

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

УПРАВЛЕНИЕ БОЛЬШИМИ СИСТЕМАМИ

***Выпуск 73
Май 2018***

**СБОРНИК
ТРУДОВ**

ISSN 1819-2467

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

Москва – 2018

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

**УПРАВЛЕНИЕ
БОЛЬШИМИ
СИСТЕМАМИ**

СБОРНИК ТРУДОВ

Выпуск 73

Москва – 2018

КООРДИНАЦИОННЫЙ СОВЕТ

Академики РАН: Васильев С.Н., Емельянов С.В., 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: Кузнецов О.П., Кульба В.В., Лотоцкий В.А., Павлов Б.В., Поляк Б.Т., Рутковский В.Ю.

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

Главный редактор: член-корр. РАН Новиков Д.А. **Зам. главного редактора:** д-р физ.-мат. наук Губко М.В. **Отв. секретарь:** канд. техн. наук Калимулина Э.Ю. **Редактор:** канд. техн. наук Квинто Я.И.

Д-ра техн. наук: проф. Алескеров Ф.Т. (ГУ ВШЭ), проф. Алчинов А.И. (ИПУ РАН), проф. Андриевский Б.Р. (ИПМ РАН), проф. Афанасьев В.Н. (МИЭМ), проф. Бахтадзе Н.Н. (ИПУ РАН), проф. Бурков В.Н. (ИПУ РАН), проф. Вишневский В.М. (ИПУ РАН), Галаев А.А. (ИПУ РАН), д-р физ.-мат. наук проф. Ерешко Ф.И. (ВЦ РАН), д-ра техн. наук Зоркальцев В.И. (ИСЭМ СО РАН), проф. Калашников А.О. (ИПУ РАН), проф. Калянов Г.Н. (ГУ ВШЭ), проф. Каравай М.Ф. (ИПУ РАН), д-р экон. наук, проф. Клочков В.В. (ИПУ РАН), д-ра техн. наук, Коргин Н.А. (ИПУ РАН), проф. Курдюков А.П. (ИПУ РАН), д-ра физ.-мат. наук, проф. Кушнер А.Г., проф. Лазарев А.А. (МФТИ), д-ра техн. наук: проф. Лебедев В.Г. (ИПУ РАН), проф. Мандель А.С. (ИПУ РАН), д-р биол. наук проф. Михальский А.И., д-р физ.-мат. наук, проф. Непейвода Н.Н. (ИПС РАН), д-р экон. наук, проф. Нижегородцев Р.М. (ИПУ РАН), д-ра техн. наук: проф. Орлов А.И. (МГТУ), д-ра физ.-мат. наук: проф. Рапопорт Л.Б. (ИПУ РАН), проф. Райгородский А.М. (МГУ), проф. Савватеев А.В. (РЭШ), д-ра техн. наук: проф. Самуйлов К.Е. (РУДН), проф. Сидельников Ю.В. (МАИ), Совлуков А.С. (ИПУ РАН) д-ра физ.-мат. наук: проф. Соловьев С.Ю. (МГУ), проф. Угольницкий Г.А. (ЮФУ), проф. Уткин В.А. (ИПУ РАН), проф. Хоботов Е.Н. (МГТУ), 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

Арзамас – д-р физ.-мат. наук проф. Пакшин П.В. **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); **Иркутск** – академик РАН Бычков И.В., д-р физ.-мат. наук, 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. Погодаев А.К., Сараев П.В. (ЛГТУ); **Самара** – д-ра экон. наук: проф. Богатырев В.Д., проф. Гераськин М.И., д-р техн. наук, проф. Засканов В.Г. (СГАУ); **Петрозаводск** – д-р физ.-мат. наук, проф. Мазалов В.В., д-р техн. наук, доц. Печников А.А. (ИПМИ КарНЦ РАН); **Санкт-Петербург** – д-р физ.-мат. наук: проф. Петросян Л.А. (СПбГУ), д-р техн. наук проф. Фуртат И.Б. (ИПМ РАН); **Старый Оскол** – д-р техн. наук, проф. Еременко Ю.И. (СТИ).

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

Адрес в интернете: ubs.mtas.ru.

СОДЕРЖАНИЕ

Системный анализ

Агасандян Г.А.

Континуальный критерий VaR и оптимальный портфель инвестора 6

Математическая теория управления

Давыдов В.А.

Метод встречных планов для определения целевого значения KPI в двухуровневой организационной системе 27

Афраймович Л.Г., Ильин С.Ю.

Многоиндексная задача построения расписания семинара..... 55

Информационные технологии в управлении

Выхованец В.С., Ду Ц., Сакулин С.А.

Обзор алгоритмов фонетического кодирования..... 67

Сетевые модели в управлении

Иванов Н.Н.

Алгоритмы управления ресурсами в обобщенных стохастических сетевых графиках 95

Захаров В.В., Мугайских А.В.

Динамическая адаптация генетического алгоритма маршрутизации транспорта на больших сетях..... 108

Управление в социально-экономических системах

Калашников П.В.

Динамическая модель оценки баланса солидарно-распределительной пенсионной системы 134

**Надежность и диагностика средств
и систем управления**

Ведешенков В.А.

Подход к организации отказоустойчивых цифровых систем со структурой минимального квазиполного графа с двумя путями между двумя абонентами (на примере графа размера 11×11) 157

**Программы и системы моделирования объектов,
средств и систем управления**

Алтыева М.Ч.

Математическое моделирование процесса нагрева сляба в методических печах толкательного типа..... 184

УДК 519.685
ББК 22.18

КОНТИНУАЛЬНЫЙ КРИТЕРИЙ VaR И ОПТИМАЛЬНЫЙ ПОРТФЕЛЬ ИНВЕСТОРА¹

Агасандян Г. А.²

(Вычислительный центр им. А.А. Дородницына
ФИЦ ИУ РАН, Москва)

Изучается оптимальное поведение инвестора на высокоразвитом рынке опционов. Вводится функция рискованных предпочтений инвестора и на ее основе определяется континуальный критерий VaR как континуальное обобщение хорошо известного обычного критерия VaR. Инвестор имеет также собственный прогноз на вероятностные свойства будущей цены базового актива опционов. Задача состоит в максимизации среднего дохода (доходности) при выполнении введенного критерия. На однопериодном теоретическом рынке с заданной картиной цен опционов строится оптимальный портфель. Метод построения иллюстрируется примером с двусторонними экспоненциальными распределениями вероятности.

Ключевые слова: базовый актив, функция рискованных предпочтений, континуальный критерий VaR, процедура Неймана–Пирсона, оптимальный портфель.

1. Введение

Инвестор, приобретая на рынке некий инструмент (портфель), получает случайный неотрицательный доход. Характер случайности оценивает (прогнозирует) сам инвестор. Оценива-

¹ Работа выполнена при финансовой поддержке РФФИ (проект 17-01-00816).

² Геннадий Аршавинович Агасандян, доктор физико-математических наук (Москва, ул. Чертановская, д. 34, тел. (495) 313-44-94).

ние результата этого действия обычно проводится по средней доходности (как правило, речь идет о ее максимизации) и риску, мерой которого чаще всего служат дисперсия или так называемый критерий VaR. На рынках с небольшим выбором инструментов этих двух мер может оказаться достаточно; более того, они вполне могут дать схожие результаты.

Однако на высокоразвитых опционных рынках с обширным инструментарием можно строить портфели в форме комбинации привычных коллов и путов с получением практически произвольного случайного дохода как функции будущей цены базового актива. И здесь могут проявиться недостатки привычных мер риска (см., например, [4, 6, 8-11]).

Так, применение критерия VaR в задаче максимизации среднего дохода сводится к получению дохода, не превышающего заложенного в критерий уровня (обычно меньше размера инвестиции) с вероятностью, близкой к единице (при частой решетке страйков), что едва ли устроит разумного инвестора.

А использование дисперсии, как в теории Марковица, приводит к значительному обеднению пространства поиска оптимума. Основанная на свойствах случайностей лишь второго порядка, она не позволяет сыграть на некоторых нюансах распределений, например, на часто наблюдаемых на рынках тяжелых хвостах распределений, для описания которых необходим учет свойств хотя бы четвертого порядка (куртозис, эксцесс).

