операционных рисков СОА выполнена в качестве надстройки для одной из достаточно распространенных программ имитационного моделирования PowerSim 8.0 (далее «среда моделирования»). Для ранжирования рисков использовался Microsoft Excel (далее «среда анализа»). Текущая версия построенной имитационной модели еще весьма ограничена.

Имитационный эксперимент выполняется согласно правилу (2). Собрана статистика по множеству типов рисков t_0, \dots, t_n , $t_n \in T$.

Следующие параметры являются входными для имитационной модели:

$N = \langle n_h, n_{IO}, n_f, n_s, n_d \rangle$ – имеющиеся данные по частоте появления каждого типа риска с разбивкой по четырем приоритетам;

d – данные по средней длительности устранения ошибки за момент времени θ_k (в имитационной модели – квартал);

w_{ij} – стоимость человеко-дня для i -й функциональности в зависимости от j -й роли и приоритета сообщения.

На основе данных входных параметров в рамках имитационного эксперимента подбираются такие показатели, как среднее время устранения ошибки и среднее время подачи заявления. Также для каждого типа риска подбирается процентное соотношение приоритетов. Так, например, на основе введенных рисков по данным нефтегазового предприятия приоритеты 1, 2, 3 и 4 соотносятся как 10%, 10%, 70%, и 10%.

В случае если статистика отсутствует, устанавливаются типы распределений, соответствующих каждому типу рисков. Эти распределения используются для генерации потока ошибок по соответствующим типам рисков. Для этого используются четыре параметра: среднее время устранения ошибки, среднее время подачи заявления, математическое ожидание и дисперсия. На основе этих параметров проводится имитация процесса устранения ошибок, т.е. генерируется поток полученных ошибок и поток устраненных ошибок.

Также проводится ранжирование множества ошибок. При этом разным типам рисков t_0, \dots, t_n соответствует последовательность оценок $r_0, \dots, r_m, m \in [1, 4]$, в соответствии с правилом (2). При генерации потока ошибок он разделяется по приоритетам в соответствии с введенным соотношением приоритетов, чтобы в дальнейшем использовать эти данные для ранжирования. Например, сгенерированный поток ошибок данных с разбивкой по приоритетам приведен в таблице 1.

Таким образом, моделируется процесс обработки ошибок информационных систем с сервис-ориентированной архитектурой. Динамика показателей полученных и исправленных ошибок приведена на рис. 3. Здесь отражены результаты моделирования на основе статистики данных и с учетом того, что среднее время исправления ошибки составляет 26 дней, среднее время подачи заявления – 20 дней.

Из проведенных имитационных экспериментов можно сделать вывод, что для минимизации рисков необходимо сокращать время подачи заявлений, время устранения ошибок и число заявлений, обрабатываемых в единицу времени. Однако необходимо также понимать, что чем меньше сроки на устранение ошибок, тем больше остается неразрешенных ошибок.

Таблица 1. Поток ошибок данных по приоритетам (разбивка по кварталам)

Время	Ошибки данных	Приоритет			
		Очень высокий	Высокий	Средний	Низкий
04.2006	5,25	0,06	0,00	2,87	1,00
06.2006	14,98	1,50	1,50	10,49	1,50
10.2006	0,01	0,00	0,00	0,00	0,00
01.2007	7,32	0,73	0,73	5,13	0,73
04.2007	35,39	3,54	3,54	24,78	3,54
06.2007	23,74	2,37	2,37	16,61	2,37
10.2007	8,63	0,86	0,86	6,04	0,86
01.2008	12,81	1,28	1,28	8,97	1,28
04.2008	6,12	0,61	0,61	4,28	0,61
06.2008	1,81	0,18	0,18	1,27	0,18
10.2008	15,64	1,56	1,56	10,95	1,56

Время	Ошибки данных	Приоритет			
		Очень высокий	Высокий	Средний	Низкий
01.2009	0,26	0,03	0,03	0,18	0,03
04.2009	41,79	4,18	4,18	29,25	4,18
06.2009	32,72	3,27	3,27	22,91	3,27
10.2009	3,08	0,31	0,31	2,15	0,31
01.2010	2,85	0,29	0,29	2,00	0,29
04.2010	1,32	0,13	0,13	0,93	0,13
06.2010	11,67	1,17	1,17	8,17	1,17
10.2010	55,63	5,56	5,56	38,94	5,56
Итого	275,77	27,57	27,57	193,05	27,57

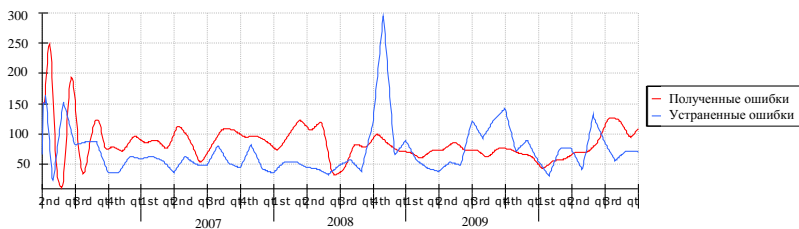


Рис. 3 Поток полученных и устраненных ошибок (вставка из среды Powersim Studio 8)

Таким образом, получаем оценки по всем типам рисков с разбивкой по приоритетам (таблица 2).

Таблица 2. Ранжирование рисков

Типы рисков	Приоритет				
	Очень высокий	Высокий	Средний	Низкий	Ранг
Ошибки ввода-вывода	15,43	15,43	108,00	15,43	2
Ошибки данных	27,577	27,577	193,039	27,577	2
Ошибки связующего ПО	7,72	7,72	54,06	7,72	4
Функциональные ошибки	51,14	51,14	358,01	51,14	1
Системные ошибки	28,31	28,31	198,15	28,31	2

Далее осуществляется перенос данных из среды моделирования в среду анализа, после чего эксперт запускает

процедуру генерации таблицы рангов. В соответствие с данной процедурой ранги генерируются автоматически с помощью разработанной автором специальной программы, использующей алгоритм ранжирования по пороговому правилу, приведенному выше.

В качестве примера приведем результаты генерации таблицы рангов для трех приоритетов и четырех критериев (таблица 3). Из приведенной таблицы видно, что, например, ранг 1 присваивается сочетанию оценок, в котором присутствуют три единицы. В соответствии с алгоритмом ранжирования такому сочетанию (или альтернативе) присваивается переходный ключ, равный трем, который преобразуется в ранг 1.

Таблица 3. Сгенерированные ранги

Ранг (ген.)	Ключ (ген.)	Сочетание числа повторений каждого приоритета (ген.)				Матрица сочетаний
		1	2	3	4	
1	3	3	0	0	0	(1, 1, 1)
2	6	2	1	0	0	(1, 1, 2)
3	9	1	2	0	0	(1, 2, 2)
4	12	0	3	0	0	(2, 2, 2)
5	18	2	0	1	0	(1, 1, 3)
6	21	1	1	1	0	(1, 2, 3)
7	24	0	2	1	0	(1, 2, 3)
8	33	1	0	2	0	(1, 3, 3)
9	36	0	1	2	0	(2, 3, 3)
10	48	0	0	3	0	(3, 3, 3)
11	66	2	0	0	1	(1, 1, 4)
12	69	1	1	0	1	(1, 2, 4)
13	72	0	2	0	1	(2, 2, 4)
14	81	1	0	1	1	(1, 3, 4)
15	84	0	1	1	1	(2, 3, 4)
16	96	0	0	2	1	(3, 3, 4)
17	129	1	0	0	2	(1, 4, 4)
18	132	0	1	0	2	(2, 4, 4)
19	144	0	0	1	2	(3, 4, 4)
20	192	0	0	0	3	(4, 4, 4)

Среда моделирования поддерживает графическое представление результатов (рис. 4), а также дает возможность загрузить документ для анализа в виде MS Excel. Результаты анализа также выгружаются в MS Excel. Они используются для дальнейшего анализа в интерфейсе, которым может быть порталное приложение или в случае прототипа – шаблон MS Excel с встроенными расчетами для применения многокритериального анализа методом порогового агрегирования, а также анализа рисков максиминным методом.

Целью создания инструмента является снижение трудоемкости процесса оценки рискованности сервис-ориентированной архитектуры на предприятии и поддержка принятия решений на основе полученного потока ошибок и ранжирования максиминным и многокритериальным анализом.

Развитие концепции инструмента предполагается на основе корпоративного хранилища данных SAP BW и порталного интерфейса.

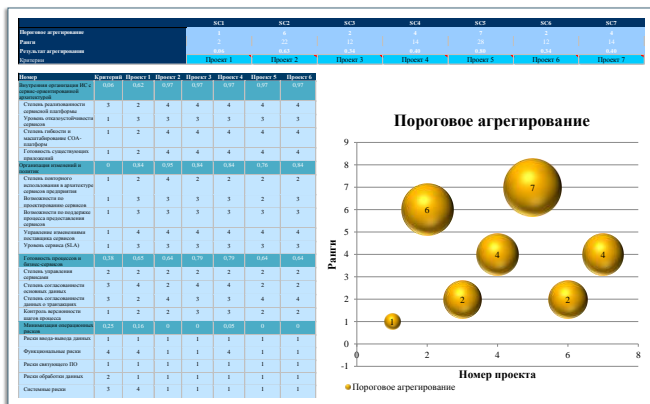


Рис. 4. Результаты ранжирования в инструменте

Пример использования предложенных методов анализа рисков и выбора оптимального проекта приведен более подробно в [12].

Сравним предложенный подход с существующими методами анализа и выбора проектов реализации информационных систем с использованием любого типа архитектуры, включая сервис-ориентированную архитектуру.

3. Известные модели выбора проектов реализации информационных систем с СОА и оценки рисков

Выбор проектов реализации информационной системы обосновывается либо эффективностью и успешностью внедрения, либо качеством системы и состоянием рынка ИС в данной области.

В силу отсутствия единого подхода к анализу рисков информационных систем, доказавшего свою эффективность и простоту использования, вопрос анализа рисков ИС в целом остается открытым. Еще более остро этот вопрос стоит при принятии решений о внедрении таких новых подходов как сервис-ориентированная архитектура. В литературе можно найти мало упоминаний о методах, показавших свою эффективность при оценке рискованности проекта с сервис-ориентированной архитектурой, поэтому обратимся к обзору более общей темы – анализу рисков ИС.

Каждый метод в виде первого шага определяет группы рисков. Одни подходы предлагают выявлять риски на основе влияния неудачи реализации СОА [23] на такие аспекты (и, соответственно, группы риска) как бизнес, планирование, стоимость и качество. Другие в данном случае ориентируются на этапы процесса разработки информационной системы [22], что дает три основные группы – риски проектирования продукта, риски среды разработки, и риски, возникающие из-за ограничений программы. В [18] для рисков уже используется более детальная классификация, включающая в себя аспекты персонала, ограничения бюджета и времени, вероятность разработки неверного функционала, меняющиеся потребности и даже внешние факторы влияния на проект.

Как объяснялось выше, риски сервис-ориентированной архитектуры тесно связаны с ресурсами или компонентами информационной системы. Для сравнения существующих

методов оценки рисков используем предложенную выше классификацию ресурсов ИС:

1. Человеческие ресурсы.
2. Прикладные ресурсы:
 - подсистема доступа;
 - прикладное ПО;
 - связующее ПО;
 - инфраструктурное ПО.
3. Технические ресурсы.

Ряд методов оценки рисков рассматривают эффект от информационной системы целиком, а некоторые подходы созданы для других областей. Поэтому добавим к классификации две группы: информационные системы в целом и применение не в ИС.

3.1. ОБЗОР МЕТОДОВ ОЦЕНКИ РИСКОВ ИНФОРМАЦИОННЫХ СИСТЕМ С СЕРВИС- ОРИЕНТИРОВАННОЙ АРХИТЕКТУРОЙ

Каждая модель оценки рисков из существующих сегодня поддерживает принятие решения с учетом определенных целей. Сравнение всех моделей оценки рисков приведено в таблице 4. Сравнение строится по степени охвата указанных типов ресурсов информационной системы, а также по возможности применения к ИС с сервис-ориентированной архитектурой в целом, ИС с другими типами архитектур (клиент-серверная, распределенная) и к другим областям, требующим оценки операционных рисков (т.е. моделям, несвязанным с информационными технологиями).

Для анализа используем пять рангов:

- – невозможность использования метода;
- ◐ – частичное использование для типа риска;
- ◑ – применимость метода в половине случаев при определенных условиях;
- ◒ – в целом метод применим за несколькими исключениями;
- – метод применим (таблица 4, см. в приложении).

Разделим модели оценки рисков на проактивные и реактивные. Проактивные модели ориентированы на определение и устранение рисков до их возникновения [22].

Реактивные модели предлагают расчет затрат или выполнение мер по устранению уже реализовавшихся рисков, или использование механизмов отслеживания приближения рисков.

Исходя из анализа, приведенного в таблице 4, видно, что предложенный автором данной работы метод является наиболее общим с точки зрения ресурсов ИС с сервис-ориентированной архитектурой по сравнению с остальными методами.

3.1.1. ЧЕЛОВЕЧЕСКИЕ РЕСУРСЫ

Данная категория используется для оценки рисков в большинстве существующих моделей. Все предложенные методы анализа рисков человеческих ресурсов (или рисков персонала) делятся на проактивные и реактивные.

Проактивные модели оценки рисков представлены в [4, 16, 30]. Они позволяют заблаговременно спланировать затраты при реализации рисков с помощью оценки вероятности риска и величины убытка соответственно [4, 30] либо на основе экспертной оценки индикаторов рисков [16], которые позволяют подсчитать суммарную величину убытков.

В [18] представлены техники проактивного и реактивного анализа рисков проекта. Функциональный метод анализа рисков проекта внедрения ИС утверждает, что влияние на проект оказывают 14 факторов (и, соответственно, категорий риска), среди которых степень распределенности модели данных, производительность системы, производительность конечных пользователей, уровень интерактивности обновления данных, легкость установки, сложность конфигурации, степень интерактивности ввода данных, сложность обработки данных, простота эксплуатации и степень поддержки изменений. К управлению человеческими ресурсами можно отнести только фактор производительности персонала. Степень влияния каждой из n -категорий выражается оценкой DI_n , которая принимает значения от 0 до 5. Общее влияние (VAF) в результате рассчитывается по формуле

$$VAF = 0,65 + 0,01 \sum_{n=1}^{14} DI_n.$$

где $DI \in [0; 5]$ – степень влияния.

Очевидно, что это лишь одна сторона и возможная область рисков при внедрении ИС. При этом подходе также не учитываются риски использования ИС, после того как она внедрена на предприятии, поэтому покрытие категории лишь на четверть.

Другим методом оценки риска, предложенным в [18], является подход Сосото II. При методе Сосото II определяется всего 6 групп проектных рисков, одна из которых – риски персонала. Суммарный риск (TR) рассчитывается как сумма уровней риска (RL) каждого типа рисков:

$$TR = \sum_{n=1}^6 RI_n .$$

Критерий риска для каждого модуля рассчитывается как $R = (TR/373) \cdot 100$. Определение уровня риска является экспертной оценкой. Недостатком данного подхода, кроме сложности оценки уровня риска, является и то, что не различаются разные приоритеты и влияния разных типов рисков.

Еще одна модель проектных рисков описана в [25]. Она предлагает реактивную модель, которая заключается в контроле индикаторов риска: изменчивость требований (RV), сложность (LGD) и эффективность (EF). Важно отметить, что в данной модели эффективность измеряется как отношение рабочего времени и времени нерабочего. Это является хорошим индикатором возможного риска персонала, поскольку полностью определяет все риски персонала и не зависит от экспертных оценок, однако есть и другие типы рисков персонала, которые данный метод не учитывает.

В [22] скорее представлены основные категории рисков информационных систем и предложены методы предотвращения появления рисков. Рассматриваются три основных модели управления рисками: управление рисками команды (TRM), постоянное управление рисками (CRM) и управление рисками программного обеспечения (SRM). TRM рассматривает как раз риски персонала, этапы их определения и

устранения. Покрытие категории персонала здесь полное, однако не предлагается методов оценки рисков.

Модель оценки рисков информационных систем с сервис-ориентированной архитектурой, предложенная автором данной статьи, так же как и в [25] ориентирована на определение и расчет рисков без использования экспертных оценок. Однако предложенная модель предлагает использовать проактивный подход оценки рисков (т.е. определение возможных рисков до их появления). Преимуществом данного подхода является то, что риски охватывают наиболее полный объем возможных затрат, в отличие от методов в [18, 22, 23], поскольку не пытаются определить выборочные типы рисков. Еще одно достоинство предложенной модели в области человеческих ресурсов – возможность определения рисков независимо от фазы информационной системы (предпроектная оценка, проект внедрения ИС, продуктивная эксплуатация ИС). Для этого, так же как и в [2], используется типы моделей распределения вероятностей. Модели определены для каждого типа рисков. Правда, в отличие от [2], предлагается определять распределение на основе ошибок ИС предыдущих лет.

3.1.2. ПРИКЛАДНЫЕ РЕСУРСЫ

Оценка прикладных ресурсов состоит из оценок типов прикладных ресурсов. Поэтому рассмотрим модели, используемые для оценки компонентов прикладных ресурсов ИС.

3.1.2.1. ПОДСИСТЕМА ДОСТУПА

Оценка рисков подсистемы доступа или интерфейса редко встречается в моделях. Среди рассмотренных моделей они определены всего в трех.

В [30] определена модель анализа рисков при дизайне ИС. Категории риска, определенные в модели, включают два типа подсистем доступа – клиентский компьютер («толстый» клиент) и веб-интерфейс. Автор предлагает определить риски, затем проранжировать и разработать стратегию предотвращения

рисков. Убыток же при реализации рисков оценивается с помощью следующего показателя:

$$ALE = SLE \times ARO,$$

где *ALE* – ежегодный ожидаемый убыток; *SLE* – единовременный ожидаемый убыток; *ARO* – годовая норма возникновения риска.

Подход дает общую оценку убытка от рисков, без выделения какой-либо группы или использования наиболее подходящей модели для каждой группы рисков. Ранжирование происходит с помощью линейной свертки, что имеет компенсаторный характер итогового ранга.

Оценка рисков подсистемы доступа частично приведена в [18]. В данном случае функциональный подход анализа рисков учитывает один из типов рисков подсистемы доступа – ввод данных через веб-интерфейс, снижающий безопасность данных.

В [22] определена категория рисков, связанных с пользовательским интерфейсом. Однако не приведен подход анализа рисков.

Модель оценки операционных рисков информационных систем с сервис-ориентированной архитектурой предлагает специализированный подход анализа операционных рисков подсистемы доступа в отличие от всех приведенных моделей. Оценка убытка на основе ошибок информационных систем предлагает комплексный подход анализа рисков этого типа.

3.1.1.2. ПРИКЛАДНОЕ ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ

В [23] рассматривается один из типов рисков прикладного программного обеспечения (ПО) – архитектурная сложность, изменение комплекса информационных систем, отсутствие дизайна и недостаточность тестирования ИС. Для данных типов определяется вероятность возникновения риска, частота и стоимость убытка. Метод может наиболее точно определить убытка только при условии, что перечислены все типы возможных рисков.

Отдельно выделенная категория рисков приложений или прикладного ПО присутствует также в методах [18, 22, 30]. Эта категория в данных методах является одной из составляющих

общего ожидаемого убытка. В случае [30] она представлена в виде общей группы прикладного ПО, в [22] – в виде рисков дизайна и разработки, а в [18] – в виде двух из 14 характеристик системы (комплексная обработка, уровень транзакций) или категории процессные риски в методе Сосото II. Таким образом, оценка анализа данной категории рисков отдельно не проводится, но тип риска влияет на определение общей величины убытка.

Элементы рисков прикладного ПО рассматриваются и в модели [25] вместе с показателем сложности (*LGD*). Она измеряется на основе формальной спецификации как $LGD = O + D + T$, где *O* – число атомарных операций; *D* – число атомарных потоков данных; *T* – число абстрактных типов данных, необходимых системе. Этот показатель затем используется как параметр распределения Вейбулла, имитирующего проект разработки прикладного ПО.

Модель оценки операционных рисков информационных систем с сервис-ориентированной архитектурой, предложенная автором данной статьи, предлагает способ оценить риски прикладного ПО на основе статистики предыдущих лет. Это позволяет сделать подход и специализированным, и общим одновременно, поскольку с одной стороны отражает риски именно в области прикладного ПО в СОА, а с другой стороны позволяет дать наиболее точную оценку потенциальных рисков. В отличие от этого существующие методы дают оценку рисков прикладного ПО либо с помощью узко специализированных методов анализа, либо с помощью более общих методов, в которых риски прикладного ПО являются одной из категорий, влияющих на общий результат.

3.1.2.3. СВЯЗУЮЩЕЕ ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ

Связующее ПО как категория рисков встречается еще реже, чем предыдущие категории. И объясняется это тем, что эффект интеграционной платформы обычно косвенный, поэтому оценку данной категории дать сложно. Один из немногих вариантов приведен в [18] в методе Сосото II, где напрямую определена группа рисков платформы. Однако не сказано, каким образом

определить события риска, относящиеся к данной группе. Другой пример приведен в методе [22]. Там определена группа рисков интеграции и тестирования и предложена модель CRM для устранения риска.

Уникальность и сильная сторона модели оценки операционных рисков информационных систем с сервис-ориентированной архитектурой, по сравнению с существующими методами, в данном случае заключается в том, что определить ошибки интеграционной платформы (иными словами – связующего ПО) значительно легче. А это дает возможность определить модель распределения вероятности связующего ПО и соответственно убыток от реализации рисков этой категории.

3.1.2.4. ДАННЫЕ ИНФОРМАЦИОННОЙ СИСТЕМЫ

Риски данных, так же как и предыдущий тип ресурсов, выделяются как категория риска редко. Группа рисков данных выделена в модели рисков при дизайне ИС [30] и в функциональной модели [18]. В первой модели рассматриваются скорее риски базы данных или сервера, на котором построена информационная система, в то время как в функциональной модели среди 14 характеристик ИС можно найти риски, связанные с характеристиками распределения данных и их обработкой. И если в первом случае определяется вероятность возникновения таких рисков, то во втором дается экспертная оценка.

В отличие от этих методов, предложенная модель оценки операционных рисков информационных систем с сервис-ориентированной архитектурой в случае с данными, так же как и в предыдущих типах ресурсов, предлагает специализированный подход оценки убытка от реализации рисков данных. При этом данный подход анализирует все или большую часть возможных рисков.

3.1.2.5. ИНФРАСТРУКТУРНОЕ ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ

Ошибки инфраструктурного ПО выделены как группа рисков в [23], где описаны риски неправильной работы инфраструктурного программного обеспечения и сбой серверного оборудования. Эти типы рисков используются для построения деревьев рисков, корнями которых являются нежелательные расходы, неопределенные требования, низкое качество и неправильное планирование. В случае если в перечисленные категории попали все возможные риски, подход определяет полную сумму убытков. Кроме того, метод предлагает проводить анализ до реализации проекта и всецело опирается на экспертную оценку.

Функциональный анализ рисков [18] включает категорию производительности. Как известно, производительность системы зависит от нескольких условий, в том числе и от правильно подобранного сервера и инфраструктурного ПО. Такой же тип риска упоминается и в [22].

Преимущество предложенной модели оценки операционных рисков информационных систем с сервис-ориентированной архитектурой в случае с инфраструктурным ПО заключается в том, что статистика, являющаяся базой для анализа рисков, обычно включает большее разнообразие ошибок инфраструктурного ПО и позволяет более четко определить типы возможных рисков. А убыток определяется исходя из каждого типа рисков.

3.1.3. ТЕХНИЧЕСКИЕ РЕСУРСЫ

Технические ресурсы практически не встречаются при анализе рисков информационных систем, хотя именно от работы сервера, частоты канала сети во многом зависит успех сервис-ориентированных приложений, постоянно обращающихся по сети к сервисам других систем.

Упоминание технических ресурсов есть только в [23], где речь идет о таких рисках, как сбой работы инструментов и потеря технического оборудования.

При использовании же предложенной модели оценки операционных рисков информационных систем с сервис-ориентированной архитектурой определяются и другие ошибки технических ресурсов, хотя ошибки ИС не всегда отражают все сбои в техническом оборудовании. Например, ошибка может говорить о временной недоступности системы, в то время как причиной ошибки может быть пожар в серверной комнате.

3.1.4. ИС С СОА

Данная группа рисков демонстрирует, что все существующие методы не были применены к оценке рисков специализировано для сервис-ориентированной архитектуры, кроме разработанного метода.

3.1.5. ИС В ЦЕЛОМ

Из таблицы 4 видно, что все рассмотренные методы, кроме [2, 19, 20], ориентированы на анализ рисков информационных систем независимо от типа архитектуры, т.е. все методы являются своего рода универсальными. Это плохо для таких архитектур информационных систем, как СОА, потому что риски возникают не только в стандартных подсистемах, но уникальных подсистемах, отличающих данный тип архитектуры.

Интересный вывод сделан в [25]: исследование и имитация нескольких проектов внедрения показали, что статистика проекта внедрения ИС имеет распределение Weibull.

Также есть ряд моделей, не связанных с типами ресурсов ИС. Так, например, при методе Coombs [18] определяются три характеристики проектных рисков: вероятность возникновения, сложность устранения и влияние на проект, если риск реализуется. Дополнительно определяется максимальное количество человеко-дней на устранение риска и максимальная стоимость устранения риска. Каждой характеристике присваивается вес (низкий – 1, средний – 3, высокий – 7). Сумма весов рисков показывает общий вес:

$$OverallWeight = \sum_{c=1}^3 W_c ,$$

где W_c – вес c -й характеристики.

Максимальная величина этого показателя будет равна 21 и это соответствует максимальной вероятности риска – 75%. Реальная частота появления риска вычисляется по формуле

$$C = \frac{OverallWeight}{OverallWeight_{max} * 0,75}.$$

Чтобы определить величину убытков, реальная частота появления умножается на максимальное количество человеко-дней на устранение риска и максимальную стоимость персонала. Преимуществами метода являются отображение влияния риска и возможность оценить стоимость риска с помощью оценки времени его устранения. Недостатком же является то, что обычно стоимость зависит не только от времени, необходимого для устранения риска, но и от приоритетности риска.

Этот недостаток устранен в модели оценки операционных рисков информационных систем с сервис-ориентированной архитектурой, предлагаемой автором.

3.1.6. ПРИМЕНЕНИЕ НЕ В ИТ

Ряд математических моделей, предложенных в [2, 19, 20], не связан с областью информационных систем. И все же заслуживает внимания, поскольку описывает наиболее популярные методы оценки рисков в других областях. Например, методология Basel II используется международными банками для того, чтобы заложить бюджет на предотвращение различных типов рисков, таких как рыночные, кредитные и операционные риски. По методологии Basel II банки часто измеряют риски, ассоциируемые с портфелем проектов [1]. Методология Basel II [1, 19] включает две методики анализа операционных рисков: стандартную и расширенную методику измерения (AMA). В первой методике общая прибыль разделяется на 8 линий бизнеса и умножается на фиксированную ставку процента, определенную по методологии Basel II. В соответствии с расширенной методикой банк должен иметь систему измерения внутренних

операционных рисков. Уровень риска определяется с учетом коэффициентов, отражающих вероятность наступления риска и его возможный размер. В [20] описываются модели операционных рисков на основе пуассоновского, биномиального и нормального распределений. Интересно, что в данной статье разные типы рисков имеют разные модели распределения:

1. Биномиальное распределение используется, чтобы описать распределение рисков, несущих значительный убыток для банка. В данном распределении вероятность наступления риска равна

$$P(n_k^{(b)} = n) = C_{N_k}^n p^n q_k^{N_k - n}, n \in \{0, \dots, N_k\}.$$

При этом $n_k^{(b)}$ – случайная величина, представляющая количество событий риска k -й группы; $C_{N_k}^n$ – число сочетаний из n по N_k ; p – вероятность того, что произойдет событие риска в каждый из n промежутков; q – вероятность того, что событие риска не произойдет; k – группа риска; N – число промежутков продолжительность τ каждый. Среднее значение величины убытка определяется по формуле $\langle Y_k^{(b)} \rangle = \langle n_k^{(b)} \rangle V_k^{(b)} = N_k p_k V_k^{(b)}$, $V_k^{(b)}$ – величина убытка от одного события риска k -й группы.

2. Пуассоновское распределение описывает события риска, которые возникают достаточно часто и требуют осуществления мероприятий, уменьшающих вероятность наступления риска. Вероятность наступления риска в данном распределении будет равняться

$$P(n_k^{(b)} = n) = \frac{(\mu_k)^n}{n!} e^{-\mu_k},$$

где μ_k – стационарное значение вероятности события риска данной группы.

Среднее значение величины убытка равно

$$\langle Y_k^{(p)} \rangle = \langle n_k^{(p)} \rangle V_k = \mu_k V_k^{(b)}.$$

3. Постоянно возникающие события риска, не приносящие таких больших убытков как предыдущие два типа, имеют

нормальное распределение. Плотность распределения в таком случае равняется

$$P(n) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(n-a)^2}{2\sigma^2}},$$

при этом среднее значение $\langle n \rangle = a$, а дисперсия $Dn = \sigma^2$.

Если говорить о предложенной модели оценки операционных рисков информационных систем с сервис-ориентированной архитектурой, ее можно применить как к информационным системам с СОА, так и с любой другой архитектурой. Однако для ее применения к другим областям деятельности, не связанным с ИТ, модель потребует замены статистической базы на другой источник статистических данных.

Стоимость убытков от информационных систем является хорошим показателем успешности проекта внедрения. Однако при принятии решения о переходе на сервис-ориентированную архитектуру и выборе правильного проекта реализации необходимо руководствоваться не только показателями рискованности проекта, но другими критериями успешности архитектуры СОА. Совокупность всех характеристик архитектуры и рисков ставит непростую задачу выбора проекта для лица, принимающего решение (ЛПР). Поэтому рассмотрим существующие методы выбора проектов.

3.2. МНОГОКРИТЕРИАЛЬНЫЙ АНАЛИЗ ИНФОРМАЦИОННЫХ СИСТЕМ С СЕРВИС- ОРИЕНТИРОВАННОЙ АРХИТЕКТУРЫ

Информационная система имеет несколько фаз жизненного цикла не зависимо от типа поддерживаемой архитектуры. В общем случае фазы ИС представлены на рис. 6.

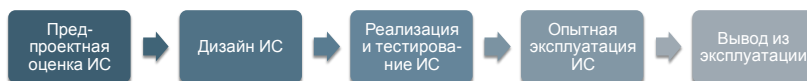


Рис. 5. Жизненный цикл ИС

Методы анализа и сравнения информационных систем применяются на разных стадиях жизненного цикла. Ряд публикаций ориентируется на выбор методологии внедрения СОА, оптимизирующей выгоды от архитектуры [4], которые рассматриваются обычно на фазе предпроектной оценки ИС. Также можно найти и описание подходов измерения успешности внедрения, т.е. анализ информационных систем на фазе опытной эксплуатации. Устоявшегося подхода сравнения проектов сервис-ориентированной архитектуры информационной системы на сегодняшний день не существует. И в научных журналах можно найти немного предложений по анализу таких проектов. Поэтому проведем более расширенный обзор существующих методов многокритериального анализа проектов информационных систем (как с сервис-ориентированной архитектурой, так и с другим типом архитектуры).

Метод выбора оптимального проекта СОА, предложенный автором в данной работе, универсален и может быть применен к ИС, находящейся на любой стадии жизненного цикла (таблица 5, см. в приложении), в том числе и в момент реализации ИС (т.е. на фазе реализации и тестирования ИС). В отличие от остальных рассмотренных методов, которые применимы либо до начала проекта реализации ИС с СОА [8, 14, 15, 21, 24, 26, 27], либо сравнивающих качество и функцию производительности уже реализованной ИС с СОА [14, 15, 17, 21, 22, 24, 27, 28, 31].

3.2.1. ПРЕДПРОЕКТНАЯ ОЦЕНКА ИС

В соответствии со стандартом [5] фаза предпроектной оценки ориентирована на определение потребностей пользователей в реализации ИС, исследовании и описании основных концепций ИС, а также их демонстрации и аттестации с целью уточнения характеристик системы, сравнения альтернативных вариантов реализации ИС.

Большинство методов сравнения и анализа эффективности поддерживают именно этот этап жизненного цикла ИС. Методы сравнения эффективности ИС предлагают принимать решение о

внедрении информационной системы независимо от того, какая архитектура внедряется на основании таких показателей (или критериев) как дисконтированный срок окупаемости инвестиций, чистый дисконтированный доход (*NPV*), индекс рентабельности инвестиций и внутренняя норма доходности (*IRR*).

В таблице 5 приведено сравнение всех рассмотренных методов по покрытию ими фаз жизненного цикла ИС. Из шести рассмотренных методов четыре ориентированы на данную фазу ИС.

В [8] предлагается использовать методологию системного анализа иерархических структур. Она основана на предположении о наличии у анализируемых альтернатив информационных систем множества общих критериев $S = \{S_1, S_2, \dots, S_m\}$, а ЛПР должно определить оптимальный набор критериев, способствующих достижению цели. Решением этой задачи является парное сравнение альтернатив на «интенсивность проявления критериев» и парное сравнение важности весов. Каждый объект оценивается по 9-балльной шкале. В результате строятся матрицы для каждого критерия, иллюстрирующие парное сравнение объектов на рассматриваемый критерий. Затем определяется максимальный собственный вектор, соответствующий максимальному собственному значению и проводится его нормализация. Сначала считаются веса объектов относительно критериев, затем относительно цели, и выбирается объект с наибольшим весом. При этом недостатком метода является невозможность достичь совершенной согласованности суждений (т.е. транзитивности и числовой согласованности), в силу неточности измерений на практике. Данный метод применяется для сравнения альтернатив проектов на фазе предпроектного анализа, когда принимается решение о будущей системе. Существенный недостаток этого метода еще и в том, что число критериев не определено заранее. Это усложняет процедуру сравнения на данной фазе, поскольку ЛПР на ранней стадии проекта не располагает подробными сведениями об архитектуре решения.

Ряд работ [14, 26, 27] посвящен оценке информационных систем на основе функции производительности. Они предлагают использовать подход экономической рациональности для оценки информационных систем с помощью ожидаемой полезности для лица, принимающего решение (ЛПР). Информационная система в данном случае сравнивается исходя из матрицы ценности (ценности информации).

Например, функция полезности [26] представляется как

$$U = U_{\pi\beta}(\eta, \alpha, \kappa) = B_{\pi\beta}(\eta, \alpha) - K_{\pi}(\kappa),$$

где $B_{\pi\beta}(\eta, \alpha) = \sum_{zya} \beta(a, z) \pi_z \eta_{zy} \alpha_{ya}$; $K_{\pi}(k) = \sum_z \pi_z \kappa_z$; x – набор входных данных ИС; y – набор исходящих данных ИС; a – действия или решения пользователя; η – функция трансформации данных; $y = \eta(x)$; π – некоторое распределение вероятностей; β – функция выгод; K – ожидаемая стоимость обработки данных; B – ожидаемые дисконтированные выгоды; $\{\alpha\}$ – набор всех трансформаций из y в a ; z – события, вышедшие из-под контроля пользователя.

Таким образом, ценность информации (V) в данном случае измеряется как

$$V_{\pi\beta}(\eta) = \max_{\alpha \in \{\alpha\}} B_{\pi\beta}(\eta, \alpha) = B_{\pi\beta}(\eta, \alpha^*)$$

где α^* – оптимальный набор.

Максимизация полезности в данном случае выглядит следующим образом:

$$\max_{\eta} U = \max_{\eta} V_{\pi\beta}(\eta) - \min_{\kappa} K_{\pi}(\kappa).$$

Этот метод прекрасно подходит для данной фазы ИС и ориентирован в основном на нее, поскольку использует по максимуму данные получаемые на этой фазе: план проекта и описание требований. Однако прогнозное значение функции полезности, конечно, может отличаться от реальных характеристик системы, полученных в ходе дальнейших фаз внедрения. Для комплексного анализа в таком случае необходимо применять систему других методов, позволяющих контролировать внедрение и использование ИС.

Другое направление [15, 21] предлагает использовать метод многокритериальной функции полезности для анализа проблем информационных систем, а затем выбрать систему, которая дает оптимальное решение функции. Автор [15] предлагает определить функцию полезности для пользователя(-лей) информационной системы. Предлагаемый подход основан на многокритериальной теории полезности, который комбинирует приведенные значения выгод. Несмотря на то, что метод основан на анализе «ex ante», т.е. до внедрения системы, ряд показателей обычно получаются уже после внедрения, поэтому для их прогнозирования используются экспериментальные оценки и значения, полученные из опросных листов и интервью. Автор [15] рассматривает следующие показатели функции полезности для информационной системы:

- Своевременность включает в себя время отклика и частоту ответа системы. Частота ответа имеет следующую функцию полезности: $U_i(t) = a_1 \exp(-b_1 t)$, где $a_1, b_1 > 0$ и a_1 служит как калибровочный коэффициент, который преобразует значение функции в цифровое значение.

- Содержание включает похожесть и уровень агрегации. Похожесть измеряется так: $U_s(s) = -a_2 \exp(-b_2 s)$, где $a_2, b_2 > 0$ и a_2 служит как калибровочный коэффициент, который преобразует значение функции U_s в цифровое значение. S основано на вероятности ошибок первого и второго типа, описанных в статье, $s = 1 - P_1 - P_2$, где P_1 – это, например, вероятность что существующая запись данных не извлечена и P_2 – вероятность того, что извлеченная запись не та, которую хотели получить.

- Формат сложно оценить с помощью каких-либо показателей в численном виде, поэтому для этого используются оценки пользователей.

Предположим, стоимость растет линейно при увеличении s и снижении t : $c = -a_3 t + b_3 s + const$, $a_3, b_3 > 0$. Объединенная функция полезности тогда $u = a_1 \exp(-b_1 t) - a_2 \exp(-b_2 s) - (-a_3 t + b_3 s) + const$. Решение данной функции должно определить оптимальную систему для организации. В случае если невозможно определить объединенную функцию

полезности, пользователь может проранжировать показатели. А затем при условии четкого доминирования одного из показателей можно выбрать оптимальную систему.

Для применения данного метода к частному случаю информационной системы с сервис-ориентированной архитектурой необходимо уточнение параметров оценки и соответственно более детальная проработка функции полезности с учетом предложенных параметров. Однако, как и в ряде предыдущих методов, оценка некоторых параметров количественно может оказаться затруднительной.

Метод оценки качественных параметров, помогающих ЛПР решать многокритериальные задачи, представлен в [24]. Здесь функция полезности определена для множества показателей следующим образом:

$$V(x) = \sum_{i=1}^n w_i v_i(x_i),$$

где n – количество показателей; i – число показателей; x_i – i -й показатель; v_i – i -я функция полезности от одного показателя (SAVF); w_i – вес i -го показателя; X – набор всех показателей; V – функция полезности от множества показателей (MAVF).

Набор показателей X не зависит от набора показателей Y . Для определения оптимального значения используется теория предпочтений. Доминирование устанавливается на основании серии вопросов, сравнивающих несколько параметров по нескольким альтернативным проектам. Далее параметры раскладываются в виде дерева анализа проблемы и матрицы принятия решения. Каждый узел дерева представляет значение атрибута, а связи – зависимости между атрибутами. Опрос помогает установить, какие предпочтения соответствуют какими вариантам. В результате часть альтернативных проектов отсеивается во время опроса. И последний шаг заключается в исключении доминируемых альтернатив. Таким образом, данный подход является последовательным, итеративным методом анализа альтернативных проектов с участием ЛПР, который помогает исключить неподходящие варианты. Недостатками

данного подхода являются неопределенность набора показателей и необходимость их определения для каждого проекта, а также необходимость участия ЛПР на протяжении практически всего анализа.

Все перечисленные проблемы устранены в предложенном методе многокритериального анализа. Он предлагает четко определенную классификацию характеристик ИС с сервис-ориентированной архитектурой. Полученные варианты реализации ранжируются с помощью многокритериального анализа методом порогового агрегирования. При этом на фазе предпроектного анализа метод позволяет проанализировать альтернативы с учетом имеющихся на этот момент деталей архитектуры.

3.2.2. ДИЗАЙН ИС

В соответствии со стандартом [5] данная фаза жизненного цикла ИС охватывает «период проектирования, создания, интеграции (сборки), тестирования и оценки системных технических средств, компьютеров, программных средств, оборудования, персонала подсистем, его обучения и объектов сопровождения».

На этом этапе часто приходится принимать решение о выборе более подходящего проекта реализации информационной системы под конкретную группу процессов, поскольку выбор по определенным причинам не был сделан на ранней фазе, поэтому выбор делается с появлением более детальной информации на этапе дизайна. Однако не так много существует методов, ориентированных на поддержку этого выбора. Один из методов, применяемых на данном этапе – анализ альтернатив проектов методом системного анализа иерархических структур [8]. Уязвимость метода, как и раньше, в неопределенности критериев принятия решения, это может привести к одностороннему анализу проекта.

В отличие от этого метода, идея предложенного метода выбора оптимального проекта СОА заключается в том, чтобы исключить приведенный недостаток за счет идентификации критериев. Критерии СОА дают возможность сравнения

альтернатив как на начальном этапе проекта, так и при дизайне ИС. С другой стороны, это делает метод четко специализированным под требования сервис-ориентированной архитектуры.

3.2.3. РЕАЛИЗАЦИЯ И ТЕСТИРОВАНИЕ ИС

В соответствии с [5] во время данной фазы спроектированную и разработанную систему адаптируют под конкретных пользователей, копируют систему на носители или аппаратные средства, готовые для промышленного использования, готовят сопроводительные документы и проводят окончательное тестирования ИС. Целями данной фазы являются «квалифицированное изготовление и поставка работоспособной и сопровождаемой системы заказчику».

Фаза реализации и тестирования обычно уже не предполагает, что решение будет заменено, поскольку ориентирована на внедрение выбранного решения. Однако в некоторых случаях данная фаза требует поддержки принятия решения при реализации дополнительных или неочевидных ранее требований. В таких случаях обычно принимается быстрое решение в зависимости от приоритетности задач. Когда возникает такая необходимость, определяется стратегическое направление подсистемы ИС, что также требует сравнения альтернативных вариантов. Как и на предыдущей фазе, это можно сделать с помощью методов системного анализа иерархических структур [8], а в случае сервисного подхода – предложенным методом выбора оптимального проекта.

В качестве проверки правильности дизайна и реализации информационной системы предлагается применять метод оценки производительности ИС. В [17] описываются методы оценки производительности информационной системы безотносительно к типу архитектуры, и предлагается подход, основанный на многокритериальном анализе дизайна информационной системы. Информационная система оценивается с точки зрения системной и пользовательской. Системная точка зрения ориентирована на показатели использования ресурсов, стоимости и эффективности. С

пользовательской точки зрения больше интересны показатели пропускной способности, надежности и времени отклика. Подход предлагает три этапа анализа: оценка системы (определяет веса по критериям для анализируемой информационной системы), оценка целей пользователей (анализирует степень достижения целей заказчика с помощью метода многоцелевого программирования (MGP)) и оценка дизайна (подытоживает общую оценку информационной системы с учетом предыдущих двух). При этом запрос пользователя характеризуется набором операций A_1, A_2, \dots, A_N . Для измерения меры достижения цели для пользователя предлагается использовать разницу между ожидаемыми целями и достигнутой производительностью системы:

$$D(i) = G(i) - \sum_{j=1}^N R_j(i),$$

Где i – число целей; j – число операций; $R_j(i)$ – оценка средней производительности шага; $G(i)$ – целевая производительность.

В результате основная формула для оценки информационной системы выглядит следующим образом:

Максимизировать $\lambda = F^+ P^+ D^+ + F^- P^- D^-$ при условии, что $R\beta + D^- - D^+ = G$, где R – матрица текущего уровня производительности; β – матрица переменной принятия решения; D^+, D^- – матрица перевыполнений и невыполнений; G – матрица целевых производительностей шагов; F, P – матрицы критериев приоритетности и весов.

Решив уравнение, получим матрицу β – производительности, необходимой, чтобы минимизировать разницу целей и достичь цели пользователя. Из-за отличий целей пользователей от конечных целей внедряемой информационной системы внедрение часто заканчивается провалом. Однако этот подход ориентирован на то, чтобы сгладить этот недостаток и предложить удовлетворительный вариант системы. Также интересно, что в данном случае каждая альтернатива информационной системы анализируется как вариант архитектуры ИС, основанной на пользовательских шагах. Каждый пользователь осуществляет разные действия и,

соответственно, использует разные сервисы и элементы аппаратного оборудования.

Недостатки метода заключаются в том, что на практике довольно сложно оценить численно все цели. А если говорить о сервис-ориентированной архитектуре, то ее основная цель – гибкость ландшафта и скорость адаптации новых требований. Данные цели довольно тяжело перевести в цифровые значения, а влияние на численные показатели лишь косвенное. К тому же еще одним недостатком является предположение о линейных зависимостях последовательности операций. В случае нелинейной зависимости или параллельно протекающих шагах метод необходимо усовершенствовать и проработать более детальное сравнение.

В отличие от этого, в предложенном методе выбора оптимального проекта цели четко определены и отражены в критериях сравнения архитектуры. Однако специфика предлагаемого метода ограничивает его применение для данной фазы.

3.2.4. ОПЫТНАЯ ЭКСПЛУАТАЦИЯ ИС

По [5] данная фаза включает «эксплуатацию, применение и использование системы пользователями».

Когда система уже используется, можно оценить насколько она соответствует требованиям пользователей и их ожиданиям. Интересно, что из шести рассмотренных методов пять можно применить на данном этапе. Часть из них специализируется на данной фазе, другие же пересекаются с методами, используемыми на этапе предпроектного анализа.

Ряд работ [28, 31] посвящен оценке качества услуг, предоставляемой информационной системы независимо от типа архитектуры. Качество услуг основано на сравнении того, что клиент/пользователь хочет получить от ИС, и того, что реально предложено. В [31] предлагается оценить качество услуг с точки зрения пяти измерений: материального (средства обслуживания, оборудование и персонал), надежности (возможность предоставлять сервис точно и в срок), ответственности (желание помочь клиенту и предоставить сервис), гарантированности

(знания и вежливость сотрудников и их возможность вызывать доверие и уверенность), симпатии (заботливое, индивидуальное внимание предоставляемое клиенту). Таким образом, качество услуг измеряется в соответствии с формулой $G = P - E$, где G – результат оценки потребления сервиса; P – восприятие; E – ожидания. Оценки собираются в виде опроса. Сравнение проводится несколько раз, в зависимости от цикла внедрения информационной системы.

Измерение качества услуг информационной системы, таким образом, используется как еще одна составляющая при оценке эффективности ИС. При этом объектом измерения результата может быть как конкретная информационная система, так и ИТ-отдел в целом, исходя из всего комплекса информационных систем, внедренного и поддерживаемого департаментом.

Все методы, ориентированные на анализ ИС на фазе предпроектной, применимы и в данном случае: метод оценки производительности ИС [17], метод оценки функции полезности [14, 26, 27], метод оценки функции полезности проблем ИС [15, 21] и многокритериальная функция полезности [24]. Все эти методы рассматривают возможности смены системы в случае необходимости.

Предложенный метод выбора оптимального проекта СОА позволяет в данном случае сосредоточиться на специализированном типе систем и провести сравнение альтернатив в этой области. Преимуществом данного метода, по сравнению с остальными, является рассмотрение критериев ИС. Это позволяет определить наиболее уязвимую область рассматриваемых альтернатив, и выбрать оптимальный проект по набору критериев. В отличие от предлагаемых подходов, которые ориентируются на более обобщенный показатель эффективности.

3.2.5. ВЫВОД ИЗ ЭКСПЛУАТАЦИИ ИС

По [5] данная фаза используется для снятия системы из эксплуатации, архивирования данных, закрытия доступа к вводу данных в систему и ограничение поддержки ее пользователей.

Этап вывода системы из эксплуатации обычно предполагает замену ИС, перевод всех необходимых данных на новую платформу и определение стратегии развития класса систем. В общем случае, на этом этапе решение уже принято. Однако практика показывает, что решение об общем направлении ИС не всегда учитывает все детали перехода и нюансы покрытия требований новой системой. В частности может оказаться, что выбранная система не покрывает каких-либо требований. В таком случае необходимо определить дополнительную подсистему, возможно на базе другой платформы. Таким образом, методы оценки производительности ИС [17], методы оценки функции полезности ИС [14, 24, 26, 27] применимы и здесь. Как в предыдущих фазах жизненного цикла ИС, чем детальнее проводится сравнение, тем лучше выбранная стратегия. А уровень определения подсистемы ИС и работы с несколькими дополнительными требованиями сильно сужает действие существующих методов, в отличие от предложенного нового метода выбора оптимального проекта ИС с сервис-ориентированной архитектурой.

3.3. МЕТОДЫ РАНЖИРОВАНИЯ И ПОРОГОВОГО АГРЕГИРОВАНИЯ

Получив большое количество разных параметров оценки архитектуры, необходимо вывести итоговое решение. Обычно для агрегирования результатов используется линейная свертка, как видно из приведенных выше методов. Поскольку в области информационных систем на сегодняшний день не применяются другие методы агрегирования для принятия решений при выборе проектов реализации ИС, обратимся к методам ранжирования и порогового агрегирования, применяемым в других областях.

В основе ранжирования по методу порогового агрегирования [16] лежит «пороговое правило»

$Wtr = \{(x, y) | [v_1(x) < v_1(y)] \text{ или } [v_1(x) = v_1(y) \text{ и } v_2(x) < v_2(y)] \text{ или } [v_1(x) = v_1(y) \text{ и } v_2(x) = v_2(y) \text{ и } v_3(x) < v_3(y)]\}$, где $v_1(x)$ – число единиц («1») в записи вектора x , $v_2(x)$ – соответственно, число

двоек («2»), а $v_3(x)$ – число троек («3»). Таким образом, отношение Wtr представляет собой множество пар векторов, для которых либо в первом сравниваемом векторе число единиц меньше, чем во втором, либо при равном числе единиц число двоек в первом векторе меньше, чем во втором. Результатом агрегирования является ранжирование векторов.

Исходя из метода [16], оценки критериев делаются по шкале $\{1, 2, \dots, m\}$, где $m \geq 3$. Функция φ считается многокритериальным пороговым правилом, если она ранжирует два n -критериальных вектора x и y следующим образом:

- пусть g – самый маленький ранг, который попадает в векторах x и y ;
- если ранг g попадает более часто в векторе x , тогда $\varphi(x) < \varphi(y)$;
- если ранг g попадает более часто в векторе y , тогда $\varphi(x) > \varphi(y)$;
- если ранг g вообще не попадает в векторах (в силу того, что все ранги одинаковы в векторах x и y), тогда $\varphi(x) = \varphi(y)$.