Предлагаемый в работе континуальный критерий VaR (CC-VaR) должен обеспечить инвестору (во всяком случае на теоретическом рынке) получение функции распределения доходов с заранее заданными им же свойствами.

Этот критерий требует, чтобы строящийся из имеющихся на рынке инструментов оптимальный портфель инвестора порождал случайный доход q , удовлетворяющий неравенствам $P\{q \geq \phi(\varepsilon)\} \geq 1 - \varepsilon$ для *всех* $\varepsilon \in [0, 1]$ ($P\{M\}$ – вероятность множества M с точки зрения инвестора). Неотрицательная, монотонно возрастающая и непрерывная функция $\phi(\varepsilon)$ называется *функцией рискованных предпочтений* (ф.р.п.) инвестора и им же задается. (Учет маржевых платежей исключает использование

в модели агрегированных позиций инвестора, допускающих отрицательные доходы.)

Типичным примером может служить функция $\phi(\varepsilon) = \varepsilon^\lambda$, $\varepsilon \in [0, 1]$, $\lambda > 0$. И чем большее значение параметра λ выбирает для себя инвестор, тем более он готов рисковать ради увеличения средней доходности (см. также [1-3, 5]).

2. Континуальный δ -рынок и его инструменты

Для простоты и из желания более полного изучения возможностей СС-VaR рассматривается *однопериодный* рынок, порожденный некоторым базовым активом (например, акцией) X . Рынок предполагается *теоретическим* в том смысле, что страйки опционов на актив могут быть любыми вещественными числами, и *идеальным*, поскольку цены продавца и покупателя предполагаются равными, комиссионные равны нулю, а торговать инструментами можно в *любых* количествах.

В начале периода цена базового актива X известна, а в конце периода она образует случайную величину X , принимающую значения x из континуального множества $X \subset \mathbb{R}_+$ (или даже \mathbb{R}). Инвестор строит для величины X свой прогноз в форме плотности вероятности $p(x)$, $x \in X$, которую называем *прогнозной*.

На рынке можно торговать любым инструментом G , порождающим доход, представимый в виде произвольной неотрицательной измеримой функции $g(x)$, $x \in X$. Ее называем *платежной функцией* инструмента и обозначаем $\pi(x; G)$, $x \in X$, т.е. $g(x) = \pi(x; G)$; в частности, $\pi(x; X) = x$. Цены инструментов формирует рынок, и они задаются на начало периода. Цена инструмента G обозначается $|G|$, а средний доход от него — $\|G\|$.

На рынке обращаются так называемые *δ -инструменты* $D(s)$, $s \in X$. Платежная функция δ -инструмента $D(s)$, $s \in X$, равна δ -функции относительно s : $\pi(x; D(s)) \equiv \delta(x - s)$. Инструмент $D(s)$ дает нулевой доход, если $x \neq s$, и бесконечный, если $x = s$, $s \in X$, притом интеграл от платежной функции по $x \in X$ равен единице.

Из инструментов $D(s)$ может быть построен фактически любой портфель. Так, портфель G с платежной функцией $g(x)$ может быть представлен в виде континуальной комбинации инструментов $D(s)$ с весами $g(s)$:

$$(1) \quad G = \int_X g(s) D(s) ds.$$

В этом смысле совокупность δ -инструментов $\{D(s), s \in X\}$ можно считать базисом в пространстве инструментов рынка, а каждый δ -инструмент – базисным. Вводя такой базис, мы рассчитываем на интуицию читателя и его определенное владение теорией обобщенных функций. Инструменты $D(s), s \in X$, служат связующим звеном континуального рынка с дискретным по страйкам рынком опционов, на которых подобную роль могли бы играть, в частности, нормированные простейшие (образованные соседними страйками) баттерфляи.

Цена и средний доход портфеля G с представлением (1) определяются соответственно формулами

$$(2) \quad |G| = \int_X g(s) |D(s)| ds = \int_X g(s) c(s) ds,$$

$$(3) \quad \|G\| = \int_X g(s) \|D(s)\| ds = \int_X g(s) p(s) ds.$$

Цена δ -инструментов $D(s)$ и их средний доход определяются соответственно формулами

$$|D(s)| = c(s), \quad \|D(s)\| = p(s), \quad s \in X.$$

Неотрицательная функция $c(x), x \in X$, должна задаваться рынком.

Рынок, на котором свободно обращаются все подобные δ -инструменты, называем δ -рынком.

Особый интерес представляют такие континуальные комбинации δ -инструментов, как *индикаторы* $H\{M\}, M \subset X$. Платежной функцией $H\{M\}$ служит характеристическая функции $\chi_M(x)$ множества M , равная единице, если $x \in M$, и нулю, если $x \notin M$. В частности, безрисковый инструмент единичного объема U является индикатором $H\{X\}$. Его платежная функция тождественно равна единице. Имеют место представления

$$H\{M\} = \int_M D(s) ds, \quad U = H\{X\} = \int_X D(s) ds.$$

Цены этих инструментов и их средние доходы:

$$(4) \quad |\mathbf{H}\{M\}| = \int_M c(s)ds = \mathbf{C}\{M\}, \quad \|\mathbf{H}\{M\}\| = \int_M p(s)ds = \mathbf{P}\{M\};$$

$$(5) \quad |\mathbf{U}| = \int_X c(x)dx = 1/r, \quad \|\mathbf{U}\| = \int_X p(x)dx = 1.$$

Первое соотношение в (5) вводит параметр r , означающий *безрисковый относительный доход* за период, при этом $(r - 1)$ – *безрисковая доходность*. Без ущерба для общности выводов полагаем далее $r \equiv 1$. Тогда интеграл от цены $c(x)$ по $x \in X$ будет равен единице, сама цена приобретает свойства плотности вероятности, которую можно считать *порожденной рынком*; ее называем *стоимостной плотностью*. Отметим, что соотношения (4) и (5) фактически вводят *стоимостную* и *прогнозную меры* $\mathbf{C}\{\cdot\}$ и $\mathbf{P}\{\cdot\}$, а (2) и (3) можно записать иначе (\mathbf{E} – знак математического ожидания):

$$|\mathbf{G}| = \mathbf{E}_{\mathbf{C}}g(\mathbf{X}), \quad \|\mathbf{G}\| = \mathbf{E}_{\mathbf{P}}g(\mathbf{X}).$$

На теоретическом рынке среди индикаторов существуют также инструменты \mathbf{U}_s^+ , \mathbf{U}_s^- – бинарные колл и пут соответственно с произвольным страйком s и платежными функциями

$$\pi(x; \mathbf{U}_s^+) = \chi_{[s, \infty)}(x), \quad \pi(x; \mathbf{U}_s^-) = \chi_{(-\infty, s)}(x), \quad x, s \in X.$$

Имеют место очевидные соотношения (формулы паритета)

$$\mathbf{U}_s^+ + \mathbf{U}_s^- = \mathbf{U}, \quad s \in X,$$

$$\mathbf{H}\{M\} = \mathbf{U}_a^+ - \mathbf{U}_b^+ = \mathbf{U}_b^- - \mathbf{U}_a^-, \quad M = [a, b), \quad a, b \in X.$$

Общеизвестные опционные инструменты колл и пут с произвольным страйком s (ценой исполнения) задаются соответственно соотношениями

$$\mathbf{C}_s = \int_{\{x \geq s\}} (x - s)\mathbf{D}(x)dx, \quad \mathbf{P}_s = \int_{\{x < s\}} (s - x)\mathbf{D}(x)dx, \quad x, s \in X.$$

Подобные представления можно дать и для прочих известных инструментов: спрэдов, баттерфляев, кондоров и др.

Методы построения оптимальных портфелей, которые излагаются в дальнейшем, должны учитывать различие между плотностями $c(x)$ и $p(x)$ (мерами $\mathbf{C}\{\cdot\}$ и $\mathbf{P}\{\cdot\}$). В связи с этим особое значение приобретает *функция относительных доходов*

$$(6) \quad \rho(x) = p(x)/c(x), \quad x \in X,$$

Такие функции под названием *функций правдоподобия* часто встречаются в математической статистике, являясь объектом приложения известной *процедуры Неймана–Пирсона* в задачах проверки статистических гипотез (см., например, [1, 2, 4]).

Для задач инвестирования на теоретическом континуальном рынке возможны разные постановки. Перечислим некоторые наиболее важные из них, включая исходную.

Задача СО (исходная). Заданы инвестиционная сумма и ф.р.п. инвестора $\phi(\varepsilon)$. Ищется портфель, доставляющий $\max E q$ при условии $P\{q \geq \phi(\varepsilon)\} \geq 1 - \varepsilon$ для всех $\varepsilon \in [0, 1]$.

Задача СВ (базовая). Инвестиционная сумма не задается, а $\phi(\varepsilon)$ – ф.р.п. инвестора. Ищется *регулярный* портфель, доставляющий $\min A$ при условии $P\{q \geq \phi(\varepsilon)\} \geq 1 - \varepsilon$ для всех $\varepsilon \in [0, 1]$.

Задача СН (однородная). Заданы инвестиционная сумма $S (> 0)$ и ф.р.п. в виде $b\phi(\varepsilon)$, где b – масштабирующий параметр. Ищутся *регулярный* портфель и значение параметра b , доставляющие $\max E q$ при условии $P\{q \geq \phi(\varepsilon)\} \geq 1 - \varepsilon$ для всех $\varepsilon \in [0, 1]$. Параметр b находится из условия $S = bA$.

Как мы отмечали ранее, исходной задаче *СО* свойственны вырожденные решения. *Задача СВ*, напротив, сразу требует построения регулярного портфеля, но минимальной инвестиционной суммой, и потому является основой для построения решений для прочих задач. *Задача СН* является примером применения задачи *СВ* к решению задачи *СО* на основе результатов леммы 9 следующего раздела в частном случае однородной по масштабу инвестиции ф.р.п..

3. Основные теоретические результаты

Процедура Неймана–Пирсона. Строится семейство множеств $\mathbf{Z} = \{Z(\tau), \tau \geq 0\}$ по правилу: $x \in Z(\tau)$ тогда и только тогда, когда для функции (6) имеет место $\rho(x) \leq \tau$. Рассматривается функция $f_P(\tau) = P\{Z(\tau)\}$, $\tau \in [\tau', \tau'']$, $\tau' = \min_x \rho(x)$, $\tau'' = \max_x \rho(x)$. Семейство \mathbf{Z} – неубывающее по τ , а потому и функция $f_P(\tau)$ – неубывающая. Кроме того, $0 \leq f_P(\tau) \leq 1$ и $f_P(\infty) = 1$. Для $\varepsilon \in [0, 1]$

находится множество X_ε из этого семейства такое, что $\mathbf{P}\{X_\varepsilon\} = \varepsilon$. Введенная функция как раз и устанавливает взаимосвязь τ и ε : $\varepsilon = f_P(\tau)$ (и $\tau = f_P^{-1}(\varepsilon)$), $\tau \in [\tau', \tau'']$, $\varepsilon \in [0, 1]$.

Если функция $f_P(\tau)$ удовлетворяет требованиям: а) $f_P(\tau) \rightarrow 1$ при $\tau \rightarrow \infty$; б) $f_P(\tau)$ непрерывна и $f_P(0) = 0$; в) $f_P(\tau)$ строго возрастает, то пару мер $(\mathbf{C}\{\cdot\}, \mathbf{P}\{\cdot\})$ называем *регулярной (невыврожденной)*, в противном случае – *выврожденной*. □

Для регулярной пары $(\mathbf{C}\{\cdot\}, \mathbf{P}\{\cdot\})$ имеет место

Лемма 1. Для любого $\varepsilon \in [0, 1]$ множество X_ε существует и единственно, при этом стоимость $\gamma(\varepsilon) = \mathbf{C}\{X_\varepsilon\}$ максимальна среди всех $\mathbf{C}\{Y\}$ с $\mathbf{P}\{Y\} = \varepsilon$, $Y \subset X$.

Доказательство можно найти, например, в [1, 2, 4].

В процедуре Неймана–Пирсона и в лемме 1 вводятся две функции: $f_P(\tau) = \mathbf{P}\{Z(\tau)\}$, $\tau \geq 0$, и $\gamma(\varepsilon) = \mathbf{C}\{X_\varepsilon\}$, $\varepsilon \in [0, 1]$. Введем еще функцию $f_C(\tau) = \mathbf{C}\{Z(\tau)\}$, $\tau \geq 0$. Функцию $f_P(\tau)$, $\tau \geq 0$, называем *прогнозной функцией*, $f_C(\tau)$, $\tau \geq 0$, – *стоимостной функцией*, $\gamma(\varepsilon)$, $\varepsilon \in [0, 1]$, – *диссонантой*.

Лемма 2. Функции $f_P(\tau)$ и $f_C(\tau)$ являются функциями распределения относительного дохода $\rho(X)$ для мер $\mathbf{P}\{\cdot\}$ и $\mathbf{C}\{\cdot\}$ соответственно, где X – будущая цена базового актива.

Доказательство сразу следует из пары цепочек равенств

$$(7) \quad f_P(\tau) = \mathbf{P}\{Z(\tau)\} = \mathbf{P}\{\rho(X) \leq \tau\} = F_{P, \rho(X)}(\tau),$$

$$f_C(\tau) = \mathbf{C}\{Z(\tau)\} = \mathbf{C}\{\rho(X) \leq \tau\} = F_{C, \rho(X)}(\tau). \quad \square$$

Теорема 3. Диссонанта $\gamma(\varepsilon)$, $\varepsilon \in [0, 1]$, является суперпозицией функций $f_C(\tau)$ и $f_P^{-1}(\varepsilon)$:

$$(8) \quad \gamma(\varepsilon) = f_C(f_P^{-1}(\varepsilon)) \left(= F_{C, \rho(X)}(F_{P, \rho(X)}^{-1}(\varepsilon)) \right), \quad \varepsilon \in [0, 1].$$

Доказательство. В соответствии с процедурой Неймана–Пирсона $f_P(\tau) = \mathbf{P}\{Z(\tau)\}$, а $Z(\tau) = X_\varepsilon$, если $\varepsilon = f_P(\tau)$. Поэтому для всех $\varepsilon \in [0, 1]$ имеет место

$$\gamma(\varepsilon) = \mathbf{C}\{X_\varepsilon\} = \mathbf{C}\{Z(\tau)\} \Big|_{\tau=f_P^{-1}(\varepsilon)} = f_C(\tau) \Big|_{\tau=f_P^{-1}(\varepsilon)} = f_C(f_P^{-1}(\varepsilon)). \quad \square$$

Важное свойство диссонанты $\gamma(\varepsilon)$ устанавливает

Лемма 4. Функция $\gamma(\varepsilon)$, $\varepsilon \in [0, 1]$, – вогнутая, $\gamma(0) = 0$, $\gamma(1) = 1$.

Доказательство. Из теоремы 3 следует, что функция $\gamma(\varepsilon)$ – монотонно неубывающая, $\gamma(0) = 0$, $\gamma(1) = 1$. Мы установим ее вогнутость, если покажем, что для любого $h > 0$ имеет место

$$(9) \quad \gamma(\varepsilon + h) - \gamma(\varepsilon) \leq \gamma(\varepsilon) - \gamma(\varepsilon - h).$$

Вводятся события

$$(10) \quad D^- = X_{\varepsilon} - X_{\varepsilon-h} = \{x \mid \tau_{\varepsilon-h} < \rho(x) \leq \tau_{\varepsilon}\},$$

$$(11) \quad D^+ = X_{\varepsilon+h} - X_{\varepsilon} = \{x \mid \tau_{\varepsilon} < \rho(x) \leq \tau_{\varepsilon+h}\}.$$

Очевидно, $\mathbf{P}(D^-) = \mathbf{P}(D^+) = h$, и потому

$$\mathbf{C}\{D^-\} \geq \mathbf{P}\{D^-\}/\tau_{\varepsilon} = \mathbf{P}\{D^+\}/\tau_{\varepsilon} \geq \mathbf{C}\{D^+\}.$$

Здесь левое неравенство следует из (10), правое – из (11). Из полученного соотношения вытекает (9). Лемма доказана. \square

Следующая теорема отчасти уточняет и усиливает лемму 4.