При этом функция φ вычисляется по формуле

$$\Phi(x) = \sum_{j=1}^m C_{n-v_j(x)+m-j-1}^{m-j} \text{ для всех } x \in X,$$

такое что $C_k^{k+1} = 0$ для всех $k \in [0; m-1]$ и $C_{-1}^0 = 1$,

где $v_j(x)$ – число рангов j в записи вектора x .

Тогда действуют следующие аксиомы:

- *Парето-доминирование*: если координаты вектора x не меньше координат вектора y , и есть хотя бы одна координата в x строго больше соответствующей координаты y , то $\varphi(x) > \varphi(y)$.
- *Попарная компенсируемость критериев*: если все координаты векторов x и y равны, кроме некоторых двух, в которых значения x и y «взаимно обратные», то $\varphi(x) = \varphi(y)$.
- *Пороговая некомпенсируемость*: для любого ранга g такого, что $1 \leq g \leq m$: $\varphi(g-1, \dots, g) < \varphi(g, \dots, g)$. Например, если хотя бы одна координата в векторе x равна 1, то $\varphi(1, \dots, n) < \varphi(2, \dots, 2)$.

• *Аксиома редукции*: Если в двух векторах x и y значения по одной из координат равны, то эту координату можно не учитывать в решении вопроса о взаимном предпочтении этих векторов.

$$\forall x, y \in A, \exists i \ x_i = y_i \Rightarrow$$

$$\Rightarrow \varphi_n(x) > \varphi_n(y) \Leftrightarrow \varphi_{n-1}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) > \varphi_{n-1}(y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_n).$$

4. Заключение

Предложенный в данной статье метод позволяет осуществить выбор проекта реализации ИС с сервис-ориентированной архитектурой, который зависит от оценок по совокупности критериев в четырех основных областях: внутренняя организация ИС с сервис-ориентированной архитектурой, организация изменений и политик, готовность процессов и бизнес-сервисов и минимизация операционных рисков. Существует большое количество разнообразных методов анализа рисков, сравнения проектов ИС и определения их эффектов. Однако не каждый метод применим без изменений к случаю СОА. В статье приведено подробное сравнение существующих методов анализа рисков ИС. Сравнение осуществлялось на основе покрытия классов рисков каждым методом. Похожий подход сравнения применен и к методам выбора проектов реализации ИС, только в данном случае проверялось покрытие различных фаз жизненного цикла ИС. С учетом достоинств и недостатков рассмотренных методов был разработан метод анализа рисков сервис-ориентированной архитектуры, предлагаемый автором в данной статье.

В качестве применения предложенного подхода приведено описание созданного прототипа инструмента. Он используется не только для имитации потока рисков при работе ИС с сервис-ориентированной архитектурой в зависимости от типа рисков, но и имеет подсистему анализа со встроенной функцией порогового агрегирования для поддержки выбора оптимального проекта.

Благодарность

Автор благодарен профессору Ф.Т. Алескерову за постановку задачи.

Литература

1. АЛЕСКЕРОВ Ф.Т., АНДРИЕВСКАЯ И.К., ПЕНИКАС Г.И., СОЛОДКОВ В.М. *Анализ математических моделей Базель II*. – М.: ФИЗМАТЛИТ, 2010. – С. 288
2. АЛЕСКЕРОВ Ф.Т., БЫКОВ А.А., КУРМАКАЕВА Е.А. *Анализ операционных убытков на основе биномиального, пуассоновского и нормального распределений* // Банковское дело. – 2011. – №6. – С. 52–57.
3. АЛЕСКЕРОВ Ф.Т., ЯКУБА В.И. *Метод порогового агрегирования трехградационных ранжировок* // Доклады Академии наук. – 2007. – Т. 413, №2. – С. 181–183.
4. БИБЕРШТЕЙН Н., БОУЗ С., ДЖОНС К., ФИММАНТ М., ША Р. *Компас в мире сервис-ориентированной архитектуры (COA): ценность для бизнеса, планирование и план развития предприятия* / Пер. с англ. — М.: КУДИЦ-ПРЕСС, 2007. – С. 256.
5. ГОСТ Р ИСО/МЭК 12207-2002. *Информационная технология. Процессы жизненного цикла программных средств. Руководство по применению*. ГОССТАНДАРТ РОССИИ. – Москва, 2001
6. ИЗБАЧКОВ Ю.С., ПЕТРОВ В.Н. *Информационные системы: Учебник для вузов*. 2-е издание. – СПб.: Питер, 2006. – 656 с.
7. КОГАЛОВСКИЙ М.Р. *Перспективные технологии информационных систем*. (Серия «ИТ-Экономика»). – М.: ДМК Пресс; Компания АйТи, 2003. – 288 с.
8. ЛОБАНОВА А.А., ЧУГУНОВА А.В. *Энциклопедия финансового менеджмента*. – М.: Альпина Бизнес Букс, 2007. – С. 878.

9. ЛОТОВ А.В., ПОСПЕЛОВА И.И. *Многокритериальные задачи принятия решений*: Учебное пособие. – М.: МАКС Пресс, 2008. – 197 с.
10. ПОДИНОВСКИЙ В.В. *Анализ задач многокритериального выбора методами теории важности критериев при помощи компьютерных систем поддержки принятия решений* // Известия РАН. Теория и системы управления. – 2008. – №2. – С. 64–68.
11. ПОДИНОВСКИЙ В.В., НОГИН В.Д. *Парето-оптимальные решения многокритериальных задач*. – М.: ФИЗМАТЛИТ, 2007. – С. 256
12. ПЫРЛИНА И.В. *Выбор эффективного проекта реализации сервисно-ориентированной архитектуры информационной системы* // Проблемы управления РАН. – 2012. – №4. – С. 59–68.
13. ПЫРЛИНА И.В. *Классификация операционных рисков при сервисно-ориентированном подходе к созданию информационной системы* // Бизнес-Информатика. – 2011. – №4(18). – С. 54–61.
14. ACKOFF R.L. *Towards a Behavioral Theory of Communication* // Management Science. – 1958. – Vol. 4, №8. – P. 218–234.
15. ANITUV N. *A Systematic Approach toward Assessing the Value of an Information System* // MIS Quarterly. – 1980. – Vol. 4, №4. – P. 61–75.
16. ALESKEROV F.T., CHISTYAKOV V.V., KALYAGIN V.A. *Multiple criteria threshold decision making algorithms*: Working paper WP7/2010/02. – Moscow: State University – Higher School of Economics. – 2010. – P. 40.
17. CHANDLER J.S. *A Multiple Criteria Approach for Evaluating Information Systems* // MIS Quarterly. – 1982. – Vol. 6, №1. – P. 61–74.
18. DASH R., DASH RAJ. *Risk Assessment Techniques for Software Development* // European Journal of Scientific Research, ISSN 1450-216X. – 2010. – Vol. 42, №4. – P. 629–636.

19. *Developments in Modeling Risk Aggregation* // Bank for International Settlements Communications CH-4002 Basel, Switzerland, ISBN 92-9197- 847-7. – 2009 – P. 32.
20. EMBRECHTS P., PUCETTI G. *Risk Aggregation. Copula Theory and its Applications* / Eds. P. Jaworski, F. Durante, W. Haerdle, T. Rychlik. Lecture Notes in Statistics / Proc. 198, Springer Berlin/Heidelberg. – 2010. – P. 111–126.
21. GROCHOW J.M. *A Utility Theoretic Approach to Evaluation of a Time-Sharing System* // Statistical Computer Performance Evaluation / In Freiburger W. – Academic Press, New York. – 1972. – P. 25–50.
22. HIGUERA R.P., HAIMES Y.Y. *Software Risk Management* // Technical Report CMU/SEI-96-TR-012 ESC-TR-96-012. – 1996. – 48 p.
23. HOODAT H., RASHIDI H. *Classification and Analysis of Risks in Software Engineering* // World Academy of Science, Engineering and Technology. – 2009. – №56. – P. 446–452.
24. KLEIN G., BECK P.O. *A Decision Aid for Selecting among Information System Alternatives* // MIS Quarterly. – 1987. – Vol. 11, №2. – P. 177–185. [Электронный ресурс]. – URL: <http://www.jstor.org/stable/249360> (дата обращения 03.09.2013).
25. LUQI J.N. *A Risk Assessment Model for Evolutionary Software Projects* // IEEE [Электронный ресурс]. – URL: <http://www.disi.unige.it/person/ReggioG/PROCEEDINGS/luqi.pdf> (дата обращения 04.09.2013).
26. MARSCHAK J. *Economics of Information Systems* // Journal of the American Statistical Association. – 1971. – Vol. 66, №333. – P. 192–219.
27. PITT L.F., WATSON R.T., KAVAN C.B. *Service Quality: A Measure of Information Systems Effectiveness* // MIS Quarterly. – 1995. – Vol. 19, №2. – P. 173–187.
28. SHOLLER D. *2008 SOA User Survey: Adoption Trends and Characteristics* // Gartner Research ID Number: G00161125. – 2008. – 17 p.

29. VERDON D., MCGRAW G. *Risk Analysis in Software Design* // IEEE computer society/IEEE security & privacy. – 2004. – Vol. 2, № 4. – P. 79–84.
30. WATSON R.T., PITT L.F., KAVAN C.B. *Measuring Information Systems Service Quality: Lessons from Two Longitudinal Case Studies* // MIS Quarterly. – 1998. – Vol. 22, №1. – P. 61–79.

SERVICE-ORIENTED ARCHITECTURE: OPERATIONAL RISKS AND CHOICE OF EFFICIENT SYSTEM

Irina Pyrlina, Higher School of Economics, Moscow, Specialist, IT Architecture Expert (Irina.pyrlina@gmail.com).

Abstract: We survey existing methods of efficient implementation projects selection for information system with service-oriented architecture (SOA), methods of risks assessment and multi-criteria analysis applied to information systems with SOA. We also suggest a new method to evaluate operational risks due to SOA and develop multi-criteria threshold algorithms to support ranking of information system implementation projects. Methods were used to design a tool consisting of two key subsystems – a simulation environment and a multi-criteria analysis module.

Keywords: multiple criteria threshold algorithm, service-oriented architecture, SOA operational risks, efficiency criteria for information system architecture, project ranking.

*Статья представлена к публикации
членом редакционной коллегии М.В. Губко*

*Поступила в редакцию 29.08.2012.
Опубликована 30.09.2013.*

ПРИЛОЖЕНИЕ.

Таблица 4. Сравнение методов оценки рисков

Типы ресурсов	Методы и модели анализа рисков ИС									
	Вероятностный анализ дерева рисков [23]	Риски при дизайне ИС [30]	Функциональный анализ рисков [18]	Socomo II [18]	Coombes et al [18]	Модель проектных рисков [25]	BASEL II [19], Агрегирование рисков [16]	Управление рисками ПО [22]	Анализ операционных убытков [2]	Оценка рисков COA
1. Человеческие ресурсы										
2. Прикладные ресурсы										
Подсистема доступа										
Прикладное ПО										
Связующее ПО										
Данные ИС										

Типы ресурсов	Методы и модели анализа рисков ИС										
	Вероятностный анализ дерева рисков [23]	Риски при дизайне ИС [30]	Функциональный анализ рисков [18]	Соотношение П [18]	Coombset all [18]	Модель проектных рисков [25]	BASEL II [19], Агрегирование рисков [16]	Управление рисками ПО [22]	Анализ операционных убытков [2]	Оценка рисков COA	
Инфраструктурное ПО											
3. Технические ресурсы											
4. ИС COA с											
5. ИС в целом											
6. Применение не в ИТ											

Таблица 5. Сравнение методов анализа и выбора проектов реализации ИС

Фазы ИС	Методы анализа и сравнения проектов реализации ИС					
	Метод системного анализа иерархич. структур [8]	Оценка производительности ИС [17]	Качество услуг [28, 31]	Функция полезности ИС [14, 26, 27]	Многокритериальная функция полезности [15, 21, 24]	Метод выбора оптимального проекта COA
Предпроектная оценка						
Дизайн ИС						
Реализация и тестирование ИС						
Опытная эксплуатация ИС						
Вывод из эксплуатации ИС						

УДК 004.52 + 004.934
ББК 22.1

**ФУНКЦИЯ МОДУЛЯ АКУСТИЧЕСКОГО
МОДЕЛИРОВАНИЯ В СИСТЕМЕ
АВТОМАТИЧЕСКОГО АНАЛИЗА
НЕСТРУКТУРИРОВАННОЙ РЕЧЕВОЙ
ИНФОРМАЦИИ**

Смирнов В. А.¹

(ООО «Спич Драйв», Санкт-Петербург),

Гусев М. Н.²

(ФГУП «НИИ «Квант», Санкт-Петербург),

Фархадов М. П.³

*(ФГБУН Институт проблем управления
им. В.А. Трапезникова РАН, Москва)*

Описывается модуль акустического моделирования – модуль системы автоматического анализа неструктурированной речевой информации, предназначенный для формирования статистического описания звуков речи – акустической модели. Дается пояснение назначения системы автоматического анализа неструктурированной речевой информации и описывается функция модуля акустического моделирования и его место в общей схеме автоматического анализа неструктурированной речевой информации; раскрывается алгоритм работы модуля и дается подробное описание каждого этапа алгоритма.

Ключевые слова: акустическое моделирование, скрытые Марковские модели, алгоритм прямого-обратного хода, пре-

¹ Валентин Александрович Смирнов, генеральный директор (*speechdrive@mail.ru*, тел. (960) 269-57-97).

² Гусев Михаил Николаевич, инженер 1-й категории, кандидат технических наук (*mgaev@kvant-rdi.spb.ru*, тел. (964) 322-43-09).

³ Маис Паша-Оглы Фархадов, заведующий лабораторией, доктор технических наук (*mais@ipri.ru*, Москва, ул. Профсоюзная, д.65, тел. (495) 334-87-10).

образование Баума-Уэлша, кластеризация по дереву регрессии, автоматический анализ неструктурированной речевой информации.

1. Введение

Система анализа неструктурированной речевой информации относится к классу систем интеллектуальной автоматизированной обработки неструктурированных данных. Такая система обеспечивает возможность значительной экономии времени при обработке данных, тем самым позволяя повысить скорость и качество принимаемых решений в динамично изменяющейся ситуации. Системы анализа неструктурированной речевой информации (далее – система АНРИ) уже более 20 лет применяются на практике тысячами коммерческих и государственных учреждений по всему миру. В производстве подобных систем лидируют зарубежные компании (www.autonomy.com), при этом российские поставщики также предлагают ряд решений (www.speech-drive.ru).

Данный класс прикладных аналитических систем путем применения технологии распознавания речи позволяет обнаруживать ключевые слова и фразы (например, «опасность», «по секрету», «не нравится») в записях речевых сообщений и тем самым определять важные аспекты для бизнеса и государственных учреждений, обеспечивать контроль качества обслуживания абонентов и т.п. Одно из основных преимуществ системы АНРИ – существенное снижение использования человеческого ресурса в механической и рутинной работе по анализу записей речевых данных и возможность комплексного объективного анализа большого потока речевой информации.

Система АНРИ представляет собой сложное программное решение, осуществляющее разноплановую обработку естественного языка. Модуль обучения акустической модели, или модуль акустического моделирования, предназначен для создания статистического описания звуков речи и является обязательным компонентом в подобных системах. Затем полученное статистическое описание звуков используется в процессе распознавания

речи для определения подобия наблюдаемых акустических явлений тому или иному элементу акустической модели.

В статье описываются основные структуры данных и алгоритмы, необходимые при акустическом моделировании: скрытые Марковские модели (СММ), типы единиц моделирования речевого потока, преобразование Баума–Уэлша и кластеризация по дереву регрессии.

2. Назначение модуля акустического моделирования и его функция в системе АНРИ

Основная задача модуля акустического моделирования (Модуль АМ) в системах распознавания речи вообще и в системе АНРИ в частности – обеспечивать формирование статистического описания звуков речи. Данное описание необходимо для дальнейшего использования в модуле декодирования для определения степени подобия наблюдаемых акустических явлений акустическим классам – элементам акустической модели.

Для того чтобы акустическое моделирование стало возможным, необходимо осуществить подготовительные работы с использованием других модулей системы, таких как Лингвистический процессор и Модуль вычисления акустических признаков. В результате формируется обучающая база данных. Отметим, что построение обучающей базы – важный подготовительный этап, от которого зависит не только будет ли успешным обучение акустической модели в целом, но и насколько качественной она получится. Критерии построения обучающей базы преимущественно определяются условиями предполагаемого использования системы. В зависимости от ее назначения обучающая база может отличаться по фонетическому составу, количеству дикторов, суммарной длительности звуковых данных, характеристикам канала записи. В современных системах АНРИ используются миллионы реализаций звуков речи, записи голосов нескольких тысяч дикторов общей длительностью несколько сотен часов.

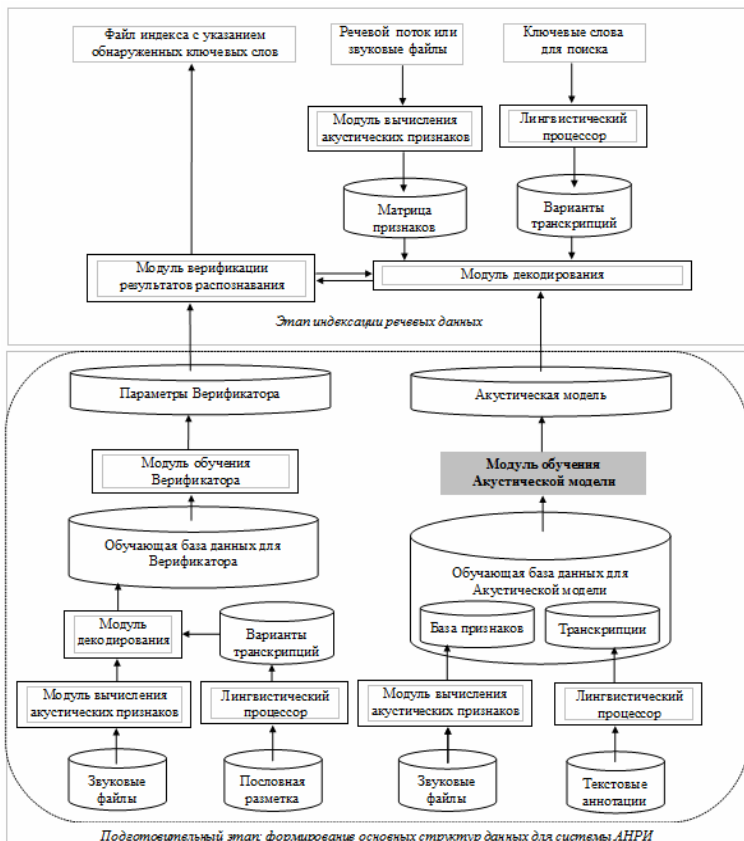


Рис. 1. Структура системы автоматического анализа речевой информации и модуль акустического моделирования

На рис. 1 представлена общая схема системы АНРИ и определены взаимосвязи Модуля АМ с другими модулями и структурами данных.

3. Единицы моделирования

Перед тем как приступить к описанию алгоритмов, предназначенных для обучения акустической модели, дадим краткий комментарий относительно базовых единиц моделирования в системе АНРИ.

Акустическая модель представляет собой набор описаний определенных единиц речи. В самом простом случае минимальной единицей описания речи является целое слово. Когда задача состоит в распознавании ограниченного количества слов, например, «да» и «нет» или последовательности цифр, это допустимо, однако в современных приложениях такие простые задачи встречаются редко. Соответственно, требуется перейти на уровень фонетического описания речи, когда единицей описания становятся отдельные звуки, которые при акустическом моделировании объединяются в единую цепь, формируя так называемую композитную модель, которая и используется в дальнейшем для оценки.

Построение акустической модели звуков включает несколько этапов, на каждом из которых применяются различные по степени подробности описания языка. На начальном этапе используются графические символы, полученные как результат работы Лингвистического процессора. Такие единицы называются монофонами. Их количество варьирует от 20 до 100 единиц в зависимости от языка и предпочтений конкретной исследовательской группы.

Однако акустическая модель, построенная на базе монофонов, недостаточно точно описывает явления звучащей речи. Это связано с тем, что в процессе порождения речи звуки, входящие в состав слов, влияют друг на друга, вызывая тем самым существенные темпоральные и акустические изменения. Монофоны являются контекстно-независимым представлением фонетического описания речи, поэтому они не могут смоделировать данную вариативность. Этот недостаток призвано исправить контекстно-зависимое представление фонем, наиболее распространенным типом которого являются трифоны. Процесс перехода от монофонов к трифонам достаточно прост и автоматизирован во всех системах построения акустических моделей. Схема преобразования транскрипций представлена на рис. 2.

Вместо * может быть использован монофон от соседнего слова или специальный символ, обозначающий начало или конец слова, например, условное обозначение тишины. Как видно, контекстно-зависимая форма представления звука учитывает влияние других звуков с обеих сторон и таким образом может

охватывать все пространство произносимых фонетических единиц. Однако из этого неизбежно вытекает следующая проблема – чрезмерно большое количество минимальных фонетических единиц. Его несложно вычислить по формуле

$$N = n^3,$$

где n – исходное число монофонов.

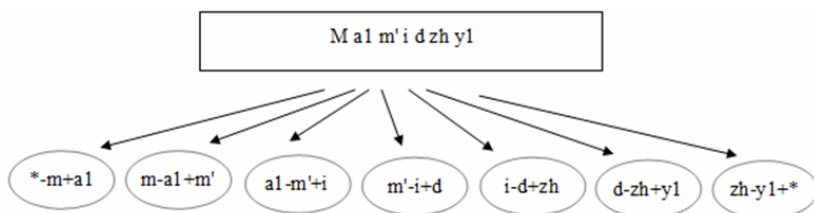


Рис. 2. Преобразование транскрипций из монофонов в трифоны (для слова «Мамиджи»)

Даже если учесть, что реальное количество трифонов несколько, но не на порядок, меньше указанной асимптоты (асимптота недостижима, поскольку ряд сочетаний монофонов в языке никогда не встречается даже с учетом стыков слов – например, три монофона h' подряд), такое число трифонов делает практически неразрешимой задачу создания речевого корпуса, содержащего достаточное число реализаций каждой фонемы для ее качественного статистического описания. Для эффективной группировки множества контекстно-зависимых единиц применяются алгоритмы кластеризации, позволяющие объединять несколько моделей описаний трифонов в одну (см. раздел «Кластеризация»).

2.1. ЗНАЧИМОСТЬ ИНВЕНТАРЯ МОНОФОНОВ

Петербургская фонетическая школа (Л.Р. Зиндер, Л.В. Бондарко [1, 2]), к числу последователей которой относят себя авторы статьи, включает в инвентарь фонем русского языка 42 единицы, среди которых 6 гласных и 36 согласных. Все вариации, возникающие в различных контекстах, считаются аллофонами, т.е. вариативными представителями одного и того же фонетического класса. Богатый опыт авторов статьи по работе с

промышленными системами показывает, что именно данный набор фонем следует использовать как отправную точку при определении состава единиц акустического моделирования.

При этом тот же опыт привел к осознанию того, что некоторые аллофоны фонем необходимо выделять в отдельные единицы акустической модели, поскольку их спектральные и темпоральные характеристики существенным образом отличаются от основного аллофона. Это особенно характерно для гласных, которые в ударных и безударных слогах значительно различаются акустически. Именно эти фонетические знания предопределили включение в наш инвентарь фонем единиц, представленных в таблице 1.

Таблица 1. Список дополнительных монофонов

Дополнительный монофон	Фонема, к которой относится монофон	Комментарий
@	a	Второй предударный и заударный аллофон фонемы «А»
a1	a	Первый предударный аллофон фонемы «А»
_a	a	Аллофон «А» после мягкого согласного
_u	u	Аллофон «У» после мягкого согласного
_o	o	Аллофон «О» после мягкого согласного
t_n	t	Аллофон «Т» перед «Н»
d_n	d	Аллофон Д перед Н
dz	c	Звонкий аллофон «Ц»
hg	h	Звонкий аллофон «Х»
\$z	\$	Звонкий аллофон «Щ»

Очевидно, что многие из перечисленных в таблице дополнительных монофонов достаточно редко встречаются в речи, и встает вопрос о целесообразности их выделения в обособленный класс. Можно было бы предположить, что подобные контекстные различия будут отражены в структуре дерева классификации трифонов (см. раздел 5 настоящей статьи). Тем не менее, при достаточно большой обучающей базе данных выделение в

отдельный класс целесообразно, поскольку кластеризация не всегда способна отразить данные отличия и в этом случае статистическое описание более редких монофонов искажается. Качественные характеристики системы АНРИ, учитывающей дополнительные монофоны, приведены в Разделе 6 настоящей статьи.

Немного выйдя за рамки проблематики настоящей статьи, отметим, что введение дополнительных аллофонов имеет большое значение не только при акустическом моделировании, но и при декодировании (распознавании речи – поиске ключевых слов) – особенно в тех случаях, когда ключевые слова или фразы короткие. Например, при поиске сочетания «вещдок» (сокращенная форма от сочетания «вещественное доказательство») использование обычного аллофона «щ» может привести к низкому акустическому подобию всей последовательности и, следовательно, к пропуску искомого слова.

4. Структура СММ

Базовым математическим аппаратом для акустического моделирования выступают скрытые Марковские модели (hidden Markov model (НММ)) [9, 14, 16, 18]. Применение данного математического аппарата для задач распознавания речи впервые было предложено в 70-е гг. такими исследователями, как Бэйкер [3] и Джелинек [12]. В дальнейшем большое число исследователей внесли свой вклад в развитие этого аппарата применительно к задачам распознавания речи. В частности, был осуществлен переход от более простых дискретных моделей к непрерывным гауссовым, предложены такие методы обучения, как дискриминативное обучение [13]. Ряд исследователей также предложили использовать при акустическом моделировании нейронные сети [5, 15]. Однако общим для всех методик остается использование скрытых Марковских моделей. Большая часть современных систем распознавания речи основана именно на данном математическом аппарате: в качестве примеров можно привести такие проекты с открытым исходным кодом, как Sphinx (<http://cmusphinx.sourceforge.net/wiki/>) и НТК (<http://htk.eng.cam.ac.uk/>), а среди наиболее известных коммерче-

ских систем следует выделить системы от таких компаний, как Google (<http://www.google.com>) и Nuance (<http://www.nuance.com>).

СММ как базовая структура представления информации об акустической стороне языка используется в системе АНРИ как в Модуле декодирования, так и в Модуле обучения акустической модели. В настоящей статье мы сконцентрируем внимание на алгоритмах, применяемых для последнего модуля. Информацию об алгоритмах, используемых модулем декодирования (в частности, об алгоритме Витерби) см. в [8, 17].

Рассмотрим более детально СММ – структуру, лежащую в его основе. Введем обозначения:

M – Марковская модель;

o_t – наблюдение (векторов акустических признаков) в момент времени t ;

O – последовательность наблюдений из которых состоит трифон: $O = o_1, o_2, \dots, o_i$;

x – состояние;

X – последовательность состояний;

a_{ij} – вероятность перехода из состояния i в состояние j ;

$b_j(o_t)$ – вероятность порождения наблюдения o в момент времени t состоянием j .

В каждый момент времени t модель находится в состоянии j ; исходя из плотности вероятностей $b_j(o_t)$, генерируется наблюдение o_t . Переход из состояния в состояние также является вероятностным и совершается под управлением отдельной вероятности a_{ij} . На рис. 3 показан пример процесса порождения цепочки наблюдений от o_1 до o_6 , где модель, состоящая из шести состояний, проходит через последовательность состояний $X = 1, 2, 2, 3, 4, 4, 5, 6$.

Дадим пояснение на простейшем примере (см. рис. 3).

Как было отмечено в разделе «Единицы моделирования», слова в системах распознавания речи обычно моделируются как последовательность фонем – монофонов или трифонов. Фонема при этом является абстрактным понятием, а в реальном мире существует бесконечное множество реализаций фонем в форме звукового потока, воспроизводимого говорящим и интерпрети-

руемого слушающим. При этом разные части реализаций фонем (начало, середина или конец) могут сильно друг от друга отличаться по акустическим характеристикам. Тогда на уровне абстрактных единиц (фонем) мы также можем выделить несколько «частей» и в упрощенном виде интерпретировать их как скрытые состояния СММ⁴.

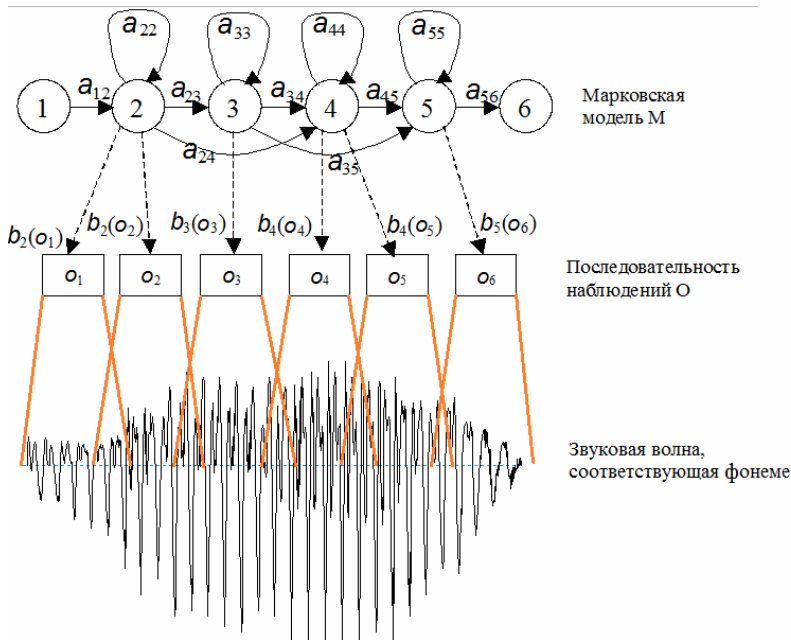


Рис. 3. Марковская модель генерации последовательности наблюдений

Наблюдения – это отрезки речевого потока (обычно – длиной в 25 мс), преобразованные в компактное векторное представление Модулем вычисления акустических признаков (см.

⁴ Речь идет именно об интерпретации в иллюстративных целях, а не о фактическом тождестве понятий «состояние» и «часть фонемы»; следует помнить, что состояния – это элементы математической модели, а фонемы – это единицы языковой системы.

рис. 1). Вероятности переходов между частями фонем – это полученные в ходе обучения акустической модели коэффициенты, показывающие, с какой вероятностью одна часть фонемы может следовать за другой. Человек, слушая речь, определяет произнесенную последовательность фонем, а система распознавания речи по аналогии с человеком пытается по последовательности наблюдений (векторам признаков) определить, какая последовательность состояний лежит в основе этих наблюдений.

Вернемся к рис. 3 и раскроем суть приведенных на нем обозначений. Кружки с цифрами 2–5 обозначают состояния, т.е. части фонемы (ее начало, середину, конец), узлы графа с цифрами 1 и 6 – это служебные состояния, используемые для построения связей с другими фонемами в композитных СММ, включающих несколько последовательных фонем. Литеры $o_1 \dots o_6$ обозначают наблюдения, т.е. векторы акустических признаков (в приведенном примере речевой сигнал был разбит на шесть отрезков длиной 25 мс, на каждом из которых были вычислены акустические параметры, используемые при распознавании). Массив значений a_{ij} – это вероятности переходов между состояниями (частями фонемы). Массив значений $b_j(o_i)$ – это вероятности того, что наблюдение o_i (вектор признаков) было порождено состоянием j (частью абстрактной единицы – фонемы).

3.1. ОСНОВНЫЕ ЭТАПЫ СОЗДАНИЯ АКУСТИЧЕСКИХ МОДЕЛЕЙ.

Работу по созданию и обучению СММ для системы распознавания речи схематично можно представить в виде последовательности из следующих этапов:

- Изучение особенностей языка, классификация фонем, выбор единиц распознавания речи, выбор типа модели для базовых единиц речи.
- Создание корпуса текстов, который должен включать слова и их сочетания, наиболее полно отражающие характеристики фонем, присутствующих в языке. Формирование наиболее полного, непротиворечивого и компактного корпуса

текстов является одной из ключевых задач при создании системы АНРИ.

- Создание на базе корпуса текстов речевой базы данных (речевого корпуса) для обучения и тестирования. Тексты произносятся разными дикторами в разных условиях и с использованием различных типов источников записи (микрофоны, телефонные аппараты и т.п.) Записи снабжаются необходимыми атрибутами (тип аудиоданных, характеристики диктора) и фонетической транскрипцией. В большинстве случаев транскрипция строится автоматически по текстам корпуса, вручную добавляются специальные символы: например, «пауза» или «шум». Часть записей речевого корпуса (обычно большая) используется для обучения моделей, другая часть – для тестирования.

- Спектральный анализ и параметризация речевых сигналов из обучающего множества записей речевой базы с целью получения векторов признаков и обучения акустических моделей, описывающих элементы речи.

- Инициализация моделей для всех элементов речи (выбор начальных параметров СММ).

- Обучение моделей, для чего последовательно выполняется переоценка параметров СММ.

- Тестирование качества обученных моделей на тестовой выборке.

5. Оценка параметров СММ и преобразование Баума–Уэлша

Известно, что не существует точного решения задачи подстройки параметров СММ. Кроме того, практически невозможно указать оптимальный способ оценки параметров на основе имеющейся обучающей базы наблюдений. При этом предложено много критериев, которые могут быть использованы для приближенного решения этой проблемы, и разработаны различные итеративные процедуры, с помощью которых можно найти параметры модели, обеспечивающие локальный максимум вероятности $P(O|M)$, где O – последовательность наблюдений, а

M – акустическая модель. Одним из наиболее распространенных оптимизационных методов для решения этой задачи является алгоритм Баума–Уэлша [4, 7], который используется во многих современных системах распознавания. Алгоритм Баума–Уэлша гарантирует, что модель с новыми параметрами будет равна или лучше предыдущей модели с точки зрения критерия максимального правдоподобия: По мнению некоторых исследователей, недостатком алгоритма Баума–Уэлша является зависимость генерируемого решения от начальной модели. Предложено несколько процедур для получения начальных оценок параметров, обеспечивающих быструю сходимость формул переоценки параметров. Наиболее простой путь – применение итеративного кластерного алгоритма k -средних. Более перспективным представляется применение генетических алгоритмов [11].

Для вычисления совместной вероятности того, что последовательность наблюдений O порождена моделью M , проходящей через последовательность состояний X , требуется вычислить произведение вероятностей переходов и вероятностей порождения. Таким образом, для последовательности состояний X , представленной на рис. 3:

$$(1) P(O, X | M) = a_{12}b_2(o_1)a_{22}b_2(o_2)a_{23}b_3(o_3)\dots$$

Однако при распознавании известна только последовательность наблюдений O , а породившая их последовательность состояний X скрыта (отсюда термин – *Скрытая*). Соответственно, вероятность того, что O была порождена моделью, вычисляется суммированием по всем возможным последовательностям скрытых состояний X .

$$(2) P(O | M) = \sum_X a_{x(0)x(1)} \prod_{t=1}^T b_{x(t)}(o_t) a_{x(t)x(t+1)},$$

где $x(0)$ – модель начального состояния, а $x(t+1)$ – модель конечного состояния.

Для того чтобы оценить значения параметров $\{a_{ij}\}$ и $\{b(o_t)\}$, существуют простые рекурсивные процедуры. Одной из таких процедур является упомянутый выше алгоритм Баума–Уэлша (Baum–Welch).

Рассмотрим основную идею рекуррентной процедуры оценки параметров НММ Баума–Уэлша, также именуемой

EM-метод (метод обнаружения максимального правдоподобия – от английского Expectation-Maximization). При всех дальнейших рассуждениях будем исходить из того, что вероятность порождения наблюдения во всех состояниях имеет непрерывную плотность распределения, соответствующую нормальному (Гауссовому) закону:

$$(3) \quad b_j(o_t) = \frac{1}{\sqrt{(2\pi)^n |\Sigma_j|}} e^{-\frac{1}{2} \frac{(o_t - \mu_j)(o_t - \mu_j)'}{\Sigma_j}},$$

Если предположить, что в НММ всего одно состояние j , то оценки максимального правдоподобия величин μ_j и Σ_j могут быть получены в результате усреднения:

$$(4) \quad \hat{\mu}_j = \frac{1}{T} \sum_{t=1}^T o_t,$$

$$(5) \quad \hat{\Sigma}_j = \frac{1}{T} \sum_{t=1}^T (o_t - \mu_j)(o_t - \mu_j)',$$

На практике, и в особенности в случае композитных СММ, одновременно используется несколько состояний, и невозможно непосредственно привязать векторы акустических признаков к отдельным состояниям. Для того чтобы корректно оценить параметры СММ, алгоритм Баума–Уэлша рассчитывает полное правдоподобие каждой последовательности наблюдений путем суммирования всех возможных последовательностей состояний. Каждый вектор наблюдения o_t вносит свой вклад в расчет значений максимального правдоподобия для каждого состояния j . Таким образом, все наблюдения связываются со всеми состояниями, при этом связь должна быть пропорциональна вероятности состояния модели при соответствующем наблюдении.

Обозначим за $L_j(t)$ вероятность пребывания в состоянии j в момент времени t . Тогда приведенные выше уравнения (4) и (5) можно записать следующим образом:

$$(6) \quad \hat{\mu}_j = \frac{\sum_{t=1}^T L_j(t) o_t}{\sum_{t=1}^T L_j(t)},$$

$$(7) \hat{\Sigma}_j = \frac{\sum_{t=1}^T L_j(t)(o_t - \mu_j)(o_t - \mu_j)'}{\sum_{t=1}^T L_j(t)},$$

где суммирование в знаменателях обеспечивает требуемую нормализацию.

Применение уравнений (6) и (7) позволяет осуществить рекуррентную оценку Баума–Уэлша для средних и дисперсий НММ. Аналогичная процедура может быть получена для вероятностей переходов из состояния в состояние.

Для применения соотношений (6) и (7) необходимо рассчитать вероятность состояния $L_j(t)$. Искомая вероятность определяется с помощью алгоритма прямого-обратного хода (Forward-Backward Algorithm). Прямая вероятность $\alpha_j(t)$ для модели M с N состояниями определена в виде:

$$(8) \alpha_j(t) = P(o_1, \dots, o_t, x(t) = j | M).$$

$\alpha_j(t)$ – это вероятность первых t наблюдений для состояния j в момент времени t , т.е. вероятность всех возможных последовательностей из $t - 1$ состояний, которые могли породить цепочку наблюдений от o_1 до o_{t-1} с учетом того, что в момент времени t мы находимся в состоянии j и наблюдаем вектор o_{t-1} . $\alpha_j(t)$, может быть рассчитана по следующей рекуррентной формуле:

$$(9) \alpha_j(t) = \left[\sum_{i=2}^{N-1} \alpha_i(t-1)a_{ij}(t) \right] b_j(o_t).$$

Формула (9) получена исходя из того, что вероятность пребывания в состоянии j в момент t при наблюдениях o_t равна сумме прямых вероятностей всех возможных предшествующих состояний, взвешенных вероятностями переходов a_{ij} . Пределы суммирования (от 2 до $N - 1$) определяются особенностью организации звуковых моделей – в них первое и последнее состояния не являются порождающими, т.е. не генерируют никаких наблюдений (ср. рис. 3). Начальные и конечные условия для (9) имеют вид:

$$(10) \alpha_j(1) = \begin{cases} 1, & j = 1, \\ a_{1j}b_j(o_1), & 1 < j < N; \end{cases} \quad \alpha_N(T) = \sum_{i=2}^{N-1} \alpha_i(T)a_{iN}.$$

Из определения $\alpha_j(t)$ следует:

$$(11) P(O|M) = \alpha_N(T).$$

Следовательно, вычисление прямой вероятности позволяет получить полное правдоподобие $P(O|M)$.

Обратная вероятность $\beta_j(t)$ определяется следующим образом:

$$(12) \beta_j(t) = P(o_{t+1}, \dots, o_T | x(t) = j, M).$$

$\beta_j(t)$ – это вероятность порождения цепочки наблюдений от o_{t+1} до o_T при условии, что в момент времени t мы находимся в состоянии j . Обратная вероятность может быть вычислена с использованием следующего рекуррентного соотношения:

$$(13) \beta_j(t) = \sum_{i=2}^{N-1} a_{ij} b_j(o_{t+1}) \beta_j(t+1).$$

Начальное и конечное условия для (13) имеют вид:

$$(14) \beta_i(T) = a_{iN}, \quad 1 < i < N; \quad \beta_1(1) = \sum_{j=2}^{N-1} a_{1j} b_j(o_1) \beta_j(1).$$

Приведенная выше прямая вероятность есть совместная вероятность, а обратная вероятность есть условная вероятность. Такое асимметричное определение позволяет определить вероятность нахождения в состоянии j как произведение этих двух вероятностей:

$$(15) \alpha_j(t) \beta_j(t) = P(O, x(t) = j | M).$$

Отсюда

$$(16) L_j(t) = P(x(t) = j | O, M) = \frac{P(O, x(t) = j | M)}{P(O | M)} = \frac{\alpha_j(t) \beta_j(t)}{P(O | M)}.$$

Формулы (6)–(16) позволяют осуществить преобразование Баума–Уэлша для последовательности наблюдений и получить новые значения средних и дисперсий. На практике для получения корректных оценок параметров необходимо использовать большое множество звуковых данных (десятки или даже сотни часов). Соответственно, преобразование Баума–Уэлша итеративно применяется ко всем файлам в обучающей базе. Отметим здесь, что оценка прямых и обратных вероятностей связана с вычислением перемножения большого количества вероятно-

стей. Это может приводить к тому, что значения оценок станут малы. В целях избежания вычислительных проблем прямые-обратные вероятности могут рассчитываться в логарифмическом масштабе.

6. Кластеризация

В разделе «единицы моделирования» мы определили трифоны как наиболее распространенный тип единиц, используемых в системе АНРИ для представления статистической модели фонетического уровня языка. При этом мы отметили ключевую трудность, возникающую в результате введения таких контекстно-зависимых единиц, – отсутствие достаточного количества данных для получения статистически обоснованных оценок. Кластеризация позволяет сократить количество элементов в АМ и тем самым обеспечить существенно большее количество наблюдений в расчете на одну единицу моделирования и избежать ситуаций, при которых количество реализаций для какого-то трифона настолько мало, что его качественное моделирование (в частности, расчет параметров Гауссовой смеси) в принципе невозможно.

Существует несколько методов кластеризации [6, 10]: «по энтропии» (когда в единый кластер объединяются наиболее частотный трифон с наиболее редким трифоном); «основанный на данных» (когда в один кластер объединяются трифоны, чьи распределения наименьшим образом друг от друга отличаются) и метод классификации по дереву регрессии (от английского Classification and Regression Trees, или CART). Именно последний метод получил наибольшее распространение в современных системах распознавания речи. Преимущество данного метода в том, что он успешно объединяет лингвистические знания (фонетический строй языка) и математический аппарат (метод минимизации среднеквадратической ошибки).

Суть метода CART состоит в том, чтобы разбить N векторов в M -мерном пространстве признаков на группы путем последовательного применения к кластеризуемым данным «функций-вопросов». Причем сами вопросы не затрагивают M признаков, описывающих вектор. Вопросы задаются относительно

других характеристик, которые могут быть сопоставлены этим векторам. В случае со звуками речи M признаков – это MFCC-или PLP-векторы, а характеристики, относительно которых задаются вопросы, это такие признаки фонем, как место образования (губной, переднеязычный, заднеязычный), способ образования (смычный, щелевой), звонкость и проч.

Алгоритм CART позволяет для каждой фонемы определить оптимальную последовательность вопросов в порядке убывания путем выявления на каждом этапе ветвления такого вопроса, для которого значение среднеквадратической ошибки минимально. На выходе алгоритм выдает дерево регрессии, в котором в качестве листов выступают итоговые кластеры, используемые в дальнейшем при акустическом моделировании. Вкратце алгоритм CART можно описать следующей последовательностью действий:

1. Вычисляется исходное значение суммарной среднеквадратической ошибки (среднеквадратическое отклонение точки от среднего в M -мерном пространстве признаков) и взвешенной среднеквадратической ошибки на всех векторах, входящих в обучающую базу. Суммарная ошибка вычисляется как сумма среднеквадратических отклонений всех векторов, взвешенная ошибка – это суммарная ошибка, деленная на количество векторов.

Далее в цикле для всех вопросов выполняются пункты 2 и 3:

2. Задаем вопрос, тем самым разбивая дерево на две «ветки»:
 - 2.1. задаем вопрос (например, «является ли правый контекст согласным») и ищем векторы характеристики которых соответствуют положительному ответу на заданный вопрос («левая ветка»);
 - 2.2. считаем по найденным векторам суммарную ошибку и взвешенную. При этом взвешивание происходит по количеству всех векторов, а не по количеству векторов, для которых ответ на вопрос положительный;
 - 2.3. задаем тот же вопрос, ищем векторы, характеристики которых соответствуют отрицательному ответу на вопрос («правая ветка»);

2.4. считаем по найденным векторам суммарную ошибку и взвешенную. При этом взвешивание происходит по количеству всех векторов, а не по количеству векторов, для которых ответ на вопрос отрицательный;

3. Вычисляем критерий разбиения, по которому и производится оптимизация. Для этого вычисляем разность исходной взвешенной ошибки и суммы взвешенной ошибки на левой ветке и взвешенной ошибки на правой ветке.

4. После того как критерий разбиения вычислен для каждого вопроса, мы выбираем вопрос с наибольшим значением критерия. Именно он соответствует наилучшему разбиению.

5. Дальнейшее разбиение веток происходит по тому же принципу, при этом на вход подаются уже только те векторы, которые располагаются в данной ветке дерева. Цель каждого разбиения состоит в том, чтобы максимально уменьшить значение ошибки.

6. Итоговое дерево вопросов сохраняется, и в соответствии с ним формируются результирующие кластеры, которые затем передаются Модулю акустического моделирования.

На рис. 4 приведен пример дерева регрессии для фонемы «а».

В результате кластеризации мы получили классифицирующее дерево регрессии, листья которого являются кластерами, используемыми при дальнейшем акустическом моделировании. Левая ветка соответствует положительному ответу на вопрос, правая ветка соответствует отрицательному ответу на вопрос. Заштрихованный узел является терминальным (который не удалось разбить на две ветки по причине невыполнения критерия останова: либо вопрос не сокращает среднеквадратическую ошибку, либо отсутствует достаточное количество реализаций). Данные терминальные узлы и есть кластеры, в совокупности составляющие акустическую модель.

При последующем моделировании все трифоны, относящиеся к кластеру, будут использованы Модулем АМ для переоценки параметров СММ данного кластера.

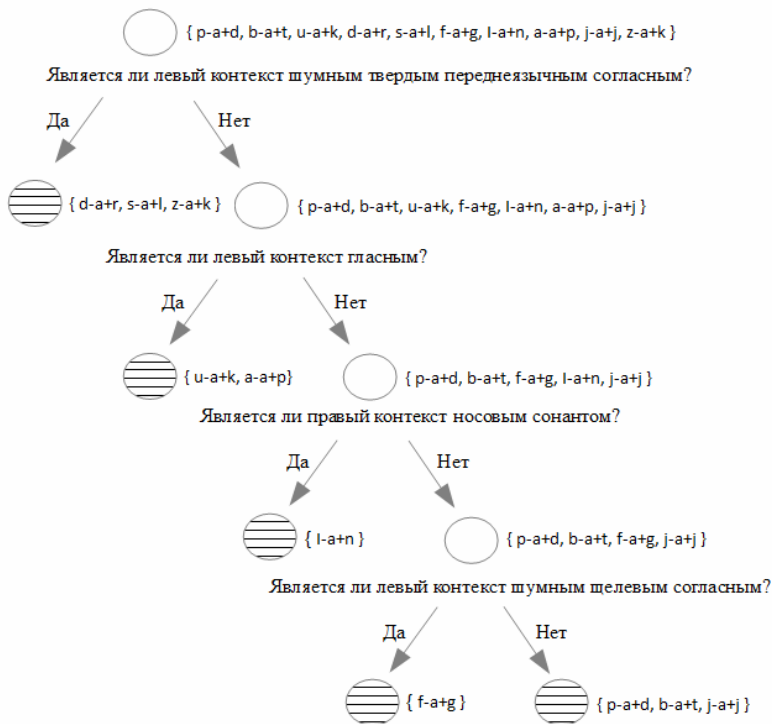


Рис. 4. Пример дерева регрессии для фонемы «а»

7. Результаты экспериментальных исследований

Оценку качества акустического моделирования следует осуществлять при помощи другого модуля системы АНРИ, а именно, модуля декодирования, отвечающего за поиск ключевых слов. Данный модуль обрабатывает тестовую выборку звуковых данных с использованием ранее созданной акустической модели. Далее результаты работы модуля декодирования сравниваются с экспертной аннотацией. Авторы используют для оценки два параметра качества – DR (Detection Rate) и FA (False Alarm).

Значение DR определяет процент правильно обнаруженных слов, и рассчитывается по формуле

$$(17) DR = 100\% \cdot N_{\text{found}}/N_{\text{all}},$$

где N_{found} – количество правильно найденных реализаций ключевых слов в тестовых данных; N_{all} – общее количество реализаций ключевых слов в тестовых данных.

Значение FA определяет количество ложных срабатываний в час и рассчитывается по формуле

$$(18) FA = (A_{\text{all}} - N_{\text{found}})/Hrs,$$

где A_{all} – общее количество всех найденных слов в тестовых данных; Hrs – длительность звучания тестовых данных в часах.

Для того чтобы показать важность различных этапов акустического моделирования, авторами статьи был проведен ряд экспериментов с подключением к модулю декодирования различных акустических моделей. Отметим, что на качество поиска ключевых слов также влияют другие компоненты системы АНРИ, такие как модуль вычисления акустических признаков или архитектура модуля верификации, а также сами обучающие базы данных (ср. рис. 1 в начале настоящей статьи). В целях обеспечения объективного сравнения все эти компоненты в ходе экспериментов не модифицировались.

Поскольку при оценке системы АНРИ используется два параметра, для сопоставления результатов экспериментов удобно определить базисное значение одной из величин, например, FA , и дальнейшее сравнение проводить уже только по одной величине. В нашем случае для каждой модели вычисляется значение DR при фиксированном значении FA , составляющем 100 шт. в час.

В таблице 2 приведены значения DR , полученные в ходе экспериментов, проведенных авторами на тестовой базе данных. В тестовую базу данных входит 100 различных ключевых слов, для каждого ключевого слова имеется в среднем 5 реализаций. Результаты экспериментов показывают, что переход к трифонам значительно повышает качество работы системы, в то же время свой вклад в качество работы системы вносит и особый инвентарь монофонов, включающий дополнительные аллофоны фонем русского языка.