Теорема 5. В условиях дифференцируемости функций $f_P(\tau)$ и $f_C(\tau)$ (вместе с плотностями $p(x)$ и $c(x)$)

$$(i) \quad l(\tau) = f'_P(\tau)/f'_C(\tau) = \tau, \quad \tau \in [\tau', \tau''],$$

$$(ii) \quad \gamma'(\varepsilon) = 1/f_P^{\leftarrow}(\varepsilon), \quad \varepsilon \in [0, 1].$$

Доказательство. Рассмотрим растущую систему множеств

$$Z_{\tau} = \{x \mid \rho(x) \leq \tau\}, \quad \tau \in [\tau', \tau''],$$

и введем приращения (для $h > 0$)

$$(12) \quad D_{\tau,h} \equiv Z_{\tau+h} - Z_{\tau} = \{x \mid \tau < \rho(x) \leq \tau + h\}.$$

Справедливы соотношения

$$\mathbf{P}\{Z_{\tau}\} = f_P(\tau), \quad \mathbf{C}\{Z_{\tau}\} = f_C(\tau),$$

$$\mathbf{P}\{Z_{\tau+h}\} = f_P(\tau + h), \quad \mathbf{C}\{Z_{\tau+h}\} = f_C(\tau + h),$$

а также

$$\mathbf{P}\{D_{\tau,h}\} = \mathbf{P}\{Z_{\tau+h}\} - \mathbf{P}\{Z_{\tau}\} = f_P(\tau + h) - f_P(\tau),$$

$$\mathbf{C}\{D_{\tau,h}\} = \mathbf{C}\{Z_{\tau+h}\} - \mathbf{C}\{Z_{\tau}\} = f_C(\tau + h) - f_C(\tau).$$

После деления первого соотношения на второе находим

$$(13) \quad (f_P(\tau + h) - f_P(\tau)) / (f_C(\tau + h) - f_C(\tau)) = \mathbf{P}\{D_{\tau,h}\} / \mathbf{C}\{D_{\tau,h}\}.$$

Переходя в этом равенстве к пределу при $h \rightarrow 0$, слева получаем отношение $l(\tau)$. С другой стороны, из непрерывности $p(x)$

и $c(x)$ на интервале $D_{\tau,h}$ для правой части равенства в соответствии с (12) можно дать при малых h приближенную оценку

$$P\{D_{\tau,h}\}/C\{D_{\tau,h}\} \approx p(x^*)/c(x^*) = \rho(x^*) \approx \tau, \quad x^* \in D_{\tau,h},$$

которая тем точнее, чем меньше h . Потому из (13) следует

$$l(\tau) = f'_P(\tau)/f'_C(\tau) = \lim_{h \rightarrow 0} P\{D_{\tau,h}\}/C\{D_{\tau,h}\} = \tau,$$

что доказывает утверждение (i) теоремы.

Теперь, учитывая представление (8) и равенство $\varepsilon = f_P(\tau)$ и применяя правила дифференцирования для суперпозиции функций и обратной функции, получаем утверждение (ii) теоремы:

$$\begin{aligned} \gamma'(\varepsilon) &= df_C(f_P^{\leftarrow}(\varepsilon))/d\varepsilon = f'_C(f_P^{\leftarrow}(\varepsilon)) \times df_P^{\leftarrow}(\varepsilon)/d\varepsilon = \\ &= f'_C(f_P^{\leftarrow}(\varepsilon))/f'_P(f_P^{\leftarrow}(\varepsilon)) = 1/l(f_P^{\leftarrow}(\varepsilon)) = 1/f_P^{\leftarrow}(\varepsilon), \quad \varepsilon \in [0, 1]. \end{aligned}$$

Теорема доказана. \square

Процедуру упорядочения, фактически, реализует однозначная функция упорядочения $w(x)$, определяемая правилом

$$(14) \quad w(x) = \varepsilon (= f_P(\tau)) \Leftrightarrow x \in \Gamma_\varepsilon = \lim_{\varepsilon' \uparrow \varepsilon} (X_\varepsilon - X_{\varepsilon'}), \quad \varepsilon \in [0, 1].$$

Здесь Γ_ε – совокупность граничных точек множества X_ε .

Обратная к $w(x)$ функция, вообще говоря, неоднозначна, и ее условно обозначаем $w^{\leftarrow}(x)$, подчеркивая такое свойство функции полужирным шрифтом:

$$\mathbf{w}^{\leftarrow}(\varepsilon) = \{x | w(x) = \varepsilon\} = \{x | x \in \Gamma_\varepsilon\}, \quad \varepsilon \in [0, 1].$$

Лемма 6. Функция упорядочения $w(x)$ имеет представление

$$w(x) = f_P(\rho(x)), \quad x \in X.$$

Доказательство. В соответствии с соотношением (14) функция упорядочения $w(x)$ является суперпозицией прогнозной функции $f_P(\tau)$ и функции относительных доходов $\rho(x)$, и потому

$$(15) \quad w(x) = f_P(\tau) \Big|_{\tau=\rho(x)} = f_P(\rho(x)), \quad x \in X. \quad \square$$

Прямым следствием утверждения леммы 6 является

Теорема 7. Случайная величина $w(X)$ равномерно распределена на отрезке $[0, 1]$.

Доказательство. Из равенств (7) и (15) следует, что для случайной величины $w(X)$ имеет место представление

$$w(X) = f_P(\rho(X)) = F_{P, \rho(X)}(\rho(X)) \in [0, 1].$$

Поэтому для функции распределения $w(X)$ при *всех* $\varepsilon \in [0, 1]$

$$\begin{aligned} F_{P, w(X)}(\varepsilon) &= P\{w(X) \leq \varepsilon\} = P\{F_{P, \rho(X)}(\rho(X)) \leq \varepsilon\} = \\ &= P\{\rho(X) \leq F_{P, \rho(X)}^{\leftarrow}(\varepsilon)\} = F_{P, \rho(X)}(F_{P, \rho(X)}^{\leftarrow}(\varepsilon)) = \varepsilon. \end{aligned}$$

Тем самым теорема доказана. \square

Таким образом, случайная величина $w(X)$ имеет общее (равномерное) для любых инвесторов распределение, не зависящее от их рискованных предпочтений и прогноза.

Теперь обратимся к рискованным аспектам инвестора.

Теорема 8. Для случайной величины $q = g(X) = \phi(w(X))$, где $\phi(\varepsilon)$, $\varepsilon \in [0, 1]$, – строго монотонно возрастающая и непрерывная функция, справедливы соотношения:

- (i) $F_{P, q}(z) = \phi^{\leftarrow}(z)$, $z \in [\phi(0), \phi(1)]$;
- (ii) $P\{q \geq \phi(\varepsilon)\} = 1 - \varepsilon$, $\varepsilon \in [0, 1]$.

(Представление функций распределения для простоты записи ограничивается лишь интервалом их роста.)

Доказательство. Из условий теоремы вытекает существование функции $\phi^{\leftarrow}(z)$, $z \in [\phi(0), \phi(1)]$. Для $z \in [\phi(0), \phi(1)]$ имеем

$$F_{P, q}(z) = P\{\phi(w(X)) \leq z\} = P\{w(X) \leq \phi^{\leftarrow}(z)\} = \phi^{\leftarrow}(z),$$

что доказывает утверждение (i) теоремы.