Таблица 2. Зависимость качества работы системы АНРИ от типа акустической модели.

Эксперимент (подключаемая акустическая модель)	DR (%)
Монофонная модель	50,25
Стандартная трифонная модель	67,02
Расширенный список монофонов	55,30
Трифонная модель на базе расширенного списка монофонов	70,42

8. Заключение

В настоящей статье был рассмотрен Модуль акустического моделирования как компонент системы автоматического анализа неструктурированной речевой информации. Дано детальное описание математического аппарата скрытых Марковских моделей применительно к задаче моделирования речи как временного ряда и определены базовые методики вычисления параметров Акустической модели – преобразование Баума–Уэлша. На иллюстративном примере рассмотрена работа алгоритма кластеризации по дереву регрессии – инструмента, используемого для объединения нескольких трифонов в одну группу, для которой становится возможной надежная статистическая оценка параметров СММ. Приведены результаты экспериментов, показывающие зависимость результатов работы системы АНРИ от типа подключаемой акустической модели.

Предложенный в статье вариант реализации АМ в целом следует общепризнанным алгоритмам. При этом отличительная особенность описанного варианта реализации АМ состоит в использовании особого инвентаря монофонов, обусловленного особенностями фонетического строя русского языка. Проведенные эксперименты показывают, что гибкая интерпретация знаний о свойствах аллофонов русских гласных позволяет повысить качество работы системы АНРИ, тем самым обеспечивая улучшенные пользовательские характеристики продуктов, построенных на базе этой системы.

Литература

1. БОНДАРКО Л.В., ВЕРБИЦКАЯ Л.А., ГОРДИНА М.В. *Основы общей фонетики*. – Предисловие. – СПб.: Просвещение, 1991. – С. 3.
2. ЗИНДЕР Л.Р. *Общая фонетика*. – М.: Высшая школа 1979. – С. 4.
3. BAKER J. *The DRAGON system-An overview* // IEEE Transactions on Acoustics, Speech, and Signal Processing. – 1975. – Vol. 23(1). – P. 24–29.
4. BAUM L.E. *An Inequality and Associated Maximization Technique in Statistical Estimation of Probabilistic Functions of a Markov Process* // Inequalities. – 1972. – Vol. 3. – P. 1–8.
5. BOURLARD H., MORGAN N. *Connectionist Speech Recognition. A Hybrid Approach* // The Kluwer International Series in Engineering and Computer Science. – 1994. – Vol. 247. – P. 100–105.
6. BREIMAN LEO, FRIEDMAN J.H., OLSHEN R.A. AND ETC. *Classification and regression trees*. – Monterey, CA, 1984. – pp. 102–116
7. DEMPSTER A.P., LAIRD N.M., UBIN B. *Maximum likelihood from incomplete data via the EM algorithm* // J. Roy. Stat. Soc. – 1977. – Vol. 39, №1. – P. 1–38.
8. FOMEY G.D. *The Viterbi algorithm* // Proc. IEEE, 61. – March, 1973. – P. 268–278.
9. GALES M. and YOUNG S. *The Application of Hidden Markov Models in Speech Recognition* // Foundations and Trends in Signal Processing. – 2007. – Vol. 1, №3. – P. 195–304.
10. HASTIE T., TIBSHIRANI R., FRIEDMAN J.H. *The elements of statistical learning: Data mining, inference, and prediction*. – New York: Springer Verlag, 2009. – 745 p.
11. HONG Q.Y., KWONG S. *A training method for hidden Markov model with maximum model distance and genetic algorithm* // Proc. IEEE International Conference on Neural Network & Signal Processing, Nanjing. – 2003. – Vol. 1. – P. 465–468.
12. JELINEK F., BAHL L., MERCER R. *Design of a linguistic statistical decoder for the recognition of continuous speech* // IEEE

- Transactions on Information Theory. – 1975. – Vol. 21 (3). – P. 250 - 256.
13. JUANG B.H., CHOU W., LEE C.H. *Statistical and discriminative methods for speech recognition* // Speech Recognition and Coding – New Advances and Trends / Eds. A.J.R. Ayuso and J.M.L. Soler. – Springer Verlag, Berlin, 1995. – P. 41–42.
 14. LEVINSON S.E., RABINER L.R., SONDHI M.M. *An introduction to the application of the theory of probabilistic function of a Markov process to automatic speech recognition* // Bell Systems Technical Journal. – Apr., 1983. – Vol. 62, №4. – P. 1035–1074.
 15. NEY H. *On the probabilistic-interpretation of neural-network classifiers and discriminative training criteria* // IEEE Trans. on Pattern Analysis and Machine Intelligence. – 1995. – Vol. 17(2). – P. 107–119.
 16. RABINER L.R. A TUTORIAL ON HIDDEN MARKOV MODELS AND SELECTED APPLICATION IN SPEECH RECOGNITION // PROC. INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS. – 1989. – VOL. 77, №2. – P. 257–286.
 17. VITERBI A.J. ERROR BOUNDS FOR CONVOLUTIONAL CODES AND AN ASYMPTOTICALLY OPTIMAL DECODING ALGORITHM // IEEE TRANS. INFORMATION THEORY, IT-13, APRIL 1967. VOL. IT-13, №2– P. 260–269.
 18. HUANG X., ACERO A., HON H.-W. *Spoken Language Processing: a guide to theory, algorithm, and system development.* – Prentice Hall, 2001 – 512 pp.

ACOUSTIC MODELING MODULE FUNCTION IN A SYSTEM FOR AUTOMATIC ANALYSIS OF UNSTRUCTURED SPEECH DATA

Valentin Smirnov, Speech Drive, Ltd. – company, St-Petersburg, General Manager, (speechdrive@mail.ru).

Michael Gusev, Research Institute “KVANT”, St-Petersburg, 1st category engineer, Ph.D. (mgaev@kvant-rdi.spb.ru).

Mais Farkhadov, Institute of Control Sciences of RAS, Moscow, Head of Lab., Doctor of Science, (Moscow, Profsoyuznaya st., 65, (495) 334-87-10), (mais@ipu.ru).

Abstract: We focus on acoustic modeling – a component of a system for automatic analysis of unstructured speech data, which aims at creating statistical description for speech sounds. In the first part of the paper we outline goals of a system for automatic analysis of unstructured speech data and explain the place and functions of acoustic modeling within the system. Then we suggest an algorithm of acoustic modeling and address each of its stages in detail.

Keywords: acoustic modeling, hidden Markov models, forward-backward algorithm, Baum-Welch re-estimation, classification and regression trees, automatic analysis of unstructured speech data.

*Статья представлена к публикации
членом редакционной коллегии М.В. Губко*

*Поступила в редакцию 29.03.2013.
Опубликована 30.09.2013.*

УДК 330.46
ББК 65.9

ИМИТАЦИОННЫЕ МОДЕЛИ КОМПЛЕКСНОГО РЕГИОНАЛЬНОГО РАЗВИТИЯ¹

Гафарова Е. А.²

*(Институт социально-экономических исследований
Уфимского научного центра РАН, Уфа)*

Представлен обзор отечественных имитационных моделей регионального развития, базирующихся на системной динамике и агент-ориентированном подходе. Показано, что универсальные модели, не учитывающие специфику развития региона и не верифицированные на его статистических данных, имеют низкую точность прогноза. Рассмотрены наиболее известные российские модели социально-экономического развития регионов, проведен их сравнительный анализ.

Ключевые слова: имитационное моделирование регионального развития, прогнозирование социально-экономического развития региона, системная динамика, агент-ориентированное моделирование.

1. Введение

Одним из наиболее эффективных экономико-математических подходов к исследованию и прогнозированию регионального развития является имитационное моделирование, позволяющее сформировать и исследовать многовариантные сценарии развития моделируемой системы, оценить социально-

¹ Работа выполнена при поддержке РГНФ (проект №13-02-00039 «Технология моделирования комплексного развития региона на основе имитационного подхода»).

² Елена Аркадьевна Гафарова, кандидат экономических наук (gafarova@mail.ru).

экономическую эффективность различных управленческих решений для региона и выбрать оптимальный вариант его развития.

Как известно, различают три подхода в имитационном моделировании: дискретно-событийное, системная динамика, агент-ориентированное моделирование. Два последних из них применяются для моделирования регионального развития. При этом наиболее востребованной для моделирования регионального развития до последнего времени считалась системная динамика. При моделировании с помощью указанной концепции абстрагируют от индивидуальных объектов системы и рассматривают только агрегированные элементы системы (например, агрегированный потребитель, агрегированный производитель). Системная динамика при компьютерном моделировании реализуется с помощью графических диаграмм потоков и накопителей, отражающих причинно-следственные связи в моделируемой системе. Прогнозная динамика основных социально-экономических показателей региона обеспечивается за счет наличия дифференциальных уравнений, рекурсивных отношений, а также зависимостей от лаговых переменных, полученных средствами эконометрики или нейроинформатики. Агент-ориентированное моделирование представляет относительно новое направление, учитывающее индивидуальное поведение активных объектов (агентов), а также их взаимодействия. В отличие от системной динамики, агентный подход позволяет оперировать не агрегированными элементами системы, а совокупностью агентов с определенным набором свойств (например, агенты – потребители, агенты – производители).

В данной статье рассмотрим эволюцию имитационных моделей социально-экономического развития регионов.

2. Универсальные имитационные модели развития регионов

Обзор имитационных моделей для прогнозирования регионального развития начнем с готовых решений, предлагаемых некоторыми фирмами для региональных и муниципальных

органов власти. Примерами могут служить программно-инструментальный комплекс «Прогноз» (ЗАО «Прогноз») [8] или система «Губернатор» (ООО «Франклин&Грант. Риск консалтинг») [5], предназначенные для мониторинга, анализа и прогнозирования регионального развития. Такие универсальные модели представляют собой, как правило, систему зависимостей, описывающих состояние региональной экономики на краткосрочную и среднесрочную перспективу. Математические модели, используемые для прогнозирования, могут настраиваться на специфику каждого моделируемого региона и его статистические данные. Обычно для прогнозирования на основе таких универсальных моделей используются статистические данные за 5–7 лет. Такой короткий ретроспективный анализ, на наш взгляд, не способствует достоверному прогнозированию. Кроме того, отсутствие специалистов по экономико-математическому моделированию среди пользователей таких систем не позволяет произвести грамотную настройку модели на специфику региона (да и не всегда это возможно). Все сказанное приводит к тому, что региональное прогнозирование на основе таких моделей приобретает формальный характер.

Поясним сказанное на примере. Проверим качество моделей, заложенных в программно-инструментальный комплекс «Прогноз», применительно к статистическим данным по республике Башкортостан. Для примера был выбран такой показатель, как численность занятых в экономике. В соответствии с методикой ЗАО «Прогноз», расчетное (прогнозное) значение этого показателя определяется в системе следующей формулой:

$$L[t] = L[t-1] \cdot \left(LM_1[t] \cdot T_VR[t] + LM_2[t] \cdot \frac{TR[t]}{TR[t-1]} \cdot 100 \right) / 100,$$

где $L[t]$, $L[t-1]$ — численность занятых в экономике (среднегодовая) в моменты времени t и $t-1$ соответственно, тыс. человек; $T_VR[t]$ — валовой региональный продукт, в % к предыдущему в сопоставимых ценах; $LM_1[t]$, $LM_2[t]$ — оцениваемые коэффициенты влияния; $TR[t]$, $TR[t-1]$ — численность

трудовых ресурсов в моменты времени t и $t - 1$ соответственно, тыс. человек.

Произведем расчеты по этой модели для статистических данных Республики Башкортостан в программе Eviews. В таблице 1 представлены оцененные значения коэффициентов LM_1 и LM_2 , значения t -статистик о проверке значимости каждого коэффициента модели, коэффициент детерминации (величина R^2), характеризующий качество подгонки, и др.

Таблица 1. Результаты оценивания регрессионного уравнения в программе Eviews на интервале 2007-2011 гг.

Зависимая переменная: L				
Метод: наименьших квадратов				
Выборка: 2007–2011				
$L = L(-1) \cdot (LM_1 \cdot T_VR + LM_2 \cdot TR_I \cdot 100)/100$				
	Коэффициент	Стандартная ошибка	t -статистика	p -уровень
LM_1	-0,027	0,075	-0,357	0,745
LM_2	1,022	0,085	11,982	0,001
R^2	0,576	Среднее зависимой переменной		1803,340
Скорректированный R^2	0,435	Стандартное отклонение зависимой переменной		40,230
Стандартная ошибка регрессии	30,232	Критерий Акаике		9,945
Сумма квадратов остатков	2741,985	Критерий Шварца		9,789
Логарифм функции правдоподобия	-22,862	Статистика Дарбина-Уотсона		1,715

Как видно из таблицы, построенное регрессионное уравнение не является адекватным, поскольку наблюдается низкое качество подгонки расчетных и наблюдаемых значений ($R^2 = 0,576$), а коэффициент LM_1 не является статистически значимым ($t = -0,357$). Для наглядности наблюдаемые и расчетные значения для численности занятых в экономике РБ, а также остатки модели представлены на рис. 1.

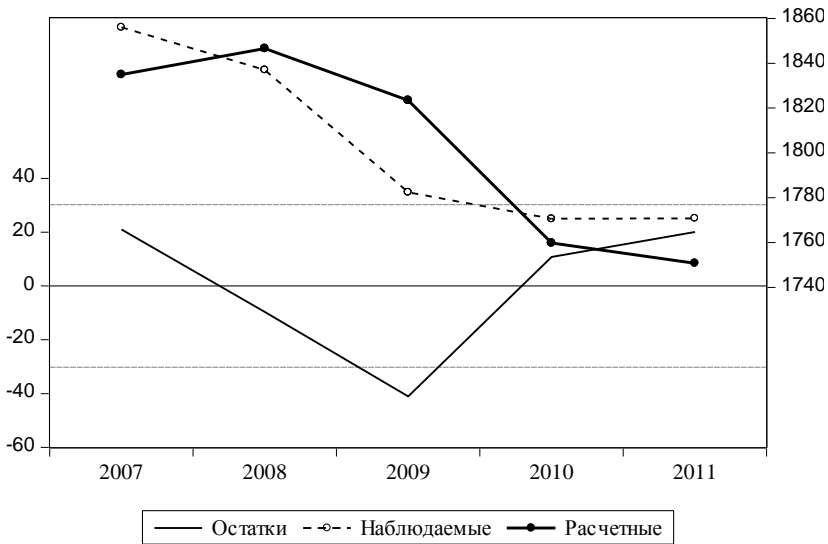


Рис. 1. График наблюдаемых, расчетных значений (левая шкала) и остатков (правая шкала)

Для настройки на специфику моделируемого региона и статистические данные программно-инструментальный комплекс «Прогноз» позволяет формировать новые зависимости или задать другие значения параметров, но для этого же уравнения. На практике пользователи занимаются многократным перебором значений параметров для заложенных в систему зависимостей. Таким образом, прогнозирование производится по неадекватному уравнению. При социально-экономическом прогнозировании РБ на 2013–2015 гг. на базе регионального сегмента государственной автоматизированной системы «Управление», разработанной компанией «Прогноз», разработчиками прогноза выбраны значения для коэффициентов $LM_1 = 0,5$ и $LM_2 = 0,5$. Наблюдаемые и расчетные значения для численности занятых в экономике РБ, рассчитанные при таких значениях параметров, показаны на рис. 2.

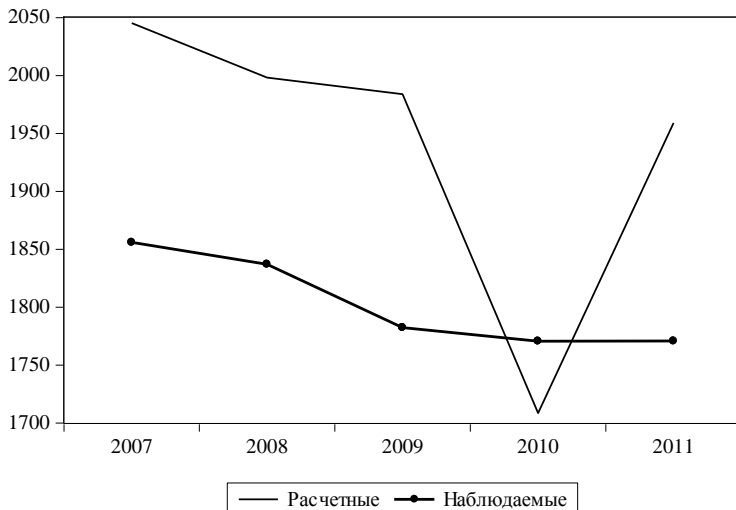


Рис. 2. График наблюдаемых (L) и расчетных (L_1) значений.

Из графиков видно, что в обоих случаях имеется существенное расхождение в расчетных и наблюдаемых значениях. Ошибка прогнозирования при этом составила порядка 10%. Безусловно, низкому качеству эконометрического уравнения и прогноза могло способствовать изменение динамики региональных показателей в кризисные годы. Однако неудовлетворительные результаты моделирования выбранного показателя на основе данного уравнения получаются и на интервалах наиболее стабильного экономического развития. В таблице 2 представлены результаты эконометрического моделирования на данных 2000–2007 гг., которые также свидетельствуют о низком качестве подгонки, незначимости коэффициента при валовом региональном продукте и, как следствие, о неадекватности рассматриваемого уравнения.

Учитывая вышесказанное, далее будем рассматривать только те модели, в которых детально исследована экономическая деятельность отдельного субъекта РФ.

Таблица 2. Результаты оценивания регрессионного уравнения в программе Eviews на интервале 2000-2007 гг.

Зависимая переменная: L				
Метод: наименьших квадратов				
Выборка: 2000–2007				
$L = L(-1) \cdot (LM_1 \cdot T_VR + LM_2 \cdot TR_I \cdot 100)/100$				
	Коэффициент	Стандартная ошибка	t -статистика	p -уровень
LM_1	0,053	5,444	0,010	0,993
LM_2	0,998	0,063	15,723	0,000
R^2	0,587	Среднее зависимой переменной		1792,923
Скорректированный R^2	0,518	Стандартное отклонение зависимой переменной		40,377
Стандартная ошибка регрессии	28,035	Критерий Акаике		9,717
Сумма квадратов остатков	4715,925	Критерий Шварца		9,737
Логарифм функции правдоподобия	-36,869	Статистика Дарбина-Уотсона		2,454

3. Имитационные модели развития региона, учитывающие его специфику

Тематика регионального моделирования сконцентрирована большей частью в научно-исследовательских институтах РАН. Исследования начаты в Сибирском отделении Академии наук. Рассмотрим прежде наиболее известные разработки российских ученых, использующих системно-динамический подход применительно к региональному моделированию (таблица 3).

Таблица 3. Системно-динамические модели регионального развития

№	Название модели	Разработчик	Структура модели	Среда ИМ
1.	Модельный комплекс СИРЕНА-2 (Синтез региональных народно-хозяйственных решений) [7, 9],	Институт экономики и организации промышленного производства СО РАН	Структура региональной подсистемы: трудовые ресурсы и население; капитальные вложения; локальные ресурсы (земельные, водные); транспорт; межрегиональные связи; социальные факторы.	Специализированные программные комплексы
2.	Территориальная автоматизированная система (ТАИС) [11]	Самарская государственная экономическая академия	Экономические агенты: хозяйствующие субъекты сектора производства товаров и рыночных услуг; хозяйствующие субъекты сектора нерыночных услуг; домашние хозяйства; региональное руководство; Федеральное руководство; внешнее окружение.	Специализированная программа Economics
3.	Аналитическая система «Моделирование социально-экономического развития региона» [2]	Государственный университет управления	Подсистемы: население; производство; непроемкая сфера; внешняя экономическая сфера; пространство; финансы; экология.	Vensim

№	Название модели	Разработчик	Структура модели	Среда ИМ
4.	Система прогнозирования развития региона [1]	Институт информатики и математического моделирования технологических процессов КИЦ РАН	Компоненты: трудовые ресурсы; добывающая промышленность (ТЭК); перерабатывающая промышленность (рыбо-промышленный комплекс).	PowerSim
5.	Модель Кировской области [6]	Вычислительный центр им. А.А. Дородницына РАН	Экономические агенты: производители; банки; домашние хозяйства; собственники; область; центральный банк; внешняя торговля.	Maple
6.	Комплексная модель обоснования среднесрочных стратегий социально-экономического развития региона [4]	Институт социально-экономических исследований УНЦ РАН	Экономические агенты: домохозяйства; совокупный производитель; финансовый сектор; государство (региональный уровень); внешний мир.	Matlab

Все модели, по утверждениям их разработчиков, позволяют строить долгосрочные и краткосрочные прогнозы основных социально-экономических и финансовых показателей региона, формировать и исследовать различные стратегии регионального развития. Исключение составляет модельный комплекс СИРЕНА-2, который является многорегиональным и поэтому позволяет проводить межрегиональные сравнения, выявлять тенденции развития нескольких регионов и др.

Что касается степени детализации, считаем, что все модели являются слишком агрегированными. Лишь в моделях №2–№4 (таблица 3) имеет место детализация по некоторым секторам или сферам экономики, что позволяет использовать их для анализа диспропорций регионального развития. Кроме того, следует отметить социальную направленность модели №3, в которой рассматривается развитие здравоохранения, образования и жилого фонда. В этой модели присутствует также детализация ресурсного потенциала региона по видам природных ресурсов. Следующим важным моментом является моделирование теневого сектора и его влияния на развитие экономики региона. В этом случае выгодно отличается модель №5, в которой исследуется теневой выпуск и теневая заработная плата. Существенным недостатком всех рассмотренных моделей является игнорирование моделирования инноваций.

Упомянутые в таблице региональные модели различаются между собой также разным экономико-математическим инструментарием, лежащим в основе имитационного моделирования (дифференциально-интегральные уравнения, нейронные сети, эконометрические уравнения и др.), обеспечением соблюдения баланса между спросом и предложением на рассматриваемых рынках. Однако разработчики, ориентируясь на широкий круг читателей, весь модельный инструментарий, как правило, не раскрывают. Так, в работе [2] нет информации о внутреннем устройстве модели №3. Поэтому сравнение моделей по этим параметрам не представляется возможным.

Ключевым моментом регионального имитационного моделирования является проведение направленных вычислительных экспериментов с целью получения информации о системе. Так,

с помощью модели №1 получены прогнозы наборов региональных индикаторов (ВРП, промышленное производство, заработная плата и др.) субъектов Сибирского федерального округа в зависимости от реализации инвестиционных проектов. Для модели №2 приводится подробное описание экспериментов, направленных на решение таких задач, как удвоение ВРП Самарской области, увеличение численности трудоспособного населения, уменьшения уровня бедности. Разработчики модели №4 представили прогнозную динамику трудовых ресурсов Мурманской области при отмене северных льгот, а также прогноз развития топливно-энергетического комплекса в зависимости от инвестиций и федерального налога на электроэнергию. На основе модели №5 проведены сценарии развития Кировской области, в которых исследовался рост ВРП и легальной наличности в зависимости от множества управляющих параметров, таких как санкции налоговой инспекции, норма выдачи кредитов, инфляция и др. Для остальных моделей описание проведенных экспериментов отсутствует.

Перейдем к рассмотрению агент-ориентированных моделей регионального развития (таблица 4). Среди них не наблюдается такого разнообразия, как среди системно-динамических моделей. В настоящее время разработки агент-ориентированных моделей большей частью сконцентрированы в Центральном экономико-математическом институте РАН.

Для первой модели (таблица 4) в работе [3] приводятся лишь концептуальные схемы взаимодействия ее агентов. Более детальные алгоритмы поведения агентов, результаты верификации и численных экспериментов для субъекта РФ авторами не представлены. Модель №3 имеет, на наш взгляд, узкую направленность. Хотя модель и предполагает расчет основных социально-экономических показателей развития г. Москвы (численность населения, ВРП, уровень безработицы, среднедушевые доходы), тем не менее, ориентирована на анализ уровня загруженности дорог и решения проблемы пробок. В силу этого, в стейтчарте, описывающем поведение агента-человека, рассматриваются такие действия, как смена работы, смена места жительства, покупка или продажа автомобиля.

Таблица 4. Агент-ориентированные модели регионального развития

№	Название модели	Разработчик	Структура модели	Среда ИМ
1.	Модель региона Российского Севера [3]	Сыктывкарский государственный университет	Уровни: экономические агенты; район; регион. Агенты: человек; предприятие; банк; собственник; район; государство.	Vensim
2.	Агент-ориентированная региональная модель «Губернатор» [10]	Центральный экономико-математический институт РАН	Агенты: физические лица; юридические лица; муниципальные районы.	Anylogic
3.	Агент-ориентированная модель г. Москвы [12]	Центральный экономико-математический институт РАН	Агенты: люди; предприятия. Среда модели: районы; дороги; общественный транспорт (метро).	Anylogic

И наконец, модель №2 разработана для апробации механизмов регионального управления с учетом человеческого фактора. В рамках этой модели предпринята попытка качественного моделирования индивидуального поведения агентов - «физических лиц»: рождение, поступление на учебу, устройство на работу, смена вида деятельности, смена места жительства, изменение отношения к труду, смерть. Модель прошла калибровку на данных о состоянии Вологодской области. С использованием модели проведены эксперименты по изменению динамики населения области, ее муниципальных районов и предприятий при разных значениях управляющих параметров. В

тоже время, считаем, что модель имеет некоторые ограничения (например, предполагает полное отсутствие безработицы) и требует дальнейшего развития.

Заканчивая обзор имитационных моделей, отметим, что создание моделей «первого поколения» требовало от исследователей прежде всего разработки программ, позволяющих создавать архитектуру модели и производить расчеты по ней. Так, для организации работы комплекса СИРЕНА-2 (таблица 3) были разработаны специализированные программы (КОМБИ для выполнения операций над матрицами, АССТРА для генерации и анализа прогнозных задач, УТРО для обработки и вывода данных в диалоговом режиме и др.). Более поздние модели создавались с использованием программ широкого профиля (Maple, Matlab и др.). Современные имитационные модели разрабатываются в специализированных программах, содержащих наборы готовых библиотек для системной динамики или агент-ориентированного моделирования. Поэтому процесс моделирования стал протекать легче и быстрее. В связи со значительным развитием программного обеспечения в последние годы появилась возможность интеграции агентного моделирования и системной динамики. Примером может служить программа AnyLogic российской компании XJ Technologies, поддерживающая все парадигмы имитационного моделирования и предоставляющая возможность их комбинации в рамках одной модели.

Таким образом, развитие агентного моделирования, а также программного обеспечения в области имитационного моделирования открывают перед учеными большие перспективы для повышения глубины проработки проблемы регионального моделирования. Так, ранее для республики Башкортостан ранее была создана имитационная модель комплексного развития региона [4], реализованная в среде Matlab. В настоящее время в Институте социально-экономических исследований УНЦ РАН ведутся разработки модели регионального развития в среде AnyLogic на основе системной динамики с детализацией ВРП по видам экономической деятельности и описанием выявленных зависимостей на основе эконометрических моделей. Кроме того, разрабатывается модель социально-экономического развития

региона средствами Anylogic на основе интеграции системной динамики и агентного моделирования.

Литература

1. ГОРОХОВ А.В., ПУТИЛОВ В.А. ШЕСТАКОВ А.А. *Системная динамика в управлении региональным развитием (на примере Мурманской области)* // Проблемы управления и моделирования в сложных системах: Тр. 6-й Международной конференции, Самара: СНЦ РАН, 2004. – С. 370–375.
2. ЛЫЧКИНА Н.Н. *Компьютерное моделирование социально-экономического развития регионов в системах поддержки принятия решений* // Материалы III Международной конференции «Идентификация систем и задачи управления» SICPRO'04 – М.: ИПУ РАН, 2004. – 25 с.
3. МИРОНОВ В.В., СМЕРНОВ А.В. *К проблеме математического моделирования имитационных моделей социально-экономического развития региона* // Корпоративное управление и инновационное развитие экономики Севера: Вестник Научно-исследовательского центра корпоративного права, управления и венчурного инвестирования Сыктывкарского государственного университета. – 2012. – № 3. – [Электронный ресурс]. – URL: <http://koet.syktsu.ru/vestnik/2012/2012-3/8/8.html> (дата обращения: 22.04.2013).
4. НИЗАМУТДИНОВ М.М., ЯМИЛОВА Л.С. *К вопросу о подходах к построению модели экономики регионов* // Проблемный анализ и государственно-управленческое проектирование. – 2010. – Т. 3, №4. – С. 52–63.
5. *Система анализа и моделирования динамики бюджета региона «Губернатор»*© / ООО «Франклин&Грант. Риск Консалтинг», 2002–2012. [Электронный ресурс]. – URL: <http://www.franklin-grant.ru/ru/services/gov-gubernator.asp> (дата обращения: 22.04.2013).
6. ОСИПОВА О.Н., ШАТРОВ А.В. *Эволюционная модель региональной экономики* // Математика. Компьютер. Образование: Сб. трудов XIV международной конференции,

2008. – [Электронный ресурс]. – URL: <http://www.mce.su/rus/archive/mce14/sect1032/doc12236/> (дата обращения: 22.04.2013).
7. Проект СИРЕНА: методология и инструментарий / А.Г. Гранберг, В.Е. Селиверстов, С.А. Суспицын и др. – Новосибирск: Наука. Сиб. отд-ние, 1991. – 256 с.
 8. *Разработка прогноза социально-экономического развития регионов с использованием комплексной имитационной модели* // ©Прогноз, 2005–2013. [Электронный ресурс]. – URL: http://prognoz.ru/ru/publication_05.php (дата обращения: 22.04.2013).
 9. СУСПИЦЫН С.А. *Взаимодействие институциональных моделей региона в информационном пространстве* // Труды ИСА РАН. – 2008. – Т. 39. – С. 435–456.
 10. СУШКО Е.Д. *Мультиагентная модель региона: концепция, конструкция и реализация* // Препринт #WP/2012/292. – М.: ЦЭМИ РАН. – 2012. – 54 с.
 11. ЦЫБАТОВ В.А. *Моделирование экономического роста* / Под ред. Г.Р. Хасаева. – Самара: Изд-во Самар. гос. экон. ун-та. – 2006. – 385 с.
 12. ФАТТАХОВ М.Р. *Агент-ориентированная модель социально-экономического развития мегаполисов (на примере г. Москвы)*: Автореф. дис. канд. эконом. наук. – М., 2011. – 30 с.

SIMULATION MODELS OF INTEGRATED REGIONAL DEVELOPMENT

Elena Gafarova, Institute of social and economic research, USC RAS, Ufa, Russia, PhD in Economics, staff scientist of economic-mathematical modeling sector (gafarova@mail.ru).

Abstract: This article provides an overview of domestic simulation models of regional development based on system dynamics and agent-based approach. It is shown that the universal models that do not take into account the specific development of the region and are not verified by the regional statistical data have low forecasting accuracy. The article describes the most famous Russian models of socio-economic regional development and provides comparative analysis of these models.

Keywords: regional development simulation, forecasting of socio-economic development of the region, system dynamics, agent-based modeling.

*Статья представлена к публикации
членом редакционной коллегии Р.М. Нижегородцевым*

*Поступила в редакцию 22.04.2013.
Опубликована 30.09.2013.*

КОНЦЕПТУАЛЬНЫЙ ПОДХОД К АНАЛИЗУ ОНЛАЙНОВЫХ СОЦИАЛЬНЫХ СЕТЕЙ¹

Губанов Д. А.², Чхартишвили А. Г.³
(ФГБУН Институт проблем управления

им. В.А. Трапезникова РАН, Москва
Лаборатория Цифрового Общества (DSL), Москва)

В статье предложен конструктивный подход к анализу социальных сетей. Предложена концептуальная модель социальной сети, поставлены базовые задачи анализа и управления в социальных сетях, определены методы и алгоритмы анализа активности в социальных сетях, а также технологии мониторинга и анализа социальных сетей.

Ключевые слова: социальная сеть, концептуальная модель, информационное управление, информационные технологии.

1. Введение

В последнее десятилетие онлайн-социальные сети прочно вошли в нашу жизнь и стали ее неотъемлемой частью. Практически у каждого сейчас есть своя страница в тех или иных онлайн-социальных сетях; наиболее популярные сети (FaceBook – 955 миллионов пользователей, ВКонтакте –

¹ Работа выполнена при поддержке гранта РФФИ № 10-07-00104.

² Дмитрий Алексеевич Губанов, кандидат технических наук, старший научный сотрудник ИПУ РАН, старший исследователь Лаборатории Цифрового Общества (DSL) (dmitry.a.g@gmail.com)

³ Александр Гедеванович Чхартишвили, доктор физико-математических наук, главный научный сотрудник ИПУ РАН, научный директор Лаборатории Цифрового Общества (DSL) (sandro_ch@mail.ru)

140 миллионов пользователей¹) пытаются обеспечить пользователям возможность общаться, обсуждать новости, вступать в группы по интересам, слушать музыку и видео, обмениваться ссылками и т.п., не покидая их пределы. В результате российский интернет-пользователь проводит в среднем в месяц в социальных сетях 10,4 часа².

В целом складывающуюся в онлайн-социальных сетях (блоги, форумы, классические социальные сети) ситуацию можно охарактеризовать следующими тезисами: быстрыми темпами растет число и интенсивность взаимодействия пользователей социальных сетей; пользователи оставляют много информационных следов в интернете (эти следы могут быть неявными); наращивание активности обсуждения может происходить «стихийно» (латентный, фрактальный характер) или по определенному поводу; на конкурентном рынке совершенствуется методический и ИТ-инструментарий поддержки взаимодействия пользователей (краудсорсинг, экспертные сети, виртуальное сотрудничество, интеллектуальные технологии).

Бурное развитие технологий и сопутствующие ему изменения в экономической, социальной, политической жизни делает актуальным разработку концептуального подхода к анализу онлайн-социальных сетей, чему и посвящена данная работа. Необходимо отметить, что существуют монографии [см. например, 2, 8], посвященные анализу онлайн-социальных сетей, но в них отсутствуют сколько-нибудь развернутые содержательные модели онлайн-социальных сетей (тем более с точки зрения постановки задач управления).

2. Подход к анализу социальных сетей

Основными задачами информационно-аналитической поддержки пассивной и активной работы с социальными сетями являются: мониторинг и анализ социальных сетей (для дости-

¹ По состоянию на июнь 2012 года (<http://pdf.secdatabase.com/700/0001193125-12-316895.pdf>) и сентябрь 2012 года (<http://vk.com/help.php?page=about>)

² ComScoreWorldMetrics, декабрь 2011 г.

жения понимания происходящих в социальных сетях процессов), прогнозирование и управление (для приведения социальной сети в требуемое состояние).

Для выполнения этих задач предлагается конструктивный подход, согласно которому:

1) определяются содержательные вопросы, на которые предполагается получить ответы;

2) определяются показатели и их системы для ответа на содержательные вопросы (отметим, что для определения показателей требуются содержательные модели, для расчета показателей – математические модели и методы);

3) разрабатываются технологии мониторинга и анализа социальных сетей;

4) разрабатываются технологии генерации отчетов и выдачи рекомендаций в зависимости от предметной области и заинтересованных лиц.

В данной работе рассматриваются первые три уровня предлагаемого нами подхода.

3. Основные элементы модели социальной сети

Основными элементами модели социальной сети предлагается считать следующие типы объектов:

- внешний узел (СМИ, информационный ресурс и т.п.; внешние узлы обычно существуют в том же виртуальном пространстве, что и онлайн-социальные сети (например, онлайн-ресурсы типа *lenta.ru*));

- внешний управляющий субъект, оказывающий те или иные воздействия на социальную сеть (политический деятель, партия, коммерческая организация и т.п.);

- социальная сеть в целом либо подсеть (множество узлов, выделенных по какому-либо признаку, и связей между ними);

- пользователь сети – член социума;

- узел сети (агент);

- информационное сообщение (пост, комментарий, сообщение);

- информационный объект – некоторая сущность, событие, персона и т.п. (информационные объекты могут быть сложными и состоять из других информационных объектов);
- информация – описание некоторых информационных объектов в информационном сообщении;
- мнение (суждение по какому-то поводу, точка зрения на какой-то объект, оценка).

Рассмотрим подробнее вышеперечисленные типы объектов.

Социальная сеть понимается нами как социальная структура, состоящая из множества агентов (пользователей, групп и сообществ) и определенного на нем множества отношений (информационного взаимодействия, знакомства, участия в группах и сообществах). В рамках этих отношений в сети происходят различные информационные процессы.

В данной модели достаточно определить явные и легко идентифицируемые отношения: отношения информационного взаимодействия (отношения post/reply, отношения цитирования, отношение подписки на сообщения); отношения знакомства (всевозможные виды контактов в социальных сетях); отношения участия в группах и сообществах. Эти виды отношений важны с точки зрения распространения информации и формирования мнений в социальных сетях. Все остальные виды отношений неявны (например, пользователи неявно связаны между собой, если они подписаны на одних и тех же пользователей, пишут на одни и те же темы и т.п.), эти виды отношений могут выявляться по мере необходимости для решения конкретных задач.

Социальная сеть не существует сама по себе, на процессы в ней могут повлиять какие-то внешние воздействия – либо события (например, выборы Президента РФ), либо какие-то целенаправленные воздействия со стороны внешних субъектов. *Внешние субъекты* (политические силы, коммерческие организации и пр.) могут быть заинтересованы в достижении сетью некоторого целевого состояния: требуемой степени связности (либо, наоборот, разобщенности) сети, формировании необходимой информированности в сети, в формировании требуемого общественного мнения в сети и т.п. Эта заинтересованность побуждает внешних субъектов осуществлять управление (рис. 1).

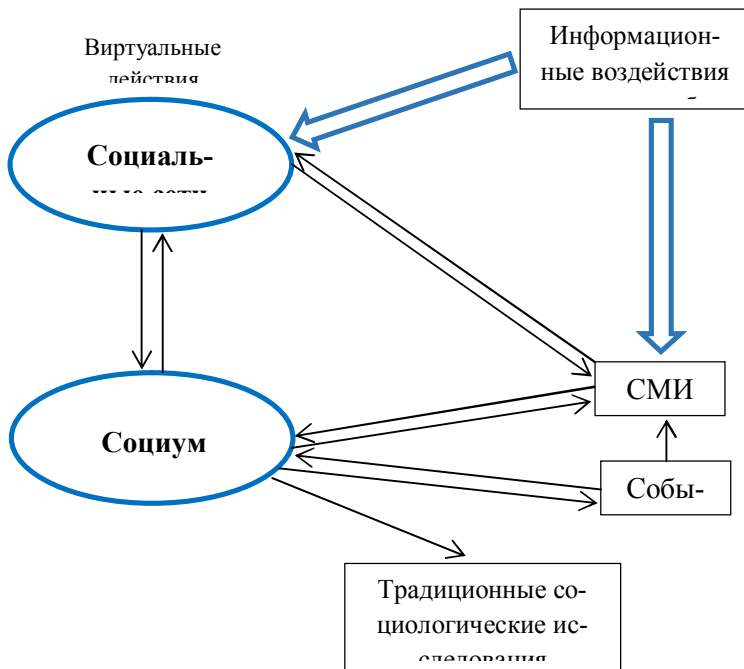


Рис. 1. Социальная сеть как открытая система

Внешнее воздействие (освещение некоторого события в новостях и др.) может повлиять на пользователя социальной сети, в результате чего пользователь может (пользуясь своим «отражением» в сети – агентом) публиковать *информационные сообщения* (посты, комментарии, записи и т.п.), которые содержат соответствующие *информационные объекты* (сущности, события, персоналии и т.п.). Публикация информационных сообщений приводит к постепенному формированию мнений и убеждений относительно информационных объектов (при этом пользователи формируют и меняют свои мнения по тем или иным вопросам под *влиянием* других членов сети, обладающих различной значимостью (авторитетностью)). Мнения и убеждение могут вызвать действия (например, резкое повышение спроса на

некий товар, стихийную массовую акцию протеста и пр.), которые, в свою очередь, могут получить отображение в СМИ.

Внешними воздействиями могут быть хакерские атаки, рекламные кампании, создание бот-сетей, воздействия на поведение пользователей социальных сетей и т.п. Внешнее воздействие на сеть можно выявить (с большой долей уверенности) по некоторым сетевым критериям «странности»/«неестественности». Например, можно предположить, что в некоторый момент на сеть оказывается внешнее воздействие посредством пользователя в каждом из таких случаев, когда пользователь: а) демонстрирует неестественную активность; б) резко меняет взгляды; в) действует согласованно с другими пользователями. Если при этом в данный момент наблюдается связанное с активностью пользователей внешнее событие и определяется круг заинтересованных внешних субъектов, то уверенность во внешнем воздействии возрастает.

Возникают две проблемы: как отличить внешнее воздействие от внутреннего (например, пользователь может поменять взгляды под влиянием своих внутренних причин) и как отличить целенаправленное воздействие (управление) от нецеленаправленного. Приведем некоторые соображения: (1) формальный анализ позволит выявить сам факт воздействия (для некоторых задач анализа этого будет достаточно); (2) формальный анализ, проводимый совместно с содержательным анализом экспертов в соответствующей предметной области (например, политологический анализ), позволит определить, является ли воздействие внешним и целенаправленным.

4. Основные подходы к моделированию социальных сетей

В данном пункте кратко описаны три взаимосвязанных подхода к моделированию социальных сетей: моделирование структуры сети, моделирование распространения информации, моделирование формирования мнений. Отметим, что если конечной целью управляющего субъекта является управление мнениями в социальной сети, то для ее достижения необходимо использовать все три подхода, поскольку структура сети во мно-

гом определяет распространение информации, а оно, в свою очередь, существенно влияет на формирование мнений.

Моделирование структуры социальной сети. Математической абстракцией социальной сети является граф. В социальных сетях вершинами графа являются социальные агенты (например, люди или организации), а ребрами – социальные взаимоотношения (например, отношения дружбы, соавторства). Реальные социальные сети являются большими и сложными (насчитывают десятки и сотни миллионов вершин), иными словами для реальных сетей нелегко ответить на вопросы относительно их структуры и, в конечном итоге, определить модели структуры сети. Поэтому необходимо определить множество ключевых признаков, характеризующих структуру больших и сложных сетей.

Исследователями уже определены десятки признаков/показателей, характеризующих структуру социальных сетей [12]: диаметр и среднее расстояние (малый диаметр социальной сети обуславливает короткую цепочку распространения воздействия по сети), распределение степеней (степенное/безмасштабное распределение в социальной сети обуславливает наличие доминирующих узлов (хабов), связанных с сотнями, тысячами и даже миллионами других узлов, в большинстве своем имеющих всего по несколько связей), показатели связанности и другие показатели.

На основе определенного множества ключевых структурных признаков и характерных их значений можно разработать модели социальных сетей, обладающих структурным сходством с реальными сетями. Долгое время все сложные сети рассматривались как полностью случайные [7], и сети моделировались соединением случайных вершин случайным образом. Однако в 1969 году Милгрэм (Milgram) обнаружил, что любых двух людей «разделяет всего шесть шагов». Появилось множество статических моделей так называемого «малого мира» [11]. Однако, несмотря на то, что статические модели дают возможность учесть сложность сети, они не объясняют причину появления безмасштабных сетей, их природу. Наиболее распространенной динамической моделью появления таких структур является модель «богатый становится богаче» [3], в которой в растущем

графе вероятность появления новых связей у вершины пропорциональна ее степени, т.е. количеству уже имеющихся связей. Существуют и другие подходы к объяснению причин появления структуры социальной сети. Например, в теоретико-игровых моделях [5] стоимость добавления новых связей влияет на механизм преимущественного присоединения новых связей (т.е. для агентов предпочтительны те или иные конфигурации связей).

Моделирование распространения информации. Распространение информации (или, если смотреть шире, активности) – это процесс, посредством которого некоторый информационный объект (информация, информационный вирус, мнение) распространяется по коммуникационным каналам во времени и в пространстве среди узлов сети [12], образуя путь воздействия в сети.

В социальной сети пользователи создают информационные сообщения (посты, комментарии, записи), которые содержат устойчивые информационные объекты (пример информационного объекта – «взрыв в минском метро»).

Информационные объекты могут распространяться по социальной сети от пользователя к пользователю, активируя их и образуя траекторию распространения информационного объекта. Пользователь становится активным по отношению к данному информационному объекту с момента использования информационного объекта в своих информационных сообщениях (совокупность состояний активности всех пользователей в сети в определенный момент времени назовем состоянием сети).

Как уже было сказано выше, на активность пользователей может повлиять внешнее воздействие, которое оказывается определенным внешним субъектом, например, средством массовой информации, освещающим события в новостях, или хакером, осуществляющим дефейс значимого веб-ресурса. В конечном итоге некоторый пользователь становится источником воздействий, используя соответствующие информационные объекты в информационных сообщениях.

Определив основные элементы информационных процессов в социальных сетях, можно разработать модели динамики активности (распространения информации) в социальных сетях. Эти модели делятся на две категории: макроуровня и микро-

уровня. Модели макроуровня (например, [6]) рассматривают сеть в целом, без учета связей между узлами. Модели микроуровня, учитывающие эти связи, в свою очередь разбиваются на следующие категории моделей: модель с порогами [10]; модель независимых каскадов (в которой каждый узел получает на некотором шаге шанс активировать другие узлы) [9]; модели пространства, базирующиеся на аналогиях с физикой, медициной и другими разделами науки [4] и т.д.

Моделирование формирования мнений. Обмен информацией пользователями сети приводит к формированию мнений (убеждений). Мнение – это суждение по какому-то поводу, точка зрения на какой-то информационный объект, оценка информационного объекта. В общем случае оно выражается в тексте сообщения или в виде численной оценки (пример – «по моему мнению, курс доллара через месяц составит 30,5 рублей»). Мнения пользователей по тем или иным вопросам могут меняться под влиянием других членов сети, обладающих различной влиятельностью. Вопросам формирования мнения, а также моделям информационного влияния, управления и противоборства посвящена монография [1].

5. Задачи анализа, прогнозирования и управления в социальных сетях

Лица, принимающие решения, заинтересованы в информационно-аналитической поддержке пассивной и активной работы с социальными сетями. В связи с этим основными задачами являются (см. рис. 2): мониторинг и анализ сети; прогнозирование состояния сети; информационное управление.

Перечислим некоторые задачи мониторинга и анализа социальных сетей: мониторинг развития существующих и появления новых актуальных обсуждаемых информационных объектов, определение характера обсуждения (критика, одобрение, призывы к действиям и т.п.), оценка естественности обсуждений, прогнозирование роста популярности информационного объекта, выявление активных групп и сообществ, социодемографический анализ сообществ, определение инициаторов

обсуждений, оценка влиятельности пользователей, деанонимизация пользователей.

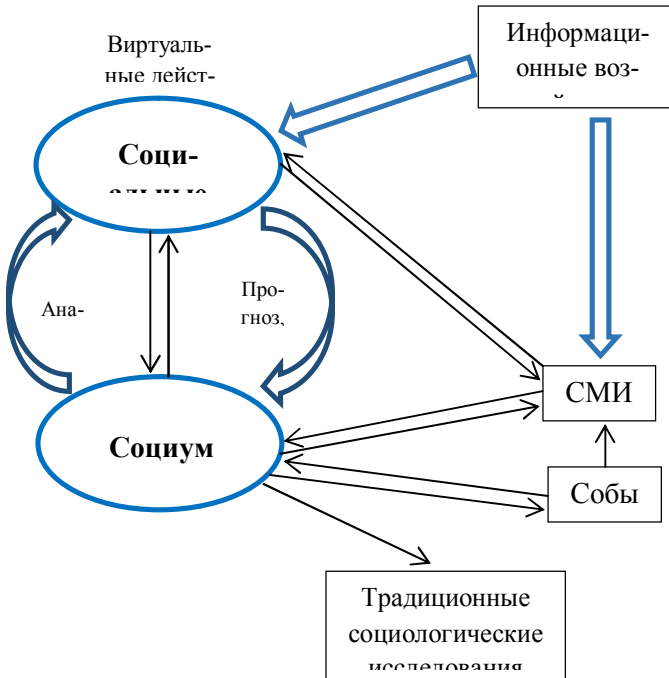


Рис. 2. Социальная сеть как объект анализа, прогнозирования и управления

Перечислим некоторые задачи прогнозирования и управления: прогнозирование развития информационных процессов, формирование общественного мнения, поддержка и повышение рейтинга виртуальных пользователей в заданных социальных группах.

6. Методы и алгоритмы анализа активности в социальной сети

В зависимости от поставленной задачи выбираются типы объектов (внешний субъект – внешний источник – сеть – агент – информационное сообщение – информационный объект – мнение).

ние). Для каждого выбранного объекта составляется набор характеристик, отражающих те или иные его свойства, важные с точки зрения моделирования. Например, агенту (узлу сети) соответствует его важность с точки зрения распространения информации, сети – активность сети по отношению к информационному объекту и т.п. Показатели являются детализацией и (или) конкретизацией характеристик объекта. Например, активность сети по отношению к информационному объекту конкретизируется в виде таких показателей, как количество описывающих информационный объект информационных сообщений, суммарное количество упоминающих информационных объектов блогеров пр.

После того как определяются объекты и их характеристики, выбирается необходимый уровень анализа данных. Выделим три уровня анализа социальных сетей.

- Расчет базовых показателей, позволяющих ответить на вопросы «Как часто пишут пользователи социальных сетей?», «Сколько друзей у пользователя социальной сети?» и т.п.

- Выявление закономерностей и паттернов, позволяющих ответить на вопросы «В какой тональности пишут пользователи?», «Насколько естественно протекает информационный процесс?» и т.п.

- Нахождение конкретных ответов на специфические вопросы «Как распространяется та или иная информация?», «Как формируются то или иное мнения в сети?» и т.п.

В зависимости от поставленных вопросов необходимо осуществить анализ на том или ином уровне. Для этого используются следующие классы методов:

- методы многомерного статистического анализа (рассматривающие на макроуровне сеть и информационные процессы, протекающие в ней);

- методы анализа сетей (учитывающие структуру связей между узлами сети и, соответственно, протекающие в рамках этих связей информационные процессы).

Следует отметить, что в некоторых случаях полезно совместно использовать как методы многомерного статистического анализа, так и методы анализа сетей.

7. Технологии мониторинга и анализа социальных сетей

Для информационно-аналитической поддержки пассивной и активной работы с онлайн-социальными сетями необходима разработка автоматизированной системы мониторинга и анализа онлайн-социальных сетей.

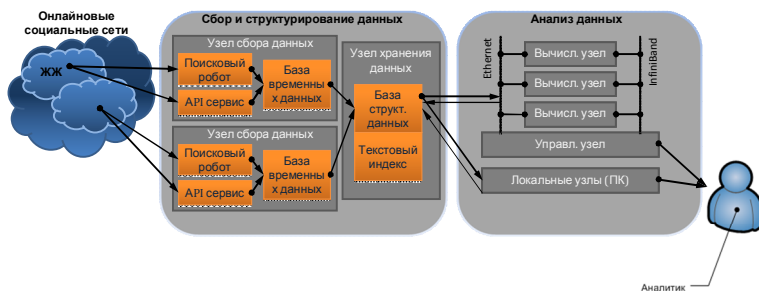


Рис. 3. Архитектура системы

Система состоит из двух больших аппаратно-программных комплексов (см. рис. 3).

Аппаратно-программный комплекс сбора и структурирования данных отвечает за сбор, структурирование и хранение первичных разнородных данных, обеспечивает эффективный доступ к ним. Комплекс состоит из узлов двух типов: узла сбора данных и узла хранения данных. Узел сбора данных осуществляет сбор данных (как при помощи поискового робота, так и при помощи программного интерфейса онлайн-социальной сети), а также их первичную структуризацию и временное хранение. Далее данные перемещаются в единое хранилище – узел хранения данных. Узел хранения данных производит их вторичную структуризацию, осуществляет резервное копирование и организует доступ к данным со стороны аналитического комплекса.

Аппаратно-программный комплекс анализа данных отвечает за анализ процессов, протекающих в социальных сетях. Комплекс ориентирован на высокопроизводительные вычисления с большими объемами входных данных. Поэтому комплекс может быть ре-

ализован на базе вычислительного кластера, узлы которого работают в режиме общей памяти. Обмен данными между узлами может осуществляться в режиме распределённой памяти.

К настоящему моменту разработан базовый вариант комплекса, осуществляющего сбор данных из русскоязычного сегмента блоговой площадки LiveJournal (<http://www.livejournal.com>). Опишем кратко системы, разработанные в рамках этого комплекса.

Система сбора данных. Сбор проводится по блоговой площадке LiveJournal, в которой пользователи публикуют посты, а также комментируют посты и комментарии других пользователей. Сбор постов осуществляется в реальном времени, сбор комментариев осуществляется регулярно (по расписанию). При этом первичные параметры постов и комментариев сохраняются в промежуточной базе метаданных, а их тексты – в файловой системе.

Система структурирования данных. По мере накопления необработанных данных на сервере их нужно актуализировать, т.е. надлежащим образом преобразовать, затем поместить в индекс и в конечную нормализованную базу данных. Такая процедура производится (при помощи специально разработанного программного обеспечения) в два этапа: актуализация данных в индексе подразумевает под собой создание текстовых индексов; актуализация постов в базе данных заключается в том, чтобы заполнить конечную базу данных, используя промежуточную базу метаданных и индекс. Индексы нужны для осуществления поиска по текстам постов и комментариев, а конечная база данных – при дальнейшей работе, для анализа онлайн-социальной сети.

Система доступа к данным. Подсистема принимает пользовательские запросы на специальном языке запросов (в частности, допускаются wildcard-символы, булевы операторы И/ИЛИ/НЕ и скобки произвольной вложенности), далее подсистема разбирает запрос и просматривает все индексы на наличие релевантных документов. В конечном итоге результаты (идентификаторы найденных документов и некоторая дополнительная информация) сохраняются в файл в определенном формате, пригодном для дальнейшего анализа.

8. Заключение

В настоящей работе определен конструктивный подход к анализу социальных сетей, рассмотрена концептуальная модель социальной сети, рассмотрены базовые задачи анализа и управления в социальных сетях, рассмотрены методы и алгоритмы анализа активности в социальных сетях, а также технологии мониторинга и анализа социальных сетей.

Перспективным направлением дальнейших теоретических и прикладных исследований является планомерное развитие и углубление методов анализа на основе сформулированной модели, а также применение этих методов для исследования реальных онлайн-социальных сетей.

Литература

1. ГУБАНОВ Д.А., НОВИКОВ Д.А., ЧХАРТИШВИЛИ А.Г. *Социальные сети: модели информационного влияния, управления и противоборства*. – М.: Изд-во физико-математической литературы, 2010. – 228 с.
2. AGGARWAL C.C. *Social Network Data Analytics*. – Kluwer Academic Publishers, 2011. – 502 p.
3. ALBERT R., BARABÁSI A.-L. *Statistical mechanics of complex networks* // *Reviews of Modern Physics*. – 2002. – Vol. 74. – P. 47–97
4. BAILEY N. *The Mathematical Theory of Infectious Diseases and Its Applications*. – New York: Hafner Press, 1975. – 413 p.
5. BALA V., GOYAL S.A. *Noncooperative Model of Network Formation* // *Econometrica*, *Econometric Society* – 2000. – Vol. 68(5). – P. 1181–1230.
6. BASS F. *A new product growth model for consumer durables* // *Management Science* – 1969. – № 15 (5). – P. 215–227.
7. ERDOS P., RENYI A. *On random graphs* // *Publicationes Mathematicae* – 1959. – №6 – P. 290–297.
8. EVERTON S.F. *Disrupting Dark Networks (Structural Analysis in the Social Sciences)*. – Cambridge University Press, 2012. – 482 p.

9. GOLDENBERG J., LIBAI B., MULLER E. *Talk of the Network: A Complex Systems Look at the Underlying Process of Word-of-Mouth* // Marketing Letters. – 2001. – №2. – P. 11–34.
10. GRANOVETTER M. *Threshold Models of Collective Behavior* // American Journal of Sociology. — 1978. – Vol. 83, №6. – P. 1420–1443.
11. NEWMAN M.E.J. *The structure and function of complex networks* // SIAM Review. – 2003. – P. 167–256. [Электронный ресурс]. – URL: <http://arxiv.org/pdf/condmat/0303516.pdf> (дата обращения 28.08.2013).
12. ROGERS E.M. *Diffusion of Innovations*. – New York, London: Free Press, 1983. – 453 p.

CONCEPTUAL APPROACH TO ONLINE SOCIAL NETWORKS ANALYSIS

Alexander Chkhartishvili, Institute of Control Sciences of RAS, Moscow, Digital Society Laboratory (DSL), Moscow, Doctor of science in mathematics (sandro_ch@mail.ru).

Dmitry Gubanov, Institute of Control Sciences of RAS, Moscow, Digital Society Laboratory (DSL), Moscow, Ph.D. (dmitry.a.g@gmail.com).

Abstract: We introduce a constructive approach to online social networks analysis. We suggest a conceptual model of a social network, describe basic problems of analysis and control in social networks, describe methods and algorithms for activity analysis in social networks, and also technologies for social networks monitoring and analysis.

Keywords: social network, conceptual model, informational control, information technologies.

*Статья представлена к публикации
членом редакционной коллегии М.В. Губко*

*Поступила в редакцию 18.12.2012.
Опубликована 30.09.2013.*

УДК 338.2
ББК 65.05.3

УПРАВЛЕНИЕ ТЕХНОЛОГИЧЕСКИМ ПОРТФЕЛЕМ ЭНЕРГЕТИЧЕСКОЙ КОМПАНИИ: СЦЕНАРНЫЙ ПОДХОД¹

Ратнер С. В.²,
(ФГБУН Институт проблем управления
им. В.А. Трапезникова РАН, Москва)

В настоящей работе на основе имитационного моделирования динамики прибыли энергетической компании/энергетического кластера в зависимости от различных стратегий инвестирования в традиционные и альтернативные технологии генерации энергии, а также в зависимости от различных сценариев роста рентабельности альтернативных технологий исследуется вопрос об оптимальном времени перехода к широкому внедрению альтернативных технологий в производственную деятельность. Рассматриваются сценарии роста эффективности технологий альтернативной энергетики, соответствующие начальному и завершающему периоду технологического разрыва.

Ключевые слова: энергетика, технологический разрыв, технологический портфель, имитационная модель, возобновляемые источники, сценарии роста.

1. Введение

С начала 2000-х годов в мировой энергетике наметилась устойчивая тенденция к росту доли возобновляемых источников в общем объеме производства и потребления энергии [12]. Этот

¹ Работа выполнена при финансовой поддержке РФФИ, проект №13-06-00169 «Моделирование стратегий развития энергетических кластеров в ситуации технологического разрыва».

² Светлана Валерьевна Ратнер, доктор экономических наук (laparat@mail.ru, тел. (495) 334-79-00).

рост происходит в основном за счет расширения использования в технологически развитых и быстроразвивающихся экономиках таких возобновляемых источников, как ветер, солнце и биомасса [4]. В качестве основных причин для сохранения и укрепления указанных тенденций в странах ЕС и США традиционно выделяют борьбу с глобальными климатическими изменениями, улучшение общей экологической ситуации, создание новых рабочих мест в высокотехнологичных секторах экономики и достижение энергетической независимости от стран – крупнейших экспортеров углеводородов. Однако еще одной важной причиной развития альтернативной энергетики, о которой не так часто говорится в официальных документах, несомненно, является завоевание устойчивого технологического лидерства в условиях формирования нового технологического уклада. Именно эта причина побуждает такие мощные транснациональные энергетические концерны как Shell, Eni S.p.A., E.ON и другие активно инвестировать средства в разработку новых технологий возобновляемой энергетики и диверсифицировать свои проектные портфели.

Одновременно с инвестициями в исследования и разработки западные энергетические концерны активно наращивают долю возобновляемых источников энергии в своем бизнес-портфолио. Так, Shell через совместные предприятия с другими компаниями уже владеет 10 ветропарками мощностью 550 МВт, из которых 8 расположены в США и 2 – в Европе¹. Итальянский концерн Eni S.p.A. с конца 2012 года реализует инновационный проект по производству биотоплива (Honeywell Green Diesel) на рафинировочном заводе в Венеции совместно с компанией UOP LLC (технология UOP/Eni Ecofining). Производство начнется в 2014 году (100 млн. галлонов ежегодно)². E.ON является одним из признанных лидеров ветровой энергетики. Например, в Великобритании он имеет 20 наземных и 3 офшорных (шельфовых) ветропарков общей мощностью 474,30 МВт и две крупные

¹ По данным, представленным на официальном сайте компании: <http://www.shell.com>.

² По данным, представленным на официальном сайте компании <http://www.uop.com>.

биогазовые станции мощностью 44 МВт. И хотя на настоящий момент доля возобновляемых источников энергии в портфеле компании составляет 10,29% от генерируемой электрической энергии и 3,07% от тепловой, она постоянно увеличивается, что отражено в стратегии развития компании.¹

Крупнейшие российские энергетические компании также начинают уделять внимание развитию технологий альтернативной энергетики, однако примеров успешной диверсификации проектных портфелей пока единицы [2]. Энергетическая стратегия России до 2020 года также не предусматривает кардинальных изменений в структуре энергобаланса страны, планируя увеличить долю возобновляемых источников энергии всего до 4–5%.

Частично это объясняется трудностью выбора между долгосрочными и среднесрочными стратегическими целями в условиях ограниченности ресурсов, а также высокой неопределенностью относительно будущего развития технологий нового уклада. Поэтому проблема прогнозирования последствий принятия решений относительно распределения инвестиционных ресурсов энергетических компаний в условиях технологического разрыва является актуальной как с теоретической, так и с практической точки зрения.

Настоящая работа является развитием подхода к моделированию инвестиционной стратегии энергетической компании, предложенного в [3] и основанного на методах системной динамики. Объектом исследования является энергетическая компания или региональный энергетический кластер, реализующие такие направления деятельности, как добыча и/или переработка и/или транспортировка нефти и/или газа, генерация электроэнергии. В России примером такой компании может быть ОАО «Лукойл», который, помимо хорошо известных направлений своей деятельности, в 2010 году создал совместное предприятие с итальянской энергетической компанией «ERG Renew», называемое «LUKERG Renew». «LUKERG Renew» в июне 2012 г. приобрел ветропарк «Черга» (Болгария) мощностью 40 МВт, а с

¹ По данным, представленным на официальном сайте компании <http://www.eonenergy.com>.

января 2013 года работает над строительством в Румынии ветропарка Inergia мощностью 84 МВт.

Основным моделируемым параметром является прибыль энергетического концерна, которая зависит от выбора стратегии инвестирования, коэффициента реинвестирования, характера и уровня снижения спроса на традиционные энергоносители, характера и уровня повышения рентабельности технологий возобновляемой энергетики. Коэффициент реинвестирования и доля инвестиционных ресурсов, направляемых на развитие инновационных технологий добычи и переработки углеводородов и технологий возобновляемой энергетики, задаются исследователем (в модели данная возможность будет предусмотрена на каждом шаге имитационного цикла).

В модели [3] рассматривается процесс параллельной реализации двух последовательностей нескольких различных проектов, в каждой из которых проекты относятся к одному технологическому классу и схожи по срокам окупаемости и средней норме рентабельности. Индивидуальные особенности проектов, принадлежащих одной и той же последовательности, никак не учитываются, а все процессы деятельности компании описываются в терминах накопителей, потоков и информации, которая определяет величину этих потоков.

Особенностью данного класса моделей является необходимость предварительного сбора, накопления и анализа больших объемов эмпирических данных о моделируемых процессах и объектах для построения уравнений перехода системы из одного состояния в другое. Прогнозирование динамики спроса на традиционные виды продукции энергетических компаний (нефть, природный газ, нефтепродукты) осуществляется в модели на основе сценария BLUE MAP Мирового энергетического агентства до 2050 года [9]. Коммерческая эффективность проектов по развитию инновационных технологий добычи, переработки и транспортировки углеводородов¹ оценена по пилотным результатам инновационной стратегии ОАО «Газпром»

¹ За исключением проектов по добыче сланцевого газа, битумной и конвенционной нефти, данные по рентабельности которых нуждаются в подтверждении.

за 2011–2013 гг. [5], которую можно принять за верхнюю оценку в силу отрицательной динамики рентабельности продукции добывающих компаний в целом по отрасли [7]. Коммерческая эффективность проектов по развитию альтернативной энергетики (в данном случае выбрана наиболее технологически зрелая отрасль возобновляемой энергетики – ветровая) оценена по динамике снижения бонусных тарифов на электроэнергию, получаемую ветропарками, заложенной в стратегии развития возобновляемой энергетики Германии до 2025 года (Renewable Energy Sources Act) [10].

Первый вариант имитационной модели был построен в предположении о линейном характере снижения цены на природный газ и нефтепродукты и динамике роста рентабельности технологий возобновляемой энергетики. Данный вариант был откалиброван на основании имеющихся эмпирических данных и показал свою работоспособность в ходе ряда численных экспериментов [3]. Однако линейные тренды не отражают реального характера роста эффективности новых технологий по мере приближения их к технологическому пределу и могут использоваться на практике лишь для получения достаточно грубых прогнозных оценок. Более точные прогнозы можно получить на основе использования логистической кривой, которая вполне адекватно отражает наиболее общие закономерности динамики поступательно-циклических процессов [1]. Рассматривая динамику формирования технологического уклада более детально, можно выделить пологие участки логистической кривой – начало и конец жизненного цикла базовых технологических нововведений, которые характеризуются относительно низкой отдачей вложений в их развитие, и участок крутого подъема кривой, на котором небольшие затраты начинают приносить значительный коммерческий эффект. В связи с этим прогнозные значения прибыли энергетической компании будут достаточно сильно различаться в зависимости от того, на каком этапе жизненного цикла находится кластер базовых технологий: на начальном (до точки перегиба логистической кривой, который может быть представлен экспоненциальным трендом) или на завершающем (после точки перегиба, моделируемый логарифмической кривой).

На основе результатов имитационного моделирования разработаны сценарии развития российских энергетических компаний (или региональных энергетических кластеров, включающих как добывающие и/или перерабатывающие компании, так и генерирующие электроэнергию) при использовании ими различных инвестиционных стратегий на различных стадиях жизненного цикла кластера базовых технологий нового технологического уклада; определены зависимости между сценариями роста рентабельности технологий альтернативной энергетики и интенсивностью инвестирования; исследованы вопросы об оптимальном времени расширения проектного портфеля за счет включения в него технологий альтернативной (ветровой) энергетики.

2. Имитационная модель развития энергетической компании и ее программная реализация: возможности и ограничения

Имитационное моделирование является практически единственным возможным способом построения прогноза развития сложной системы в том случае, когда известны (точно или приблизительно) лишь закономерности перехода ее из одного состояния в другое. В таком случае, задавая набор экзогенных для изучаемой системы параметров на «входе» модели и представляя последовательность изменения эндогенных параметров в виде определенного алгоритма действий по преобразованию эндогенных и экзогенных параметров на каждом шаге имитационного цикла, можно рассчитать значения интересующих исследователя параметров системы на «выходе». Множественность вычислений, проводимых при имитации поведения сложной системы, обуславливает необходимость использования компьютерных технологий и программирования.

Вопросы моделирования динамики развития энергетических рынков в зависимости от поведения основных игроков и мер, предпринимаемых национальными и международными регуляторами, в настоящее время привлекают внимание как отечественных, так и зарубежных исследователей. При этом в большинстве исследований основное внимание уделяется влия-

нию на решение инвестора относительно реализации того или иного проекта различного рода неопределенностей. В отличие от работ [6, 8, 11], в которых исследуется вопрос оптимального инвестиционного выбора в условиях, когда экономические параметры всех возможных технологий генерации энергии известны, а цены на различные источники энергии могут изменяться случайным образом в связи с неопределенностью экономической политики государства или международных регулирующих структур, в настоящей работе ставится в некотором смысле обратная задача: динамика цен предполагается известной (заданной определенными сценариями развития мировой экономики), в то время как технико-экономические параметры доступных технологий колеблются в пределах определенного интервала. Кроме того, в отличие от вышеперечисленных работ, в которых ресурсы различных инвесторов предполагаются неограниченными в связи со множественностью потенциальных участников рынка, в данной работе объем инвестиционного ресурса, которым располагает компания/кластер в каждый момент времени зависит от объема инвестиционного ресурса в предыдущие моменты времени и того, насколько успешным оказался инвестиционный выбор компании на всех предыдущих шагах имитационного цикла. Поэтому традиционный в случае моделирования различных видов неопределенностей метод Монте-Карло, используемый в [6, 8, 11], для решения данной задачи неприменим.

Состояние системы (энергетической компании или регионального энергетического кластера) предлагается описывать с помощью трех параметров: 1) объема прибыли, полученной за счет продажи традиционной углеводородной продукции и продукции (электроэнергии), произведенной на основе технологий ветровой энергетики; 2) доли прибыли, направляемой на развитие инновационных технологий всех групп (коэффициент интенсивности инвестирования); 3) доли инвестиционных ресурсов, направляемой на развитие технологий каждой группы.

Используем следующие уравнения перехода [3]:

$$(1) \quad \Pi(i) = (P_g(i) - A) \cdot V_g + (P_o(i) - B) \cdot V_o + \text{ПП}(i) + \Pi_ITT(i),$$

$$(2) \quad I(i) = k \cdot \Pi(i-1) = ITT(i) + IAT(i),$$

$$(3) \quad ITT(i) = k_1 \cdot I(i), \quad LAT(i) = k_2 \cdot I(i),$$

$$(4) \quad \Pi_{-ITT}(i) = RIT(i) = 2,4 \cdot ITT(i - 3),$$

$$(5) \quad \Pi(i) = (0,2 + 0,073 \cdot (i - 1)) \cdot IAT(i),$$

где

$P_g(i), P_o(i)$ - цена на природный газ и нефтепродукты соответственно на i -м шаге имитационного цикла (имитационный цикл – весь период работы модели, определяемый исследователем; один шаг приблизительно равен одному году);

A, B – их текущая себестоимость;

V_g, V_o - объем производства газа и нефтепродуктов соответственно, для простоты принятые константами, соответствующими текущему уровню объемов производства той или иной компании;

$\Pi\Pi(i)$ - потенциальная прибыль от внедрения технологий нового технологического уклада на i -м шаге имитационного цикла;

$\Pi\Pi_{-ITT} = \Delta A \cdot V_g + \Delta B \cdot V_o$ – прибыль от инвестиций в разработку инновационных решений по технологиям традиционной энергетики;

$I(i)$ – инвестиции компании на i -м шаге имитационного цикла;

k – коэффициент интенсивности реинвестирования (доля прибыли, направляемой на развитие инновационных технологий всех рассматриваемых типов);

k_1 – доля инвестиционных ресурсов, направляемая на реализацию проектов на основе инновационных технологий углеводородной энергетики;

k_2 – доля инвестиционных ресурсов, направляемая на реализацию проектов на основе инновационных технологий возобновляемой (ветровой) энергетики;

$ITT(i)$ – инвестиции компании в разработку и внедрение технологий традиционной энергетики на i -м шаге имитационного цикла;

$LAT(i)$ – инвестиции компании в разработку и внедрение технологий возобновляемой (ветровой) энергетики на i -м шаге имитационного цикла.

Формула (5) задает динамику роста рентабельности технологий возобновляемой энергетики и в представленном виде соответствует линейному тренду. Однако, как уже было отмечено, линейный тренд не всегда адекватно отражает характер изменения эффективности (в частности, рентабельности) технологий. На начальной стадии жизненного цикла более точные результаты дает экспоненциальный тренд, тогда как на завершающей – логарифмический. Построение экспоненциального и логарифмического трендов, удовлетворяющих верхним и нижним ограничениям рентабельности, выведенным на основе анализа сценария развития возобновляемой энергетики Германии, выполнено на основе решения систем алгебраических уравнений.

В результате уравнение перехода (5) в случае, если настоящий момент можно считать начальным этапом жизненного цикла технологий возобновляемой (ветровой) энергетики, принимает следующий вид:

$$(5') \text{ПП}(i) = 0,187133 \exp(0,0665 \cdot i) \cdot \text{IAT}(i).$$

В том же случае, если настоящий начальный шаг имитационного цикла приходится на завершающую стадию жизненного цикла технологий возобновляемой энергетики (их эффективность приближается в технологическому пределу), то уравнение (5) должно быть заменено на следующее:

$$(5'') \text{ПП}(i) = (0,815131 \ln i + 0,2) \cdot \text{IAT}(i).$$

Формулы (1)-(5) полностью задают алгоритм преобразований эндогенных и экзогенных параметров модели при переходе от одного ее состояния к другому. Программная реализация модели выполнена на языке FPK Object Pascal. Программный комплекс состоит из программы-установщика и основной программы.

Основная программа производит расчет прогнозных оценок прибыли концерна в динамике при использовании инвестиционных стратегий различного типа - традиционной, альтернативной или комбинированной, которые определяются по отношению инвестиционных ресурсов, направляемых на развитие традиционных технологий добычи и переработки углеводородов и альтернативных энергетических технологий (95/5,

80/20, 50/50). Пользователь задает значение коэффициента реинвестирования и выбирает сценарий развития отрасли: линейный, экспоненциальный или логарифмический рост рентабельности технологий альтернативной энергетики (рис. 1).

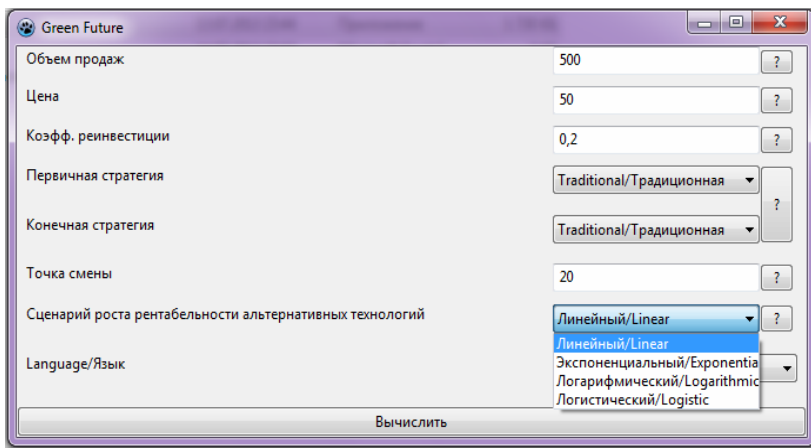


Рис. 1. Фрагмент работы программного модуля

Кроме того, программа позволяет осуществить переход от изначально избранной стратегии инвестирования к некоторой другой на определенном (также задаваемом пользователем) шаге имитационного цикла.

Доля нераспределенной прибыли, инвестируемой в разработку технологий всех типов (коэффициент интенсивности реинвестирования k , $0 < k \leq 1$), может варьироваться на «входе модели», но сохраняется постоянной на протяжении всего периода имитации. Теоретически компания может инвестировать всю прибыль в развитие, однако, исходя из анализа эмпирических данных, представленных в отчетности крупных нефтегазовых компаний, значение коэффициента интенсивности реинвестирования больше 0,5 маловероятно.

Следует отметить, что модель позволяет определить качественные закономерности развития компаний и оценить относительные показатели роста в динамике, но не их абсолютные величины. Результаты, полученные на «выходе», нуждаются в

осторожной интерпретации: полученные количественные оценки имеют смысл лишь в сравнении с их первоначальными значениями (на первом шаге имитационного цикла).

3. Анализ чувствительности модели к сценариям роста рентабельности альтернативных технологий

Результаты имитационного моделирования полностью подтвердили авторскую гипотезу о зависимости как динамики (рис. 2-3), так и величины общей за период моделирования прибыли энергетической компании от сценария роста эффективности технологий возобновляемой энергетики даже *при одинаковых* нижнем и верхнем порогах рентабельности (определенным приблизительно по имеющимся эмпирическим данным) (рис. 4).

Как и при линейном сценарии роста рентабельности альтернативных технологий, резкий скачок показателей прибыли компании на интервале между 3 и 5-м шагом имитационного цикла с момента осуществления первоначальных инвестиций связан с достижением периода окупаемости инновационных решений, принадлежащих группе технологий традиционной энергетики. Подобные, но менее резкие скачки, также наблюдаются между 8–9-м и 12–13-м шагами с момента осуществления первоначальных инвестиций. В последующий период происходит медленное снижение показателей прибыли при традиционной и комбинированной стратегиях инвестирования за счет снижения спроса на ископаемые источники энергии и цены на продукцию компании.

Сравнивая общую прибыль компании за весь имитационный цикл, можно сделать вывод о том, что экспоненциальный рост, соответствующий начальной стадии жизненного цикла кластера технологий, является наименее предпочтительным с точки зрения максимизации прибыли компании за весь период моделирования, а логарифмический – наиболее предпочтительным. При этом разница между значениями общей прибыли, полученной по различным сценариям роста, тем существенней,

чем больше коэффициент реинвестирования и чем большая доля инвестиционных ресурсов направляется на проекты по развитию технологий возобновляемой энергетики (рис. 4).

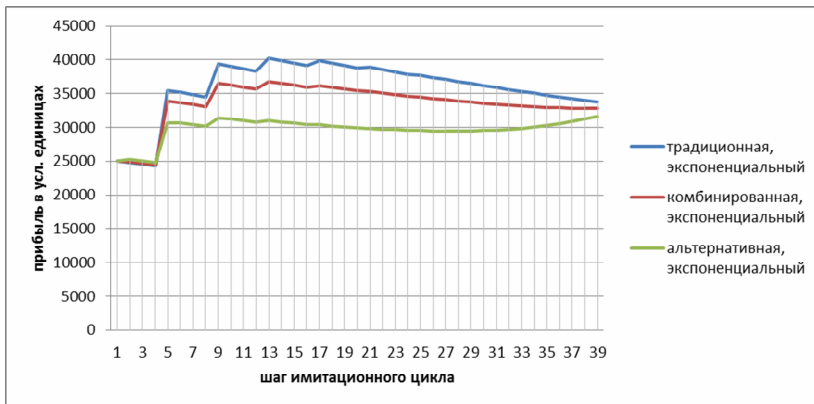


Рис. 2. Динамика общей прибыли при экспоненциальном сценарии роста рентабельности альтернативных технологий (коэффициент $k = 0,2$)

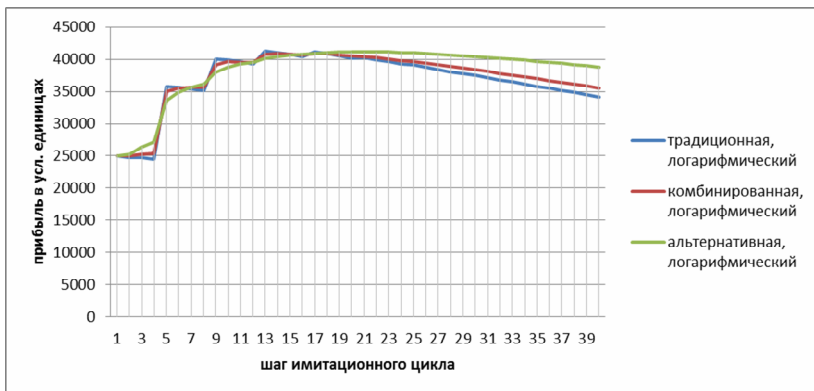


Рис. 3. Динамика общей прибыли при логарифмическом сценарии роста рентабельности альтернативных технологий ($k = 0,2$)

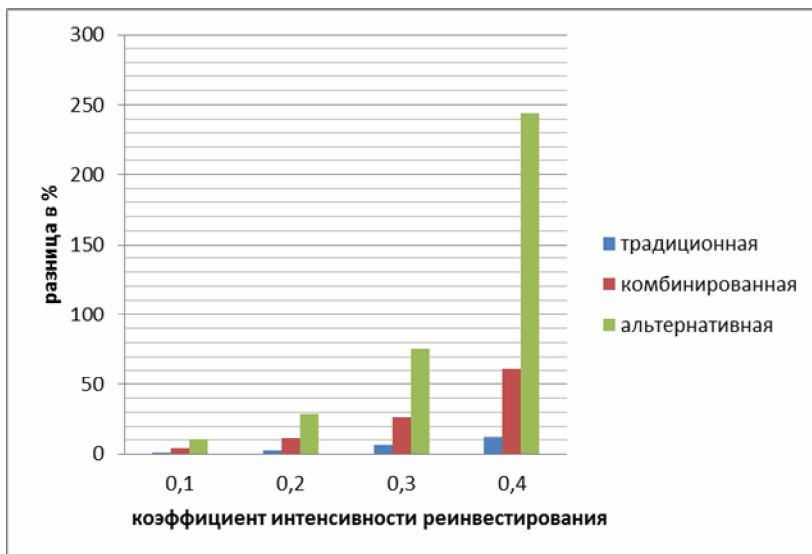


Рис. 4. Разница (в процентах) между общей прибылью компании при экспоненциальном и логарифмическом сценариях роста рентабельности технологий возобновляемой энергетики

Такой результат моделирования объясняется тем, что при традиционной стратегии инвестирования, когда основная доля инвестиционных ресурсов направляется на реализацию проектов по внедрению инновационных технологий углеводородной энергетики, прибыль компании в большей степени зависит от рентабельности данных технологий (которая принята в модели постоянной на всем протяжении имитационного цикла) и спроса на углеводороды (который линейным образом снижается в соответствии с прогнозами сценария BLUE MAP). По мере того как вклад альтернативных технологий в общую прибыль компании возрастает (при использовании комбинированной и альтернативной стратегий инвестирования), возрастает и чувствительность основного моделируемого параметра (прибыли компании) к характеру роста рентабельности этих технологий.

Аналогичным образом объясняется чувствительность прибыли к значению коэффициента интенсивности

реинвестирования. При низких значениях коэффициента интенсивности реинвестирования основная доля прибыли компании формируется за счет продажи углеводородной продукции, а, следовательно, в большей степени зависит от спроса на нее, нежели от других параметров. При увеличении коэффициента интенсивности реинвестирования все более существенная доля прибыли формируется за счет внедрения инновационных технологий, как углеводородной, так и возобновимой энергетики), рентабельность которых становится все более значимым фактором.

На рис. 3 показана динамика прибыли компании при различных сценариях роста рентабельности технологий возобновляемой энергетики в случае использования альтернативной стратегии инвестирования на всем протяжении имитационного цикла и при значении коэффициента интенсивности инвестирования 0,3. Нетрудно заметить, что потеря прибыли компании при экспоненциальном сценарии относительно линейного и логарифмического сценариев происходит в основном за счет первой фазы имитационного цикла (шаги 1-17).

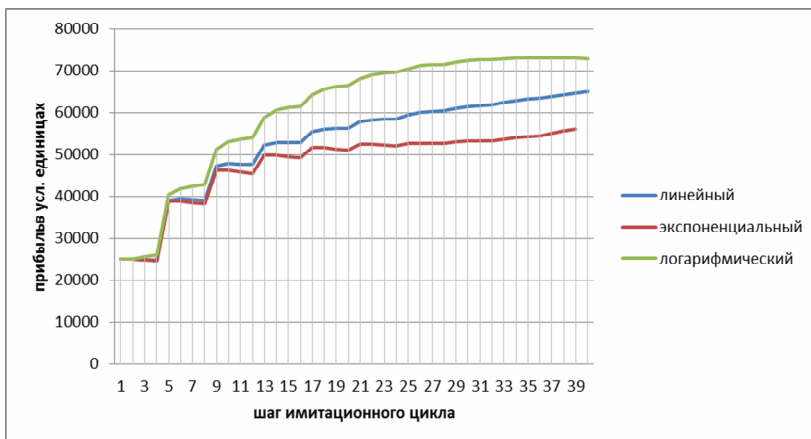


Рис. 5. Динамика прибыли энергетической компании при различных сценариях роста рентабельности технологий возобновляемой энергетики (альтернативная стратегия, $k = 0,3$)

Это свидетельствует о том, что по мере приближения кластера новых технологий к его предполагаемому технологическому пределу освоение данных технологий «с нуля» обходится компаниям все дороже.

Следует заметить, что данный результат свидетельствует о правильности калибровки модели и, следовательно, о корректности ее работы, а также о том, что ее можно использовать для построения прогнозов развития компаний.

4. Выбор стратегии инвестирования в зависимости от сценария роста рентабельности альтернативных технологий

Серия численных экспериментов, описанная в работе работе [3] и проведенная в предположении о линейном характере роста рентабельности альтернативных технологий, показала, что несмотря на явную предпочтительность динамики изменений прибыли при альтернативной стратегии инвестирования (прибыль компании не снижается по сравнению с первоначальными значениями), суммарная прибыль за исследуемый период остается наибольшей при использовании традиционной стратегии инвестирования. С увеличением коэффициента интенсивности реинвестирования сначала разница между суммарной прибылью компании при традиционной стратегии инвестирования и суммарной прибылью при альтернативной стратегии инвестирования увеличивается, а потом, наоборот, сокращается. Так, разница между значениями суммарной за период прибыли традиционной и комбинированной стратегиями при коэффициенте интенсивности реинвестирования $k = 0,3$ составляет 5,68%, а между традиционной и альтернативной – более 14%. При $k = 0,4$ традиционная стратегия превосходит комбинированную по суммарному значению прибыли за период на 7,8%, а альтернативную – почти на 17,5%.

При коэффициенте интенсивности реинвестирования $k = 0,5$ разница между значениями суммарной за период прибыли традиционной и комбинированной стратегией составляет 5,9%, а между традиционной и альтернативной – 7,5%. И только при

увеличении коэффициента интенсивности реинвестиций до 0,6 суммарное значение прибыли за период при традиционной стратегии инвестирования уступает аналогичному показателю при комбинированной стратегии на 2,5%, а при альтернативной стратегии – на 35% (рис. 6).

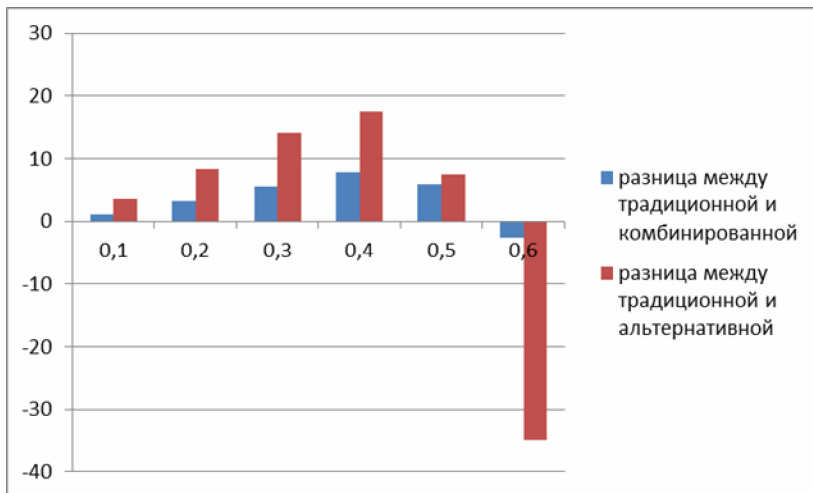


Рис. 6. Разница в значениях суммарной за период прибыли между традиционной, комбинированной и альтернативной стратегиями

Изменяя характер роста рентабельности альтернативных технологий, получим результаты, достаточно сильно отличающиеся от предыдущих. Так, даже при низких значениях коэффициента интенсивности реинвестирования при логарифмическом сценарии роста рентабельности альтернативных технологий, комбинированная и альтернативная стратегии инвестирования по показателю суммарной прибыли за весь имитационный цикл становятся более предпочтительными, нежели традиционная (рис. 7). Однако при экспоненциальном сценарии потеря в общей прибыли компании за весь имитационный цикл при использовании данных стратегий будет также более существенной.

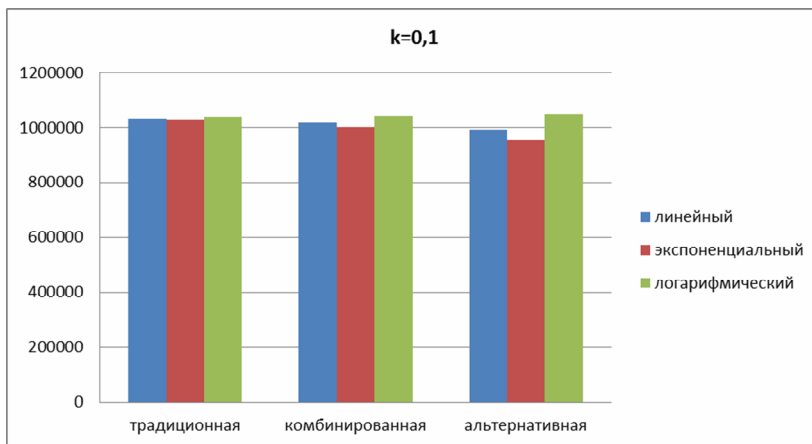


Рис. 7. Суммарная прибыль компании при использовании различных стратегий инвестирования в условиях реализации исследуемых сценариев роста рентабельности альтернативных технологий ($k = 0,1$)

Таким образом можно утверждать, что на стадии жизненного цикла кластера альтернативных технологий, описываемой логарифмическим сценарием роста эффективности, распределение инвестиционных ресурсов между технологиями углеводородной энергетики и технологиями альтернативной энергетики в отношении 50/50 является заведомо более предпочтительным, нежели в отношении 95/5 или 80/20 даже при относительно низких объемах инвестиций.

С увеличением значения коэффициента интенсивности реинвестирования, как было показано в предыдущем параграфе, чувствительность основного моделируемого параметра к характеру роста рентабельности альтернативных технологий возрастает. На рис. 8 представлены значения общей прибыли компании за весь имитационный цикл в условиях реализации линейного, экспоненциального и логарифмического сценариев роста эффективности технологий возобновляемой энергетики при использовании различных стратегий инвестирования в

случае, когда коэффициент интенсивности реинвестирования равен 0,4.

В силу того, что логарифмический и экспоненциальный характер роста эффективности альтернативных технологий гораздо более вероятны на практике, нежели линейный, энергетическим компаниям следует иметь в виду тот факт, что чем больший объем ресурсов компания направляет на развитие инновационных технологий, тем более существенны ее потери (упущенная выгода) в случае неверного выбора стратегии инвестирования.

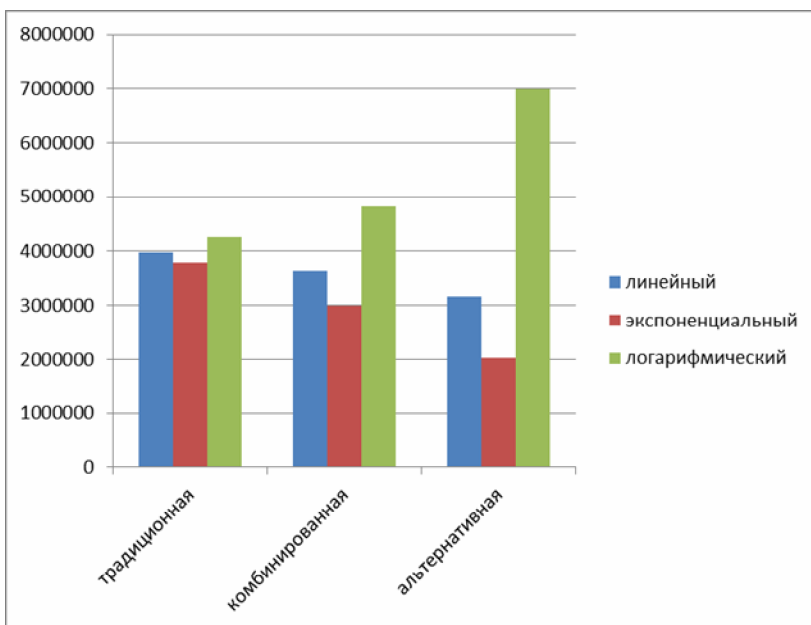


Рис. 8. Суммарная прибыль компании при использовании различных стратегий инвестирования в условиях реализации исследуемых сценариев роста рентабельности альтернативных технологий ($k = 0,4$)

5. Выбор временного интервала для смены инвестиционной стратегии

Анализ динамики прибыли компании при использовании различных стратегий инвестирования в условиях реализации линейного сценария роста эффективности альтернативных технологий, проведенный в работе [3], позволил выдвинуть гипотезу о том, что добиться максимальных значений суммарной за период прибыли компании возможно путем смены инвестиционной стратегии от традиционной или комбинированной к альтернативной на определенном шаге имитационного цикла. Проверка данного предположения путем проведения серии численных экспериментов показала, что при небольших значениях коэффициента интенсивности реинвестиций оптимальным временем перехода компании от традиционной инвестиционной стратегии к альтернативной является 20–30 шаг от начала имитационного периода, что соответствует 2030–2040 году.

При более высоких значениях коэффициента интенсивности реинвестирования оптимальное время перехода к альтернативной инвестиционной стратегии достаточно резко смещается ближе к началу имитационного периода. Так, при $k = 0,5$ переход к альтернативной стратегии позднее 8-го шага (что соответствует 2018 году) приводит к резкому увеличению упущенной выгоды (потерь в значениях суммарной прибыли).

Изменяя сценарий роста эффективности кластера новых технологий на экспоненциальный при $k = 0,1$, получим смещение оптимального времени перехода к альтернативной инвестиционной стратегии к самому концу имитационного цикла, что соответствует 2050 году (рис. 9).

При логарифмическом сценарии роста и том же значении коэффициента интенсивности реинвестирования оптимальное время смены стратегии инвестирования с традиционной на альтернативную смещается к 16–17-м шагам имитационного цикла, что соответствует 2028–2030 гг. (рис. 10).

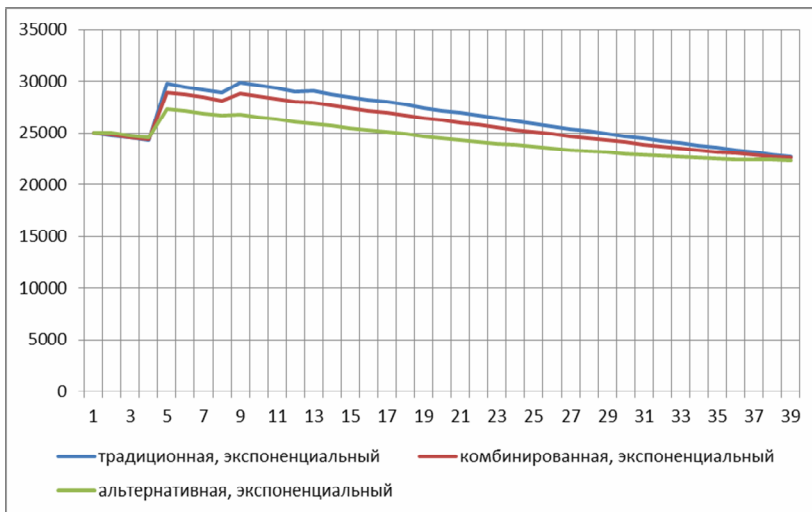


Рис. 9. Динамика прибыли компании при экспоненциальном сценарии роста эффективности альтернативных технологий ($k = 0,1$)

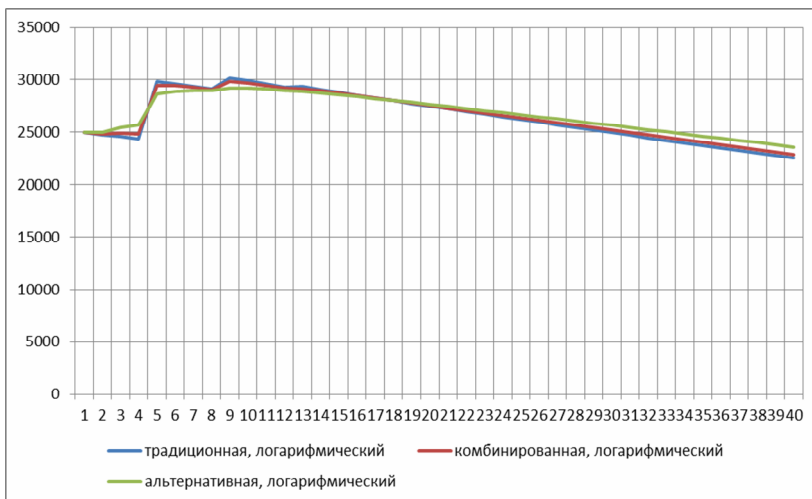


Рис. 10. Динамика прибыли компании при логарифмическом сценарии роста рентабельности альтернативных технологий ($k = 0,1$)

При $k = 0,2$ и экспоненциальном сценарии роста рентабельности альтернативных технологий время смены стратегии инвестирования с традиционной на альтернативную смещается за пределы исследуемого периода (рис. 2), так как на всем протяжении имитационного цикла традиционная и комбинированная стратегии инвестирования, несмотря на непрерывное снижение общей прибыли начиная с 17-го шага, предпочтительнее, нежели альтернативная. При логарифмическом сценарии роста альтернативная стратегия становится явно предпочтительнее, начиная с 17-го шага (рис. 3). Для того чтобы определить, какая из трех стратегий является предпочтительнее на начальных шагах имитационного цикла, рассчитаем суммарную прибыль компании за первые 16 шагов (рис. 11).

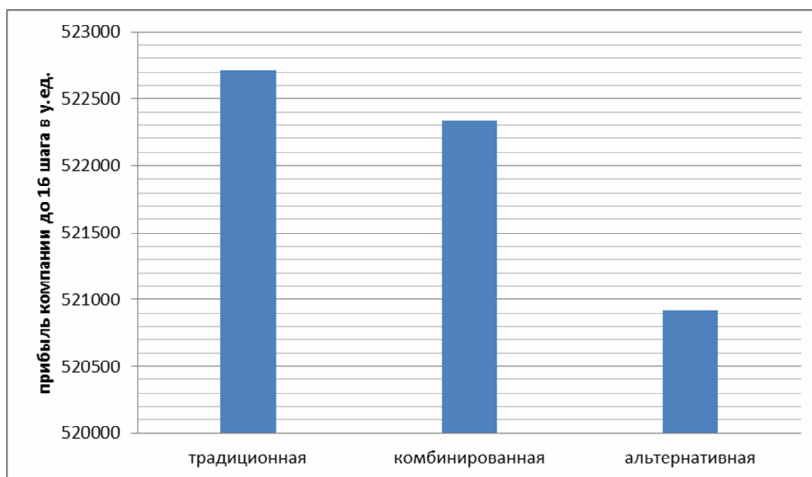


Рис. 11. Сравнение общей прибыли компании за первые 16 шагов имитационного цикла при логарифмическом сценарии ($k = 0,2$)

Из диаграммы, представленной на рис. 11 видно, что до 16-го шага имитационного цикла более предпочтительной является традиционная стратегия инвестирования. Таким образом, при $k = 0,2$ и логарифмическом сценарии оптимальным временем перехода от традиционной стратегии к альтернативной

является 16-й шаг имитационного цикла, что соответствует 2028 году.

При дальнейшем увеличении коэффициента интенсивности реинвестирования смена стратегий на протяжении всего имитационного цикла при нелинейном характере роста рентабельности альтернативных технологий не дает увеличения общей прибыли компании за весь период моделирования. При экспоненциальном сценарии роста эффективности кластера новых технологий традиционная стратегия является однозначно более предпочтительной на всем протяжении имитационного цикла. При логарифмическом сценарии роста более предпочтительной однозначно является альтернативная стратегия инвестирования.

6. Заключение

В настоящее время энергетические компании всего мира (в том числе и нефтегазовые) находятся в ситуации так называемого «технологического разрыва». Существующие технологии, основанные на использовании в качестве первичных источников энергии топливно-энергетических полезных ископаемых (технологии *замещаемой* группы), все меньше удовлетворяют растущим требованиям восполнения общественных потребностей с минимальными издержками по затратам и ресурсам, снижения негативного воздействия на окружающую среду и сохранения целостности экосистем. Однако для большинства новых технологий возобновляемой энергетики (технологии *замещающей* группы) коммерческий потенциал еще до конца не ясен, а стоимость перехода на новые технологии чрезвычайно высока. Определение оптимального времени перехода на замещающие технологии является фундаментальной научной проблемой, решение которой позволяет энергетическим компаниям и региональным энергетическим кластерам повысить свою конкурентоспособность в будущем за счет достижения высокого уровня технологической сложности своих производственных процессов, выхода на новые рынки и установления своего лидерства на них, в том числе и посредством введения новых стандартов энергоэффективности.

Для определения оптимальной степени диверсификации проектного портфеля энергетических компаний/региональных энергетических кластеров и необходимой для будущего технико-технологического развития доли проектов, направленных на внедрение технологий возобновляемой энергетики, в настоящей работе проведено исследование на основе имитационного моделирования. Отличительной особенностью построенной имитационной модели от предложенной ранее является то, что в ней предусмотрена возможность выбора сценария роста эффективности технологий замещающей группы: линейный, экспоненциальный (соответствующий начальной фазе жизненного цикла технологий замещающей группы) или логарифмический (соответствующий завершающей фазе жизненного цикла замещающих технологий, когда их эффективность приближается к технологическому пределу).

В результате проведенного исследования получены следующие основные выводы:

1. Как динамика, так и величина общей за период моделирования прибыли энергетической компании существенно зависят от сценария роста эффективности технологий возобновляемой энергетики даже *при одинаковых* нижнем и верхнем порогах рентабельности.