Для доказательства утверждения (ii) в случае регулярной пары мер $(C\{\cdot\}, P\{\cdot\})$ достаточно воспользоваться цепочкой равенств, последнее из которых следует из теоремы 7,

$$P\{q \leq \phi(\varepsilon)\} = P\{\phi(w(X)) \leq \phi(\varepsilon)\} = P\{w(X) \leq \phi^{\leftarrow}(\phi(\varepsilon))\} = \varepsilon.$$

Теорема полностью доказана. \square

Устанавливая четкую взаимосвязь функции распределения доходов инвестора с его ф.р.п., именно теорема 8 позволяет инвестору выбором ф.р.п. формировать по своему желанию распределение своих инвестиционных доходов. Непосредственным следствием утверждения (ii) теоремы 8 является

Лемма 9. Случайный доход q портфеля

$$(16) \mathbf{G} = b \int_{\mathcal{X}} \phi(w(x)) \mathbf{D}(x) dx, \quad b > 0,$$

для всех $\varepsilon \in [0, 1]$ удовлетворяют равенствам

$$P\{q \geq b\phi(\varepsilon)\} = 1 - \varepsilon. \quad \square$$

Строящиеся портфели можно сравнивать и по привычным интегральным характеристикам. Для них используются также специальные обозначения: R – *средний доход*, A – *стоимость* (инвестиционная сумма), y – *доходность*. Имеем

$$(17) A = |\mathbf{G}| = \int_{\mathcal{X}} g(s) |\mathbf{D}(s)| ds = \int_{\mathcal{X}} g(s) c(s) ds,$$

$$(18) R = \|\mathbf{G}\| = \int_{\mathcal{X}} g(s) \|\mathbf{D}(s)\| ds = \int_{\mathcal{X}} g(s) p(s) ds, \quad y = R/A - 1.$$

Часто бывает более удобным анализировать эти показатели в терминах переменных ε или τ . Фактически в представлениях (17) и (18) для оптимальной весовой функции $g(s)$ портфеля речь идет о замене переменной s переменными ε или τ .

Однако обратная к $g(s)$ функция часто бывает неоднозначной, и хотя такая замена разбиением на монотонные ветви осуществима, ее формальное представление громоздко. Мы предложим иной способ, связанный с дискретизацией задачи и последующим переходом к предельной континуальной схеме.

Для дискретного случая задаемся числом n интервалов разбиения множеств изменения параметров ε , γ и ϕ . Нетрудно видеть, что последовательное применение процедуры Неймана–Пирсона для вероятностных уровней ε_i и соответствующих им уровней γ_i и ϕ_i , $i \in I = \{1, \dots, n\}$, дает следующие представления для портфеля (16) при $b = 1$, его стоимости и среднего дохода, а также для требований многоступенчатого критерия VaR:

$$\mathbf{G}^{(n)} = \sum_{i=1}^n (\phi_i - \phi_{i-1}) \mathbf{H}\{\mathbf{X} - X_{\varepsilon_i}\},$$

$$A^{(n)} = \sum_{i=1}^n (\phi_i - \phi_{i-1})(1 - \gamma_i) = \sum_{i=1}^n \phi_i (\gamma_i - \gamma_{i-1}), \quad \phi_0 = 0, \gamma_0 = 0,$$

$$R^{(n)} = \sum_{i=1}^n (\phi_i - \phi_{i-1})(1 - \varepsilon_i) = \sum_{i=1}^n \phi_i (\varepsilon_i - \varepsilon_{i-1}), \quad \phi_0 = 0, \varepsilon_0 = 0,$$

$$P\{q \geq \phi_i\} \geq 1 - \varepsilon_i \quad \text{для всех } i \in I.$$

Заменяя в этих формулах разности дифференциалами по ε , γ и ϕ , а суммы – интегралами, для континуального случая имеем:

$$G = \int_0^1 \mathbf{H} \{ \mathbf{X} - X_\varepsilon \} d\phi(\varepsilon),$$

$$A = \int_0^1 (1 - \gamma(\varepsilon)) d\phi(\varepsilon) = \int_0^1 \phi(\varepsilon) d\gamma(\varepsilon),$$

$$R = \int_0^1 (1 - \varepsilon) d\phi(\varepsilon) = \int_0^1 \phi(\varepsilon) d\varepsilon,$$

$$\mathbf{P} \{ q \geq \phi(\varepsilon) \} \geq 1 - \varepsilon \quad \text{для всех } \varepsilon \in [0, 1].$$

Таким образом, справедлива

Теорема 10. Средний доход R и стоимость A портфеля (16) при $b = 1$ определяются формулами

$$(19) \quad R = \int_0^1 \phi(\varepsilon) d\varepsilon, \quad A = \int_0^1 \phi(\varepsilon) d\gamma(\varepsilon) = \int_0^1 \phi(\varepsilon) \gamma'(\varepsilon) d\varepsilon. \quad \square$$

Из этой теоремы с помощью определений функций $f_P(\tau)$ и $f_C(\tau)$ и утверждений леммы 2 и теорем 3, 5 можно получить эквивалентные представления в терминах параметра τ . В этом случае замена $\varepsilon = f_P(\tau)$ осуществляется без труда. Имеет место

Теорема 11. При $b = 1$ справедливы соотношения

$$R = \int_0^1 \phi(\varepsilon) d\varepsilon = \int_{\tau'}^{\tau''} \phi(f_P(\tau)) df_P(\tau) = \int_{\tau'}^{\tau''} \phi(f_P(\tau)) f'_P(\tau) d\tau,$$

$$A = \int_0^1 \phi(\varepsilon) d\gamma(\varepsilon) = \int_{\tau'}^{\tau''} \phi(f_P(\tau)) df_C(\tau) = \int_{\tau'}^{\tau''} \phi(f_P(\tau)) f'_C(\tau) d\tau.$$

Доказательство. Первое из соотношений является очевидным результатом замены $\varepsilon = f_P(\tau)$, а второе – прямое следствие теоремы 3, поскольку при такой замене в соответствии с представлением (8) $\gamma(\varepsilon) = f_C(\tau)$. \square

Наконец, нам остается установить оптимальность сконструированного портфеля.

Теорема 12. Построенное в соответствии с процедурой Неймана–Пирсона семейство $\mathbf{Z} = \{X_\varepsilon, \varepsilon \in [0, 1]\}$ доставляет минимум инвестиционной сумме A , определяемой формулой (19).

Доказательство. Пусть $\mathbf{V} = \{Y_\varepsilon, \varepsilon \in [0, 1]\}$ – произвольное иное континуальное семейство, такое что $\mathbf{P}\{Y_\varepsilon\} = \mathbf{P}\{X_\varepsilon\} = \varepsilon$ для всех $\varepsilon \in [0, 1]$. Будем считать, что $\chi(\varepsilon) = \mathbf{C}(Y_\varepsilon)$ – функция с ограниченной вариацией (чего вполне достаточно для приложений), а также $\chi(0) = 0, \chi(1) = 1$. По лемме 1 для всех $\varepsilon \in [0, 1]$ справедливо неравенство $\chi(\varepsilon) \leq \gamma(\varepsilon)$.

Обозначая через A_V инвестиционную сумму для V , сравниваем A и A_V . Интегрируя по частям с учетом (19), получаем

$$\begin{aligned} A_V - A &= \int_0^1 \phi(\varepsilon) d\chi(\varepsilon) - \int_0^1 \phi(\varepsilon) d\gamma(\varepsilon) = \\ &= \int_0^1 (\gamma(\varepsilon) - \chi(\varepsilon)) d\phi(\varepsilon) \geq 0, \end{aligned}$$

что и доказывает теорему. \square

4. Алгоритм оптимизации и иллюстративный пример

Здесь подытоживается содержание предыдущего раздела в отношении построения оптимального по CC -VaR портфеля и перечисляются наиболее важные введенные там формальные агрегаты, образующие ядро алгоритма оптимизации. При этом речь идет о решении основной задачи CB . Приводится и иллюстративный пример со вторым распределением Лапласа.