2. Сравнивая общую прибыль компании за весь имитационный цикл, можно сделать вывод о том, что экспоненциальный рост, соответствующий начальной стадии жизненного цикла кластера технологий, является наименее предпочтительным с точки зрения максимизации прибыли компании за весь период моделирования, а логарифмический - наиболее. При этом разница между значениями общей прибыли, полученной по различным сценариям роста, тем существенней, чем больше коэффициент реинвестирования и чем большая доля инвестиционных ресурсов направляется на проекты по развитию технологий возобновляемой энергетики.

3. При низких значениях коэффициента интенсивности реинвестирования ($k = 0,1$ и $k = 0,2$) добиться максимальных значений суммарной за период прибыли компании возможно путем смены инвестиционной стратегии от традиционной или комби-

нированной к альтернативной на определенном шаге имитационного цикла.

4. В условиях реализации экспоненциального сценария роста эффективности технологий замещающей группы при $k = 0,1$ оптимальное время перехода от альтернативной инвестиционной стратегии наступает к самому концу имитационного цикла, что соответствует 2050 году. При логарифмическом сценарии роста и том же значении коэффициента интенсивности реинвестирования оптимальное время смены стратегии инвестирования с традиционной на альтернативную смещается к 16–17-м шагам имитационного цикла, что соответствует 2028–2030 гг.

5. При $k = 0,2$ и экспоненциальном сценарии роста рентабельности альтернативных технологий время смены стратегии инвестирования с традиционной на альтернативную смещается за пределы исследуемого периода. При $k = 0,2$ и логарифмическом сценарии оптимальным временем перехода от традиционной стратегии к альтернативной является 16-й шаг имитационного цикла, что соответствует 2028 году.

6. При увеличении коэффициента интенсивности реинвестирования и нелинейном характере роста эффективности замещающих технологий смена стратегий инвестирования не дает эффекта увеличения общей прибыли компании. При экспоненциальном сценарии роста эффективности кластера новых технологий традиционная стратегия инвестирования является однозначно более предпочтительной на всем протяжении имитационного цикла, в то время как при логарифмическом сценарии более предпочтительной является альтернативная стратегия.

7. Полученные результаты свидетельствуют о том, что, несмотря на достаточно прозрачную логику процесса смены технологических укладов, даже при известных технологических пределах замещаемой и замещающей групп технологий выбор оптимальной стратегии инвестирования, позволяющей максимизировать суммарную прибыль компании за весь исследуемый период, является сложным, а простого алгоритма принятия правильного решения не существует. Необходимость учета на практике многих факторов, таких как доля нераспределенной

прибыли, направляемой на технологическое развитие, структура распределения инвестиционных ресурсов, характер роста эффективности технологий замещающей группы и технологические пределы технологий каждой из групп, актуализирует разработку специализированных программных инструментов для моделирования реальной ситуации и построения прогнозов развития компании.

8. Дальнейшим направлением исследований может быть уточнение значений верхнего и нижнего порога рентабельности технологий обеих групп (традиционной и альтернативной энергетики) по данным о реализации новых проектов, а также введение поправочных коэффициентов для учета таких социально-экономических эффектов, оказывающих влияние на рентабельность, как эффект обучения, эффект масштаба и эффект технологического спиловера.

Литература

1. НИЖЕГОРОДЦЕВ Р.М. *Основы теории инноваций*. – М.: Доброе слово, 2011. – 88 с.
2. РАТНЕР С.В., МИХАЙЛОВ В.О. *Диверсификация проектного портфеля нефтегазовых корпораций как способ поддержания стратегической конкурентоспособности* // Экономический анализ: теория и практика. - 2012. - №13(268). - С. 11-20.
3. РАТНЕР С.В., МИХАЙЛОВ В.О. *Управление развитием энергетических компаний в ситуации технологического разрыва* // Управление большими системами. – 2012. - №37. - С. 180-207.
4. *Регулирование возобновляемой энергетики*. - ERA, 2011. – 113 с. [Электронный ресурс]. – URL: http://www.fstrf.ru/about/activity/inter/2/erra/2/RES_textbook_FINAL_rus.pdf (дата обращения: 08.09.2013).
5. ФЕДОРЧУК А.А. *Моделируемые и немоделируемые риски энергетической компании* // Национальные интересы: приоритеты и безопасность. - 2012. - №9. - С. 42-47.

6. FRAYER J., ULUDERE N.Z. *What Is It Worth? Application of Real Options Theory to the Valuation of Generation Assets* // The Electricity Journal. – 2011. - Vol. 14, №8. - P. 40-51.
7. GUILFORD H., CLEVELAND C. *A New Long Term Assessment of Energy Return on Investment (EROI) for U.S. Oil and Gas Discovery and Production* // Sustainability. - 2011. - №3. - P. 1866-1887.
8. LUNDMARK R., PETTERSON F. *The Economic of Power Generation Technology Choice and Investment Timing in the Presence of Policy Uncertainty* // Low Carbon Economy. – 2012. - Vol. 3, №1. - P. 1-10.
9. *The Perspectives of Energy Technologies: Strategies and Scenario until 2050*. - Russian Translation. International Energy Agency. – Paris, France, 2010. - 25 p.
10. *The Wind Energy Industry in Germany* // Germany Trade and Invest. – Issue 2010/2011. – Berlin, 2010. — 16 p.
11. YANG M., BLYTH W., BRADLEY R. et al. *Evaluating the Power Investment Options with Uncertainty in Climate Policy* // Energy Economics. – 2008. - Vol. 30, №4. - P. 1933-1950.
12. *2011 Renewable Energy Data Book*. - U.S. Department of Energy, 2011. - 128 p.

MANAGEMENT OF TECHNOLOGICAL PORTFOLIO OF ENERGY COMPANIES: SCENARIO APPROACH

Svetlana Ratner, Institute of Control Science, RAS, Moscow,
Doctor of Science (lanarat@mail.ru).

Abstract: On an example of an energy company we study the optimal timing for the company to widely implement alternative technologies to production activity. We employ the method of simulation of the energy company profit with respect to different strategies of investment into traditional and alternative energy technologies to test the hypothesis of profit dependence on the scenario of alternative technologies' profitability dynamics. We consider the scenarios of alternative energy technology efficiency growth corresponding to the initial and the final stage of a technological gap.

Keywords: energy industry, technology gap, technology portfolio, simulation model, renewables, growth scenarios.

*Статья представлена к публикации
членом редакционной коллегии Р.М. Нижегородцевым*

*Поступила в редакцию 19.07.2013.
Опубликована 30.09.2013.*

УДК 336.71.078.3

ББК 78.34

УПРАВЛЕНИЕ УСТОЙЧИВОСТЬЮ БАНКОВСКОЙ СЕТИ С УЧЕТОМ ОТРАСЛЕВЫХ РИСКОВ

Стежкин А. А.¹

(Московский физико-технический институт, Москва)

Рассматривается подход к оценке вероятных потерь в ситуациях дефолта, основанный на разделении банков в зависимости от преобладающей кредитуемой области в их кредитном портфеле. Поставлена и проанализирована задача управления по определению оптимальной величины ссуды в отрасль, а также задача оптимизации, доставляющая минимум совокупной вероятности невыплаты банками по обязательствам в случае кризиса в отрасли при условии выполнения банками норматива достаточности капитала, а также достаточности средств для деятельности компании. Рассмотрена модель оценки возможных потерь в результате стрессовых ситуаций в банковской сети и определения системной значимости банков, основанная на векторе Сноу, для которой разобран ряд модельных примеров. Описанные в статье подходы ранее не применялись в российской практике в силу того, что существенная необходимость оценки системного риска возникла в российской банковской системе относительно недавно – в связи с изменениями в международных консультативных документах.

Ключевые слова: множественный дефолт, задача оптимизации, logit-модель, норматив достаточности собственных средств (капитала).

¹ Александр Александрович Стежкин, студент (alex79216@mail.ru).

1. Введение

Задача оценки вероятных потерь в ситуации дефолта чрезвычайно актуальна как в России, так и во всем мире. В банковской практике разработано и опробовано большое число различных моделей, использующих информацию самого разного уровня.

На фоне событий, происходивших в России в ходе и после кризиса 2007-2009 годов и значительным образом затронувших многие отрасли хозяйства, представляется разумным рассматривать определенную отрасль в отдельности с точки зрения вероятных потерь, которые могут понести как сами предприятия, так и банковский сектор, связанный с этой отраслью. Действительно, стрессовая ситуация в отрасли приводит к невыплате по обязательствам компаний этой отрасли и, следовательно, значительному увеличению потерь активов банков. В результате банки испытывают затруднения с выплатой по собственным обязательствам другим кредитным организациям. Возникает лавинообразный процесс, «эффект домино». Именно в таких случаях чрезвычайно важным оказывается решение задачи управления для банков по определению оптимальной ссуды предприятиям – такой, которая, с одной стороны, в случае невозврата не ставила бы под угрозу деятельность кредитной организации, а с другой стороны была бы приемлемой для ведения бесперебойной деятельности компаний отрасли. Причем потребность в решении такой задачи управления существует как у самой кредитной организации, так и у надзорного органа (Центрального Банка), в чьих интересах не допустить развитие стрессовых ситуаций в экономике страны.

В настоящее время распространены модели по оценке вероятных потерь банков, основанные на внутренних данных компаний-заемщиков – состоянии баланса, котировках акций и кредитном рейтинге. К сожалению, такая информация зачастую оказывается недоступной, особенно в случае средних и небольших предприятий. Еще один распространенный класс моделей основывается на исторических данных о частоте дефолтов и средней величине потерь компаний и банков и, как правило, сопряжен со значительными вычислительными трудностями.

Предлагаемые в данной работе задачи используют теорию оптимального управления и итерационные алгоритмы для оценки вероятных потерь банков, определения их системной значимости в рамках данной отрасли хозяйства, а также для принятия управленческих решений о кредитовании предприятий в условиях стрессовых ситуаций в экономике. При этом в задачах не предполагается использование труднодоступной информации и сложных вычислительных процедур, что является несомненным преимуществом.

2. Выбор индикатора

Для того чтобы установить обоснованность подхода к оценке вероятных потерь банков через рассмотрение отрасли, которую они преимущественно кредитуют, нужно убедиться, что в случае стрессовых ситуаций эти банки ведут себя сходным образом, т.е. можно делать общие выводы относительно всех банков внутри выбранной отрасли. С точки зрения анализа наличия этой связи необходимо выбрать такой показатель деятельности, который являлся бы хорошим индикатором финансового состояния банка в зависимости как от внешних, так и от внутренних условий

Центральный Банк Российской Федерации, выступая в роли регулятора, устанавливает ряд нормативов, которым должна соответствовать деятельность всех кредитных организаций. Нормативы эти описаны в Инструкции Банка России «Об обязательных нормативах банков» №139-И от 3 декабря 2012 года (в предыдущей редакции – инструкция №110-И от 16 января 2004 года) [3]. Одним из них является норматив достаточности собственных средств (капитала) банка (Н1), который регулирует (ограничивает) риск несостоятельности банка и определяет требования по минимальной величине собственных средств (капитала) банка, необходимых для покрытия кредитного, операционного и рыночного рисков. Норматив Н1 определяется как отношение размера собственных средств (капитала) банка и суммы его активов, взвешенных по уровню риска. Н1 является наиболее комплексным показателем, учитывающим множество факторов. Примем его в качестве искомого показателя.

Вообще говоря, в зависимости от структуры кредитного портфеля юридических лиц, банки могут быть многоотраслевые и обслуживающие преимущественно одну из отраслей. В России в настоящее время преобладают многоотраслевые банки. Однако существуют и целые группы банков, специализирующихся на кредитовании определенных областей хозяйства, которые, таким образом, не диверсифицируют свои риски, в результате чего оказываются подвержены большим потенциальным потерям в случае ситуации кризиса в обслуживаемой ими области.

Рассмотрим в качестве примера три отрасли: сельское хозяйство, нефтегазовую отрасль и металлургическую промышленность. В целях анализа были выделены отдельные группы банков, занимающихся кредитованием конкретно одной из перечисленных выше отраслей [1].

1) Промышленность.

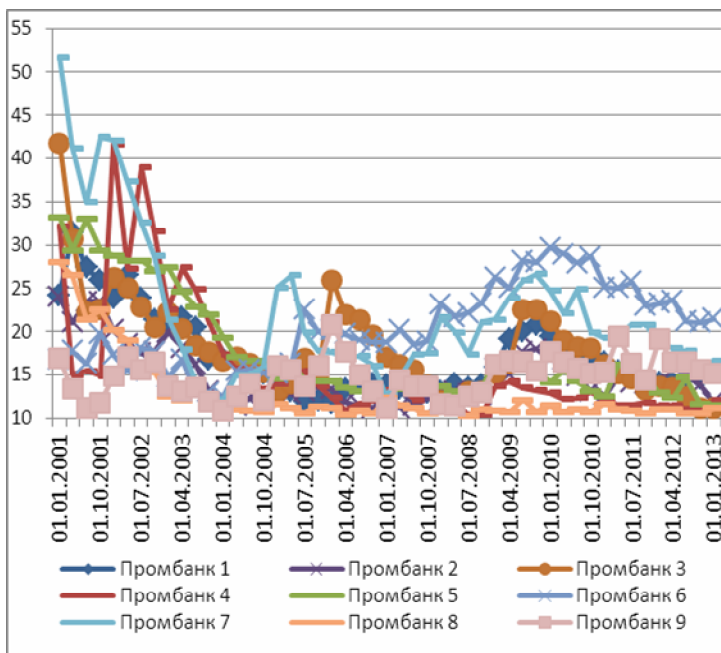


Рис. 1. Изменение норматива HI банков, занимающихся кредитованием предприятий металлургической промышленности

Как видно из рис. 1, норматив Н1 банков металлургической промышленной отрасли согласованно рос в период 2010 года, а затем снижался.

2) Нефтегазовая отрасль.

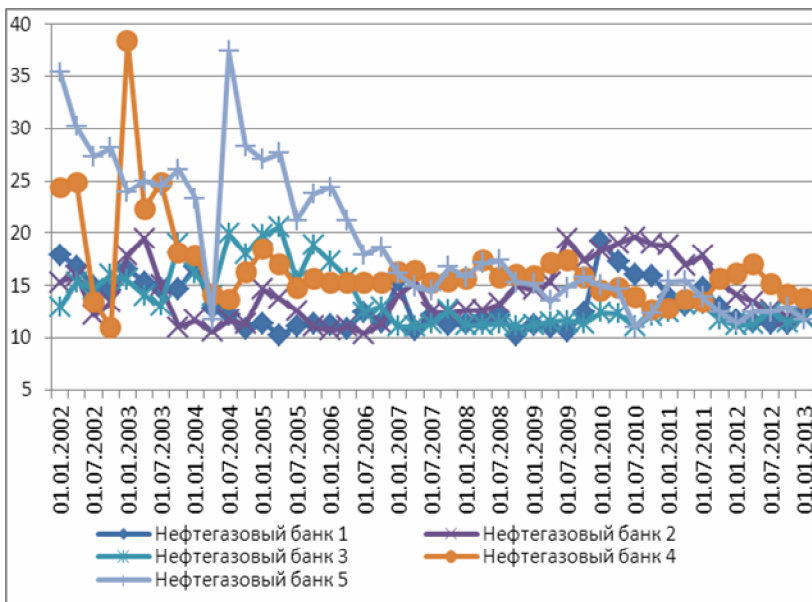


Рис. 2. Изменение норматива Н1 банков, занимающихся кредитованием предприятий нефтегазовой отрасли

Из рис. 2 видно довольно стабильное состояние в период кризиса (значение Н1 на уровне 13%), затем краткосрочный рост в 2010 году и вновь стабильное состояние, но уже на менее высоком уровне. Заметим, что по сравнению с банками промышленных отраслей, банки, занимающиеся кредитованием предприятий нефти и газа, более стабильны и не испытывали резкого роста или падения на рассматриваемом горизонте.

3) Сельское хозяйство

На рис. 3, помимо падения норматива Н1 за 2011-2012 годы отчетливо видна также сезонность. Норматив Н1 проявляет тенденцию изменения поздней осенью-зимой, когда закончен сбор урожая и кредиты либо возвращаются (хороший урожай),

либо выплаты задерживаются (неурожай). Именно сезонностью обслуживаемого бизнеса можно объяснить скачкообразное поведение графиков.

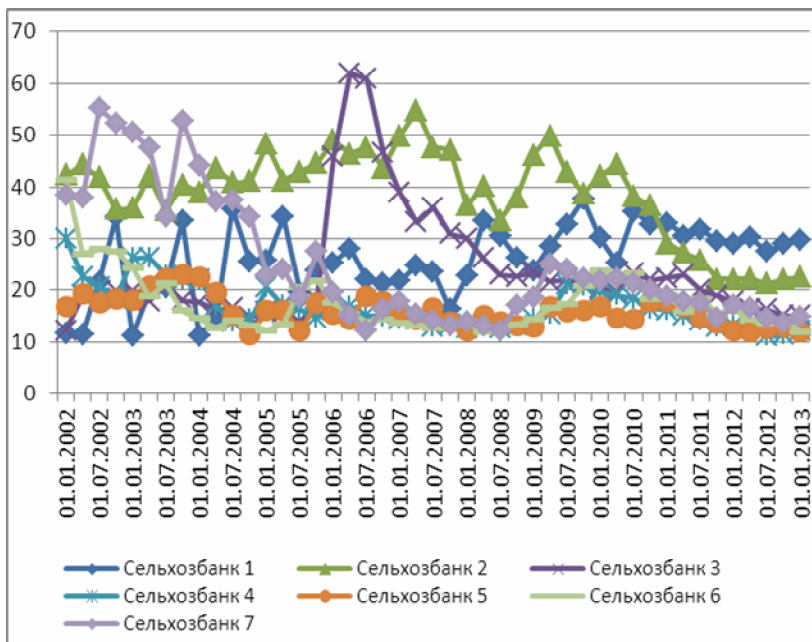


Рис. 3. Изменение норматива Н1 банков, занимающихся кредитованием сельскохозяйственных предприятий

Попробуем формально проанализировать такое, достаточно синхронное поведение норматива Н1 для банков одной отрасли. Рассмотрим 2 ситуации.

1) Отрасль испытывает спад.

Это означает, что предприятия ограничивают свои заимствования у банков и испытывают проблемы с погашением уже имеющихся займов. Оба фактора негативно сказываются на прибыли банка (если банк по итогам периода несет убытки, то уменьшается размер его собственных средств), влияют на величину расчетного риска (кредитный риск, операции с повышенным коэффициентом риска и т.д.), а также существенно воздействуют на величину создаваемых резервов (непогашенные

кредиты переводятся банком в более низкие категории качества, что сопровождается формированием большего резерва под них). Таким образом, в этой ситуации в показателе Н1 мы имеем возможное снижение числителя (за счет убытков) и увеличение знаменателя (за счет роста формируемых резервов), что приводит к уменьшению норматива.

2) Отрасль испытывает подъем.

Здесь складывается обратная ситуация. Бизнес компаний растет. Это приводит к росту объема кредитов, а также к уменьшению случаев просрочки или невозврата этих кредитов. В результате банк получает прибыль, что увеличивает размер его собственных средств. Кроме того, хороший кредитный портфель банка означает меньшие риски и величину формируемого резерва под ссуды, что уменьшает величину знаменателя норматива Н1. В этой ситуации мы наблюдаем рост норматива Н1 за счет роста числителя (капитала) и уменьшения знаменателя (риски и резервы).

Разумеется, банки стремятся диверсифицировать свои кредитные портфели путем ссудирования нескольких областей, чтобы снизить риски невыплаты. Однако, часты случаи, когда компании имеют договоренности с определенными банками, то есть невозможен вариант привлечения новых компаний-клиентов для банков, желающих диверсифицировать свои портфели (поскольку их обслуживают другие банки). В таких случаях единственным приемлемым вариантом является расширение портфеля кредитов физическим лицам. В результате, портфель кредитов юридическим лицам остается неизменным с преобладанием определенной отрасли, а риски в целом снижаются благодаря увеличению объемов кредитов физическим лицам.

3. Статистическая значимость

Итак, на примере трех отраслей было высказано предположение, что норматив Н1 у банков, специализирующихся на кредитовании определенной отрасли хозяйства, изменяется во времени согласованно. Для того чтобы исключить случайность таких изменений, проведем регрессионный анализ, а именно установим существенность зависимости между величиной

норматива и величиной вклада в отрасль (в абсолютном выражении или в долях от совокупного кредитного портфеля).

Для этого используем метод наименьших квадратов, минимизирующий сумму квадратов отклонений реально наблюдаемых Y от их оценок \hat{Y} :

$$(1) \sum_{k=1}^M (Y_k - \hat{Y}_k)^2 \rightarrow \min$$

где M – объем выборки.

Хорошим индикатором служит множественный коэффициент корреляции (множественный R) – квадратный корень из коэффициента детерминации. Этот коэффициент определяют тесноту связи между зависимой переменной и фактором.

Иногда показателям тесноты связи можно дать качественную оценку по так называемой шкале Чеддока.

Таблица 1. Шкала Чеддока

Количественная мера тесноты связи	Качественная характеристика силы связи
0,1–0,3	Слабая
0,3–0,5	Умеренная
0,5–0,7	Заметная
0,7–0,9	Высокая
0,9–0,99	Очень высокая

Рассмотрим выборку банков, динамика норматива $H1$ во времени для которых была построена ранее, и проведем регрессионный анализ для них на предмет зависимости норматива $H1$ от величины ссуд в отрасль. Анализ был проведен на уровне значимости 95%. Это значит, что в случае если p -value превысит 0,05, делается вывод об отсутствии зависимости между значениями $H1$ и ссудами в отрасль.

Таким образом, по случайной выборке из рассматриваемого перечня банков видно, что норматив $H1$ находится в зависимости, от умеренной до высокой, от величины ссуд в отрасль (значения множественного R от 39% и выше), при этом от 15% до 57% вариации норматива $H1$ объясняется вариацией величин ссуд в отрасль.

Итак, динамика норматива H_1 в большой степени зависит от динамики вкладов в отрасль, несмотря на то, что H_1 представляет собой величину, зависящую от многих параметров.

Таблица 2. Результаты регрессионного анализа

Банк	Множественный R	R -квадрат	Стандартная ошибка	Значимость F (p -value)
Нефтегазовый банк 1	75,98%	57,73%	1,88	0,00000046
Нефтегазовый банк 2	44,11%	19,46%	1,97	0,0115
Промбанк 1	41,86%	18,86%	3,02	0,0338
Промбанк 2	46,18%	20,02%	2,94	0,0117
Промбанк 3	60,99%	37,20%	3,06	0,0002
Сельхозбанк 1	39,00%	15,21%	2,42	0,0273
Сельхозбанк 2	39,15%	15,32%	11,58	0,0267

4. Задача определения оптимальной величины ссуды в отрасль.

Существующие методы оценки вероятных потерь банков в результате дефолтов рассматривают главным образом парные связи банка с другим банком или компанией и данные о финансовом состоянии компаний. Кроме того, многие методы предполагают использование значительных массивов статистической информации. В российских реалиях использование таких методов затруднено в силу отсутствия актуальной информации по компаниям, не принадлежащим к числу крупнейших, а также ограниченности выборки статистических данных по стрессовым ситуациям в прошлом. Предлагаемые задачи оптимизации, напротив, позволяют проанализировать ситуацию на основе текущей информации с использованием небольшого объема требуемых данных по компаниям и банкам, а также учитывают множественные связи банков с компаниями и другими банками.

Применение теории оптимального управления в целях решения задач российской банковской практики ранее отмечено не было. В зарубежных работах чаще всего можно встретить применение теории оптимального управления к организации

потоков активов и обязательств, нежели чем к оценке рисков, в том числе системных (одни из последних подобных работ – см. [16, 18]). Использование показателя достаточности собственных средств (капитала) в качестве индикатора деятельности банков непосредственно и отраслей хозяйства опосредованно не было отмечено ни в российской, ни в зарубежной практике.

Начнем с простейшего формального описания задачи управления, стоящей перед нами.

Пусть у нас есть n банков, кредитующих определенную отрасль;

K_1, K_2, \dots, K_n – размер собственных средств (капитала) этих банков;

$\eta_1, \eta_2, \dots, \eta_n$ – значение норматива Н1 банков;

C_1, C_2, \dots, C_n – величины вложений банков в отрасль.

Упрощенно запишем Н1 как $\eta_i = K_i / AR_i$, где AR_i – величина взвешенных по уровню риска активов.

Кроме того, рассмотрим целиком группу банков. Помимо кредитования отрасли эти банки имеют обязательства (часто взаимные) друг перед другом. Таким образом, можно построить граф, в узлах которого находятся сами банки, а ребра – обязательства. С точностью до процентной ставки можно произвести процедуру неттинга, что значительно облегчит дальнейшее рассмотрение задачи.

Итак, обозначим KR_{ij} – величина кредита, который банк i выдал банку j .

(2) для любой пары банков i и j существует величина

$$KR_{ij}^* = KR_{ij} - KR_{ji}.$$

Если $KR_{ij}^* > 0$, то банк i по отношению к банку j выступает нетто-кредитором. Иначе – нетто-заемщиком.

В результате кризиса и сопутствующего невозврата ссуды норматив банка упадет до величины

$$(3) \quad \eta_i^* = \frac{(K_i - C_i)}{(K_i / \eta_i) - C_i}.$$

Дальше есть 2 пути развития ситуации.

$$1) \eta_i^* < 10\% .$$

То есть пусть часть банков потеряли средства, их норматив Н1 опустился ниже 10%, и выплаты по обязательствам они приостановили. Обозначим это множество банков через J_{do} .

Общая величина потерь всей системы банков в результате этого, очевидно, может быть определена следующим выражением:

$$(4) \sum_{i \in J_{do}} \sum_{j: KR_{ji}^* > 0} KR_{ji}^* .$$

$$2) \eta_i^* > 10\%$$

Несмотря на потери, банку удалось сохранить значение норматива на уровне выше минимума, установленного регулятором ($i \in J / J_{do}$).

Здесь можно оценить величину средств, которую банк еще может вернуть заемщику, не нарушив норматив. Определяется это из условия

$$(5) \frac{K_i - C_i - X_i}{(K_i / \eta_i) - C_i - X_i} = 10\% ,$$

где X_i – искомая величина.

Отсюда получаем:

$$(6) X_i = \frac{10}{9} \cdot (K_i - 0,1 \cdot (K_i / \eta_i) - 0,9C_i) .$$

Отметим, что возврат суммы обязательств свыше X_i может быть установлен регулятивным органом, как и величина Н1.

Таким образом, нашу задачу можно интерпретировать как задачу управления – определения оптимальной величины ссуды в отрасль с точки зрения концентрации риска.

Рассмотрим один упрощенный иллюстративный пример.

Пример 1. На отчетную дату норматив Н1 Банка равен 11,1%. Собственные средства (капитал) Банка – 150 млрд. руб. Банк кредитует предприятия Отрасли на общую сумму 10 млрд. руб., а также имеет обязательства по отношению к другим банкам, специализирующимся на кредитовании Отрасли, на общую сумму 40 млрд. руб. В случае невозврата Банку кредитов предприятиями Отрасли, какую сумму обязательств он будет в состоянии выплатить?

Итак, согласно произведенным выше рассуждениям, получим, что в случае невозврата кредитов новый норматив Π_1 Банка равен

$$\eta^* = \frac{150 - 10}{150/11,1\% - 10} = 10,4\% .$$

Норматив достаточности собственных средств (капитала) Банка снизился, однако по-прежнему остался выше установленного минимума.

Искомая величина обязательств, которую Банк сможет выплатить, определяется следующим образом:

$$X = \frac{10}{9} \cdot (150 - 0,1(150/0,111) - 0,9 \cdot 10) = 6,5 \text{ млрд.руб.}$$

Таким образом, Банк сможет выплатить лишь 16,25% своих обязательств по отношению к другим банкам Отрасли.

Вернемся к нашей общей задаче управления. Мы хотим найти такой набор (C_1, C_2, \dots, C_n) ссуд компаниям, при которых система банков останется в равновесии и не испытает стресса. То есть компании должны эти ссуды выплатить, а банки в случае невыплаты сохранить своё значение норматива Π_1 выше минимума.

Разумно добавить к этой задаче вероятности невыплаты компаниями ссуд. Первым делом используем для моделирования *logit*-модель [7]. Пусть u – величина ссуды, выданной компании, тогда в рамках модели функция вероятности выплаты ссуды имеет вид [10]:

$$(7) \quad \tilde{F}(u) = \frac{2e^{-u}}{1 + e^{-u}} I_{\{u \geq 0\}} ,$$

где коэффициент 2 введен для того, чтобы выполнялось $F(0) = 1$.

Вероятность невыплаты ссуды может быть, в таком случае, задана такой функцией:

$$(8) \quad F(u) = 1 - \tilde{F}(u) = \frac{1 - e^{-u}}{1 + e^{-u}} I_{\{u \geq 0\}} .$$

Видим, что $F(0) = 0$, $F(+\infty) = 1$, что согласуется со смыслом функции вероятности.

Добавим также, что в формулу функции вероятности, в силу экспоненциального характера её роста, можно добавить нормировочный фактор, который сглаживал бы быстрый рост

функции распределения и уход её на единичный уровень. Фактор зависит от величин ссуд. Таким образом, в общем виде функцию вероятности невыплаты можно записать так:

$$(9) \quad F(u) = \frac{1 - e^{-u/k}}{1 + e^{-u/k}} I_{\{u \geq 0\}},$$

где k – упомянутый фактор. Условие нормировки функции плотности при этом соблюдается.

В условиях рассмотренной выше ситуации одной и той же компании выдают ссуды сразу несколько банков определенной нами группы.

Пусть C_{ij} – величина ссуды i -го банка для j -й компании, при этом у нас n банков и m компаний. Заметим, что величина C_{ij} может быть нулевой. У каждой компании есть потребность в финансировании. Обозначим D_j минимальную сумму, которую должна привлечь j -я компания для продолжения своей деятельности.

Задача 1. Задача минимизации вероятности невыплаты ссуд.

Целью задачи является нахождение такого набора ссуд банков компаниям определенной отрасли, который минимизирует совокупную по всем компаниям вероятность невыплаты ссуд при условии, что выданных ссуд хватит для организации деятельности компаний и при этом, в случае возможной невыплаты этих ссуд банкам, последние не нарушат норматив достаточности капитала.

Математическая постановка.

$$(10) \quad \left\{ \begin{array}{l} P = \sum_j P_j = \sum_j \frac{1 - e^{-\sum_i C_{ij}}}{1 + e^{-\sum_i C_{ij}}} I_{\{\sum_i C_{ij} \geq 0\}} \rightarrow \min_{C_{ij}}, \\ K_i - \sum_j C_{ij} \geq 10\%((K_i / \eta_i) - \sum_j C_{ij}), \\ \sum_i C_{ij} \geq D_j. \end{array} \right.$$

Если поставленная задача является разрешимой, т.е. множество решений не пусто, то ограничения на величины средств для компаний являются равенствами. Задачу можно переписать так:

$$(11) \left\{ \begin{array}{l} P = \sum_j P_j = \sum_j \frac{1 - e^{-D_j}}{1 + e^{-D_j}} \rightarrow \min_{C_{ij}}, \\ K_i - \sum_j C_{ij} \geq 10\%((K_i / \eta_i) - \sum_j C_{ij}), \\ D_j = \sum_i C_{ij}. \end{array} \right.$$

То есть для минимизации совокупной вероятности невыплаты ссуд необходимо выдавать минимально необходимые компаниям ссуды, соблюдая при этом норматив достаточности.

Проанализируем данную задачу с точки зрения теории оптимизации [8]. Это задача Лагранжа.

Строим функцию Лагранжа:

$$(12) L = \lambda_0 \cdot \sum_{j=1}^m P_j + \sum_{i=1}^n \lambda_1^i (0,1((K_i / \eta_i) - \sum_j C_{ij}) - (K_i - \sum_j C_{ij})) + \sum_{j=1}^m \lambda_2^j (D_j - \sum_i C_{ij})$$

Необходимые условия локального минимума:

1) Стационарность:

$$\frac{\partial L}{\partial C_{ij}} = \lambda_0 \cdot \frac{\partial P_j}{\partial C_{ij}} - \lambda_1^i - \lambda_2^j = 0;$$

для любого $i \in \{1, \dots, n\}$, для любого $j \in \{1, \dots, m\}$.

2) Дополняющая нежесткость:

$$\lambda_1^i (0,1((K_i / \eta_i) - \sum_j C_{ij}) - (K_i - \sum_j C_{ij})) = 0;$$

$$\lambda_2^j (D_j - \sum_i C_{ij}) = 0;$$

для любого $i \in \{1, \dots, n\}$, для любого $j \in \{1, \dots, m\}$.

3) Неотрицательность:

$$\lambda_0 \geq 0, \lambda_1^1 \geq 0, \dots, \lambda_1^n \geq 0, \lambda_2^1 \geq 0, \dots, \lambda_2^m \geq 0.$$

Решение задачи нельзя свести к минимизации по каждому участнику в отдельности из-за ограничений как касательно компаний, так и банков. Функция P – монотонная по каждому аргументу, непрерывная, без особенностей, а множество, задаваемое ограничениями, представляет собой компакт. Отсюда по теореме Вейерштрасса следует, что P достигает своих экстремальных значений на множестве, т.е. задача имеет решение [9].

Поскольку $F(C_j) = \frac{1 - e^{-\sum_i C_{ij}}}{1 + e^{-\sum_i C_{ij}}}$, $\sum_i C_{ij} \geq 0$, – функция выпук-

лая, а сумма выпуклых функций есть функция выпуклая, то минимизируемая P – выпуклая. Функции, задающие ограничения равенств и неравенств, линейные. Пусть для любого j $\bar{C}_j = (C_{1j}, \dots, C_{nj})$ и $\bar{\lambda} = (\lambda_0, \lambda_1^1, \dots, \lambda_1^n, \lambda_2^1, \dots, \lambda_2^m)$ таковы, что выполняются необходимые условия локального минимума 1)–3).

Отсюда, используя теорему из теории оптимизации, задающую достаточные условия такой задачи с ограничениями типа равенств и неравенств, делаем вывод, что есть решение 1)–3) существует, то это решение и есть искомое, т.е. минимум. Мы показали существование решения задачи математически.

Поставленная выше задача являлась упрощенной и не учитывала заинтересованность банков в размещении как можно большей величины средств с целью получения дохода (в виде процентов по выдаваемым ссудам).

Рассмотрим другую задачу.

Задача 2. Задача оптимизации ожидаемых потерь и ожидаемой доходности.

Эта задача ставит цель нахождения такого набора ссуд, выдаваемых компаниям банками, который бы одновременно минимизировал ожидаемые потери банков в результате невозврата ссуд, а также максимизировал ожидаемую доходность банков при тех же ограничениях, что и в предыдущей задаче. То есть в данном случае учитывается не только желание банка уменьшить вероятность потери своих активов, но и стремление заработать на них, выдав ссуды компаниям настолько много, насколько возможно.

Математическая постановка.

$$(13) \quad \begin{cases} ES = \sum_j (P_j \cdot \sum_i C_{ij}) \rightarrow \min_{C_{ij}}, \\ ER = \sum_j ((1 - P_j) \cdot \sum_i C_{ij} \cdot k_j) \rightarrow \max_{C_{ij}}, \\ K_i - \sum_j C_{ij} \geq 10\%((K_i / \eta_i) - \sum_j C_{ij}), \\ \sum_i C_{ij} \geq D_j. \end{cases}$$

Здесь первая (минимизируемая) функция, есть функция ожидаемых потерь банков, а вторая (максимизируемая) есть ожидаемая доходность от выдаваемых ссуд, где k_j – средняя ставка размещения ресурсов банков. В такой постановке упомянутое выше тривиальное решение задачи не подходит. Для поиска решения нужно применить разобранный выше математический аппарат для каждой подзадачи в отдельности. Решение в подзадачах существует в силу монотонности функций и компактности множеств.

Итак, использование теории оптимизации позволяет банкам принимать управленческие решения относительно величины ссуд компаниям определенной отрасли, приемлемых как с точки зрения риска потерь, так и с точки зрения ожидаемой доходности.

Наконец, рассмотрим одну модель, позволяющую оценить системную значимость банка в банковской сети, основываясь на величине совокупных потерь, которые понесет система в результате дефолта этого банка.

5. Вектор Сноу

Метод анализа сетевых взаимосвязей на межбанковском кредитном рынке, позволяющий охарактеризовать негативный финансовый эффект в случае множественного дефолта. Вектор Сноу – по сути своей вектор размерности числа банков в системе, каждый элемент которого есть совокупная величина потерь, которую дефолт конкретно взятого банка принесет системе. Вычисление значения вектора основывается на переборе всех возможных комбинаций распространения кредитного шока. Один сценарий представляет собой результаты расчета по одному банку. Число сценариев соответствует численности участников межбанковского рынка.

Сам по себе вектор Сноу близок к еще одному методу оценки системного риска – вектору Шепли. Данный метод имеет широкое применение в зарубежной практике для оценки системного риска (см. [13, 14]). В России этот способ был не столь актуален в силу лишь недавно появившейся необходимости определять системную значимость кредитных организа-

ций с целью оценки вероятных потерь. Алгоритмы и идеи, лежащие в основе как оригинальных методов вектора Сноу, так и описанные в указанных трудах, были переработаны, максимально упрощены и адаптированы к российским реалиям в настоящей работе.

Итак, алгоритм расчета состоит из нескольких этапов. На первом этапе определяются последствия дефолта рассматриваемого банка. По каждому участнику межбанковского рынка выдвигается гипотеза, что он может выступить стороной, не выполняющей долговые обязательства. Посредством имитационного моделирования рассчитываются потери в результате реализации кредитного риска по каждому из контрагентов дефолтера. После определения величины потерь его банки-контрагенты проверяются на предмет способности выполнения ими обязательств перед другими участниками межбанковского рынка. Дефолт исходного контрагента влияет на финансовое положение участников по двум направлениям: возникновение убытков, которые влияют на достаточность капитала, а также потеря ликвидности в результате невыполнения платежного графика. Для определения, выступит ли, в свою очередь, контрагент исходного дефолтера неплательщиком по обязательствам, в векторе Сноу используется формальный критерий (можно использовать и несколько): снижение фактического значения норматива достаточности капитала Н1 (ниже 11% при минимуме в 10%). Если хотя бы один из критериев выполняется, банк считается потенциально уязвимым для «эффекта домино» и в последующем рассматривается как объявивший дефолт. На втором этапе по всем дефолтерам, включая исходного контрагента, просчитываются убытки, навязываемые ими банковскому сектору, и определяются новые дефолтеры на второй рабочий день. На третьем этапе идентифицируются новые дефолтеры на третий рабочий день и т.д.

В качестве иллюстрации рассмотрим несколько примеров банковских групп и соответствующих ситуаций распространения дефолта в них.

В каждом из следующих примеров банковская сеть исследуется на устойчивость к дефолту одного из участников. Алгоритм, таким образом, позволяет оценить относительную значи-

мость конкретных банков для устойчивости всей системы в целом, а также величину потерь в абсолютном выражении.

Пример 2. Имеется следующая банковская сеть.



Рис. 4. Банковская сеть для Примера 2

Исходные данные по банкам выглядят так:

Таблица 3. Финансовые показатели банков для Примера 2

Банк	Капитал	Н1
1	1000000	12%
2	500000	13%
3	380000	14%
4	1200000	11%

Рассматриваем ситуации дефолта у всех банков по очереди.

1) Пусть дефолт объявил Банк №1.

Таблица 3.1. Потери банков в условиях Примера 2 в случае дефолта Банка №1

Банк	Потери	Н1
2	0	13,0%
3	0	14,0%
4	80000	10,3%

Поскольку Банк №1 был должен только Банку №4 (см. схему), то и пострадает в результате его дефолта только этот банк.

А поскольку Банк №4 в свою очередь никому не был должен, то на его дефолте «эффект домино» закончится.

2) Пусть дефолт объявил Банк №2.

Таблица 3.2. Потери банков в условиях Примера 2 в случае дефолта Банка №2

Банк	Потери	Н1	Потери	Н1
1	300000	8,4%	x	x
3	0	14,0%	0	14,0%
4	0	11,0%	80000	10,3%

Здесь ситуация более сложная. Сначала объявляет дефолт Банк №1 из-за неисполненных обязательств перед ним Банка №2. Затем, уже вследствие дефолта Банка №1, обанкротится Банк №4.

3) Пусть дефолт объявил Банк №3.

Таблица 3.3. Потери банков в условиях Примера 2 в случае дефолта Банка №3

Банк	Потери	Н1
1	200000	9,6%
2	0	13,0%
3	0	14,0%
4	120000	9,9%

Здесь в результате невыплаты по обязательствам банкротятся также два банка.

По итогам рассмотрения всех ситуаций получаем такую картину.

Таблица 4. Итог по Примеру 2

Банк	Потери в результате дефолта	Число дефолтов	Число связей	Объем операций
1	80000	1	1	80000
2	380000	2	1	300000

Банк	Потери в результате дефолта	Число дефолтов	Число связей	Объем операций
3	400000	2	2	320000
4	0	0	0	0

Итак, мы видим, что дефолт Банка №3 вызвал наибольшие потери в сети. Однако Банк №2 хоть и привел к чуть меньшим потерям, но характеризовался меньшим числом связей с банками, что свидетельствует о более явном «эффекте домино» в случае его дефолта.

Рассмотрим еще один пример.

Пример 3. Имеется следующая банковская сеть.



Рис. 5. Банковская сеть для Примера 3

Исходные данные по банкам выглядят так.

Таблица 5. Финансовые показатели банков для Примера 3

Банк	Капитал	Н1
5	200000	24%
6	5000000	32%
7	760000	19%

Рассматриваем ситуации дефолта у всех банков по очереди.

1) Пусть дефолт объявил Банк №5.

Таблица 5.1. Потери банков в условиях Примера 3 в случае дефолта Банка №5

Банк	Потери	Н1	Потери	Н1
6	0	32,0%	1010000	25,5%
7	400000	9,0%	x	x

В результате дефолта Банка №5 не сможет выполнять обязательства и Банк №7 (см. схему).

2) Пусть дефолт объявил Банк №6.

Таблица 5.2. Потери банков в условиях Примера 3 в случае дефолта Банка №6

Банк	Потери	Н1	Потери	Н1
5	150000	6,0%	x	x
7	0	19,0%	400000	9,0%

Если объявляет дефолт Банк №6, то вся сеть перестает существовать, так как остальные два банка в таком случае также нарушают норматив достаточности собственных средств (капитала).

3) Пусть дефолт объявил Банк №7.

Таблица 5.3. Потери банков в условиях Примера 3 в случае дефолта Банка №7

Банк	Потери	Н1
5	0	24,0%
6	1010000	27,3%

Дефолт Банка №7 сам по себе не вызывает других дефолтов.

По итогам рассмотрения всех ситуаций получаем такую картину (см. таблицу 6).

В данной сети, как уже упоминалось, самым уязвимым является Банк №6, так как в результате его дефолта дефолт объявят и два других банка. Отметим, что дефолт Банка №5 приводит также к значительной величине потерь.

Таблица 6. Итог по Примеру 3

Банк	Потери в результате дефолта	Число дефолтов	Число связей	Объем операций
5	1410000	1	1	400000
6	1560000	2	1	150000
7	1010000	0	1	1010000

Таким образом, на простых примерах можно убедиться, что алгоритм позволяет смоделировать «эффект домино» в банковской сети, опираясь на соответствие нормативу достаточности собственных средств (капитала). Здесь же можно убедиться, является ли банк системно значимым или нет, в зависимости от того, к какой величине потерь приводит его дефолт.

Возникающий в отрасли кризис приводит к невыплатам компаний по обязательствам. Располагаемые средства банков сокращаются, риски растут. Поскольку банки, кредитующие одну отрасль, зачастую связаны между собой межбанковскими кредитами, то кризис отрасли провоцирует приостановки выплат по обязательствам самих банков, что, в конечном итоге, приводит к «эффекту домино» и усилению кризиса не среди компаний отрасли, но и среди банков, кредитующих эту отрасль.

6. Заключение и перспективы

Итак, нами были рассмотрены подходы к оценке вероятных потерь банков в результате стрессовых ситуаций, в том числе ситуации множественного дефолта. Общий подход основан на разделении банков согласно кредитуемым ими отраслям и использовании норматива достаточности собственных средств (капитала) в качестве индикатора стабильности как банков, так и соответствующей отрасли, что было подтверждено регрессионным анализом. На основе полученных предпосылок была поставлена, формализована с помощью logit-модели и исследована на существование и единственность решения задача нахождения оптимальной ссуды в отрасль, то есть такой, при кото-

рой компании не испытывают недостатка средств, а банки не несут существенных потерь. Было показано, что теория оптимального управления является эффективным инструментом для принятия решений о величине ссуды. В продолжении темы оценки потерь был проанализирован алгоритм нахождения вектора Сноу, позволяющий оценивать системную значимость банков на основе тестов-ситуаций множественного дефолта, а также абсолютную величину потерь всей системы в результате невыплаты по обязательствам отдельных банков. Изложенный подход к оценке потерь и оптимальных ссуд может быть усовершенствован в дальнейшем путем дополнительного учета межотраслевых балансов, т.е. добавления условия на связи между отраслями и возможностей распространения кризиса от одной отрасли к другой.

Литература

1. *Бюллетень банковской статистики* // ЦБ РФ. – 2008–2013. – С. 35-68 с.
2. ГАЛЕЕВ Э.М., ТИХОМИРОВ В.М. *Оптимизация*. – М.: «Эдиториал УРСС», 2000. – 312 с.
3. *Инструкция Центрального Банка РФ «Об обязательных нормативах банков» №139-И от 3 декабря 2012 года*. – 33 с.
4. КЕНДАЛЛ М., СТЬЮАРД А. *Статистические выводы и связи*. – М.: Наука, 1973. – 350 с.
5. КРАМЕР Г. *Математические методы статистики* – М.: Мир, 1975. – 648 с.
6. ЛОБАНОВ А.А., ЧУГУНОВА А.В. *Энциклопедия финансового риск-менеджмента*. – М.: «Альпина Паблишер», 2003. – 760 с.
7. МАГНУС Я.Р., КАТЫШЕВ П.К., ПЕРЕСЕЦКИЙ А.А. *Эконометрика. Начальный курс*. – М.: «Дело», 2004. – 575 с.
8. МИЛЮТИН А.А., ДМИТРУК А.В., ОСМОЛОВСКИЙ Н.П. *Принцип максимума в оптимальном управлении*. – М.: Механико-математический факультет МГУ, 2004. – 256 с.

9. ПОЛОВИНКИН Е.С., БАЛАШОВ М.В. *Элементы выпуклого и сильно выпуклого анализа*. – М.: «Физматлит», 2004. – 415 с.
10. СТЕЖКИН А.А. *Применение теории оптимального управления к анализу ситуаций множественного дефолта в сделках через центрального контрагента* // Финансовая аналитика: проблемы и решения. – 2013. – №22(160). – С. 52–60.
11. ALTMAN E. *Corporate Financial Distress and Bankruptcy*. – John Wiley and Sons, 2005. – 33 p.
12. BLACK F., SCHOLES M. *The pricing of options and corporate liabilities*». – The Journal of Political Economy. – 1973. – Vol. 81, №3. – P. 637–654.
13. BLUHM M., KRAHNEN J. *Default Risk in an Interconnected Banking System with Endogeneous Asset Markets*. Goethe Universität, 2010. – 31 p.
14. CAO Z. *Multi-CoVaR and Shapley value: A Systemic Risk Measure*. – Manchester University, 2013. – 45 p.
15. GREENE W. *Econometric Analysis*. – New York University, 2003. – 51 p.
16. HUGHES J. *Efficiency in Banking: Theory, Practice, and Evidence*. – The Wharton School, 2008. – 30 p.
17. JONES E.P., MASON S.P., ROSENFELD E. *Contingent Claims Analysis of Corporate Capital Structure: An Empirical Investigation*. – Journal of Finance. – 1984. – №39. – P. 611–625.
18. KUROSAKI T., KIM Y.S. *Mean-CoAVaR optimization for global banking portfolios*. – Investment Management and Financial Innovations. – 2013. – Vol. 10, Issue 2. – P. 15–20.
19. WOOLBRIDGE J. *Econometric Analysis of Cross Section and Panel Data*. – Cambridge: MIT Press, 2002. – 737 p.

BANKING NETWORK STABILITY MANAGEMENT TAKING INTO ACCOUNT INDUSTRY-SPECIFIC RISKS

Alexander Stezhkin, Moscow Institute of Physics and Technology, Moscow, student (alex79216@mail.ru).

Abstract: We consider an approach to loss evaluation in banks under various scenarios of market crash, which is based on a classification of banks with respect to the dominant industry sector credited. We formulate and solve the problem of optimal loan size to each industrial sector, and the probability minimization problem of a loan default in the case of a crisis in a branch of industry under constraints of capital adequacy ratio and a sufficient level of current assets. A model is suggested of possible loss evaluation for stress situations in a banking sector and a routine for banks ranking based on Snow vector approach; several examples are considered. All approaches considered in this paper were not applied to the Russian banking sector before, as the problem of systemic risk evaluation emerged recently as a result of changes in international regulations.

Keywords: multiple default, optimization problem, logit-model, capital adequacy ratio.

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

*Поступила в редакцию 26.06.2013.
Опубликована 30.09.2013.*

УДК 519.816
ББК 60.050.9(2)

ИСПОЛЬЗОВАНИЕ ДВУХ СЕТЕЦЕНТРИЧЕСКИХ ПРИНЦИПОВ МОДИФИКАЦИИ ЭКОНОМИЧЕСКИХ ЦЕЛЕЙ И СТРАТЕГИЙ В КРИЗИСНОЙ СИТУАЦИИ

Трахтенгерц Э. А.¹

*(ФГБУН Институт проблем управления
им. В.А. Трапезникова РАН, Москва)*

Рассматриваются два принципа сетецентрической концепции – сосредоточения усилий и максимального использования информации, которые могут быть применены при модификации экономических целей и стратегий в кризисной ситуации

Ключевые слова: цели, стратегии, сетецентрические методы, сетецентрические принципы, кризисные ситуации, сосредоточение усилий, использование информации.

1. Введение

Слегка перефразируя известную поговорку, можно сказать, что «благими целями вымощена дорога в ад». Это в первую очередь относится к целям, генерируемым или модифицируемым в кризисных ситуациях, а также стратегиям (сценариям) их реализации. В настоящее время темпы конкурентных и технологических изменений настолько стремительны, что востребованность новых или модифицированных экономических целей и стратегий функционирования фирмы может быть вызвана как катастрофами и рисками, так и быстроменяющейся обстановкой во внешнем мире (появившимися новыми продуктами на рынке, новыми технологиями, изменяющимися качеством и цену традиционных товаров, изменени-

¹ Эдуард Анатольевич Трахтенгерц, профессор, доктор технических наук, заслуженный деятель науки РФ, главный научный сотрудник (Москва, ул. Профсоюзная, д. 65, тел. (495) 334-88-40, E-mail: tracht@ipu.rssi.ru).