4.1. АЛГОРИТМ ОПТИМИЗАЦИИ

Основанный на применении процедуры Неймана–Пирсона алгоритм оптимизации использует результаты предыдущего раздела, представленные в нем в форме утверждений лемм и теорем с привлечением необходимых обозначений. Имеем

$$\begin{aligned} X_\varepsilon &= \{\rho(x) \leq \tau\}, \quad \tau \in [\tau', \tau''], \\ f_P(\tau) &= P\{\rho(x) \leq \tau\} = F_{P;\rho(x)}(\tau), \\ f_C(\tau) &= C\{\rho(x) \leq \tau\} = F_{C;\rho(x)}(\tau), \\ \gamma(\varepsilon) &= f_C(f_P^{\leftarrow}(\varepsilon)) = F_{C;\rho(x)}(F_{P;\rho(x)}^{\leftarrow}(\varepsilon)), \quad \varepsilon \in [0, 1], \\ \gamma'(\varepsilon) &= 1/f_P^{\leftarrow}(\varepsilon), \\ w(x) &= f_P(\tau)|_{\tau=\rho(x)} = f_P(\rho(x)), \\ g(x) &= \phi(w(x)), \quad x \in X, \\ G &= \int_X g(x) D(x) dx, \\ q &= g(X) = \phi(w(X)), \end{aligned}$$

$$R = \int_0^1 \phi(\varepsilon) d\varepsilon = \int_{\tau'}^{\tau''} \phi(f_P(\tau)) f'_P(\tau) d\tau \quad (=E_P q),$$

$$A = \int_0^1 \phi(\varepsilon) d\gamma(\varepsilon) = \int_{\tau'}^{\tau''} \phi(f_P(\tau)) f'_C(\tau) d\tau \quad (=E_C q),$$

$$y = R/A - 1,$$

$$F_{P,q}(z) = \phi^{\leftarrow}(z), \quad z \in [\phi(0), \phi(1)],$$

$$P\{q \geq \phi(\varepsilon)\} \equiv 1 - \varepsilon \quad \text{для всех } \varepsilon \in [0, 1].$$

4.2. ВТОРОЕ РАСПРЕДЕЛЕНИЕ ЛАПЛАСА И ЕГО СВОЙСТВА

Второе распределение Лапласа (его называют также двусторонним экспоненциальным распределением), которое будем обозначать $\text{Exp}(\mu, \alpha)$, имеет плотность

$$\frac{1}{2\alpha} \exp\left(-\frac{|x-\mu|}{\alpha}\right), \quad x \in \mathfrak{R},$$

с математическим ожиданием $\text{EX} = \mu$, дисперсией $\text{DX} = 2\alpha^2$ и куртозисом, равным 6 (эксцессом -3).

Выбор данного типа распределений обусловлен исключительно нашим желанием довести исследование примера до аналитического решения. В силу относительной сложности алгоритма в подобных задачах это сделать весьма затруднительно. Хотя и некоторые свойства говорят в пользу естественности такого выбора: функции плотности унимодальны, а эксцесс выше аналогичного параметра, например, для нормального распределения, что отвечает ныне «модным» в рыночных кругах «тяжелым» хвостам распределений.

Положим $p(x) \sim \text{Exp}(\mu, \alpha)$, $c(x) \sim \text{Exp}(\mu, \beta)$ и $\beta \neq \alpha$, т.е. предполагается, что центры симметрии обеих плотностей одинаковы и равны μ . В силу симметрии задачи полагаем $\mu = 0$, т.е. под x в формулах понимаем не будущую цену базового актива, а ее отклонение от среднего.

Поскольку

$$p(x) = \frac{1}{2\alpha} \exp\left(-\frac{|x-\mu|}{\alpha}\right), \quad c(x) = \frac{1}{2\beta} \exp\left(-\frac{|x-\mu|}{\beta}\right), \quad x \in \mathfrak{R},$$

функция относительных доходов при $\mu = 0$ принимает вид

$$\rho(x) = \frac{\beta}{\alpha} \exp\left(\frac{\alpha-\beta}{\alpha\beta} |x|\right).$$

Параметризация $\rho(x) = \tau$, где $0 \leq \tau \leq \beta/\alpha$ при $\alpha \leq \beta$ и $\tau > \beta/\alpha$ при $\alpha > \beta$, порождает две обратные к $\rho(x)$ функции (в парных равенствах индексу «1» слева отвечает нижний знак справа)

$$\rho_{1,2}^{\leftarrow}(\tau) = \pm \rho^{\leftarrow}(\tau), \quad \text{где } \rho^{\leftarrow}(\tau) = \frac{\alpha\beta}{\beta-\alpha} \ln\left(\frac{\beta}{\alpha\tau}\right).$$

Поведение функции $\rho(x)$ определяется значениями α и β . Параметр $\kappa = \alpha/\beta$ порождает два нетривиальных случая: $\kappa < 1$ и $\kappa > 1$. В первом случае $\rho(x)$ возрастает при $x < 0$, убывает при $x > 0$ и в точке $x = x^* = 0$ имеет максимум, во втором – убывает при $x < 0$, возрастает при $x > 0$ и в точке $x = 0$ имеет минимум.

В первом случае прогнозная дисперсия меньше рыночной, т.е. инвестор рассчитывает на более стабильное поведение базового актива, чем рынок в целом. Во втором – рынок и инвестор меняются ролями. По терминологии рынка в первом случае инвестор продает волатильность («изменчивость»), во втором – покупает. В обоих случаях принимаем $\phi(\varepsilon) = \varepsilon^\lambda$, $\varepsilon \in [0, 1]$, $\lambda > 0$. При рассмотрении случая с продажей волатильности будет удобно еще использовать обозначение $\theta = \lambda/\alpha$.

4.3. ПРОДАЖА ВОЛАТИЛЬНОСТИ

В случае $\kappa < 1$ инвестор считает рынок менее волатильным, чем об этом свидетельствуют цены опционов. Применение алгоритма порождает следующие симметричные конструкции:

$$Z(\tau) = \left\{ x \mid |x| \geq \rho^{\leftarrow}(\tau) = \frac{\alpha\beta}{\beta-\alpha} \ln\left(\frac{\beta}{\alpha\tau}\right) \right\},$$

$$X_\varepsilon = \left\{ x \mid |x| \geq x_\varepsilon \right\}, \quad -x'_\varepsilon = x''_\varepsilon = x_\varepsilon = \rho^{\leftarrow}(\tau) \geq 0,$$

$$\tau = \rho(x'_\varepsilon) = \rho(x''_\varepsilon) = \rho(x_\varepsilon),$$

$$P\{X_\varepsilon\} = 2 \int_{x_\varepsilon}^{\infty} p(x) dx = \alpha^{-1} \int_{x_\varepsilon}^{\infty} e^{-x/\alpha} dx = \exp(-x_\varepsilon/\alpha) = \varepsilon,$$

$$x_\varepsilon = -\alpha \ln \varepsilon - (1 - \varepsilon/2)\text{-квантиль распределения } \text{Exp}(0, \alpha),$$

$$\Gamma_\varepsilon = \{x'_\varepsilon, x''_\varepsilon\} = \{-x_\varepsilon, x_\varepsilon\} - \text{множество граничных точек } X_\varepsilon.$$

Прогнозную функцию (для $0 \leq \tau \leq \kappa^{-1}$) и обратную к ней, а также стоимостную функцию дают соотношения

$$f_p(\tau) = \exp(-\rho^{\leftarrow}(\tau)/\alpha) = \exp\left(-\frac{\beta}{\beta-\alpha} \ln\left(\frac{\beta}{\alpha\tau}\right)\right) = \left(\frac{\alpha\tau}{\beta}\right)^{\frac{\beta}{\beta-\alpha}} = (\kappa\tau)^{\frac{1}{1-\kappa}},$$

$$f_p^{\leftarrow}(\varepsilon) = \kappa^{-1} \varepsilon^{1-\kappa},$$

$$f_C(\tau) = \exp(-\rho^{\leftarrow}(\tau)/\beta) = \exp\left(-\frac{\alpha}{\beta-\alpha} \ln\left(\frac{\beta}{\alpha\tau}\right)\right) = \left(\frac{\alpha\tau}{\beta}\right)^{\frac{\alpha}{\beta-\alpha}} = (\kappa\tau)^{\frac{\kappa}{1-\kappa}}.$$

Функция упорядочения с двумя симметричными ветвями, весовая функция и диссонанта с ее производной имеют вид

$$u_{1,2}(\varepsilon) = \{x'_\varepsilon, x''_\varepsilon\} = \pm x_\varepsilon = \mp \alpha \ln \varepsilon,$$

$$w(x) = \{u_1^{\leftarrow}(x), x < 0; u_2^{\leftarrow}(x), x > 0\} = \exp\left(-\frac{|x|}{\alpha}\right),$$

$$g(x) = \phi(w(x)) = e^{-\theta|x|}, \quad g'(x) = \mp \theta e^{-\theta|x|}, \quad g''(x) = \theta^2 e^{-\theta|x|},$$

$$\gamma(\varepsilon) = \mathbf{C}\{X_\varepsilon\} = \beta^{-1} \int_{-\alpha \ln \varepsilon}^{\infty} e^{-x/\beta} dx = \varepsilon^\kappa,$$

$$\gamma'(\varepsilon) = \kappa \varepsilon^{\kappa-1} \quad (= 1/f_p^{\leftarrow}(\varepsilon)).$$

Основные показатели инвестиции:

$$A = \int_0^1 \phi(\varepsilon) \gamma'(\varepsilon) d\varepsilon = \kappa(\kappa + \lambda)^{-1},$$

$$R = \int_0^1 \phi(\varepsilon) d\varepsilon = (1 + \lambda)^{-1}, \quad y = \lambda(1 - \kappa) / (\kappa(1 + \lambda)) > 0.$$

Оптимальный портфель одновременно в пут- и колл-спрэдах дается смешанным представлением (естественного происхождения: при $x \leq 0$ используются путы, при $x \geq 0$ – коллы)

$$G = \left(U - \theta \int_{-\infty}^0 e^{\theta x} dP(x) + \theta \int_0^{\infty} e^{-\theta x} dC(x) \right),$$

а в терминах самих путов и коллов – представлением

$$G = U - \theta(P(0) + C(0)) + \theta^2 \int_{-\infty}^0 e^{\theta x} P(x) dx + \theta^2 \int_0^{\infty} e^{-\theta x} C(x) dx.$$

Платежная функция этого портфеля однозначно определяется параметром θ , сконцентрировавшим в себе как рисковые предпочтения инвестора (числитель λ), так и его прогноз (знаменатель α). Чем меньше α (больше расхождение прогноза и рынка) и больше λ (больше склонность инвестора к риску), тем более острым выглядит пик функции $g(x)$ при $x = 0$ и больше доходность y .

Отметим, что построенный портфель является некоторым континуальным усложнением длинного баттерфляя, реализующего на реальном рынке ту же продажу волатильности.

4.4. ПОКУПКА ВОЛАТИЛЬНОСТИ

В случае $\kappa > 1$ инвестор считает рынок более волатильным, чем об этом свидетельствуют цены опционов. В данном случае упорядочение функции $\rho(x)$ по величине также очевидно, а семейство \mathbf{Z} состоит из множеств, дополнительных к случаю $\kappa < 1$. Сообразно этому формулы видоизменяются, и мы имеем

$$Z(\tau) = \left\{ x \mid |x| \leq \rho^{\leftarrow}(\tau) = \frac{\alpha\beta}{\beta-\alpha} \ln\left(\frac{\beta}{\alpha\tau}\right) \right\},$$

$$X_\varepsilon = \{x \mid |x| \leq x_\varepsilon\}, \quad -x'_\varepsilon = x''_\varepsilon = x_\varepsilon = \rho^{\leftarrow}(\tau),$$

$$\tau = \rho(x'_\varepsilon) = \rho(x''_\varepsilon) = \rho(x_\varepsilon),$$

$$P\{X_\varepsilon\} = 2 \int_0^{x_\varepsilon} p(x) dx = \alpha^{-1} \int_0^{x_\varepsilon} e^{-x/\alpha} dx = 1 - \exp(-x_\varepsilon/\alpha) = \varepsilon,$$

$$x_\varepsilon = -\alpha \ln(1 - \varepsilon) - \quad (1 + \varepsilon)/2\text{-квантиль} \quad \text{распределения}$$

$\text{Exp}(0, \alpha)$,

$$\Gamma_\varepsilon = \{x'_\varepsilon, x''_\varepsilon\} = \{-x_\varepsilon, x_\varepsilon\} - \text{множество граничных точек } X_\varepsilon.$$

Прогнозная функция (для $\tau \geq \kappa^{-1}$) и обратная к ней, а также стоимостная функция имеют вид

$$f_P(\tau) = 1 - \exp(-\rho^{\leftarrow}(\tau)/\alpha) = 1 - \exp\left(-\frac{\beta}{\beta-\alpha} \ln\left(\frac{\beta}{\alpha\tau}\right)\right) = 1 - (\kappa\tau)^{\frac{1}{1-\kappa}},$$

$$f_P^{\leftarrow}(\varepsilon) = \kappa^{-1} (1 - \varepsilon)^{1-\kappa},$$

$$f_C(\tau) = 1 - \exp(-\rho^{\leftarrow}(\tau)/\beta) = 1 - \exp\left(-\frac{\alpha}{\beta-\alpha} \ln\left(\frac{\beta}{\alpha\tau}\right)\right) = 1 - (\kappa\tau)^{\frac{\kappa}{1-\kappa}},$$

Функция упорядочения с двумя симметричными ветвями, весовая функция портфеля и диссонанта с производными даются представлениями

$$u_{1,2}(\varepsilon) = \{x'_\varepsilon, x''_\varepsilon\} = \pm x_\varepsilon = \mp \alpha \ln(1 - \varepsilon),$$

$$w(x) = \left\{ u_1^{\leftarrow}(x), x < 0; u_2^{\leftarrow}(x), x > 0 \right\} = 1 - \exp\left(-\frac{|x|}{\alpha}\right),$$

$$g(x) = \phi(w(x)) = \left(1 - \exp\left(-\frac{|x|}{\alpha}\right)\right)^\lambda,$$

$$g'(x) = \pm \frac{\lambda}{\alpha} e^{-|x|/\alpha} \left(1 - e^{-|x|/\alpha}\right)^{\lambda-1},$$

$$g'(\pm 0) = 0, \pm \alpha^{-1}, \pm \infty \quad \text{при } \lambda > 1, = 1, < 1,$$

$$g''(x) = \frac{\lambda^2}{\alpha^2} e^{-|x|/\alpha} \left(1 - e^{-|x|/\alpha}\right)^{\lambda-2} \left(e^{-|x|/\alpha} - \frac{1}{\lambda}\right),$$

$$\gamma(\varepsilon) = \mathbf{C}\{X_\varepsilon\} = \beta^{-1} \int_0^{-\alpha \ln(1-\varepsilon)} e^{-x/\beta} dx = 1 - (1-\varepsilon)^\kappa,$$

$$\gamma'(\varepsilon) = \kappa(1-\varepsilon)^{\kappa-1}.$$

Основные показатели инвестиции:

$$A = \int_0^1 \phi(\varepsilon) \gamma'(\varepsilon) d\varepsilon = \frac{\Gamma(1+\alpha/\beta)\Gamma(1+\lambda)}{\Gamma(1+\alpha/\beta+\lambda)}, \quad R = \int_0^1 \phi(\varepsilon) d\varepsilon = (1+\lambda)^{-1},$$

$$y = R/A - 1 = \frac{\Gamma(1+\alpha/\beta+\lambda)}{(1+\lambda)\Gamma(1+\alpha/\beta)\Gamma(1+\lambda)} - 1 = \frac{\Gamma(1+\alpha/\beta+\lambda)}{\Gamma(1+\alpha/\beta)\Gamma(2+\lambda)} - 1 > 0,$$

$$\Gamma(\nu) = \int_0^\infty x^{\nu-1} e^{-x} dx.$$

Оптимальный портфель одновременно в пут- и колл-спрэдах (также естественного происхождения) дается смешанным представлением

$$\mathbf{G} = \int_{-\infty}^0 |g'(x)| d\mathbf{P}(x) - \int_0^\infty g'(x) d\mathbf{C}(x).$$