ем законодательства и т.п.). Модификация экономических целей и стратегий при резком изменении обстановки, как правило, принимается в условиях возникновения новых, часто еще недостаточно определенных ситуациях, иногда при отсутствии устоявшихся методов их генерации, четких критериев их оценки и дефиците времени принятия решения.

Важно отметить, что дефицит времени в данном случае не означает, что в поиске решения не надо тратить времени на анализ обстановки и формирование новых или модифицированных целей и стратегий различного типа. Обычно в таких случаях быстро реализуют оперативные решения, а затем в процессе ликвидации кризисных явлений ищут методы выхода из кризиса. Так было во время «великой депрессии» в США, мирового кризиса 2007–2009 г.г. и других экономических потрясений. В этих условиях, наряду с принятием оперативных воздействий, целесообразно помимо экспертных предложений по модификации экономических целей и стратегий воспользоваться сетецентрическими принципами.

Поскольку сетецентрическая концепция – понятие относительно новое, необходимо дать комментарий. Она появилась в военной среде и широко используется главным образом в военной литературе. Концепция возникла в результате развития технических и информационных возможностей, позволивших объединить все имеющиеся боевые, вспомогательные и даже невоенные средства в единую распределенную надежную сетецентрическую систему. Сетецентрическая система включает в себя все уровни и направления управления. Главное содержание сетецентрической концепции заключается не в новых видах и формах ведения военных действий, а в изменении способа управления войсками [5].

Концепция была введена во второй половине 90-х годов [17], вызвала не только широкую дискуссию, но и легла в основу программ развития и совершенствования вооруженных сил ряда стран, в том числе США и Китая. Она позволила впервые в истории военного искусства преодолеть пространственный, временной и информационный разрыв между войсками и органами управления [12]. Поскольку между управлением военными действиями и руководством преодоления кризисных ситуаций

можно провести аналогию, использование двух сетецентрических принципов – сосредоточения усилий для выполнения поставленной задачи и максимального использования всей информации для формирования решений – в нашем случае позволяет расширить и сделать более эффективными методы компьютерной модификации экономических целей и стратегий.

ПОСТАНОВКА ЗАДАЧИ

Цель работы – попытка наряду с традиционными методами использовать некоторые идеи сетецентрических методов при модификации экономических целей и стратегий в условиях резкого изменения обстановки. В соответствии с сетецентрической концепцией подготовка к возможной модификации целей и стратегий должна начинаться по возможности задолго до возникновения кризиса. Однако опыт показывает, что как ни готовься к кризисам, «зима всегда наступает неожиданно». Одним из средств борьбы с последствиями наступившего кризиса является модификация экономических целей и стратегий. В работе основное внимание уделяется особенностям сетецентрических методов, осуществляющих принципы сосредоточения усилий и широкого использования всей доступной информации. Рассматриваются:

- компьютерный мониторинг с анализом состояния фирмы;
- использование сетецентрических принципов максимальной реализации доступной информации;
- анализ причин, требующих модификации целей и стратегий;
- проверка эффективности сосредоточения усилий для реализации модифицированных экономических целей и стратегий;
- использование сетецентрических принципов при модификации и генерации целей и стратегий, позволяющих ликвидировать или ослабить последствия кризисов.

Работа не претендует на охват всех аспектов модификации экономических целей и стратегий. В ней рассматриваются только компьютерные методы их формирования в свете сетецентрических принципов сосредоточения усилий и максимального использования информации.

2. Особенности сетецентрических методов мониторинга и анализа изменения экономической обстановки

Мониторинг – это систематическое накопление и обработка данных о состоянии и динамике изменения параметров анализируемого объекта или процесса и представление результатов в удобном для руководителя или эксперта виде. Задачей мониторинга при комплексном управлении ликвидацией последствий различного вида кризисов является сбор и анализ данных о нанесенном ущербе и характере произведенных воздействий на общественное мнение.

Особенности сетецентрической системы мониторинга в процессе модификации целей и стратегий заключаются в:

- анализе всех видов информации: внутренней, архивной, от других организаций и полученной из СМИ;
- предоставлении информации, обеспечивающей сосредоточение усилий всех имеющихся в распоряжении руководства возможностей;
- доведении всей необходимой информации по мере ее получения до адресатов в реальном времени или близком к нему.

Такой подход позволяет резко улучшить понимание сложившейся ситуации руководителями всех степеней и повысить уровень их взаимодействия. Использование такого диапазона данных изменяет логику анализа сложившейся обстановки.

2.1. ОПРЕДЕЛЕНИЕ МОМЕНТА РЕЗКОГО ИЗМЕНЕНИЯ ОБСТАНОВКИ

Необходимость модификации целей и стратегий может произойти от резкого изменения обстановки. Поэтому одной из важнейших задач мониторинга является определение начала такого изменения.

В каждой области – экономике, социологии, метеорологии и др. существуют константы нормального состояния: средняя температура на этот день года, средняя доходность акций фирмы за последние годы, количество ежемесячно продаваемой продукции данного вида и т.д. В процессе компьютерного мониторинга отслеживаются показатели, являющиеся в данной

области важными, и в случае отклонения их от нормы определяется необходимость модификации целей и стратегий. Например, на рынке появилась продукция с качественно новыми характеристиками, которые могут оказать большое влияние на продажу некоторых наборов товаров, или возникло большое количество конкурентов. Они могут появиться в самых инновационных обществах.

Одна из задач мониторинга – определение момента изменения тенденции анализируемого процесса, который может потребовать модификации управленческих решений. В приведенном ниже конкретном примере – это резкое увеличение производства вновь появившимися конкурентами сапфировых подложек. Российская фирма «Монокристалл» выращивает синтетический сапфир и делает из него специальные подложки, используемые в производстве светодиодов. В своем сегменте она является мировым лидером и занимает 28% мирового рынка этого вида продукции. Объемы продукции, выручка и доходы росли. Увидев перспективное промышленное направление, Китай поспешно вложил только за один год миллиарды долларов в строительство сапфировых заводов. В результате сапфиры подешевели в 5 раз. Акции крупнейшей в этой области американской фирмы Rubicon Technology подешевели в 4,5 раза. Одним из выходов из возникшего в этой области кризиса фирма «Монокристалл» видит в производстве пластин диаметром 8–10 дюймов для бурного развивающегося рынка смартфонов (сейчас ходовой размер 2 дюйма), т.е. в модификации одной из целей, реализация которой потребовала частичной модернизации производства [8].

Рассмотрим несколько простых методов, определяющих момент изменения ситуации по соотношениям (1) и (2) [15].

Пусть в базе данных на момент t хранятся данные $x(t)$ (например, величины продаж некоторого товара) за интервалы времени τ , фиксированные в N моментов времени. Определим скользящее среднее, т.е. среднее значение x за последнее τ моментов, считая для простоты, что интервалы между фиксируемыми моментами времени равны:

$$(1) \quad M_{\tau}(t) = \frac{1}{\tau} \sum_{i=t-\tau}^{t-1} x(i). \quad t = \tau + 1, \dots, N.$$

Легко показать, что если значения $x(t)$ возрастают, то $M_\tau(t) < x(t)$, а если убывают, то $M_\tau(t) > x(t)$. Перемена в отношении $M_\tau(t)$ к $x(t)$ и показывает изменение тенденций анализируемого процесса, демонстрируемое компьютерной системой руководителям и экспертам.

Аналогичным инструментом анализа может являться «момент фондов» на интервале τ . Он определяется по соотношению

$$R_\tau(t) = \frac{x(t) - x(t - \tau)}{\tau}; \quad t = \tau + 1, \dots, N,$$

скользящее среднее моментов определим в виде

$$R_\tau^*(t) = \sum_{i=t}^{t+\tau-1} \frac{x(i) - x(i - \tau)}{\tau}, \quad t = \tau + 1, \dots, N.$$

Экономические системы помимо цены или изменчивости характеризуются также физическими массами различного рода запасов. Пусть величина $V(t)$ – объем некоторого запаса с ценой $x(t)$ в момент t . Обобщенный момент F_τ на интервале τ , как и в физике, может быть определен в виде

$$(2) \quad F_\tau(t) = \frac{V(t)}{(V(t))_\tau} \times \frac{x(t) - x(t - \tau)}{\tau} = m(t) \frac{\Delta x}{\Delta t}, \quad t = \tau + 1, \dots, N,$$

где $(V(t))_\tau = \sum_{i=t}^{\tau} V(i)$.

То есть вводится обобщенная масса $m(t)$, зависящая от времени. Анализ изменения во времени запасов (обобщенных масс) аналогично анализу функций (1) и (2) может помочь определению необходимости изменения целей и стратегий.

Таким образом, система мониторинга определяет не только состояние объекта или процесса в момент наблюдения, но и изменения параметров за заданные промежутки времени, а также тренд их развития. Есть, конечно, более глубокие методы.

Пусть на основании анализа мониторинга этими или другими методами установлено, что увеличился импорт продукции, аналогичной выпускаемой фирмой ААА, которая оказалась привлекательнее или ее оказалось очень много, и спрос на продукцию фирмы ААА резко упал. Эта фирма в дальнейшем будет рассматриваться в качестве иллюстративного примера. Такой

фирмой могла быть и упоминавшаяся выше фирма «Монокристалл», но мы не будем рассматривать конкретную фирму.

2.2. СЕТЕЦЕНТРИЧЕСКИЙ ПОДХОД К АНАЛИЗУ СОСТОЯНИЯ ФИРМ, ПОДВЕРГШИХСЯ КРИЗИСНЫМ ВОЗДЕЙСТВИЯМ

Одна из задач сетецентрического анализа заключается еще и в том, чтобы определить изменение состояний не только фирмы AAA, но и фирм, характеристики которых аналогичны характеристикам этой фирмы. В процессе мониторинга при сетецентрическом подходе в докризисный период помимо анализа текущего состояния система должна проанализировать архивные данные и найти условия, характеризующие ситуации, аналогичные потребовавшим в настоящее время принятия незапланированных решений. Источники информации могут касаться самых различных областей, связанных с особенностями управляемого процесса или объекта. На основании мониторинга архивных данных компьютерная система ищет фирмы, сумевшие преодолеть трудности, вызванные резкими изменениями ситуации, аналогичной текущей. Характеристики этих фирм должны быть аналогичны характеристикам фирмы, попавшей в кризисную ситуацию, по характеру деятельности (производства), объему выпускаемой продукции, численности сотрудников (возможно по категориям), объему оборотных средств, величине уставного капитала и т.д. Компьютерная система должна проанализировать данные по официальным отчетам этих фирм, текущей обстановке того времени и проанализировать прессу (пока автоматически это сделать нельзя).

Будем считать, что по архивным данным компьютерная система определила, что в недалеком прошлом три фирмы BCD, EFG и HIJ, производившие продукцию, по своей номенклатуре близкую к продукции фирмы AAA, оказались в кризисном положении, аналогичном этой фирме в текущий момент. Они сумели преодолеть кризис, правда, с разным уровнем успеха.

Анализ состояния фирм, как всегда, начнем с определения набора критериев, по которым оно будет оцениваться. Выбор набора критериев, да еще для сравнительной оценки функционирования фирм в прошлом и настоящем, достаточно сложен,

поэтому эксперты решили воспользоваться каким-либо известным набором критериев. Используя компьютерную процедуру голосования, эксперты выбрали так называемые «пять сил Портера» [18] плюс оценка средств массовой информации. Методика Портера для анализа положения и выработки стратегии бизнеса была разработана в 1979 г. За прошедшие годы резко возросла роль СМИ, поэтому дополнительно они включены в качестве «шестой силы». Достоинство «пяти сил Портера» заключается в том, что для части «сил» (критериев) можно определить их реальные значения в анализируемый период, и субъективизм экспертов проявляется только при переходе к лингвистическим или балльным оценкам значений критериев.

Рассмотрим возможные методы определения значений критериев Портера, которые в дальнейшем могут быть использованы для определения близости фирм BCD, EFG и HIJ в прошлой ситуации к текущей обстановке фирмы AAA [3], а значит, могут быть использованы для модификации целей и стратегий.

1. Угроза появления товаров-заменителей. Оценка по этому критерию может быть дана по архивным данным и по текущему сравнению цен и/или качества традиционного продукта и его заменителей по анализу информации о ведущихся новых разработках и торговой политике этих стран. При этом необходимо оценивать стоимость перехода на товар-заменитель.

2. Угроза появления новых игроков. Система анализирует политику игроков в других странах и возникающих игроков на своем рынке. Барьеры их входа определяются законодательством, условиями защиты старых игроков, тарифной политикой, стартовыми затратами новых игроков, доступом к каналам распределения и т.п. По параметрам критериев 1 и 2 специалисты обычно используют общепринятые оценки.

3. Рыночная власть поставщиков. Оценка по 3-му критерию можно делать по различным показателям, например, по функции совокупного предложения производственного сектора [1]:

$$(3) \quad \psi_0(p) = \left\{ y^0 : y^0 = \sum_{s=1}^S y^s, \quad y^s = \psi_s(p), \quad s = \overline{1, S}, \quad y^0 \in Y \right\},$$

где y^s – предложение s -го производителя; $\psi_s(p)$ – функция предложения продукции фирмами S , т.е. фирмами выпускающими

продукцию, аналогичную продукции анализируемой фирмы по ценам p ; ψ_s – технологическое множество товаров (объемы, цены, номенклатура), выпускаемые фирмой s .

Каждый s -й производитель характеризуется технологическим множеством y^s . Вектор $y^s \in Y$ называется общеэкономическим вектором затрат выпуска. Параметры функции (3) могут достаточно резко меняться. Так в примере с сапфировыми подложками все эти параметры резко поменялись в короткий срок, отчего поменялись и значения в формуле (3).

4. Рыночная власть потребителей. Она может быть оценена функциями спроса и предложения.

Функцию многозначного спроса на продукцию фирмы при векторе цен p на продукцию фирмы можно представить в виде [1]

$$(4) \quad \Phi(p) = \left\{ x : x = \sum_{j=1}^J x^j, x^j \in \Phi_j(p), j = \overline{1, J} \right\},$$

где x^j – спрос на товары при цене p со стороны j -го потребителя может быть оценен по архивным данным и текущей статистике; Φ_j – технологическое множество товаров, потребляемое фирмой j .

В процессе анализа архивов компьютерная система может определить значения аргументов и функций (3) и (4).

5. Уровень конкуренции. Она может оцениваться, например, размером прибыли по соотношению [1]

$$(5) \quad \prod(\bar{x}) = \bar{p}_g f(\bar{x}) - \bar{p}\bar{x} \rightarrow \max, \quad ,$$

где p_g – вектор цен производимых фирмой товаров; \bar{x} – вектор затрат; \bar{p} – вектор цен факторов производства; $\bar{p}_g f(\bar{x})$ – доход; $\bar{p}\bar{x}$ – издержки производства.

Данные для оценки 3–5 критериев могут быть получены по архивной и текущей статистике.

6. Компьютерная оценка воздействий на фирму средствами массовой информации. Она достаточно специфична и подробно рассмотрена, например, в [9], поэтому здесь методы ее оценки не рассматриваются.

Теперь система должна оценить близость условий функционирования различных фирм в прошлом к условиям функ-

ционирования фирмы (в нашем случае фирмы ААА) после того как произошло резкое изменение обстановки. Для этого система высвечивает на дисплее экспертов данные, полученные по критериям Портера для фирмы ААА, находящейся в кризисной ситуации, и трех фирм – BCD, EFG и HIJ, обнаруженных в процессе архивного поиска до того как они оказались в кризисе. Эти три фирмы некоторое время тому назад оказались в кризисной ситуации, по своим параметрам близкой к текущей, и сумели выйти из нее с разной степенью успеха. Фирма ААА в тот период проблем со сбытом не испытывала.

Каждый эксперт поставил в таблице 1 на своем дисплее субъективные лингвистические оценки значений критериев, используя оценки критериев, полученные по алгоритмам, рассмотренным выше. Затем система согласовала оценки экспертов, используя один из алгоритмов согласования, например, рассмотренных в [10]. Будем считать, что в таблице 1 показаны уже согласованные оценки. Теперь надо определить фирмы, значения критериев которых наиболее близки фирме ААА.

Естественно, можно (а при оценке реальной ситуации нужно) ввести более тонкие оценки, возможно, с учетом отношений спроса и предложения на каждый вид товара. Поскольку каждая оценка таблице 1 должна учитывать несколько факторов, влияющих на каждую «силу», для ее оценки можно использовать средневзвешенные значения [7]:

$$(6) \quad h_i = \frac{1}{m \times n} \sum_{i=1}^m k_i \sum_{j=1}^n c_{ij},$$

где c_{ij} – балльная оценка j -го эксперта значения i -го фактора, влияющего на оценку; k_i – коэффициент важности i -го фактора; m – число рассматриваемых факторов; n – количество экспертов.

Перевод значений h_j формулы (6) в лексические или балльные оценки таблицы 1 компьютерная система может сделать по заранее составленной таблице (в работе не показана).

Таблица 1.

№№ и наименование критериев	Значения критериев фирм в докризисный период			
	фирмы AAA	фирмы BCD	фирмы EFG	фирмы HIJ
1. Угроза появления продуктов-заменителей	отсутствует (5)	отсутствовала (5)	была небольшая (4)	существовала (3)
2. Угроза появления новых игроков	незначительна (4)	существовала (3)	сильная (2)	существовала (3)
3. Рыночная власть поставщиков	не повышают цены (5)	не повышали цены (5)	незначительно повышали цены (4)	незначительно повышали цены (4)
4. Рыночная власть потребителей с учетом формул (2.3) и (2.4)	снижают цены (3)	снижают цены незначительно (4)	снижают цены (3)	сильно снижают цены (2)
5. Уровень конкуренции (размер прибыли по формуле (2.5))	высокий (4)	средний (3)	средний (3)	низкий (2)
6. Оценка средств массовой информации	положительная (4)	положительная (4)	положительная (4)	резко критическая (2)

Примечание 1. Оценки по 4-му критерию давались в зависимости от следующих соотношений: $\Phi(\bar{p}) \gg \Psi(\bar{p})$ – отлично;

$\Phi(\bar{p}) \approx \Psi(\bar{p})$ – хорошо; $\Phi(\bar{p}) < \Psi(\bar{p})$ – удовлетворительно;

$\Phi(\bar{p}) \ll \Psi(\bar{p})$ – плохо. 2. Оценки по 5-му критерию могут производиться, исходя из отношения величины прибыли, определяемой соотношением (5) к величине капитализации фирмы. Шкалу лингвистических значений этих оценок определяют эксперты.

Сравнивая значения параметров работы анализируемой фирмы в текущий момент со значениями этих же параметров для фирм, оказавшихся в похожей кризисной ситуации в прошлом, система определяет, насколько похожи положения этих фирм. В сложных случаях модификации целей и стратегий, например, при крупном пожаре, неожиданно возникшем дефиците или избытке товаров, цунами и т.п., как правило, одного какого-либо признака, например, пламени при пожаре, увеличения объ-

ема нераспроданной продукции бывает недостаточно для принятия мер по ликвидации пожара или исправления экономической ситуации. В тех случаях, когда сравниваемых параметров немного, можно просто сравнивать значения характеристик анализируемого события или процесса в прошлом с характеристиками процесса или события в настоящий момент. При этом часть параметров может совпадать полностью, например, типы сравниваемых пожаров, характеристики их мест, а, скажем, площади пожаров могут значительно отличаться.

Если руководитель или эксперт хочет сопоставить, например, изменения во внешней среде крупной корпорации в некоторые моменты прошлого с настоящим, то ему придется сравнивать около 30 параметров [2]. Поскольку список критериев, приводимых в экономических работах и в работах по другим областям деятельности, обычно достаточно велик, анализ и сравнение ситуаций в прошлом и настоящем становится непростой задачей. Поэтому в общем случае система управления высвечивает на дисплеях экспертов и руководителей список критериев и просит отметить те, которые, по их мнению, могут быть использованы для сравнения ситуаций. Список может дополняться экспертами. Затем система по одному из алгоритмов, приведенных, например в [10], согласует список критериев и представляет его результат на утверждение руководителю. В нашем иллюстративном примере такой список представлен в таблицы 1. Затем, например, методом распознавания образов [6] система определяет близость сложившейся ситуации с бывшими ранее.

Для определения меры близости между текущей ситуацией и ситуацией, анализируемой по архивным данным, в n -мерном векторном пространстве значений критериев вводится метрика, удовлетворяющая обычным аксиомам расстояний: среднеквадратичное расстояние между векторами значений критериев текущей и архивными ситуациями:

$$L(w, W_i) = \sqrt{\sum_{j=1}^n k_j (x^j - x_i^j)^2},$$

где k_j – «вес» j -го критерия; x^j – значение j -го критерия фирмы ААА, а x_i^j – значение j -го критерия i -й фирмы, в нашем примере одной из трех перечисленных выше фирм.

Наиболее близкой считается одна из перечисленных в таблице 1 фирм W_i , $i = 1, \dots, m$, для которой

$$(7) \quad L(w, W_i) = \min_{i \in I} L(w, W_i),$$

где I – множество фирм W_i , показатели которой аналогичны текущим показателям фирмы w .

Для определения «весов» критериев система просит экспертов дать свои оценки «веса», используя, например, один из методов, рассмотренных в [10] или какой-нибудь другой по каждому критерию таблицы 1. Будем считать, что результаты обработки данных показали, что по оценкам экспертов значения «весов» для всех критериев близки, поэтому $k_j = 1, j = 1, \dots, n$.

Теперь система определяет меру близости между значениями критериев Портера фирмы ААА и фирмами, данные о которых взяты из архивов.

$$L(w, W_1) =$$

$$= \sqrt{(5-5)^2 + (4-3)^2 + (5-5)^2 + (3-4)^2 + (4-3)^2 + (4-3)^2} = 1,7,$$

$$L(w, W_2) =$$

$$= \sqrt{(5-4)^2 + (4-2)^2 + (5-4)^2 + (3-3)^2 + (4-4)^2 + (4-3)^2} = 2,6,$$

$$L(w, W_3) =$$

$$= \sqrt{(5-3)^2 + (4-3)^2 + (5-4)^2 + (3-2)^2 + (4-2)^2 + (4-2)^2} = 3,9,$$

где W_1, W_2, W_3 – вектора значений критериев таблицы 1 для фирм BCD, EFG и HIJ.

Таким образом, в соответствии с формулой (7) самыми близкими оказались фирмы ААА и BCD, самые большие различия – у фирм ААА и HIJ. Поэтому одним из источников данных для формирования цели фирмы ААА может стать фирма BCD, но в дальнейшем для сравнения мы будем рассматривать также показатели фирм EFG и HIJ.

2.3. СЕТЕЦЕНТРИЧЕСКИЙ ПОДХОД К АНАЛИЗУ ПРИЧИН, ТРЕБУЮЩИХ МОДИФИКАЦИИ ЦЕЛЕЙ И СТРАТЕГИЙ

Ряд специалистов считает, что прежде чем формулировать цель модификации, целесообразно проанализировать причины ее необходимости. Эффект сосредоточения усилий, возникающий за счет взаимодействия отдельных факторов, показывает увеличение или уменьшение эффективности совместно проводимых действий. В нашем случае можно рассматривать как аналогию синергетического эффекта сосредоточения усилий. К этому принципу мы еще будем возвращаться. Анализ причин начнем с оценки эффекта сосредоточения усилий, вызванных причинами, требующими модификации использовавшихся целей. Сначала система формирует список причин необходимости модификации. Он берется из базы данных и при необходимости корректируется экспертами или формируется ими известными для формирования списков методами. Такой список представлен в таблице 2.

Компьютерная система представляет на дисплеи экспертов таблицу типа таблицы 2, в которой экспертов просят оценить попарное влияние (усиление или противоречие) причин необходимости модификации целей. Знаком «+» эксперт должен указать, что i -я причина, указанная в строке, увеличивает эффект сосредоточения усилий, вызванный причиной, указанной в j -м столбце, 0 – если эффекта нет и знак «-», если они противоречивы. Эти оценки могут быть большей размерности, например, пяти- или десятибалльными. Но формально методы таких оценок, дающих адекватные результаты предложить сложно, а правильность субъективных экспертных оценок априорно проверить крайне трудно, если это вообще возможно. Поэтому в данном случае кажется разумной достаточно грубая оценка по трехбалльной системе: «+», 0, «-», поскольку даже такая оценка позволяет ранжировать причины, требующие модификации целей по рассматриваемому критерию.

Таблица 2

Причины, требующие модификации использовавшихся целей до кризиса	№№ причин, требующих модификации целей и их оценки сосредоточения усилий									Баллы оценок
	1	2	3	4	5	6	7	8	9	
1. Проведение анализа для выбора направления модернизации	0	+	+	+	+	+	+	-	+	6
2. Техническое обоснование модернизации производства	+	0	+	+	+	+	+	-	+	6
3. Экономическое обоснование производства	+	+	0	+	+	+	+	+	-	6
4. Использование высокотехнологического оборудования	+	+	+	0	+	+	+	+	+	8
5. Концентрация сил и средств на относительно узком виде деятельности	+	+	+	+	0	+	+	-	+	6
6. Перераспределение средств	+	+	+	+	+	0	+	-	+	6
7. Выпуск более дешевой или более качественной продукции	+	+	+	+	+	+	0	+	+	8
8. Завоевание новых секторов рынка	-	-	-	+	-	-	+	0	+	-2
9. Борьба с конкурентами	+	+	+	+	+	+	+	+	0	8

Естественно, что оценки, представленные разными экспертами, могут не совпадать. Одним из вариантов согласования может быть голосование по принципу большинства: считается согласованным тот знак, который набрал большинство голосов экспертов. Возможны, конечно, другие правила. Например, считая знак «+» за «+1», а знак «-» за «-1», находить взвешенное среднее значение с учетом «веса» эксперта (если «веса» определялись). Каким бы методом эксперты не пользовались, будем считать, что согласование произведено и представлено в таблице 2.

Баллы каждой причины необходимости модификации цели могут оцениваться по формуле

$$(8) S_i = \sum_j \alpha_{ij},$$

где α_{ij} – значение знака, стоящего на пересечении i -й строки и j -го столбца:

$$\alpha_{ij} = \begin{cases} 1, & \text{если знак "+"}, \\ 0, & \text{если знак "0"}, \\ -1, & \text{если знак "-"} \end{cases}$$

Эксперты решили что, если сумма баллов причины в последнем столбце таблицы 2 отрицательная, то такую причину следует исключить из списка причин, требующих модификации целей, поскольку она не соответствует принципу сосредоточения усилий. Из таблицы 2 видно, что все причины, кроме причины 8, соответствуют сетецентрическому принципу сосредоточения усилий. Балл причины 8 «Завоевание новых секторов рынка» – отрицательный. На этом основании система предлагает исключить причину 8 из списка причин необходимости модификации целей и стратегий, т.е. не ориентировать фирму на захват новых рынков сбыта.

3. Использование сетецентрических принципов при модификации целей в условиях резкого изменения обстановки

Сетецентрическая концепция предполагает максимальное использование всей доступной информации, в частности, широкие горизонтальные связи между руководителями на всех уровнях и направлениях управления, а также доступность необходимой информации. Поэтому совершенно естественно, что руководство фирмы ААА, оказавшись в кризисной ситуации, захотело выяснить, как действовало и действует руководство фирм, близких по параметрам к фирме ААА, в аналогичных критических ситуациях.

3.1. ОЦЕНКА ЦЕЛЕСООБРАЗНОСТИ ИСПОЛЬЗОВАНИЯ ДОКРИЗИСНЫХ ЦЕЛЕЙ В КРИЗИСНОЙ СИТУАЦИИ

Исходя из сетецентрического принципа максимального использования информации, во многих случаях оказывается полезным установить, от каких целей, реализуемых раньше, стоит отказаться, а какие по-прежнему надо стремиться реализовать. Определение эффективности использования докризисных целей,

как всегда, начнем с формирования списка критериев оценки целей.

Выбор набора критериев зависит от многих внутренних и внешних для фирмы причин. Поэтому на различных этапах процесса компьютерной модификации целей и стратегий могут выбираться различные наборы критериев, определяющих, по мнению экспертов, эффективность принимаемых решений.

Список критериев зависит от характера деятельности фирмы. Так, в работе [2] рассматривается следующий список критериев оценки изменения во внешней среде, влияющий на крупную фирму: тенденции мирового рынка (протекционизм и свободная торговля); возрастание роли правительства в качестве заказчика; развитие «общего рынка»; деловые связи с другими странами; экономические и политические тенденции в развивающихся странах; валютные тенденции; инфляционные тенденции; политика транснациональных корпораций; технология как средство конкуренции; размер фирмы как средство конкуренции; достижение пределов роста; появление новых отраслей промышленности; технологические прорывы; рост сектора услуг; наличие богатых потребителей; изменение возрастного состава потребителей; сбыт продукции неподатливым потребителям; социальные установки в отношении бизнеса; государственный контроль; давление со стороны потребителей; отношения с профсоюзами; влияние требований охраны окружающей среды; сокращение жизненного цикла изделий; столкновение интересов транснациональных корпораций с национальными интересами; недоверие к бизнесу; сужение горизонта прогнозирования; неожиданности стратегического характера; конкуренция со стороны развивающихся стран; дефицит стратегических ресурсов; требование поддержания уровня занятости. Надо еще учитывать набор критериев, характеризующие внутреннее состояние фирмы. Он тоже достаточно большой [11].

Из приведенного перечня видно, что список таких критериев, приводимых в экономических работах, обычно достаточно велик, что, конечно, затрудняет оценки целей и стратегий, но для того чтобы продемонстрировать идею проверки целесообразности использования ранее реализуемых целей в новой об-

становке, в рассматриваемом иллюстративном примере будем обходиться всего несколькими критериями.

Компьютерная система высвечивает на дисплеях экспертов и руководителей извлеченный из базы данных список критериев оценки целей (приведенный выше или другой). Каждый эксперт и представитель руководства может его дополнить или вычеркнуть какие-то критерии. Система обрабатывает результаты действий экспертов и руководителей (например, в списке остаются только те критерии, с которыми согласны все) и представляет результат для утверждения руководителю. Если представленный список утверждается, он считается согласованным.

Будем считать, что результатами компьютерной процедуры выбора критериев, по которым будут оцениваться цели, стали: 1) специализация в тех областях деятельности, в которых фирма достигла лучших результатов; 2) процент изношенности оборудования; 3) квалификация сотрудников фирмы и их способности; 4) реакция средств массовой информации. Для определения эффективности цели компьютерная система высвечивает на дисплеях экспертов и руководителей таблицу типа таблицы 3.1 и просит проставить экспертные значения критериев для каждой цели, которые фирма ААА стремилась реализовать до кризиса. Оценки необходимо дать для двух вариантов использования: реализовать цель в создавшихся новых условиях или отказаться от нее. После того как оценки проставлены, компьютерная система согласовывает их одним из известных способов.

Для определения эффективности использования цели система применила метод Борда. По этому методу при сравнении объектов оценивается число объектов, которые превосходят данный. В нашем случае

$$r_k^u = \begin{cases} 1, & \text{если оценка использования цели } k\text{-му критерию} \\ & \text{превосходит оценку отказа от нее,} \\ 0, & \text{в противном случае;} \end{cases}$$

r_k^i имеет противоположные значения.

Затем подсчитываются суммы r^u и r^i – использования и не использования рассматриваемой цели в изменившейся обстановке:

$$r^l = \sum_{k=1}^4 r_k^l, \quad l = \text{и, н.}$$

Если $r^{\text{и}} > r^{\text{н}}$ – цель рекомендуется использовать, в противном случае – нет. В таблице 3 система проставляет значения $r_k^{\text{и}}$ и $r_k^{\text{н}}$ в соответствии с согласованными оценками, проставленными экспертами. В ней же показан результат оценки эффективности использования целей, реализовывавшихся фирмой до кризиса, в соответствии с согласованными оценками экспертов. На основании этих оценок система дает рекомендации по использованию целей (последний столбец таблицы 3).

Таким образом, компьютерная система не рекомендовала руководству фирмы использовать список целей в том виде, который фирма стремилась реализовать до кризиса. Поэтому руководство фирмы, оказавшись в кризисном состоянии, решило, что до принятия решения о модификации списка необходимо выяснить, какие цели для выхода из кризиса преследовали фирмы, оказавшиеся в аналогичном кризисе в прошлом и успешно вышедшие из него.

Таблица 3

Наименование целей	Целесообразность использования цели	Значения критериев				r^l	Рекомендация
		1	2	3	4		
Применение разработанных или освоенных технологий	не использовать	высокая (1)	снизится (1)	достаточная (1)	хорошая (1)	4	не использовать
	использовать	низкая (0)	высокий (0)	недостаточная (0)	отрицат. (0)	0	
Максимизация прибыли	не использовать	высокая (1)	удовл. (1)	удовл. (0)	средняя (1)	3	не использовать
	использовать	низкая (0)	увеличится (0)	удовл. (0)	отрицат. (0)	0	
Увеличение продаж	не использовать	удовл. (1)	не влияет (0)	хорошая (1)	средняя (0)	2	не использовать
	использовать	низкая (0)	не влияет (0)	низкая (0)	средняя (0)	0	

Примечание. В скобках проставлены значения $r_k^{\text{и}}$ и $r_k^{\text{н}}$.

3.2. МОДИФИКАЦИЯ ЦЕЛЕЙ С МАКСИМАЛЬНЫМ ИСПОЛЬЗОВАНИЕМ ДОСТУПНОЙ ИНФОРМАЦИИ

Затем система анализирует действия фирм, которые успешно вышли из похожих кризисных ситуаций. Эти оценки делаются на основании архивных данных о функционировании фирм, в частности, по значениям функций (3)–(6). Поскольку свои цели фирмы афишируют далеко не всегда, то цели фирм, сумевших выйти ранее из кризиса, система определяет по результатам их деятельности или по данным, предоставленным самими фирмами в порядке обмена горизонтальной информацией. Для этого компьютерная система выбирает из архивов данные, определяющие результаты деятельность фирм BCD, EFG и HIJ и/или использует данные, представленные руководством фирм. То есть фактически сравниваются результаты, достигнутые этими фирмами после кризиса, с цифровыми значениями критериев целей, которых хотело достигнуть руководство фирмы AAA до резкого изменения обстановки. Собранные данные система представляет экспертам и руководителям, предлагая дать балльные или лингвистические оценки по каждой цели и фирме. Полученные оценки от каждого эксперта сводятся в таблицу типа таблицы 4 вместе с докризисными целями фирмы AAA.

Таблица 4

№.№ п/п	Наименование целей	Цели фирмы AAA	Результаты фирмы BCD	Результаты фирмы EFG	Результаты фирмы HIJ
1.	% модернизации производства	0	90	40	0
2.	Балльная оценка применения разработанных или освоенных технологий	10	9	6	1
3.	% максимизации прибыли	10	8	3	2
4.	% увеличения продаж	5	6	4	1

Примечание. Оценка цели 2 дается по десятибалльной шкале.

Согласовывать значения оценок таблицы 4 за исключением цели 2 нет необходимости, так как это конкретные данные отчет-

тов и планов. Оценки строки 2 могут быть согласованы одним из известных методов.

На таком упрощенном примере можно «на глаз» определить результаты какой фирмы ближе к целям фирмы ААА. В реальности, когда показателей много, ранжирование степени близости показателей может оказаться сложной задачей.

Поэтому проведем формальное сравнение их близости, используя функцию $K(s_i, s_j)$, значение которой увеличивается по мере того, как значения показателей сближаются. Одна из таких функций имеет вид:

$$(9) \quad K(s_i, s_j) = \frac{2 \sum_{i=1}^I \min(x_{i1}, x_{i2})}{\sum_{i=1}^I x_{i1} + \sum_{i=1}^I x_{i2}},$$

где x_{i1}, x_{i2} – значения i -го показателя фирмы s_1 и s_2 соответственно.

В соответствии с таблицей 4 компьютерная система получает:

$$K(s_1, s_2) = \frac{2[\min(0,90) + \min(10,9) + \min(10,8) + \min(5,6)]}{(0+10+10+5) + (90+9+8+6)} = 0,319,$$

$$(10) \quad K(s_1, s_3) = \frac{2[\min(0,40) + \min(10,6) + \min(10,3) + \min(5,4)]}{(0+10+10+5) + (40+6+3+4)} = 0,333,$$

$$K(s_1, s_4) = \frac{2[\min(0,1) + \min(10,1) + \min(10,2) + \min(5,1)]}{(0+10+10+5) + (0+1+2+1)} = 0,1,$$

где s_1, s_2, s_3, s_4 – индексы фирм ААА, ВСD, ЕFG и НIJ.

Таблица 3 показывает, что после резкого изменения рыночной обстановки реализовывать цели, сформулированные фирмой ААА при изменившейся рыночной обстановке, без их модификации неэффективно. Из соотношений (10) видно, что хотя значения критериев оценки целей фирмы ААА ближе всего к значениям фирмы ВСD, но степень сходства небольшая. С другой стороны, по критериям Портера (таблица 1) фирмы ААА и ВСD находились в примерно одинаковых условиях. Система также определяет, по каким целям у фирм ААА и ВСD отличия самые существенные. В нашем примере – это цель «Модернизация производства», т.е. различия возникли после резкого изме-

нения обстановки. Исходя из этого, система проверяет необходимость реализации фирмой ААА этой цели.

Анализ цели «Модернизация производства» руководство предлагает начать с выяснения причин необходимости введения этой цели. Список причин приведен в таблице 2. В сетевых методах для сосредоточения усилий предполагается всесторонний анализ принятия решений с учетом всех имеющихся данных, в том числе и выяснение причин, требующих принятия данного решения. Список, если он есть в базе данных, высвечивается на дисплеях экспертов, а если он отсутствует – его составляют эксперты. Каждый эксперт имеет право вычеркнуть причину из списка или дополнить. В нашем случае будем использовать список причин таблицы 2.

Оценку соответствия причин необходимости модификации целей система предлагает произвести с помощью таблицы типа таблицы 5, используя все тот же метод Борда. Заметим, что эти критерии отличаются от критериев, по которым производились оценки в таблице 3, т.е. критерии могут меняться в зависимости от характера принимаемых решений.

В работе [14] утверждается, что фирмы и корпорации для оценки экономических целей и стратегий их реализации чаще всего используют следующие критерии:

- сбалансированность бизнеса;
- синергетика, т.е. достижение суммарной эффективности большей, чем сумма эффективностей каждой стратегии и/или отдельного подразделения или предприятия, входящего в фирму;
- компетентность коллектива, определяемая уровнем знаний и умений сотрудников фирмы;
- специализация в тех целях, стратегиях или областях деятельности, в которых фирма достигла наилучших результатов;
- обеспеченность реализации целей и стратегий, возможно, и фирмы необходимыми средствами;
- минимизация риска потерь или даже краха фирмы за счет диверсификации областей риска;
- рост капитализации фирмы.

Эти критерии могут быть использованы и при коррекции целей и стратегий.

В качестве критериев оценки соответствия причин необходимости модификации целей фирмы эксперты путем компьютерного голосования выбрали:

1. Минимизация потерь или даже краха фирмы за счет диверсификации областей риска.

2. Компетентность коллектива, определяемая уровнем знаний и умением сотрудников.

3. Обеспеченность реализации целей и стратегий необходимыми средствами.

4. Возможность использования СМИ (помимо прямой рекламы) продвижения своей продукции на рынок.

Последний критерий эксперты добавили.

Система предлагает экспертам заполнить таблицу 5, проставив знаки «+» или «-» в столбцах 3–6. Ответы экспертов отражают их отношение к причинам необходимости реализации этой цели по каждому критерию.

Таблица 5.

№№ и наименования причин необходимости введения цели «модернизация производства»	Наименование оценки	Соответствие причин критерию					Соответствие причин критериям модификации цели
		1	2	3	4	r	
1	2	3	4	5	6	7	8
1. Проведение анализа для выбора направления модернизации	соответствует	+	+	-	+	3	соответствует
	не соответствует	-	-	+	-	1	
2. Техническое обоснование модернизации производства	соответствует	+	-	+	+	3	соответствует
	не соответствует	-	+	-	-	1	
3. Экономическое обоснование производства	соответствует	+	+	+	+	4	соответствует
	не соответствует	-	-	-	-	0	
4. Расширение рынка сбыта	соответствует	+	-	+	+	3	соответствует
	не соответствует	-	+	-	-	14	
6. Концентрация сил и средств на узком виде деятельности	соответствует	+	+	+	-	3	соответствует
	соответствует	-	-	-	+	1	
6. Перераспределение средств	соответствует	+	+	+	+	4	соответствует
	не соответствует	-	-	-	-	0	

№№ и наименования причин необходимости введения цели «модернизация производства»	Наименование оценки	Соответствие причин критерию					Соответствие причин критериям модификации цели
		1	2	3	4	r	
1	2	3	4	5	6	7	8
7. Выпуск более дешевой и более качественной продукции	соответствует	+	-	+	+	3	соответствует
	не соответствует	-	+	-	-	1	
9. Борьба с конкурентами	соответствует	+	-	+	+	3	соответствует
	не соответствует	-	+	-	-	1	

Примечание. 1. Причина 8 в таблице 2 не рассматривается, так как она не соответствует принципу сосредоточения усилий. 2. В этой таблице знаки «+» и «-» означают то же, что в таблице 3 «1» и «0».

После заполнения таблицы экспертами система производит согласование знаков. Таблица 5 показывает, что цель «Модернизация производства» соответствует всем причинам необходимости модификации целей за исключением причины 8 таблицы 2. Так как эта причина не соответствует сетцентрическому требованию сосредоточения усилий, она в таблице 5. не рассматривается.

Корректировку оценок параметров система может сделать на любом этапе модификации, если система управления обнаружит расхождение получаемых результатов с планируемыми.

Компьютерная система согласовывает оценки столбца 8 таблицы 5, определяет возможную стоимость модернизации, имеющиеся финансовые ресурсы, состояние рынка и т.п. (здесь эти вопросы не рассматриваются), и предлагает модернизировать производство фирмы AAA на 80%. Тогда в соответствии с формулой (9) и скорректированными данными таблицы 4 (% модернизации фирмы AAA – 80%, а не 0) получаем:

$$k(s_i, s_j) = \frac{2[\min(80, 90) + \min(10, 9) + \min(10, 8 + \min(5, 6))]}{(80 + 10 + 10 + 5) + (90 + 9 + 8 + 6)} = 0,936, \quad (11)$$

$$k(s_i, s_j) = \frac{2[\min(80, 40) + \min(10, 6) + \min(10, 3) + \min(5, 4)]}{(80 + 10 + 10 + 5) + (40 + 6 + 3 + 4)}$$

где s_1, s_2, s_3 – индексы фирм AAA, BCD и EFG.

Теперь близость между значениями критериев фирм ААА и BCD очень высокая. Соотношение (11) показывает, что фирме ААА целесообразно ввести новую цель «Модернизацию производства». Значения параметров целей, которых хотела достигнуть фирма ААА до резкого изменения обстановки (таблица 3), близки к параметрам этих же целей фирмы BCD, успешно вышедшей из кризиса. Поэтому эти значения критериев целесообразно учитывать при модификации целей. Таким образом, опираясь на сетцентрический принцип использования всей имеющейся информации (в нашем случае о фирмах BCD и EFG) и анализ причин обеспечения концентрации усилий, компьютерная система получила рекомендации по модернизации целей. Система предлагает модифицировать список целей фирмы ААА, рассматриваемых в таблице 3 (до резкого изменения обстановки) путем включения в него цели «Модернизация производства».

Появление цели «Модернизация производства» – это возникновение нового, можно сказать, фундаментального проекта, очень важного для любой фирмы. То есть фирма ставит перед собой задачи, вызванные причинами модернизации, указанных во всех пунктах таблицы 5. Все они соответствуют принципу сосредоточения усилий.

3.3. СОСРЕДОТОЧЕНИЕ УСИЛИЙ ЦЕЛЕЙ

Сетцентрический принцип сосредоточения усилий требует концентрации усилий всех имеющих в распоряжении руководства средств на выполнение поставленной задачи за счет маневрирования и наиболее эффективного их использования. Задача сосредоточения усилий – получить максимальный синергетический эффект. Для этого из целей таблицы 4, которые используют все рассмотренные фирмы, система строит таблицу 6. Она строится с использованием формулы (8) по тому же алгоритму, что и таблица 2. По тем оценкам, которые дали эксперты в таблице 6, цель 2 «Применение разработанных и освоенных технологий» не отвечает требованию сосредоточения усилий. Но в ряде случаев к этой оценке надо подойти осторожно. Все зависит от процента модернизации оборудования. В примере разд. 2.2 фирма «Монокристалл» не отказалась полностью от

подложек для светодиодов малого размера, хотя спрос на них резко упал. Это решение объясняется сложившейся обстановкой: подложки больших размеров потребуются в ближайшем будущем. Сегодня предприятие вынуждено выпускать продукцию, которая стала пользоваться меньшим спросом, но все-таки продается.

Этот пример показывает, что результаты таблицы 6 требуют дополнительного анализа. В нашем случае цель 2 стала «уходящей» целью. Ее еще приходится реализовывать, но в отличие от, например, цели «Модернизация производства», она не цель развития, а скорее цель выживания. Тем не менее одной из задач формирования целей является установление объема производства, выпущенного в соответствии с целями 1 и 2 и объемом затрат, который необходимо при этом произвести.

Таблица 6.

№№ и наименование целей	№№ целей				Баллы оценок
	1	2	3	4	
1. Модернизация производства	0	0	+	+	2
2. Применение разработанных и освоенных технологий	0	0	–	–	-2
2. Максимизация прибыли	+	–	0	+	1
4. Увеличение продаж	+	–	+	0	1

Для этого в экономике введено понятие производственной функции ПФ (production function – PF), выражающей количественную зависимость выпуска продукции от затрат. ПФ применяют для анализа влияния различных сочетаний факторов на объем выпуска в определенный момент времени (статический вариант ПФ), а также прогнозирования соотношения характеров факторов и объемов выпуска в разные моменты времени (динамический вариант ПФ) на разных уровнях экономики [1]. Вид производственной функции, устанавливающей явную зависимость объема производства продукции от наличия или потребления ресурсов, называется функцией выпуска. Наряду с подобными ПФ используются обратные к ним функции производственных затрат. Они характеризуют зависимость затрат ресурсов от объема выпуска продукции. Частными случаями ПФ можно

считать функцию издержек (связь объема продукции и издержек производства), инвестиционную функцию (зависимость потребных капиталовложений от производственной мощности модернизируемого предприятия) и др.

Пусть Δx_i – приращение затрат на i -й вид ресурсов, $\Delta f(x) = f(x + e_i \Delta x_i) - f(x)$ – приращение выпуска продукции. Тогда отношение $\Delta f(x)/\Delta x_i$ показывает прирост выпуска продукции на единицу затрат i -го вида ресурсов.

Возникновение теории производственных функций принято относить к 1928 г., когда появилась статья американских ученых Д. Кобба и П. Дугласа «Теория производства», в которой (для конкретной задачи) они построили производственную функцию вида:

$$Q = AK^\alpha L^\beta,$$

где K – капитал; L – трудовые ресурсы; Q – объем выпущенной продукции; A, α, β – числовые параметры ($A > 0, \alpha, \beta \geq 0, \alpha + \beta = 1$). С тех пор функцию этого вида называют ПФ Кобба–Дугласа. Имеются и другие типы ПФ [1]. Используя различные виды производственных функций, компьютерная система может рассчитать параметры реализации 1-й и 2-й цели так, чтобы реализация этих двух целей минимально противоречила принципу сосредоточения усилий.

4. Использование сетцентрических принципов модернизации стратегий после резкого изменения обстановки

4.1. ОЦЕНКА СООТВЕТСТВИЯ СТРАТЕГИЙ ПРИНЦИПУ СОСРЕДОТОЧЕНИЯ УСИЛИЙ

При резком изменении обстановки возникают задачи генерации списка стратегий для новых целей и модификации списка стратегий для целей, сформированных и реализуемых до кризиса, но используемых в процессе ликвидации его последствий, если это необходимо. При выборе набора стратегий целесообразно оценить, насколько они соответствуют причинам, вызвавшим необходимость модификации стратегий, и насколько

эффективно будут соответствовать сетцентрическому принципу сосредоточения усилий.

Набор стратегий, принимаемых в повторяющихся кризисных ситуациях, для каждого типа ситуации достаточно хорошо известен. Фрагмент такого набора показан в списке 1. Из всего большого списка таких стратегий, высвечиваемого на дисплеях экспертов, компьютерная система просит указать те, которые соответствуют утвержденным причинам необходимости модификации целей и стратегий фирмы ААА и, если требуется, дополнить его.

Для этого на дисплее каждого эксперта выводится таблица типа таблицы 7, в которой №№ причин модернизации соответствуют номерам таблицы 5. Система предлагает экспертам проставить знак «+», если стратегия по данному критерию соответствует данной причине, и знак «-» в противном случае. Оценки экспертов могут не совпадать. Желательно, чтобы их согласование проходило по компьютерным алгоритмам без участия экспертов. Согласование знаков можно осуществлять методом компьютерного голосования, например, согласованным считается знак, поставленный в соответствующей клетке таблицы 7 большинством экспертов. Стратегия может считаться утвержденной, если она набрала в согласованной таблице типа 7 не менее половины знаков «+» по всем стратегиям (могут быть, конечно, и другие оценки).

Список 1.

1. Организационно-технологический аудит для выбора направления модернизации.
2. Техничко-экономическое обоснование модернизации производства.
3. Закупка, поставка и установка высокотехнологического оборудования.
4. Усиление специализации в успешных видах деятельности.
5. Приобретение фирм, занимающихся теми же видами деятельности.
6. Приобретение фирм-поставщиков и / или посредников.
7. Приобретение фирм, с далекой специализацией.
8. Расширение области применяемых технологий.
9. Изменение товарного ассортимента.

10. Введение методов ценообразования через анализ рынка.
11. Повышение удобств использования продукции фирмы.
12. Изменение технологии функционирования продукции и услуг.
13. Разработка нового дизайна производимой продукции.

По критерию соответствия стратегии причинам необходимости модификации и ранжированию по правилам таблицы 7 стратегии 1–4, 10 – наиболее соответствуют причинам необходимости модернизации, несколько меньше – стратегия 11 и еще меньше стратегия 12. Все они соответствуют не менее половине причин необходимости модернизации.

Таблица 7.

№№ стратегий	Соответствие стратегии «Модернизация производства» причинам, показанным в таблице 11								Число знаков «+»
	1	2	3	4	5	6	7	9	
1.	+	+	+	+	+	+	+	+	8
2.	+	+	+	+	+	+	+	+	8
3.	+	+	+	+	+	+	+	+	8
4.	+	+	+	+	+	+	+	+	8
5.	–	+	–	–	+	+	–	–	3
6.	–	+	–	–	–	+	+	–	3
7.	–	–	–	–	–	–	–	–	0
8.	–	+	+	–	–	–	–	–	2
9.	–	+	+	+	–	–	–	–	3
10.	+	+	+	+	+	+	+	+	8
11.	–	+	+	+	+	–	+	+	6
12.	–	–	+	+	–	–	+	+	4
13.	–	–	–	+	–	–	+	+	3

Примечание. 1. Число знаков «+» соответствует числу причин необходимости реализации стратегии. 2. Причина 8 не учитывается, так как цель «Модификация производства» не вызвана этими причинами.

В соответствии с сетцентрической концепцией система проверяет сосредоточение усилий стратегий. Для этого система высвечивает на дисплеях экспертов таблицы 8 и так же, как в таблице 2, просит экспертов оценить попарное синергетическое влияние стратегий. Знаком «+» эксперт должен указать, что стратегия, указанная в *i*-м столбце, увеличивает эффект страте-

гии, указанной в j -й строке, 0 – отсутствие эффекта и знаком « \rightarrow », если эта пара противоречива.

Суммарные баллы могут определяться по формуле (8). Эксперты решили что, если сумма баллов стратегии положительная, то такую стратегию целесообразно включить в рекомендуемый список стратегий.

Таким образом, цели «Модернизация производства» по анализируемому сетцентрическому принципу, рассмотренному в разд. 2.4 и 3.3, соответствуют стратегии с номерами 1–4, 10–12. Аналогично проверяются стратегии, которые должны реализовать остальные цели. Будем считать, что им соответствуют те же стратегии плюс стратегия 13. Ранжирование стратегий по таблице 7 и 8 так же, как и по последующим таблицам, будет использовано для генерального ранжирования стратегий. Но, естественно, для оценки модернизируемых стратегий этого недостаточно. Необходимо использовать еще и экономические показатели.

Таблица 8.

№№ стратегий	Эффект сосредоточения усилий стратегий													Сумма баллов
	1	2	3	4	5	6	7	8	9	10	11	12	13	
1.	0	+	+	+	-	+	-	+	+	+	0	0	0	5
2.	+	0	+	+	-	+	-	-	+	+	+	-	+	4
3.	+	+	0	+	+	-	-	-	+	+	+	+	+	6
4.	+	+	+	0	-	+	-	-	+	+	+	+	+	6
5.	-	-	+	-	0	-	-	-	-	-	+	-	+	-6
6.	+	-	-	+	-	0	-	-	-	+	+	-	+	-2
7.	-	-	-	-	-	-	0	-	-	-	-	-	-	-12
8.	+	-	-	-	-	-	-	0	+	+	-	+	+	-2
9.	+	+	+	-	-	-	-	+	0	+	-	-	+	0
10.	+	+	+	+	-	+	-	+	+	0	-	-	+	4
11.	0	+	+	+	+	+	-	-	+	+	0	+	+	5
12.	0	-	+	+	-	+	-	+	+	-	+	0	+	3
13.	0	-	+	+	+	+	-	-	-	-	+	+	0	0

4.2. ОЦЕНКИ СТРАТЕГИЙ ПО ЭКОНОМИЧЕСКИМ ПОКАЗАТЕЛЯМ С ИСПОЛЬЗОВАНИЕМ СЕТЕЦЕНТРИЧЕСКИХ ПРИНЦИПОВ

Эффективность реализации стратегии определяется изменением выгод в результате проекта минус изменение затрат [4]. Пример оценки эффективности реализации стратегий по некоторым возможным экономическим критериям показан ниже.

В настоящее время наиболее употребимы следующие критерии [4]:

1. Чистая текущая стоимость (net present value) NPV .
2. Индекс прибыльности (Profitability index) PI .
3. Внутренняя норма доходности или прибыльность проекта (internal rate of return) IRR .
4. Период окупаемости (payback period) PP .

Введем обозначения: B_t – выгоды проекта в год t ; C_t – затраты проекта в год t ; $t = 1, \dots, n$, – годы жизни проекта; r – норма дисконта. Компьютерная система, используя сетецентрический принцип реализации всей доступной информации по анализу архивных данных и/или по сведениям, предоставленным руководством фирм, вышедших из кризиса, может рассчитать эти величины по следующим соотношениям:

$$NPV = \frac{B_1 = C_1}{1+r} + \frac{B_2 = C_2}{(1+r)^2} + \dots + \frac{B_n = C_n}{(1+r)^n}.$$

Индекс прибыльности:

$$PI = \frac{NPV}{C_0},$$

где C_0 – первоначальные затраты.

Значение критерия IRR можно найти по формуле

$$\frac{IRR - r_1}{r_2 - r_1} = \frac{NPV_1}{NPV_2 - NPV_1}.$$

Значение IRR можно трактовать как нижний гарантированный уровень прибыльности инвестиционного проекта. Критерии NPV , IRR , PI , часто используемые в инвестиционном анализе, связаны друг с другом.

Сетецентрические принципы предполагают широкое использование всех доступных данных. В соответствии с ним в

нашем случае для оценки стратегий могут оказаться полезными оценки фирмы BCD после выхода из кризиса, близкой по своим характеристикам фирме AAA, полученные из архивов или сообщенные руководством фирмы. Эти данные будут использованы при оценке стратегий фирмы AAA.

Компьютерная система рассчитывает значения NPV , IRR , PP и PI для всего множества стратегий, реализующих цели фирмы BCD. Для формирования таблицы типа таблицы 9 система высвечивает на дисплеях экспертов результаты расчетов значений критериев и просит экспертов поставить балльные оценки критериев для всех стратегий, считая лучшим баллом 1, худшим – 10.

Для цели «Модернизация производства» балльные оценки стратегий по показателям NPV , IRR , PB , PI показаны в таблице 9. Ранжирование стратегий проведено по сумме рангов и по методу турнирной таблицы.

Алгоритмы выбора по турнирной таблице основаны на максимизации числа показателей объекта, превосходящих показатели других объектов. Для этого строится матрица такая, что

$$\forall x, y \in A, S = \{n(x, y)\}, \quad n(x, y) = \{l | P_l(x) < P_l(y) + \varepsilon_l\} \text{ или} \\ n(x, y) = \{l | P_l(x) > P_l(y) + \varepsilon_l\}, \quad w(x) = \sum_{x \neq y} n(x, y).$$

В нашем случае A – множество стратегий; l – идентификаторы параметров; P_l – l -й параметр оценки стратегии; ε_l – параметр чувствительности, соответствующий характеристике l ; $n(x, y)$ – число «выигрышей» стратегии x у стратегии y ; $w(x)$ определяет общее число «выигрышей» стратегии x у других стратегий.

Согласовывать ранжирование стратегий в таблице 9 и аналогичных таблицах для других целей нужно, так как хотя все переменные (коэффициенты) формул расчета экономических показателей были получены от фирмы BCD, но балльные оценки дали эксперты. Согласование балльных оценок в таблице 9 произведено по известным алгоритмам. Таким образом, произведено ранжирование стратегий цели «Модернизация производства» по экономическому принципу. Аналогичное ранжирование может быть проведено и по другим целям. Результаты ран-

жирования с учетом субъективных оценок экспертов используется при формировании стратегий фирмы ААА.

Таблица 9.

№ и наименование стратегий	Балльные значения критериев стратегий, реализующих цель «Модернизация производства»					Место		Сумма мест	Ранг стратегии
	NPV	IRR	PP	PI	Сумма баллов	по сумме баллов	по турнирной таблице		
1. Организационно-технологический аудит для выбора направления модернизации	1	4	3	2	10	1	1	2	1
2. Технико-экономическое обоснование модернизации производства	2	5	2	1	10	1	1	2	1
3. Закупка, поставка и установка высоко-технологического оборудования	3	1	4	5	13	3	4	7	3
4. Усиление специализации	4	3	5	4	16	4	3	7	3
10. Введение методов ценообразования через анализ рынка	7	4	6	3	20	5	6	11	6
11. Повышение удобств использования продукции	9	2	5	7	23	7	6	13	7
12. Изменение технологии функционирования продукции	4	6	1	9	20	5	5	10	5

Примечание. 1. Здесь указаны только принятые к реализации стратегии. 2. В таблице даются балльные оценки, а не физические показатели.

4.3. СЕТЕЦЕНТРИЧЕСКИЙ ПРИНЦИП В ОЦЕНКЕ СТРАТЕГИЙ МЕТОДОМ ПРЕДПОЧТЕНИЯ

Сетецентрический принцип предполагает, что должна быть использована вся доступная информация, поэтому наборы критериев разд. 4.1 и 4.2 недостаточны. Необходим еще набор критериев, который бы сильно зависел от создавшейся обстановки и субъективных предпочтений руководителей и экспертов. Поэтому в отличие, скажем, от оценки экономических показателей, этот набор критериев нельзя считать стандартным. В соответствии с сетецентрическими принципами система включает в список критериев для рассмотрения экспертов не только сугубо внутренние, но и, если так можно сказать, критерии «широкого взгляда», часть которых перечислена в разд. 3.1. Необходимость такого подхода видна даже из короткого описания ситуации для фирмы «Монокристалл».

Компьютерная система высвечивает на дисплеях экспертов и руководителей извлеченный из базы данных список критериев оценки стратегий (приведенный, например, в работе [2] или какой-либо другой). Каждый эксперт может его дополнить или вычеркнуть какие-то критерии. Система обрабатывает результаты действий экспертов (например, в списке остаются только те критерии, с которыми согласны все эксперты) и представляет результат для утверждения руководству. Если представленный список утверждается, он считается согласованным. В нашем примере будем считать согласованным следующий список критериев:

- 1) инвестиционная составляющая;
- 2) возможность реализации стратегии при существующем уровне квалификации сотрудников;
- 3) борьба с иностранными корпорациями за традиционный для фирмы рынок сбыта;
- 4) соответствие политики государства в отрасли;
- 5) соответствие мировым производственным тенденциям;
- 6) изменение курса акций фирмы.

Следующий шаг системы – определение «веса» (значимости) каждого критерия для всех указанных в таблице 10 стратегий. Методы определения «весов» известны. Они показаны, на-

пример, в работах [10, 11]. Поэтому, не останавливаясь на этих методах, будем считать, что они выполнены каждым экспертом.

Теперь система показывает экспертам таблицу 10 и предлагает им проставить «веса» каждой стратегии, рассмотренной в таблице 9 по каждому критерию. Система согласовывает оценки экспертов и по заданному алгоритму согласования, например, рассмотренному в [11], определяет «веса» стратегий по критерию, указанному в соответствующем столбце таблицы 10. Будем считать значения критериев в таблице 10 согласованными.

Таблица 10.

№ и наименование целей	№ и наименование стратегий	«Веса» критериев каждой стратегии y_{ij}					
		1	2	3	4	5	6
1. Модернизация производства	1. Организационно-технологический аудит для выбора направления модернизации	4	5	4	4	5	3
	2. Техничко-экономическое обоснование модернизации производства	4	5	4	4	5	3
	3. Закупка, поставка и запуск в производство высокотехнологического оборудования	4	4	4	4	5	3
2. Применение разработанных или освоенных технологий	4. Усиление специализации в успешных видах деятельности	5	5	4	5	4	4
3. Максимизация прибыли	10. Введение методов ценообразования через анализ рынка	3	4	4	4	4	3
4. Увеличение продаж	11. Повышение удобств использования продукции	3	3	4	4	4	3
	12. Изменение технологии функционирования продукции и услуг	3	3	4	3	3	3

Для критериального анализа ситуации введем в рассмотрение в пространстве значений критериев два подмножества S и D . S – это подмножество, в котором руководителю желательно иметь значения критериев, характеризующих объект после выполнения решения (реализации цели и/или стратегии). D – это подмножество, определяющее текущее состояние объекта или процесса.

Балльные оценки значения i -го критерия в областях D и S обозначим соответственно K_i^D и K_i^S . Балльную оценку значения i -го критерия K_i определим функцией:

$$(12) \quad K_i = \gamma_i F_i(K_i^D, K_i^S).$$

Конкретным видом функции F_i может быть, например, разность K_i^S и K_i^D , показывающая насколько надо улучшить создавшееся положение, или их частное, показывающее во сколько раз его надо улучшить.

Коэффициент γ_i определяется на основе опыта и знаний руководителя или эксперта. Для балльной оценки значения по формуле (12) используем значения K_i^S и K_i^D как фирмы ААА (верхняя строка каждой стратегии таблицы 11), так и фирмы ВСД (нижняя строка). Данные нижней строки могут быть получены либо анализом состояния фирмы ВСД до и после кризиса либо непосредственно от руководства фирмы в порядке реализации «горизонтальных» связей между организациями. Эти данные могут быть использованы руководством фирмы ААА для оценок скорректированных балльных значений стратегий фирмы ААА в таблице 11. В таблице 11 они оказали влияние на «веса» стратегий 1, 3 и 11.

Теперь надо определить наиболее эффективные стратегии по методу предпочтений. Для этого введем нормированную оценку i -ой стратегии

$$\beta_i = \frac{\sum_j y_{ij} \gamma_i K_i}{\sum_i \sum_j y_{ij} \gamma_i K_i},$$

где y_{ij} – «вес» соответствующей i -й стратегии по j -му критерию таблицы 10; $\gamma_i K_i$ – скорректированная балльная оценка значения i -й стратегии, показанная в последнем столбце таблицы 11.

Используя таблицы 10 и 11, получаем значения нормированных оценок по методу предпочтений:

$$(13) \quad \beta_1 = \frac{4 \times 25}{457} = 0,22, \quad \beta_2 = \frac{4 \times 25}{457} = 0,22, \quad \beta_3 = \frac{3 \times 24}{457} = 0,16, \\ \beta_4 = \frac{3 \times 27}{457} = 0,18, \quad \beta_{10} = \frac{1 \times 22}{457} = 0,05, \quad \beta_{11} = \frac{3 \times 21}{457} = 0,14,$$

$$\beta_{12} = \frac{1 \times 19}{457} = 0,04.$$

На основании оценок по формулам (13) система ранжирует стратегии, сводя результаты в таблице 12.

Таблица 11

№ и наименование стратегий	Желательное значение оценки K_i^S	Текущее значение оценки K_i^D	Балльная оценка значения i -й стратегии K_i	Скорректированное балльное значение i -й стратегии $\gamma_i K_i$
1. Организационно-технологический аудит для выбора направления модернизации	очень высокое (5)	низкое (2)	3	4
	очень высокое (5)	очень низкое (1)	4	
2. Техничко-экономическое обоснование модернизации производства	очень высокое (5)	низкое (2)	3	4
	очень высокое (5)	низкое (2)	3	
3. Закупка, поставка и запуск в производство высокотехнологического оборудования	хорошее (4)	среднее (3)	1	3
	хорошее (4)	очень низкое (1)	3	
4. Усиление специализации в успешных видах деятельности	очень высокое (5)	среднее (3)	2	3
	очень высокое (5)	среднее (3)	2	
10. Введение методов ценообразования через анализ рынка	хорошее (4)	среднее (3)	1	1
	хорошее (4)	хорошее (4)	0	
11. Повышение удобств использования продукции	очень высокое (5)	среднее (3)	2	3
	высокое (4)	очень низкое (1)	3	
12. Изменение технологии функционирования продукции и услуг	среднее (3)	среднее (3)	0	1
	среднее (3)	среднее (3)	0	

Таблица 12.

№№ стратегий	Ранг стратегий
1	1
2	1
4	3
3	4
11	5
10	6
12	7

4.4. ИНТЕГРАЛЬНОЕ РАНЖИРОВАНИЕ

Естественно, руководство не может реализовать все стратегии. Обычно рассматриваются только стратегии, имеющие больший интегральный ранг. Поэтому система ранжирует стратегии с учетом результатов по всем рассмотренным трем группам критериев (разд. 4.1–4.3). Ранжирование производится по сумме рангов и сводится в таблицу 13.

Таблица 13.

№№ стратегий	Ранги стратегий по рассмотренным ранее методам			Сумма ранга	Интегральный ранг стратегий
	по соответствию причинам модернизации	по экономическим показателям	по методу предпочтения		
1	1	1	1	3	1
2	1	1	1	3	1
3	1	3	4	8	3
4	5	3	3	11	4
10	1	6	6	13	5
11	6	7	5	18	6
12	7	5	7	19	7

В интегральном ранжировании рассматриваются только 7 лучших стратегий. Учитывались только те стратегии, которые не были отброшены при ранжировании по отдельным критериям, как это сделано в разд. 4.1, но отказ от стратегий мог произойти и в отборе по другим критериям. При ранжировании в таблице 13 система считала, что все методы имеют одинаковый «вес» (значение). Если бы эксперты посчитали, что «веса» разные, то они определили бы эти «веса» одним из рассмотренных выше способов или каким-нибудь другим и в соответствии с

полученными «весами» изменили бы ранги стратегий в таблице 13.

Таким образом, система сформировала и проранжировала модернизированные стратегии, используя некоторые сетечентрические методы.

Заметим, что в нашем случае если руководство отказывается от реализации стратегий с минимальным интегральным рангом, например, стратегий 11, 12, то оно отказывается от цели «Увеличение продаж» (см. таблицу 10).

5. Заключение

Использование двух рассмотренных сетечентрических принципов расширяет методы модификации целей и стратегий фирмы при резком изменении обстановки, повышает правильность принимаемых решений и позволяет реализовать концепцию сосредоточения усилий, максимально используя информацию из самых различных источников.

Искренне благодарен Н.И. Злобинской за помощь в этой работе.

Литература

1. АЛЬСЕВИЧ В.В. *Введение в математическую экономику*. – М.: URSS, 2007. – 256 с.
2. АНСОФ И. *Стратегическое управление*. – М.: Экономика, 1989. – 519 с.
3. ВИССЕЛК Х. *Стратегический менеджмент и предпринимательство: возможность для будущего предпринимательства*. – Пер. с англ. М.: ФИНПРЕСС, 2000. – 262 с.
4. ВОЛКОВ И.М., ГРАЧЕВА М.В., АЛЕКСАНОВ Д.С. *Критерии оценки проектов*. [Электронный ресурс]. – URL: <http://bre.ru/risk/911.html> (дата обращения 10.09.2013).
5. ГОРБАЧЕВ Ю.Е. *Сетечентрическая война: миф или реальность?* // Военная мысль. – 2006. – №1. – С. 66–76.
6. ГОРЕЛИК Л.Д., СКРИПКИН В.А. *Методы распознавания*. – М.: Высшая школа, 2004. – ?? с.

7. *Методика исследования конкуренции на рынке.* [Электронный ресурс]. – URL: <http://www.advertology.ru/article20560.htm> (дата обращения 10.09.2013).
8. ПОЛТЕВ К. *Чудны дела. Чем расстроила и чем порадовала российская экономика в 2012 г.* // Итоги. – 2012. – №52–53. [Электронный ресурс] URL: <http://www.itogi.ru/delo/2012/52/185429.html> (дата обращения 30.09.2013)
9. ТРАХТЕНГЕРЦ Э.А. *Компьютерные технологии манипулирования общественным мнением.* – М.: СИНТЕГ, 2011. – 296 с.
10. ТРАХТЕНГЕРЦ Э.А. *Компьютерные методы реализации экономических и информационных управленческих решений.* – М.: СИНТЕГ, 2009. – 172 с.
11. ТРАХТЕНГЕРЦ Э.А., СТЕПИН Ю.П. *Методы компьютерной поддержки формирования целей и стратегий в нефтегазовой промышленности.* – М.: СИНТЕГ, 2007. – ?? с.
12. ХАМЗАТОВ М.М. *Влияние концепции сетецентрической войны на характер современных операций* // Военная мысль. – 2006. – №7. – [Электронный ресурс] URL: <http://www.milresource.ru/Hamzatov-article-1.html> (дата обращения 30.09.2013)
13. ШУМИЛОВ В.В. *Многокритериальная оценка инвестиционных проектов* // Техника машиностроения. – 2000. – №6(28). – С. 93.
14. AMBROSINI V., BOWMAN C. *Managerial consensus and corporate strategy* // European Management Journal. – 2003. – Vol. 21, №2. – P. 213–221.
15. AUSLOOS M., IVANOVA K. *Mechanistic approach to generalized technical analysis of share prices and stock market indices* // The European Physical Journal B. – 2002. – Vol. 27. – P. 177–187.
16. BOYLE H.F., SCHENCK G.K. *Investment Analysis: US Oil and Gas Producers Score High in University Survey* // Proc. Hydrocarbon Econom. Evaluat. Symp. Dallas., 1985.
17. *Department of defense* // The implementation of Network – Centric Warfare. – Washington D.C. – 2005.

18. PORTER M.E. *Competitive advantage techniques for analyzing industries and competitors*. – New-York: The Free Press, 1985.

USE OF TWO NETWORK-CENTRIC PRINCIPLES OF MODIFICATION OF ECONOMIC TARGETS AND STRATEGY IN CRISIS SITUATIONS

Eduard Trahtengerts, Institute of Control Sciences of RAS, Moscow, Doctor of Science, Professor, Honoured Scientist of Russia, Chief Researcher (Moscow, Profsoyuznaya st., 65, (495)334-88-40, E-mail: tracht@ipu.rssi.ru).

Abstract: We consider two principles of a network-centric paradigm – those of efforts concentration and the maximum use of information – which can be used to modify economic goals and strategies in a crisis situation.

Keywords: objectives, strategies, computer methods, network-centric principles, crisis, concentration of efforts, use of information.

Статья представлена к публикации членом редакционной коллегии В.В. Кульбой

*Поступила в редакцию 11.03.2013.
Опубликована 30.09.2013.*

УДК 519
ББК 32.817

МОДЕЛЬ ПОВЕДЕНИЯ ЛЕЧЕБНО-ПРОФИЛАКТИЧЕСКОГО УЧРЕЖДЕНИЯ НА ТЕРРИТОРИАЛЬНОМ РЫНКЕ МЕДИЦИНСКИХ УСЛУГ

Мирзоян Г. Л.¹

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

Для одного лечебно-профилактического учреждения (ЛПУ), функционирующего на территориальном рынке, строится и исследуется модель принятия им решений о цене и качестве предоставляемых медуслуг с учетом внутренних механизмов экономической мотивации его сотрудников.

Ключевые слова: лечебно-профилактическое учреждение, механизмы управления, мотивационное управление, оптимальное распределение ресурса.

1. Введение

Результаты теоретического и имитационного исследования механизмов *управления организационными системами* (ОС) находят свое применение при решении широкого круга практических задач управления в самых разных прикладных областях [7, 13]:

- предприятия, корпорации и регионы [4];
- проекты и программы;
- образовательные системы [11, 12];
- информационная и др. безопасность;
- социальные системы;

¹ Гагик Левонович Мирзоян, аспирант (mirzoyangl@yandex.ru).

- организационно-технические системы [13];
- эколого-экономические системы.

Тем не менее, такому массовому классу объектов управления как *системы здравоохранения* в теории управления ОС пока не было уделено должное внимание исследователей. С другой стороны, существует ряд направлений, так или иначе использующих различные разделы прикладной математики для построения и изучения моделей систем здравоохранения.

Экономика здравоохранения – наука о правилах распределения ограниченных ресурсов, об оптимизации соотношения выгод и издержек в процессе предоставления медицинских услуг. Представление о современном состоянии российских исследований в области экономики здравоохранения можно получить из учебников [1, 6, 16 и др.]. Отдельно можно выделить классы работ, посвященных проблемам: оплаты труда [5], финансирования [3, 15, 17, 24 и др.] и неформальных экономических отношений, а также информатизации систем здравоохранения (не говоря о системах обязательного и добровольного медицинского страхования, которые также представляют самостоятельный предмет активных исследований). К экономике здравоохранения тесно примыкает *менеджмент систем здравоохранения* как раздел современного менеджмента [1, 2, 14, 15 и др.].

За рубежом существуют такие развитые разделы экономической теории, как:

- экономика здравоохранения (Health Economics) [см. обзоры в 23, 26 и др.], отличающаяся, кроме прочего, активным использованием экономико-математических моделей и аппарата эконометрики [21, 25], в том числе при моделировании конкуренции на рынке медицинских услуг [18, 20, 21].

- *экономика здоровья* – см. пионерские работы К. Эрроу, Г. Беккера и М. Гроссмана, а также обзор развития этого направления в [23].

Настоящая статья является попыткой построить часть «моста» между теорией управления ОС и экономикой здравоохранения в рамках программы исследований, намеченной в

[10]. Для этого рассматривается *территориальный рынок*, характеризуемый спросом на медицинские услуги, локализованным транспортной доступностью. Для простоты считается, что на территории функционирует единственное ЛПУ, оказывающее единственную медицинскую услугу (т.е. потенциальный пациент, проживающий на рассматриваемой территории, выбирает: обратиться ему в данное медучреждение или вообще не обращаться за медицинской помощью). Предположим, что *спрос* $A(Q, \lambda)$ на эту услугу (т.е. объем услуг) со стороны населения территории в общем случае зависит от *качества* услуги $Q \geq 0$ (например, доли ранее обратившихся в ЛПУ пациентов, получивших адекватную медицинскую помощь) и ее *цены* $\lambda > 0$, причем спрос не убывает по первой переменной и не возрастает по второй. ЛПУ потенциально способно удовлетворить любой спрос.

Оба ключевых параметра (и цена, и качество) для коммерческих услуг, оказываемых коммерческим или бюджетным ЛПУ, в существенной мере зависят от него самого; для деятельности же бюджетного ЛПУ в рамках программ медстрахования цена, фактически, задана экзогенно, а качество (и, следовательно, спрос) частично зависит от действий ЛПУ.

Таким образом, перед ЛПУ стоит задача выбора цены и качества оказываемой им медуслуги, а также системы стимулирования своих сотрудников, побуждающих их удовлетворить «индуцируемый» (данными ценой и качеством) спрос. Критерием экономической эффективности функционирования ЛПУ будем считать его «прибыль».

Описав поведение ЛПУ (см. второй раздел), можно анализировать зависимость выбираемых им значений цены и качества от экзогенных (например, ограничения на цену, удельную себестоимость) и эндогенных (например, квалификация персонала) параметров модели – см. третий раздел и заключение.

2. Модель выбора ЛПУ цены и качества медуслуги

В пионерской статье [18] рассматриваются две простые модели. В первой предприятие выбирает цену на свою продукцию и объем инвестиций в рекламу этой продукции (инвестиции в «качество» продукции и/или *репутацию* предприятия) таким образом, чтобы максимизировать свою прибыль (зависимость спроса, точнее – возможного объема продаж, от обоих выбираемых параметров считается известной). Во второй модели предприятие выбирает цену на свою продукцию и качество этой продукции (условно говоря, выбирает свою репутацию), а известной полагается зависимость спроса от цены и качества. При этом рассуждения в упомянутой работе ведутся на уровне качественных зависимостей между переменными (монотонность и т.п.), т.е. при минимальных предположениях получаются адекватно общие выводы.

В дальнейшем эти модели получили активное развитие в экономико-математической литературе применительно к системам здравоохранения – обзор моделей конкуренции нескольких ЛПУ (ценовой, за счет качества оказываемых услуг, с учетом решений принимаемых пациентами [19], с учетом пространственного расположения ЛПУ [18] и др.), качественное (обоснование монотонного роста равновесного качества с увеличением конкуренции) их исследование и результаты эконометрической идентификации можно найти в [21, 23]. Однако детального рассмотрения структуры экономических интересов ЛПУ при этом не производилось, что, в том числе, не позволяет проводить анализ сравнительной статистики, и тем более не дает возможности в дальнейшем ставить и решать формальные задачи управления территориальными системами здравоохранения.

Пусть *целевая функция ЛПУ* имеет вид

$$(1) F(Q, \lambda) = A_0(1 + \tau(\delta Q - \gamma\lambda))(\lambda - \ell) - \sum_{z=1}^m \sigma_z(Q, \lambda) - kQ^b - c_0,$$

где $A = A_0(1 + \tau(\delta Q - \gamma\lambda))$ – суммарный спрос на медицинскую услугу; $\tau \geq 0$ – безразмерная константа; $\delta \geq 0$ и $\gamma \geq 0$ – размер-

ные константы (пропорциональные производной функции $A(Q, \lambda)$ по качеству и цене в «точке» A_0); ℓ – удельная (в расчете на одного пациента) себестоимость оказываемой ЛПУ медуслуги (без учета затрат на стимулирование); $\sum_{z=1}^m \sigma_z$ – суммарные затраты ЛПУ на стимулирование (мотивацию) врачей; m – фиксированное число врачей, σ_z – вознаграждение z -го врача, $k \geq 0$ – размерная константа (соответствующий «максимальному» члену – степени $\beta > 2$ – разложения функции затрат kQ^β на обеспечение качества); c_0 – постоянные затраты ЛПУ.

Все слагаемые в выражении (1) имеют стоимостное измерение: первое слагаемое представляет собой произведение спроса на разность между ценой медуслуги и ее постоянной удельной себестоимостью (не включающей затраты на оплату труда), второе – его затраты на оплату труда (мотивацию врачей), третье – затраты ЛПУ на обеспечение качества (например, развитие материально-технического оснащения, маркетинг и т.п.), четвертое – постоянные затраты ЛПУ.

Что касается второй составляющей целевой функции ЛПУ (мотивации врачей), то ниже мы воспользуемся результатами, полученными в [9] при исследовании внутренних *механизмов экономической мотивации*. В качестве примера (остальные механизмы мотивации рассматриваются аналогично) будем анализировать *механизм отчислений* [9], согласно которому вознаграждение z -го врача $\sigma_z = \lambda s_z x_z$ пропорционально количеству принятых им пациентов x_z с коэффициентом пропорциональности, равным произведению *норматива отчислений* $s_z \in [0; 1]$ на цену медуслуги λ . При использовании механизма отчислений целевая функция ЛПУ примет вид

$$(2) \quad F(Q, \lambda) = A_0(1 + \tau(\delta Q - \gamma\lambda))(\lambda - \ell) - \lambda \sum_{z=1}^m s_z x_z - kQ^\beta - c_0.$$

Условие «сбалансированности» спроса имеет вид

$$(3) \quad \sum_{z=1}^m x_z = A_0(1 + \tau(\delta Q - \gamma\lambda)).$$

Прежде чем решать задачу максимизации целевой функции ЛПУ (2) по цене и качеству, необходимо найти оптимальный механизм экономической мотивации врачей, т.е. такие значения нормативов отчислений, которые минимизировали бы затраты ЛПУ на стимулирование и обеспечивали бы сбалансированность спроса (3):

$$(4) \quad \begin{cases} \lambda \sum_{z=1}^m s_z x_z \rightarrow \min_{\{s_z \in [0;1]\}} \\ \sum_{z=1}^m x_z = A_0(1 + \tau(\delta Q - \gamma\lambda)). \end{cases}$$

В рамках теоретико-игровых моделей материального стимулирования, принятых в теории управления ОС [13] и теории контрактов, врачи, в свою очередь, при заданной системе мотивации будут стремиться выбирать действия (количество пациентов), максимизирующие разность между их вознаграждением и затратами (см. подробности в [9]). Если функция затрат z -го врача типа Кобба–Дугласа имеет вид

$$(5) \quad c_z(x_z) = \frac{1}{\alpha} x_z^\alpha r_z^{1-\alpha}, \alpha > 1,$$

где r_z – *тип* [13] врача (параметр, отражающий, например, его квалификацию), то, как показано в [9], в механизме отчислений оптимальным с точки зрения z -ого врача будет действие $x_{z,s}^* = r_z (s_z \lambda)^{\frac{1}{\alpha-1}}$. Подставив это действие в выражение (4), получим задачу

$$(6) \quad \begin{cases} \lambda^{\frac{\alpha}{\alpha-1}} \sum_{z=1}^m s_z^{\frac{\alpha}{\alpha-1}} r_z \rightarrow \min_{\{s_z \in [0;1]\}} \\ \lambda^{\frac{1}{\alpha-1}} \sum_{z=1}^m s_z^{\frac{1}{\alpha-1}} r_z = A_0(1 + \tau(\delta Q - \gamma\lambda)). \end{cases}$$

Обозначим $H = \sum_{z=1}^m r_z$ и, решая задачу (6), найдем оптимальные значения нормативов отчислений:

$$(7) \quad s_z = s^* = \lambda^{-1} \left(\frac{A_0(1 + \tau(\delta Q - \gamma\lambda))}{H} \right)^{\alpha-1}.$$

Суммарные затраты на стимулирование врачей с учетом выражения (7) будут выглядеть следующим образом:

$$(8) \quad \sum_{z=1}^m \sigma_z = (A_0)^\alpha (1 + \tau(\delta Q - \gamma\lambda))^\alpha (H)^{1-\alpha}.$$

Итак, в рассматриваемой модели оптимально использованные унифицированного (одинакового для всех врачей) норматива отчислений s^* , при этом коллектив врачей «эквивалентен» одному врачу с типом H (правая часть выражения (8) может интерпретироваться как функция «мотивационных затрат» ЛПУ в целом – ср. (5) и (8)). Отметим, что данное свойство (так называемая «идеальная агрегируемость») оптимального решения задач типа (4) и (6) характерно для многоагентных систем, свойства которых описываются степенными производственными функциями типа Кобба–Дугласа (см. [13]). Идеальная агрегируемость является существенным плюсом с точки зрения возможности получения аналитического решения в иерархических системах. Но она обладает и своими минусами: аддитивность типов агентов в выражениях вида (8) приводит к компенсации низких значений типов одних агентов их количеством или высокими типами других агентов. От этого эффекта удастся избавиться, введя постоянные издержки на привлечение агентов, однако при этом теряется свойство идеальной агрегируемости (см. подробности в [13]).

Подставляя (8) в (2), получим следующее выражение для зависимости целевой функции ЛПУ от цены и качества медуслуги:

$$(9) \quad F(Q, \lambda) = A_0(1 + \tau(\delta Q - \gamma\lambda))(\lambda - \ell) - (A_0)^\alpha (1 + \tau(\delta Q - \gamma\lambda))^\alpha (H)^{1-\alpha} - kQ^\beta - c_0.$$

Таким образом, в рамках рассматриваемой модели ЛПУ будет, используя оптимальную систему стимулирования (7), стремиться при известной зависимости спроса от цены и качества

выбирать такие значения этих параметров, при которых достигается максимум его «прибыли» (9), т.е.

$$(10) \quad A_0(1 + \tau(\delta Q - \gamma\lambda))(\lambda - \ell) - (A_0)^\alpha (1 + \tau(\delta Q - \gamma\lambda))^\alpha (H)^{1-\alpha} - kQ^\beta - c_0 \rightarrow \max_{Q \geq 0, \lambda > 0}.$$

Исследуем зависимость решений задачи (10) от параметров модели.

3. Зависимость оптимальных цены и качества от параметров модели

В качестве иллюстрации построим график целевой функции ЛПУ (9) – см. рис. 1 с выделенной областью значений цены и качества, при которых значение целевой функции ЛПУ (9) положительно, используя следующие значения параметров: $Q = 0 : 100$; $\lambda = 0 : 3000$; $\gamma = 0,05$; $\beta = 3$; $k = 0,9$; $A_0 = 1000$; $\tau = 0,01$; $c_0 = 100000$; $\delta = 0,7$; $\ell = 490$, $\alpha = 2$; $H = 10$. Напомним, что во введении было введено предположение, что ЛПУ способно удовлетворить любой спрос. Данное предположение не является обременительным или нереалистичным, так как условие неотрицательности целевой функции ЛПУ все равно накладывает косвенные ограничения на «индуцируемый» им спрос.

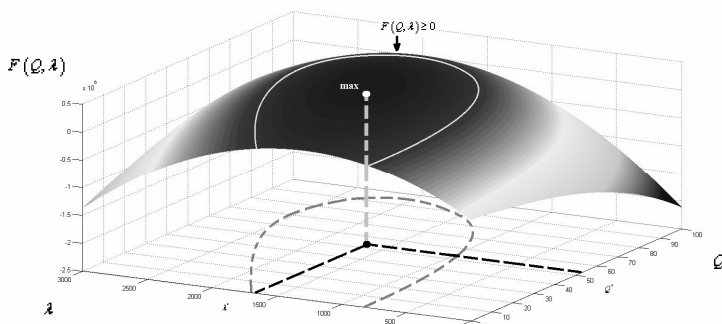


Рис. 1. Значения целевой функции ЛПУ (9)

Предположим, что ЛПУ имеет возможность выбирать самостоятельно произвольные (!) неотрицательные значения цены и качества. Для исследования этой ситуации, продифференцировав целевую функцию ЛПУ (9), найдем оптимальное значение цены λ^* :

$$1 + \tau\delta Q - 2\tau\gamma\lambda + \gamma\tau\ell + \tau\gamma\alpha \left((1 + \tau(\delta Q - \gamma\lambda)) \frac{A_0}{H} \right)^{\alpha-1} = 0.$$

Из данного выражения не представляется возможным аналитически выразить значение оптимальной цены для общего случая. В качестве примера рассмотрим случай $\alpha = 2$. Тогда

$$(11) \quad \lambda^* = \frac{\frac{1}{2}\ell + \left(\frac{A_0}{H} + \frac{1}{2\tau\gamma} \right) (1 + \tau\delta Q)}{\left(1 + \tau\gamma \frac{A_0}{H} \right)}.$$

Целевую функцию ЛПУ (9) с учетом выражения (11) при $\alpha = 2$ можно представить следующим образом:

$$(12) \quad F(Q, \lambda^*(Q)) = A_0 \left(1 + \tau \left(\delta Q - \gamma \frac{\frac{1}{2}\ell + \left(\frac{A_0}{H} + \frac{1}{2\tau\gamma} \right) (1 + \tau\delta Q)}{\left(1 + \tau\gamma \frac{A_0}{H} \right)} \right) \right) \times$$

$$\times \left(\frac{\frac{1}{2}\ell + \left(\frac{A_0}{H} + \frac{1}{2\tau\gamma} \right) (1 + \tau\delta Q)}{\left(1 + \tau\gamma \frac{A_0}{H} \right)} - \ell \right) - kQ^\beta - c_0 -$$

$$- \left(A_0 \left(1 + \tau \left(\delta Q - \gamma \frac{\frac{1}{2}\ell + \left(\frac{A_0}{H} + \frac{1}{2\tau\gamma} \right) (1 + \tau\delta Q)}{\left(1 + \tau\gamma \frac{A_0}{H} \right)} \right) \right) \right)^2 (H)^{-1}.$$

Используя значения параметров, приведенных ранее в примере (см. рис. 1), можно представить оптимальное значение цены следующим образом:

$$(13) \lambda^* = 1281 + 7,35Q.$$

Тогда целевая функция ЛПУ (12) будет иметь вид:

$$(14) F(Q, \lambda^*(Q)) = -0,9Q^3 + 23,16Q^2 + 5018,4Q + 171800.$$

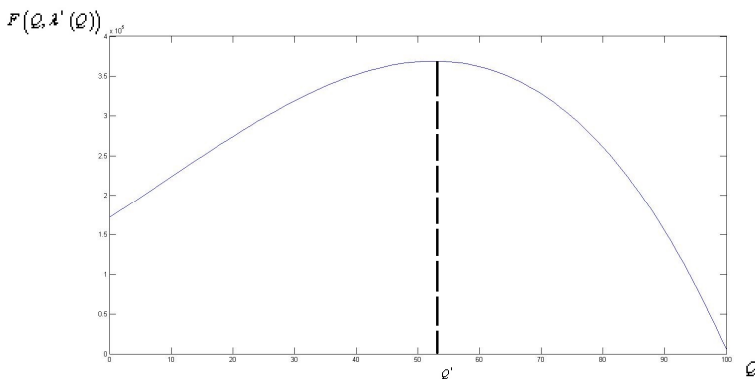


Рис. 2. Значения целевой функции ЛПУ $F(Q, \lambda^*(Q))$

Продифференцировав целевую функцию ЛПУ (14) и приравняв ее к нулю, найдем оптимальное значение качества: $Q^* \approx 53$, затем подставим полученное Q^* в выражение (13) и найдем оптимальное значение цены: $\lambda^* \approx 1671$ (данные значения цены и качества изображены на рис. 1).

4. Заключение

В предложенной модели затраты на стимулирование учитываются в явном виде, но также их можно включать в себестоимость (см. [8]). Тогда целевую функцию ЛПУ (9) можно представить следующим образом:

$$F(Q, \lambda) = A_0(1 + \tau(\delta Q - \gamma \lambda))(\lambda - \ell_0) - kQ^\beta - c_0,$$

где $\ell_0 \geq 0$ – общая удельная (в расчете на одного пациента) себестоимость оказываемой медицинской услуги (включает в себя, в том числе, затраты на стимулирование сотрудников).

Существенно, что оптимальные значения параметров Q^* и λ^* для построенных целевых функций ЛПУ в данной работе и в

[8] совпадают. Исходя из оптимальных значений цены λ^* и качества Q^* , можно рассчитать общую удельную (в расчете на одного пациента) себестоимость оказываемой медицинской услуги ℓ_0 (включает в себя, в том числе, затраты на стимулирование сотрудников), используемую в модели, описанной в [8], через удельную себестоимость оказываемой ЛПУ медуслуги ℓ (без учета затрат на стимулирование) и наоборот. Таким образом, модель настоящей работы и модель, описанная в [8], в этом смысле «эквивалентны» – можно для конкретных значений параметров учитывать затраты на стимулирование как в явном виде, так и включать их в себестоимость (см. [8]). Например, пусть $\ell_0 = 600$ (см. [8]), а для значений параметров, рассмотренных в примере (см. рис. 1), из оптимальных значений цены λ^* получим:

$$\ell_0 \approx 0,95\ell + 130,51 \approx 596,01,$$

а из оптимальных же значений качества Q^* получим:

$$\ell_0 \approx 2063,65 - \frac{1}{186,67} \left(2,18\sqrt{75010 - 32,4\ell} - 4,36 \right)^2 \approx 582,83.$$

Поэтому в дальнейшем для моделирования и исследования ситуаций, когда несколько ЛПУ конкурируют по цене, качеству (репутации), цене и качеству, можно использовать как данную модель, так и модель, представленную в [8].

Литература

1. ВЯЛКОВ А.И. КУЧЕРЕНКО В.З., РАЙЗБЕРГ Б.А. и др. *Управление и экономика здравоохранения: учеб. пособие*. 3-е изд., доп. – М.: ГЭОТАР-Медиа, 2009. – 664 с.
2. ДРОЗДОВА Е.А. *Менеджмент в здравоохранении*. – Благовещенск: АГМА, 2008. – 32 с.
3. ДУГАНОВ М.Д. *Оценка эффективности расходов на здравоохранение на региональном и муниципальном уровнях*. – М.: ИЭПП, 2007. – 112 с.
4. ЗАЛОЖНЕВ А.Ю. *Модели и методы внутрифирменного управления*. – М.: Сторм-Медиа, 2004. – 320 с.

5. КАДЫРОВ Ф.Н. *Стимулирующие системы оплаты труда в здравоохранении (материальное стимулирование специалистов различных подразделений и учреждений здравоохранения)*. – М.: ГРАНТ, 1998. – 336 с.
6. КОЛОСНИЦЫНА М.Г., ШЕЙМАН И.М., ШИШКИН С.В. *Экономика здравоохранения*. – М.: Издательский дом ГУ-ВШЭ, 2009. – 479 с.
7. *Механизмы управления* / Под ред. Д.А. Новикова. – М.: Ленанд, 2011. – 192 с.
8. МИРЗОЯН Г.Л. *Макромодель лечебно-профилактического учреждения* // Управление большими системами: материалы X Всероссийской школы-конференции молодых ученых. Том 2. – Уфа: УГАТУ, 2013. – С. 162 – 166.
9. МИРЗОЯН Г.Л. *Модели мотивационного управления в лечебно-профилактическом учреждении* // Системы управления и информационные технологии. – 2013. – №1.1(51). – С. 207 – 210.
10. МИРЗОЯН Г.Л. *Модель территориальной системы здравоохранения и классификация задач управления* // Управление большими системами: материалы IX Всероссийской школы-конференции молодых ученых. Том 1. – Липецк: Липецкий государственный технический университет, 2012. – С. 203 – 205.
11. НОВИКОВ Д.А., ГЛОТОВА Н.П. *Модели и механизмы управления образовательными сетями и комплексами*. – М.: ИУО РАО, 2004. – 142 с.
12. НОВИКОВ Д.А. *Теория управления образовательными системами*. – М.: Народное образование, 2009. – 416 с.
13. НОВИКОВ Д.А. *Теория управления организационными системами: учебник*. – 3-е изд. – М.: Издательство физико-математической литературы, 2012. – 608 с.
14. СУНГАТОВ Р.Ш. *Современные технологии управления здравоохранением как экономической системой*. – Казань: КГМА, 2005. – 211 с.

15. *Управление ЛПУ в современных условиях* / Под ред. академика РАМН В.И. Стародубова. – М.: Издательский дом «Менеджер здравоохранения», 2009. – 416 с.
16. ЩЕПИН О.П., МЕДИК В.А. *Общественное здоровье и здравоохранение: учебник*. – М.: ГЭОТАР-Медиа, 2011. – 592 с.
17. ШИШКИН С.В. и др. *Организация финансирования и управления здравоохранением в регионах России. Общая характеристика*. – М.: Независимый институт социальной политики, 2005. – 31 с.
18. BREKKE K., SICILIANI L., STRAUME O. *Hospital Competition and Quality with Regulated Prices*. CESifo Working Paper Series 2635. – Munich: CESifo Group, 2009. – 19 p.
19. CAPPS C., DRANOVE D., SATTERTHWAITE M. *Competition and Market Power in Option Demand Markets* // RAND Journal of Economics. – 2003. – Vol. 34. – №4. – P. 737–763.
20. DUBOVIK A., MAARTEN J. *Oligopolistic Competition in Price and Quality*. Discussion Paper TI 2008-068/1. – Amsterdam: Tinbergen Institute, 2008. – 36 p.
21. GAYNOR M., Town R. *Competition in Health Care Markets*. Working Paper №12/282. – Bristol: University of Bristol, 2012. – 153 p.
22. GROSSMAN M. *The Human Capital Model of the Demand for Health*. – NBER Working Paper № 7078, 1999. – 102 p.
23. *Handbook of Health Economics*. Vol. 2. – Amsterdam: Elsevier, 2012. – 1126 p.
24. *Implementing Health Financing Reform (Lessons from Countries in Transition)* / Eds. Kutzin J., Cashin C., Jakab V. – Copenhagen: WHO, 2010. – 411 p.
25. JONES A. *Health Econometrics*. – NY.: University of New York, 1998. – 81 p.
26. WAGSTAFF A., CULYER A. *Four Decades of Health Economics through a Bibliometric Lens* // Policy Research Working Paper 5829. – Washington DC: The World Bank. 2011. – 72 p.

THE MODEL OF MONOPOLISTIC TREATMENT-AND-PROPHYLACTIC INSTITUTION ON TERRITORIAL HEALTH-CARE MARKET

Gagik Mirzoyan, Institute of Control Sciences of RAS, Moscow, post-graduate student (mirzoyangl@yandex.ru).

Abstract: We state and study the model of decision-making on pricing and quality for the monopolist treatment-and-prophylactic institution on the regional market. The model accounts for intra-firm employees' motivation mechanisms.

Keywords: treatment-and-prophylactic institutions, control mechanisms, motivational control, optimal resource allocation.

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

Поступила в редакцию 09.04.2013.

Опубликована 30.09.2013.

УДК 004.724.2+004.272.43

ББК 3.9.7.3.02

РАСШИРЕННЫЙ ОБОБЩЕННЫЙ ГИПЕРКУБ КАК ОТКАЗОУСТОЙЧИВАЯ СИСТЕМНАЯ СЕТЬ ДЛЯ МНОГОПРОЦЕССОРНЫХ СИСТЕМ

Каравай М. Ф.¹, Подлазов В. С.²
(ФГБУН Институт проблем управления
им. В.А. Трапезникова РАН, Москва)

Рассматривается новая структура системной сети для высокопроизводительных многопроцессорных вычислительных систем. Рассматривается системная сеть в виде расширенного обобщенного гиперкуба, в строках которого связи с топологией полного графа заменены на связи с топологией квазиполного графа, имеющего много меньше ребер.

Ключевые слова: высокопроизводительные многопроцессорные вычислительные системы, системные сети, прямые каналы, неблокируемые самомаршрутизируемые сети, идеальные сети, распределенные полные коммутаторы, обобщенные гиперкубы, сервер PERCS.

1. Введение

Одно из направлений построения высокопроизводительных многопроцессорных вычислительных систем (суперкомпьютеров) предполагает использование многопроцессорных и многоядерных (тяжелых) процессорных узлов [8, 9]. Такие процессорные узлы используются в тесной связке с многопортовыми связными узлами. Множество портов в них требу-

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

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

ется для возможно большего распараллеливания системной сети, объединяющей связные узлы. Здесь возникает задача эффективного использования заданного множества портов или даже задания этого множества. В работе [1] предложено одно решение этой задачи на основе использования системных сетей с прямыми каналами [1, 7]. Оно основывается на маршрутно-инвариантном расширении таких простейших сетей как кольца и полные коммутаторы и осуществляется посредством замены топологии связей полного графа на квазиполный (ор)граф [2, 3]. Этот подход привел к построению распределенных полных коммутаторов и некоммутируемых мультиколец, пропускная способность которых пропорциональна квадрату числа портов связных узлов сети.

В данной работе вышеупомянутый подход применяется для построения системной сети со структурой расширенного обобщенного гиперкуба. В обычном обобщенном гиперкубе узлы в каждой строке (столбце) любого измерения имеют связи с топологией полного графа. В расширенном обобщенном гиперкубе связи каждого измерения имеют топологию квазиполного (ор)графа. Полный и квазиполный графы имеют одинаковые маршрутные свойства на перестановочном трафике, но квазиполный граф содержит много меньше ребер. Это свойство позволяет многократно увеличивать пропускную способность и отказоустойчивость системной сети при сохранении числа узлов или увеличивать число узлов и отказоустойчивость при сохранении числа каналов в сети. При этом фактически сохраняется диаметр сети и, как следствие, задержка передачи данных по сети.

2. Обобщенный гиперкуб

Обобщенный гиперкуб [10, 13] является «кубическим» аналогом многокаскадной сети Клоза, так же как гиперкуб является «кубическим» аналогом многокаскадной сети Бенеша. Как сеть Бенеша является двоичной сетью Клоза, так и гиперкуб является двоичным обобщенным гиперкубом.

Обобщенный гиперкуб обычно задается как d -мерный s -ичный гиперкуб, который имеет $V = s^d$ узлов степени $d(s - 1)$

каждый, размещенных в строках (столбцах) по s узлов, задающих «ребра» d -мерного простого куба. «Ребра» здесь понимаются не в графовом, а в геометрическом смысле. Наоборот, в графовом смысле «ребро» есть полный граф, т.е. все s узлов одного геометрического «ребра» связаны $s(s - 1)$ графовыми ребрами.

Будем обозначать обобщенный d -мерный s -ичный гиперкуб как $ОГК(V, d, s)$. На рис. 1 приведен пример $ОГК(16, 2, 4)$.

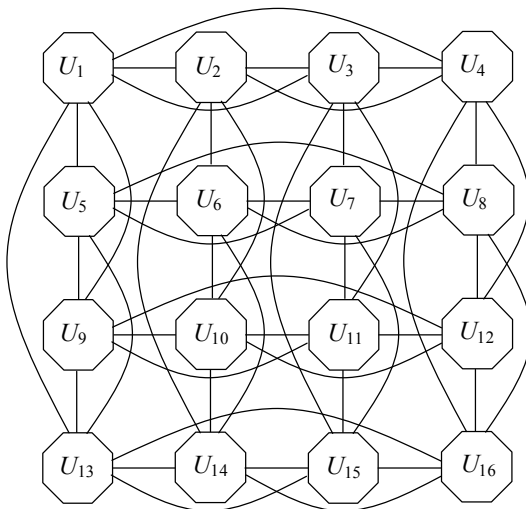


Рис. 1. Структура связей $ОГК(16, 2, 4)$

Диаметр $ОГК(V, d, s)$ равен $D = d$ и величина бисекции $B = V$.

Для реализации системной сети с топологией $ОГК(V, d, s)$ каждый узел должен иметь схемную реализацию, представленную в самом общем виде на рис. 2. Узел U_i содержит абонента (процессорный узел) A_i , «коммутатор измерений» $g \times g$ ($g = d + \pi$) и d коммутаторов каналов $1 \times m$ ($m = s - 1$) для реализации ребер полного графа в каждой строке (каждом столбце). Здесь π задает число каналов между процессорным и связным узлом. Для простоты в схеме на рис. 2 не показаны входные-выходные буферные очереди между коммутатором измерений и коммутаторами каналов, которые несущественны при описании

топологии системной сети. В суперкомпьютерах [8, 9] коммутаторы измерений и каналов входят в состав связного узла.

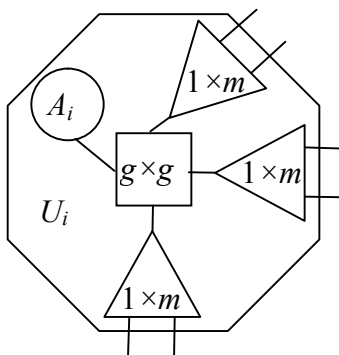


Рис. 2. Схемная реализация узла в ОГК($V, 3, m + 1$)

3. Идеальная сеть и распределенный полный коммутатор

Рассмотрим однородный двудольный граф, каждую долю которого составляют N узлов степени m . Значение m выбирается минимальным, при котором любые два узла в одной доле связны σ путями длины 2 через разные узлы в другой доле. Такой граф мы называем минимальным квазиполным графом [2]. Если он существует, то его параметры связаны соотношением $N = m(m - 1)/\sigma + 1$.

В данной статье предполагается, что узлами одной доли являются полные коммутаторы $m \times m$, а другой доли — m -портовые абоненты (связные узлы). Каждый путь между абонентами проходит через один коммутатор, и разные пути проходят через разные коммутаторы. Пример такого графа приведен на рис. 3 для $m = 4$, $N = 7$ и $\sigma = 2$. На рис. 3 толстыми линиями выделены пути между абонентами, выделенными одинаковой заливкой — их два для каждой пары абонентов.

Здесь возникает вопрос о существовании минимальных квазиполных графов и об их параметрах. Оказывается, что он уже давно решен в комбинаторике. Такие графы описываются на

языке неполных уравновешенных блок-схем, в частности, симметричных блок-схем [1–4].

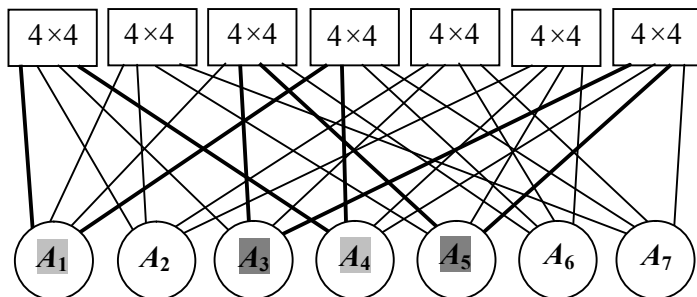


Рис. 3. Минимальный квазиуполный граф с $m = 4$, $N = 7$ и $\sigma = 2$.
Простейшая сеть ПС(7, 4, 2)

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 $N = m(m - 1)/\sigma + 1$.

Любая блок-схема описывается таблицей, в которой строки 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. При этом все абоненты связаны между собой прямыми каналами (через коммутаторы), как в полном графе. В отличие от полного графа рассматриваемый граф может иметь σ независимых путей между любой парой вершин, не являясь при этом мультиграфом, поскольку эти пути

не параллельны. Вся блок-схема интерпретируется как минимальный квазиполный граф, одна доля которого состоит из абонентов, а другая – из коммутаторов. Он описывает «простейшую» [2, 7] системную сеть с σ -кратным резервированием каналов. Задающая блок-схему таблица описывает схему межсоединений абонентов и коммутаторов. На рис. 3 приводится пример ПС(7, 4, 2), в таблице 1 – описание $B(7, 4, 2)$ и ПС(7, 4, 2).

Таблица 1. Схема межсоединений в ПС(7, 4, 2)

Блоки 4 × 4	$B(7, 4, 2)$ ПС(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

ПС(N, m, σ) является «идеальной» сетью, которая имеет возможность использования прямых каналов (без промежуточной буферизации пакетов) для бесконфликтной реализации произвольной перестановки пакетов данных между узлами [2, 3].

Введение в ПС(N, m, σ) коммутаторов $1 \times m$ дуплексных каналов (разветвителей/объединителей каналов – РОК m) превращает ее в распределенный полный коммутатор РК(N, m, σ), на который у авторов имеется патент [4]. На рис. 4 приводится схема РК(7, 4, 2), состоящая из коммутаторов 4×4 и РОК4.

По построению распределенный полный коммутатор РК(N, m, σ) является неблокируемым самомаршрутизируемым коммутатором $N \times N$ как и исходный коммутатор $m \times m$. Это означает, что произвольная перестановка пакетов данных между абонентами может осуществляться в нем бесконфликтно по прямым (без промежуточной буферизации пакетов) каналам. Каждый абонент прокладывает свой канал самостоятельно – независимо от других абонентов. Обычно предполагается, что прокладка канала осуществляется динамически посредством червячной маршрутизации. Однако возможна и

статическая маршрутизация по заранее составленным маршрутным таблицам, которые необходимо хранить в каждом коммутаторе $m \times m$ [7].

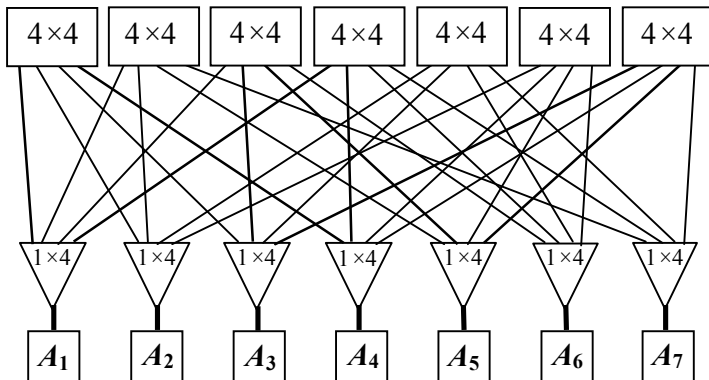


Рис. 4. Схема $PK(7, 4, 2)$ из коммутаторов 4×4 и $POK4$

С формальной точки зрения диаметр $PC(N, m, \sigma)$ и $PK(N, m, \sigma)$ равен 2 ($D = 2$). Однако диаметр можно выражать и в числе «скачков» D^* (передач по прямым каналам без промежуточной буферизации пакетов). Такой диаметр $D^* = 1$.

В $PC(N, m, 1)$ и $PK(N, m, 1)$ величина N является нечетным числом. Поэтому величину бисекции B определим как минимальное число дуплексных каналов «точка-точка» между множествами из $\lceil N/2 \rceil$ и $\lfloor N/2 \rfloor$ абонентов и тогда $B = N + 1$. Если же определять величину бисекции B^* в числе прямых каналов, то и в этом случае $B^* = N + 1$, так как прямой канал является симплексным каналом и два ребра, составляющие путь между абонентами, содержат два встречных прямых канала.

$PC(N, m, 1)$ имеет топологию квазиполного графа и содержит $w = Nm$ дуплексных каналов «точка-точка» между коммутаторами и абонентами. Сеть с топологией полного графа, содержащая N узлов, имеет $W = N(N - 1)$ дуплексных каналов «точка-точка» между абонентами. Легко проверить, что $W/w > \sqrt{N - 1} - 1$, т.е. имеет место сокращение числа каналов в $\sim \sqrt{N}$ число раз. $PK(N, m, 1)$ содержит $N(m + 1)$ каналов за счет использования POK_m .

Квазиполный граф существует не при любых значениях параметров m и σ . Эта проблема может быть разрешена двояко. При $\sigma = 1$ можно использовать квазиполный орграф, который существует при любых значениях m и имеет $N = m^2$ [1, 7]. Дополнительным ограничением для его использования в качестве ПС($N, m, 1$) служит невозможность использования дуплексных портов. При $\sigma \geq 1$ можно использовать 1-расширенные квазиполные графы, которые удалось построить для всех экспериментально проверенных значений m и σ [5, 7]. В них число узлов N^* каждой доли задается выражением $N^* = N - \delta$, где $\delta < m$. В 1-расширенном квазиполном графе малая часть узлов одной доли связаны $\sigma + 1$ путями длины 2, а остальные – σ путями длины 2. В матрице смежности такого графа номера узлов, связанных $\sigma + 1$ путями, размещаются на 2δ диагоналях. Системная сеть с топологией 1-расширенного квазиполного графа обозначается как ПС($N^*, m, \sigma | \sigma + 1$). Дополнительно в ней за счет выбора значения δ можно задавать четность N^* .

Перечисленные выше свойства показывают, что на перестановочном трафике системная сеть с топологией квазиполного графа по пропускной способности и задержкам практически не уступает сети с топологией полного графа, имея много меньшую канальную сложность.

4. Сетевые характеристики простейшей сети ПС(N, m, σ)

По другому может обстоять дело при трафике общесетевого вида, когда несколько источников могут параллельно обращаться к одному приемнику. В сети с топологией полного графа каждый приемник может принять $N - 1$ параллельных пакетов, а в сети с топологией квазиполного графа с $\sigma = 1$ – только m , т.е. в $\sim \sqrt{N}$ меньше, что приведет к уменьшению пропускной способности и к увеличению задержек передачи. Это уменьшение объяснимо, поскольку в случае полного графа нет конфликта доступа к абоненту при любом распределении адресов назначения. Для квазиполного графа конфликты могут возникать в локальном коммутаторе при одновременном обращении через него

разных источников к одному и тому же приёмнику. Возникает вопрос – во сколько раз падает пропускная способность и растет задержка передачи? Эти характеристики можно оценить имитационным моделированием на простой модели с трафиком, состоящим из пакетов одинаковой длины со случайными адресами приемников.

При моделировании каждый источник генерирует пакет с адресом приемника, который (адрес) распределен по степенному закону, т.е. с вероятностью p_i выбора i -го приемника, задаваемой как $p_i = g(a, N) (i/N)^{a-1}$ ($1 \leq i \leq N$), где $g(a, N)$ – нормировочный множитель, который определяется соотношением $g(a, N) \sum_{i=1}^N p_i = 1$. С увеличением a все большая часть источников

адресуются к N -му приемнику. Это распределение включает равномерное распределение – $a = 1$, $g(1, N) = 1$, линейное распределение – $a = 2$, $g(2, N) = 2/(N + 1)$; параболическое распределение – $a = 3$, $g(3, N) = 6/[(N + 1)(2N + 1)]$; и т.д.

В модели случайный адрес при заданном a находится как $\max(u_1, \dots, u_i, \dots, u_a)$, где u_i ($1 \leq i \leq a$) – случайное целое с равномерным распределением на $[0, N - 1]$.

Каждый источник может иметь пакет не более чем к одному приемнику, т.е. моделируется случай $\sigma = 1$ с одним путем между любой парой абонентов. В модели все источники действуют синхронно по тактам, передавая в каждом такте по одному пакету. Если несколько источников адресуются к одному приемнику через один и тот же коммутатор, то пакет передает только один из них, а остальные задерживают передачу до следующего такта. После каждого такта источники, которые не имеют пакетов, заново их генерируют с заданным распределением адресов приемников.

Исследовались два базовых режима: с постоянным распределением адресов (с постоянным a – режим ПР) по источникам, со случайным выбором закона распределения одинаковым для **всех** источников на каждом такте (с равномерным распределением a на $[1, N]$ – режим СР) и со случайным выбором закона распределения для **каждого** источника на каждом такте (с равномерным распределением a на $[1, N]$ – режим СР*). В каждом ре-

жиме еще предусматривается перемешивание адресов приемников в виде случайного смещение максимума распределения на каждом Δ -м такте (случайное смещение – СМД). Он осуществляется как сдвиг назначенных адресов на одинаковый случайный шаг с равномерным распределением на $[0, N-1]$. При $\Delta = 0$ изменений адресов приемников фактически не производится, при $\Delta = 1$ оно осуществляется на каждом такте. По мнению авторов, режим СР*СМ1 соответствует условиям плохого (случайного) пространственно-временного размещения данных по узлам сети. Подобное размещение данных по оперативной памяти вычислительной системы с кэш-памятью приводит к сильной деградации её производительности.

В сети с топологией полного графа за каждый такт передается $N - 1$ пакетов, а в сети с топологией квазиполного графа – случайное число пакетов η ($1 \leq \eta \leq N$). Измерялось среднее значение η^* за большое число тактов в установившемся режиме. Результаты моделирования в режиме ПР для малых m в представлены в графиках на рис. 5 и рис. 6. Ось абсцисс задает значение показателя степени a , а ось ординат – отношение $\rho = \eta^*/N$. Рис. 5 задает графики режима ПРСМ0, а рис. 6 – режима ПРСМ1. Разрывы на графиках возникли потому, что для представленных m нет значений a с выбранной дискретностью представления. Выбор малых m объясняется тем, что для них легко построить простейшие и 1-расширенные простейшие сети ПС($N, m, 1$) и ПС($N^*, m, \sigma | \sigma + 1$).

Если обозначить пропускную способность сети с топологией полного графа как W , а сети с топологией квазиполного как w , то рис. 5 и рис. 6 показывают, что в режиме ПРСМ0 при малых m $w/W = \rho > 0,2$, а в режиме ПРСМ1 $w/W = \rho \geq 0,69$.

В [5, 11] и в разделе 7 рассматриваются сети с $m = 7$ и с $m = 38$. На рис. 7 представлены результаты моделирования в режиме ПРСМД для $m = 38$. Они показывают, что варианты с $\Delta = 0$ и с $\Delta \gg 1$ (две нижние кривые) практически совпадают.

Для задержек передачи картина выглядит не столь оптимистично В каждом такте измеряется задержка передачи δ (в тактах) для каждого источника. Случайная величина δ усредняется в каждом такте по источникам, осуществившим передачу пакета

в данном такте, а затем – по всем тактам. В результате формируется средняя задержка передачи τ .

Сначала рассмотрим режим с постоянным распределением адресов приемников для всех источников – ПРСМД.

На рис. 8 и рис. 9 представлены графики зависимости $\tau(a)$ задержки передачи от степени распределения в случае малых m . Эти графики соответствуют графикам зависимости $\rho(a)$ на рис. 5 и рис. 6.

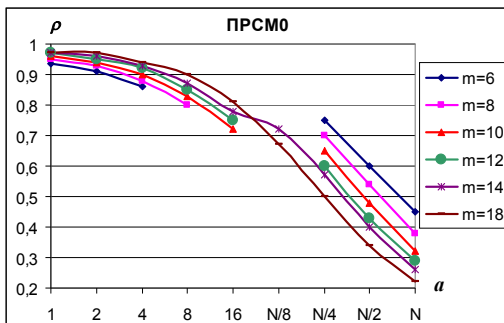


Рис. 5. Отношение w/W пропускных способностей сетей с топологией квазиполного и полного графов в режиме ПРСМ0

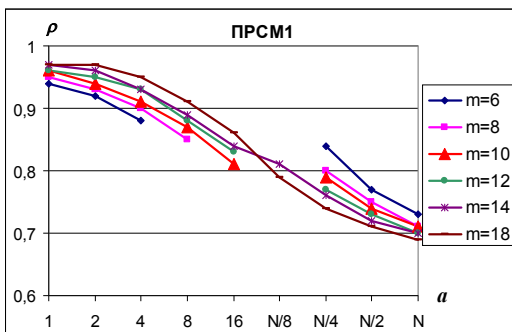


Рис. 6. Отношение w/W пропускных способностей сетей с топологией квазиполного и полного графов в режиме ПРСМ1

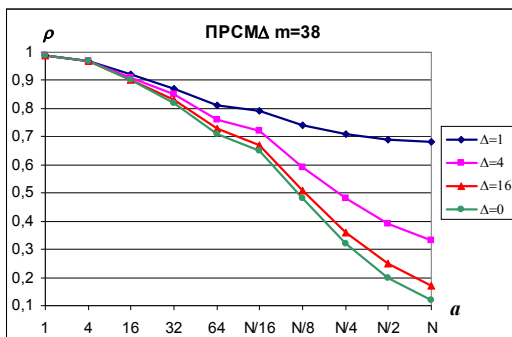


Рис. 7. Отношение w/W пропускных способностей сетей с топологией квазиполного и полного графов в режиме ПРСМА для $m = 38$

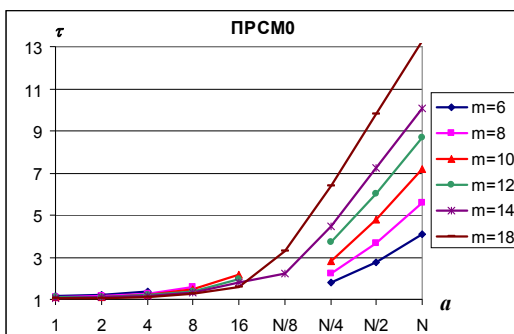


Рис. 8. Задержки сетей с топологией квазиполного графа в режиме ПРСМО

Здесь надо иметь в виду, что $\tau(a) = 1$ для сетей с топологией полного графа. Отметим три вывода из представленных графиков. Задержки в режиме ПРСМО существенно меньше, чем в режиме ПРСМО. Задержки в режиме ПРСМО растут с ростом a , при том что пропускная способность остается достаточно высокой при всех значениях a : $\rho(a) > 0,69$ (рис. 6). В режиме ПРСМО $\tau(a) \leq 1,6$ при $a \leq 8$, а в режиме ПРСМО – при $a \leq 16$. Эта область может считаться рабочей, в которой задержки остаются еще приемлемыми.

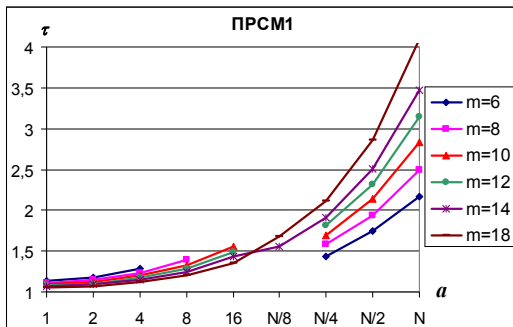


Рис. 9. Задержки сетей с топологией квазишполного графа в режиме ПРСМ1

На рис. 10 представлены графики зависимости $\tau(a)$ для режима ПРСМД при $m = 38$. Здесь рабочая область расширяется до $a \leq 32$.

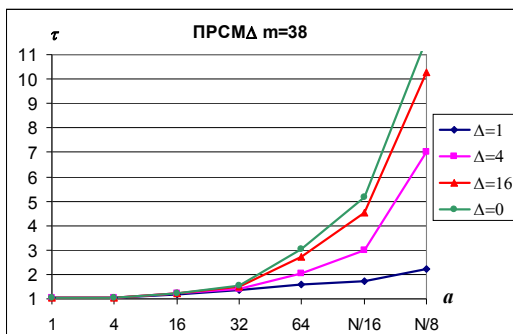


Рис. 10. Задержки сетей с топологией квазишполного графа в режиме ПРСМД для $m = 38$

На рис. 11 и рис. 12 представлены результаты моделирования в режиме СРСМД (случайное распределение a одинаковое для всех источников) для разных m и разных максимальных значений степени a ($\max a$). Режимы с $\Delta = 0$ представляются пунктирными графиками.

На рис. 11 в режиме СРСМ1 для всех проверенных m ($6 \leq m \leq 38$) режим СРСМ1 при $\max a = N$ имеет место соотно-

шение $\rho > 0,7$. В то же время, в режиме СРСМ0 при $\max a = N$ имеет место вырождение пропускной способности, как и в режиме ПРСМ0. На рис. 11 приводятся также графики при $\max a = N/8$ и $\max a = N/16$, для которых $\rho > 0,8$ и $\rho > 0,9$ соответственно.

На рис. 12 приводятся графики $\tau(a)$ для тех же значений $\max a$. Видно, что для $\max a = N/16$ при всех t $\tau(a) < 1,5$, т.е. находится в рабочей области.

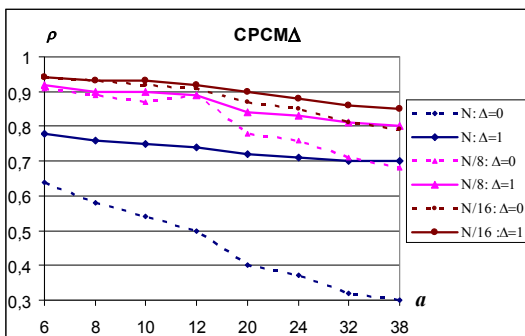


Рис. 11. Пропускная способность в режиме СРСМΔ

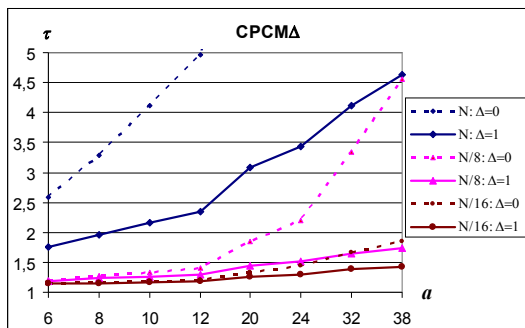


Рис. 12. Задержки в режиме СРСМΔ

На рис. 13 представлены результаты моделирования в режиме СР*СМΔ (случайное распределение a для каждого источника). Здесь значения $\rho(a)$ для режимов с $\Delta = 1$ и $\Delta = 0$ различаются мало: $\rho \geq 0,85$ и $\rho \geq 0,79$ соответственно. Для зависимости

$\tau(a)$ различия больше, однако для $\Delta = 1$ $\tau < 1,5$, т.е. находится в рабочей области. Отметим, что при $m \leq 24$ имеет место еще более узкий диапазон $\tau < 1,3$.

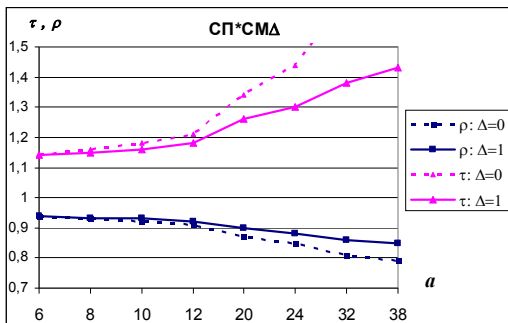


Рис. 13. Характеристики сети с топологией квазиполного графа в режиме СР*СМΔ.

Характеристики сети в режиме СР*СМ1 достаточно близки к характеристикам сети при равномерном распределении ($a = 1$) адресов приемников. Они одинаковы во всех режимах и представлены на рис. 14. Из него следует, что пропускные способности и задержки передачи сетей с топологией полного и квазиполного графов практически совпадают.

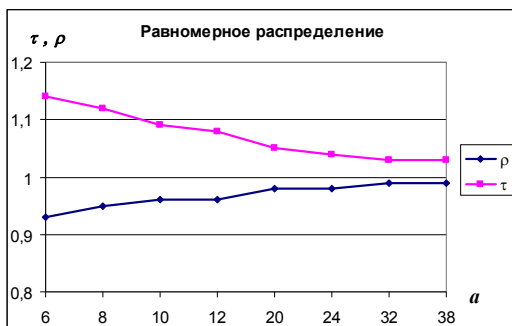


Рис. 14. Характеристики сети с топологией квазиполного графа при $a = 1$ для разных m

5. Расширенный обобщенный гиперкуб

Как упоминалось в начале статьи (рис. 1), обобщенный гиперкуб имеет по строкам и столбцам топологию связей полного графа. Расширение обобщенного гиперкуба можно осуществить за счет замены топологии связей строки (столбца) каждого измерения – с топологии полного графа на топологию квазиполного графа.

Пусть имеется ОГК(V, d, s). Добавим к каждому узлу d коммутаторов $m \times m$ ($m = s - 1$). Для объединения узлов в каждой строке (столбце) будем использовать ПС(N, m, σ), построенную за счет использования одного коммутатора $m \times m$ при каждом узле. ПС(N, m, σ) позволяет объединить в строке N узлов, где $N = m(m - 1)/\sigma + 1$. Увеличим число узлов в строке каждого измерения до этой величины и объединим их посредством ПС(N, m, σ). Тем самым удаётся «расширить» ОГК с s -ичности до N -ичности.

В результате получим расширенный обобщенный гиперкуб с $R = N^d$ узлами, в котором узлы любой строки (столбца) каждого измерения связаны σ прямыми каналами. Такой гиперкуб будем называть расширенным обобщенным d -мерным N -ичным гиперкубом и обозначать РОГК(R, d, N, m, σ). На рис. 15 для примера показан РОГК(49, 2, 7, 3, 1).

В описанном построении ПС(N, m, σ) может быть заменена на 1-расширенную ПС($N^*, m, \sigma | \sigma + 1$) (см. раздел 3), и тогда расширенный гиперкуб обозначается как РОГК($R, d, N^*, m, \sigma | \sigma + 1$).

Напомним, что 1-расширенная простейшая сеть ПС($N^*, m, \sigma | \sigma + 1$) порождается при уменьшении числа узлов сети ПС(N, m, σ) до значения $N^* = N - \delta$, где $\delta < m$. При этом являются узлы одной доли, связанные $\sigma + 1$ путями длины 2, а остальные – σ путями длины 2. Сеть ПС($N^*, m, \sigma | \sigma + 1$) удаётся построить для любых m и σ .

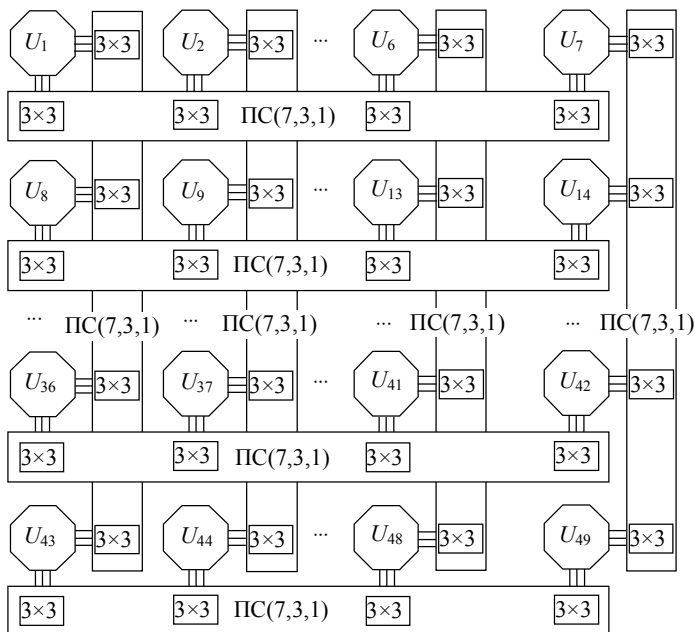


Рис. 15. Структура связей РОГК(49, 2, 7, 3, 1)

Эти дополнительные пути потребуются в следующем разделе 4 для формирования отказоустойчивой конфигурации РОГК по связным узлам. Далее всюду где упоминается РОГК($R, d, N^*, m, \sigma | \sigma + 1$) имеется в виду он или РОГК(R, d, N, m, σ), если последний существует для заданных m и σ .

Сравним некоторые характеристики трехмерных обобщенных гиперкубов ОГК($V, 3, s$) и РОГК($R, 3, N^*, m, \sigma | \sigma + 1$). Сначала оценим фактор R/V увеличения числа узлов при одинаковых параметрах узла, который приводится в таблице 2 при малых m и σ .

Таблица 2 показывает, что за счёт увеличения «-ичности» РОГК($R, 3, N^*, m, \sigma | \sigma + 1$) по сравнению с ОГК($V, 3, s$) может иметь во много раз большее число узлов и/или в σ раз большую пропускную способность каждого измерения.

Таблица 2. Фактор R/V при $m = s - 1$

$\sigma \setminus m$	4	6	8	10	12
1	17,6	96	254	566	1071
2	2,7	9,8	27,0	56	114
3	1,0	3,9	9,4	18,3	36,2
4	–	1,5	4,6	8,0	16,4

В таблице 3 показана степень узла s по каждому измерению в ОГК(V, d, s) при $V = R$ из РОГК($R, d, N^*, m, \sigma | \sigma + 1$). Таблица 3 показывает, что РОГК($R, d, N^*, m, \sigma | \sigma + 1$) по сравнению с ОГК(V, d, s) может иметь во много раз меньшее число портов в каждом узле и/или в несколько раз большую пропускную способность каждого измерения.

Таблица 3. Значение s при $R = V (s = N^*)$

$\sigma \setminus m$	4	6	8	10	12
1	13	32	57	91	133
2	7	15	27	42	63
3	5	11	19	29	43
4	–	8	15	22	33

6. Отказоустойчивость расширенных обобщенных гиперкубов

Полная однородность каждой строки (столбца) в РОГК($R, d, N^*, m, \sigma | \sigma + 1$) по узлам обеспечивает его κ -отказоустойчивость по процессорным узлам [6] при любом $1 \leq \kappa \leq \lfloor N/2 \rfloor$. Наличие в каждой строке σ независимых путей между любой парой процессорных узлов абонентов обеспечивает $(\sigma - 1)$ -отказоустойчивость по каналам и коммутаторам $m \times m$ каждой ПС($N^*, m, \sigma | \sigma + 1$) [6, 7]. Однако 1-скачковая связь любых строк (столбцов) разных измерений только через один связной узел не обеспечивает отказоустойчивости РОГК($R, d, N^*, m, \sigma | \sigma + 1$) по связным узлам.

Для обеспечения 1-отказоустойчивости по узлам можно использовать большую избыточность РОГК($R, d, N^*, m, \sigma | \sigma + 1$)

по числу узлов следующим образом. Разобьем узлы каждой строки на пары четный–нечетный. Каждая такая пара должна входить в состав строки каждого измерения. Это приведет к уменьшению вдвое числа строк всех измерений кроме первого, но обеспечит два варианта перехода между строками разных измерений – через четный и нечетный узлы. Такой 1-отказоустойчивый расширенный обобщенный гиперкуб будем означать $\text{РОГК}(R^*, d, N^*, m, \sigma | \sigma + 1, 1)$. Он содержит $(N^*/2)^d$ пар узлов, т.е. $R^* = 2(N^*/2)^d = N^* (N^*/2)^{d-1}$ узлов. В принятых ранее обозначениях $\text{РОГК}(R, d, N^*, m, \sigma | \sigma + 1)$ обозначается как $\text{РОГК}(R, d, N^* m, \sigma | \sigma + 1, 0)$.

На рис. 16 показан 2-мерный $\text{РОГК}(18, 2, 6, 3, 1, 1)$. Для построения $\text{ПС}(6, 3, 1)$ использован 1-расширенный гиперкуб, в котором пары абонентов (i, j) с номерами $j = (i \pm 3) \bmod 6$ связаны двумя путями, а остальные – одним путем. На рис. 17 этот гиперкуб расширен до трех измерений – $\text{РОГК}(54, 3, 6, 3, 1, 1)$.

На этих рисунках толстыми линиями показаны наборы из 3 ребер. Если заменить их на наборы из 4 ребер и использовать $\text{ПС}(6, 4, 2)$ для объединения узлов в строках, то эти же рисунки зададут $\text{РОГК}(18, 2, 6, 3, 2, 1)$ и $\text{РОГК}(54, 3, 6, 3, 2, 1)$ соответственно. В $\text{ПС}(6, 4, 2)$ пары абонентов (i, j) с номерами $j = (i \pm 1) \bmod 6$ связаны тремя путями, а остальные – двумя путями.

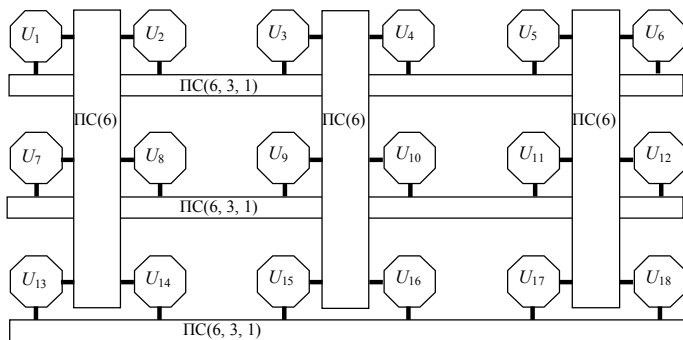


Рис. 16. Топология связей в 2-мерном $\text{РОГК}(18, 2, 6, 3, 1, 1)$

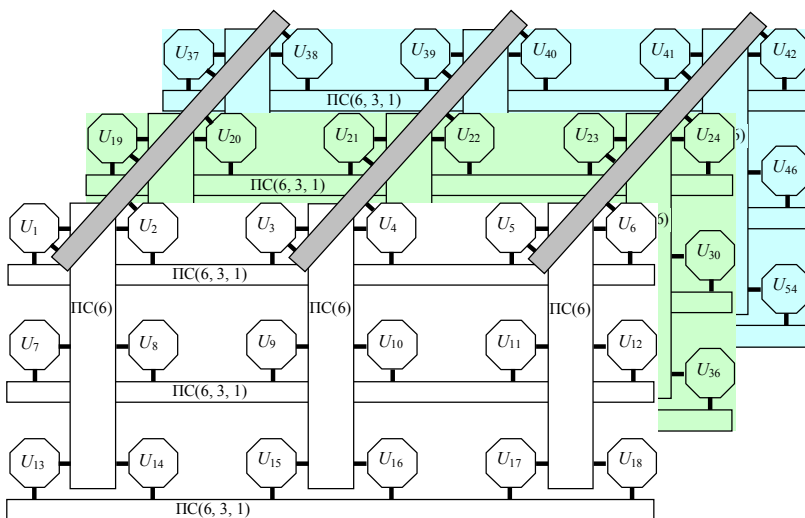


Рис. 17. Топология связей в 3-мерном РОГК(54, 3, 6, 3, 1, 1)

Аналогичным образом можно строить и μ -отказоустойчивые расширенные обобщенные гиперкубы РОГК(R^* , d , N^* , m , $\sigma | \sigma + 1$, μ). Для этого строки должны разбиваться на группы по $(\mu + 1)$ узлов, а гиперкуб будет содержать $R^* = N^* (N^*/(\mu + 1))^{d-1}$ узлов. В этом случае отказоустойчивость можно разминивать на пропускную способность между измерениями.

6. Системная сеть на основе расширенного обобщенного гиперкуба

В данном разделе рассматривается возможность расширения системной сети суперкомпьютера, разработанного IBM для проекта Blue Waters [9]. Теперь его принято именовать как сервер PERCS (Productive, Easy-to-use, Reliable Computer System) или система P775 [12]. Одна попытка такого расширения была предпринята авторами в [1]. Для возможностей последующего сравнения изложим кратко ее суть.

Системная сеть в [9] имеет топологию двухуровневого полного графа (рис. 18).

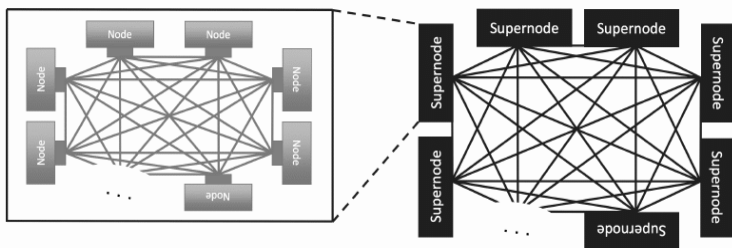


Рис. 18. Узлы, суперузлы и связи между ними

Каждый узел связи (node) имеет межузловые каналы трех видов: 7 высокоскоростных каналов K_1 с пропускной способностью V_1 ; 24 канала низкоскоростных K_2 с пропускной способностью $V_2 = V_1/5$ и 16 среднескоростных каналов K_3 с пропускной способностью $V_3 = 2V_2$. Каналы K_1 выполнены медным кабелем, а каналы K_2 и K_3 – оптическим кабелем. На рис. 19 приведена структура связей этой системной сети.

32 узла связи образуют суперузел (supernode), в котором узлы связаны по схеме полного графа каналами K_1 и K_2 . Среди них выделяются 4 группы по 8 узлов, связанных каналами K_1 . Остальные узлы связаны каналами K_2 . Суперузлы связаны каналами K_3 также по схеме полного графа.

Каждый суперузел имеет 512 каналов K_3 . В максимальной конфигурации суперкомпьютера каждый такой канал используется для связи с другим суперузлом по схеме полного графа. В этом случае он содержит 513 суперузлов. Передача пакета между любыми двумя узлами занимает не более трёх смен каналов с промежуточной буферизацией пакетов (скачков).

В [1] был рассмотрен подход, основанный на замене топологии связей полного графа в узлах и суперузлах на топологию квазиполного графа. В нем оставались неизменными следующие параметры сети: число портов в узле связи и максимальное число скачков между любыми узлами сети.

Для модификация сети внутри суперузла (левая часть рис. 18) к каждому узлу добавлялся коммутатор 7×7 каналов K_1 и узлы связывались сетью с топологией 1-расширенного квазиполного графа ПС(39, 7, 1|2). В результате образовывался су-

перузел из 39 узлов, связанных прямыми высокоскоростными каналами K_1 . Каналы K_2 не использовались. Параметры полученной сети представлены в таблице 4. В ней фактор f_1 оценен по результатам моделирования, представленным на рис. 13.

Таблица 4. Параметры системной сети суперузла

Параметры	Новая сеть	Исходная сеть
Коммутатор каналов K_1	7×7	–
Число узлов	39	32
Фактор межузловой пропускной способности f_1	$3 < f_1 < 5$	$f_1 = 1$
Каналы K_2	0	24
Фактор энергопотребления узла e_1	$1 \approx e_1 < 2$	$e = 1$

Для модификации сети между суперузлами (правая часть рис. 18) к каждому суперузлу добавлялось 16 коммутаторов 38×38 каналов K_3 и суперузлы связывались 16 сетями ПС(1407, 38, 1). Параметры полученной сети представлены в таблице 5. В ней фактор f_3 оценен по результатам моделирования (рис. 8).

В результате предложенной модификации получена расширенная системная сеть с большим числом узлов, с большей пропускной способностью и с многократной отказоустойчивостью 3-скачковых путей. В ней остались неиспользованными 24 канала K_2 , которые оказалось невозможно использовать в описанной топологии связей.

Однако их можно использовать в топологии трехмерного расширенного обобщенного гиперкуба – точнее 3-мерного параллелепипеда, в каждом измерении которого узлы имеют разные степени узлов. В измерении X она равна 7, в измерении Y – 16 и в измерении Z – 24.

Пусть строки в измерении X составляют рассмотренные выше модифицированные суперузлы с коммутаторами 7×7 каналов K_1 при каждом узле. Тогда строки в измерении Y могут составлять «новые» суперузлы с коммутаторами каналов K_2 при каждом узле, а строки в измерении Z могут составлять «новые» суперузлы с коммутаторами каналов K_3 при каждом узле. Общая топология такой системной сети представлена на рис. 19.

Таблица 5. Параметры системной сети между суперузлами

Параметры	Новая сеть	Исходная сеть
Коммутаторы каналов K_3	38×38	–
Число коммутаторов в суперузле	16	0
Число суперузлов	1407	513
Число узлов всей сети	54873	16416
Число путей между суперузлами	16	1
Отказоустойчивость 3-скачковых путей	есть	нет
Фактор пропускной способности между суперузлами f_3	$13 < f_3 < 15$	$f_3 = 1$
Фактор энергопотребления суперузла e_3	$1 < e_3 < 2$	$e_3 = 1$

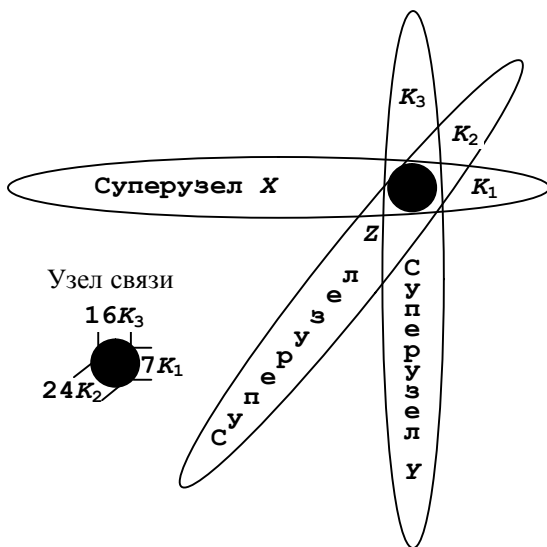


Рис. 19. Суперузлы разных измерений

В рассматриваемой топологии возможны различные варианты построения суперузлов каждого измерения. Первый вариант – это использование малых коммутаторов 8×8 в суперузлах измерений Y и Z . Для построения суперузлов измерения Y к каждому узлу необходимо добавить 2 таких коммутатора, а измерения Z – 3 таких коммутатора. В данном случае в суперузлах

каждого измерения удобно использовать топологию квазиполного орграфа [1 – 3].

Таблица 6. Параметры 1-отказоустойчивой по связным узлам системной сети с топологией 3-мерного параллелепипеда

Параметры	Измерение X: каналы K_1	Измерение Y: каналы K_3	Измерение Z: каналы K_2	Общее число узлов
Число и вид коммутаторов при узле	1 – 7 × 7	2 – 8 × 8	3 – 8 × 8	$R^* =$ $= N_X N_Y N_Z / 4 =$ $= 50\ 176$
Число путей в суперузле	1	2	3	
Число узлов в суперузле	$N_X = 48$	$N_Y = 32$	$N_Z = 32$	

В результате узлы в суперузлах измерения X связываются ПС(49, 7, 1), в суперузлах измерения Y – двумя независимыми ПС(64, 8, 1) и в суперузлах измерения Z – тремя независимыми ПС(64, 8, 1). Это позволяет построить 1-отказоустойчивую по узлам системную сеть, параметры которой приведены в таблице 6.

В другом варианте можно иметь более высокую пропускную способность системной сети в суперузлах за счет использования топологии квазиполного графа с $\sigma > 1$. Для этого в узлах для создания суперузлов измерений Y и Z потребуется использовать коммутаторы 16 × 16 и 24 × 24 соответственно. Конкретно: узлы суперузлов измерения X свяжем 1-расширенной ИС(20, 7, 2|3), узлы суперузлов измерения Y свяжем 1-расширенной ПС(60, 16, 4|5) и узлы суперузлов измерения Z свяжем 1-расширенной ПС(70, 24, 8|9). Количество каналов в суперузлах выбиралось из соображения приблизительного равенства суммарной пропускной способности между узлами суперузлов. В результате можно построить 1-отказоустойчивую по узлам и по каналам системную сеть, параметры которой приведены в таблице 7.

Таблица 7. Параметры 1-отказоустойчивой по связным узлам системной сети с топологией 3-мерного параллелепипеда с повышенной пропускной способностью

Параметры	Измерение X: каналы K_1	Измерение Y: каналы K_3	Измерение Z: каналы K_2	Общее число узлов
Число и вид коммутаторов при узле	1 – 7 × 7	1 – 16 × 16	1 – 24 × 24	$R^* =$ $= N_X N_Y N_Z / =$ $= 21\ 000$
Число путей в суперузле	2	4	8	
Число узлов в суперузле	$N_X = 20$	$N_Y = 60$	$N_Z = 70$	

Если не повышать пропускной способности сети в суперузлах измерения X, то получим вариант системной сети, представленной в таблице 8. При этом этот вариант является ориентировочным, так как 1-расширенные квазиполные графы для ПС(80, 16, 3|4) и ПС(94, 24, 6|7) еще не построены. Здесь значение числа узлов в суперузлах после построения может измениться на несколько единиц.

Таблица 8. Параметры 1-отказоустойчивой по связным узлам системной сети с топологией 3-мерного параллелепипеда с повышенной пропускной способностью

Параметры	Измерение X: каналы K_1	Измерение Y: каналы K_3	Измерение Z: каналы K_2	Общее число узлов
Число и вид коммутаторов при узле	1 – 7 × 7	1 – 16 × 16	1 – 24 × 24	$R^* =$ $= N_X N_Y N_Z / 4 =$ $= 73\ 320$
Число путей в суперузле	1	3	6	
Число узлов в суперузле	$N_X = 39$	$N_Y = 80$	$N_Z = 94$	

7. Заключение

В работе предложена новая топология системной сети в виде расширенного обобщенного гиперкуба. Эта топология разрабатывалась с ориентацией на суперкомпьютер фирмы IBM, разрабатывавшийся в рамках проекта Blue Waters [9]. Теперь этот суперкомпьютер именуют как сервер PERCS или система P775 [12]. Эта система первая пробилла «стену памяти», т.е. показала выдающиеся характеристики в условиях плохой пространственно-временной локализации данных в памяти. В частности, на тестах «Graph 500» [11] она в сильно усеченном варианте имела лучшую удельную производительность. Эти характеристики достигнуты, в частности, за счет высокой параллельности и малой глубины системной сети. В данной работе указанные свойства системной сети сохранены, но повышены ее пропускная способность, масштабируемость и отказоустойчивость, за счет использования системной сети в виде расширенного обобщенного гиперкуба в форме 3-мерного параллелепипеда. В нем строки (столбцы) разных измерений имеют не только разные простейшие сети, объединяющие их узлы, но и разные скорости передачи по каналам этих сетей.

Литература

1. КАРАВАЙ М.Ф., ПОДЛАЗОВ В.С. *Топологические резервы суперкомпьютерного интерконнекта* // Управление большими системами. – 2012 - №40. – С. 395–423.
2. КАРАВАЙ М.Ф., ПОДЛАЗОВ В.С. *Метод инвариантного расширения системных сетей многопроцессорных вычислительных систем. Идеальная системная сеть* // Автоматика и телемеханика. – 2010. – №10. – С. 166–176.
3. КАРАВАЙ М.Ф., ПОДЛАЗОВ В.С. *Распределенный полный коммутатор как «идеальная» системная сеть для многопроцессорных систем* // Управление большими системами. - 2011 - №34. –С. 92–116.
4. КАРАВАЙ М.Ф., ПОДЛАЗОВ В.С., СОКОЛОВ В.В. *Способ построения неблокируемого самомаршрутизируемого рас-*

- ширенного коммутатора* // Патент на изобретение №2435295 РФ от 06.09.2009. Зарегистрирован 03.08.2011.
5. КАРАВАЙ М.Ф., ПОДЛАЗОВ В.С. *Расширенные блок-схемы для идеальных системных сетей* // Проблемы управления. – 2012 - №4. – С. 45–51.
 6. КАРАВАЙ М.Ф., ПОДЛАЗОВ В.С. *Сетецентрический подход к обеспечению отказоустойчивости многопроцессорных систем реального времени* // Четвертая всероссийская мультikonференция по проблемам управления (МКПУ-2011), Дивноморское. Т. 1. – С. 305–308.
 7. ПОДЛАЗОВ В.С., КАРАВАЙ М.Ф. *Системные сети с прямыми каналами для многопроцессорных вычислительных систем – идеальные системные сети* // Palmarium Academic Publishing. – 2012. – 168 с. – [Электронный ресурс]. – URL: <http://www.ipu.ru/sites/default/files/publications/18125/3747-18125.pdf> (дата обращения: 14.09.2013).
 8. ALVERSON R., ROWETH D., KAPLAN L. *The Gemini System Interconnect* // 18th IEEE Symposium on High Performance Interconnects, 2009. – P. 83–87.
 9. ARIMILI B., ARIMILI R., CHUNG V., ET AL. *The PERCS High-Performance Interconnect* // 18th IEEE Symposium on High Performance Interconnects, 2009. – P. 75–82.
 10. BHUYAN L.N. AND AGRAWAL D.P. *Generalized Hypercube and Hyperbus Structures for a Computer Network* // IEEE Transaction on Computers. – 1984 - Vol. C-33, №4. – P. 323–333.
 11. *Graph 500 List* // June, 2012. – [Электронный ресурс]. – URL: http://www.graph500.org/results_june_2012 (дата обращения: 14.09.2013).
 12. HRUSKA J. *After Years of Work IBM, NCSA Cancel «BlueWaters» Supercomputer* // August, 2011. - [Электронный ресурс]. – URL: <http://hothardware.com/News/After-Years-of-Work-IBM-NCSA-Cancel-Blue-Waters-Supercomputer> (дата обращения: 14.09.2013).
 13. ZIAVRAS S.G. AND KRISHNAMURTHY S. *Evaluating the communications capabilities of the generalized hypercube interconnection network* // Concurrency: Practice and Experience. – 1999 - Vol. 11, №6. – P. 281–300.

EXTENDED GENERALIZED HYPERCUBE AS FAIL-SAFE SYSTEM NETWORK FOR MULTIPROCESSOR SYSTEMS

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 suggest a novel system network structure for high-performance multiprocessor computers. The system network has the structure of an extended generalized hypercube where links of complete graph topology in rows are replaced with links of quasi-complete graph topology with much less number of links.

Keywords: massive parallel multiprocessor computer, system area network, direct channel, nonblocking self-routing network, ideal network, distributed full switch, generalized hypercube, server PERCS.

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

*Поступила в редакцию 14.03.2013.
Опубликована 30.09.2013.*

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 рецензирования статьи - 3-4 месяца. Плата с авторов за публикацию рукописей не взимается.