Характер оптимального портфеля из самих путов и коллов зависит от параметра λ , определяющего поведение функции $g(x)$ в нуле. При $\lambda > 1$, $\lambda = 1$, $\lambda < 1$ получаем соответственно

$$\mathbf{G} = \int_{-\infty}^0 g''(x) \mathbf{P}(x) dx + \int_0^\infty g''(x) \mathbf{C}(x) dx,$$

$$\mathbf{G} = \frac{1}{\alpha} (\mathbf{P}(0) + \mathbf{C}(0)) - \frac{1}{\alpha^2} \int_{-\infty}^0 e^{-|x|/\alpha} \mathbf{P}(x) dx - \frac{1}{\alpha^2} \int_0^\infty e^{-|x|/\alpha} \mathbf{C}(x) dx,$$

$$\mathbf{G} = \int_{-\infty}^0 g''(x) (\mathbf{P}(x) - \mathbf{P}(0)) dx + \int_0^\infty g''(x) (\mathbf{C}(x) - \mathbf{C}(0)) dx.$$

При $\lambda < 1$ функция $g(x)$ имеет в нуле особенность типа «острие»; чем меньше λ , тем больше инструмент походит на безрисковый актив. При $\lambda = 1$ функция $g(x)$ имеет в нуле излом и напоминает платежную функцию короткого баттерфляя. При $\lambda > 1$ функция $g(x)$ в нуле непрерывна и $g'(0) = 0$; она напоминает платежную функцию кондора – комбинации длинного и короткого стрэнглов с ближе расположенными к нулю страйками длинного стрэнгла. В этом случае также легко прослеживаются качественные аналогии с реальным рынком опционов. (Спрэды, баттерфляя, стрэнглы, кондоры – специальные линейные комбинации, т.е. портфели, коллов и путов.)

5. Заключение

Предлагаемая теория оптимизации портфеля на рынке опционов при соблюдении требований $CC-VaR$ носит универсальный характер и она должна работать для любой пары распределений. Здесь принципиально не затрагиваются вопросы формирования цен на рынке, равно как и вопросы прогнозирования будущих цен базового актива. Первые инвестор принимает за данность, какова бы она ни была. За решение вторых инвестор вне зависимости от источника прогноза несет ответственность сам. В работе предлагается лишь инструментарий, позволяющий инвестору совместить оба распределения к своей выгоде, притом со своими же рисковыми предпочтениями.

Если отвлечься от распределений иллюстративного примера, а также некоторых других, допускающих аналитическое исследование лишь частично, то обычно приходится применять численные методы, связанные с дискретизацией задачи. Тем более если речь идет о реальных рынках опционов, на которых складывается дискретная по страйкам решетка цен. Но это – тема уже для других работ.

Также придется отказываться и от допущений, связанных с нашим использованием терминов «теоретический» и «идеальный», например, тестируя портфель на ценах покупателя и продавца реального рынка с учетом их спреда. В любом случае для приложений результаты теории следует рассматривать как первое приближение к оптимуму. В этом качестве она полезна уже в том, что позволяет избегать грубых ошибок с точки зрения выбранного критерия и вероятностного прогноза инвестора.

Литература

1. АГАСАНДЯН Г.А. *Финансовая инженерия и континуальный критерий VaR на рынке опционов* // Экономика и математические методы, – 2005. – Т. 41, №4. – С. 88–98.
2. АГАСАНДЯН Г.А. *Применение континуального критерия VaR на финансовых рынках.* – М.: ВЦ РАН, 2011. – 299 с.
3. АГАСАНДЯН Г.А. *Континуальный критерий VaR на многомерных рынках опционов.* – М.: ВЦ РАН, 2015. – 298 с.
4. КАСИМОВ Ю.Ф. *Основы теории оптимального портфеля ценных бумаг.* – М.: Филинь, 1998.
5. КРАМЕР Г. *Математические методы статистики.* – М.: Наука, 1975. –750 с. (Перевод с англ.: Cramer H. *Mathematical methods of statistics.* – Princeton University Press, 1946.)
6. МАРШАЛЛ Дж. Ф., БАНСАЛ В.К. *Финансовая инженерия.* – М.: ИНФРА-М, 1998. – 784 с.
7. AGASANDYAN G.A. *Optimal Behavior of an Investor in Option Market* // Int. Joint Conference on Neural Networks. The 2002 IEEE World Congress on Computational Intelligence (Honolulu, Hawaii, Mai 12-17, 2002). – P. 1859–1864.
8. ARTZNER P., DELBAEN F., EBER J.-M., HEATH D. *Coherent measures of risk* // *Mathematical Finance.* – 1999. – No. 9. – P. 203–229.
9. MARKOWITZ H. *Portfolio Selection* // *Journal of Finance.* – 1952. – No. 7. – P. 77–91.
10. ROCKAFELLAR R.T., URYASEV S. *Optimization of conditional value at risk* // *Journal of Risk.* – 2000. – No. 2. – P. 21–41.
11. TOBIN J. *The Theory of Portfolio Selection* // *Interest Rates / Hahn F., Breechling F., eds.* – London: Macmillan, 1965.

CONTINUOUS VAR-CRITERION AND INVESTOR'S OPTIMAL PORTFOLIO

Gennady A. Agasandyan, Dorodnicyn Computing Centre, FRC CSC RAS, Moscow, Doctor of Science, leading fellow (Moscow, Chertanovskaya st., 34, (495) 313-44-94).

Abstract: The problem of optimal behavior of an investor in the high-developed option market is studied. The investor's risk-preferences function is introduced, and the continuous value at risk-criterion as a continuous generalization of well-known common VaR-criterion is introduced on its basis. Moreover, the investor has his own forecast for probability properties of future behavior of basic option price. The problem is to maximize an average income (or yield) under introduced criterion. The optimal portfolio on the theoretical one-period option market with the given prices picture is constructed. The method is illustrated by an example with two-sided exponential probability distributions.

Keywords: underlier, risk-preferences function, continuous VaR-criterion, Newman-Pearson procedure, optimal portfolio.

Статья представлена к публикации членом редакционной коллегии Ф.И. Ерешко.

*Поступила в редакцию 28.02.2017.
Опубликована 31.05.2018.*

УДК 338.984 + 336.051

ББК 65.290

МЕТОД ВСТРЕЧНЫХ ПЛАНОВ ДЛЯ ОПРЕДЕЛЕНИЯ ЦЕЛЕВОГО ЗНАЧЕНИЯ КРІ В ДВУХУРОВНЕВОЙ ОРГАНИЗАЦИОННОЙ СИСТЕМЕ

Давыдов В. А.¹

*(Московский Институт электроники и математики
НИУ ВШЭ, Москва)*

Рассматривается трехпараметрическая модель стимулирования агента организационной системы (ОС), использующая модификацию метода встречных планов. Для предложенной модели вводятся ограничения на параметры, при которых выбор агента, равный его собственному прогнозу, является доминантной стратегией (ДС) агента. Доказывается, что такой выбор для всех агентов ОС является равновесием в ДС. Приводится пример применения модели для стимулирования подразделений по работе с проблемными активами (ПРПА) в Северо-Западном банке Сбербанка РФ. Отмечается более точное установление планов ПРПА по предложенной модели по сравнению с традиционным методом планирования «сверху вниз».

Ключевые слова: метод встречных планов, стимулирование, КРІ, проблемные активы банков.

1. Введение

В последние годы показатели КРІ (англ. Key Performance Indicators) стали неотъемлемой частью систем планирования и мотивации современного предприятия [7, 9]. Правильным переводом термина КРІ будет «ключевой показатель результата деятельности», так как результат деятельности содержит в себе и степень достижения, и затраты на получение результата [10].

¹ Вячеслав Анатольевич Давыдов, Советник МИЭМ НИУ ВШЭ
(v.davydov@hse.ru).

Кроме задач правильного выбора КРІ, увязки их со стратегией предприятия, систем контроля исполнения КРІ и т.п. одной из важнейших нерешенных задач является задача определения целевого значения КРІ, рассмотренная в [1].

Одним из возможных вариантов решения задачи по определению целевого показателя КРІ могут быть подходы, аналогичные подходам к составлению бюджетов. Это подходы «снизу вверх», «сверху вниз», а также итеративный подход.

При использовании подхода «снизу вверх» КРІ вышестоящего подразделения составляется исходя из КРІ подразделений и проектов, которые передаются наверх для определения по ним агрегированных результирующих показателей. При подходе «сверху вниз» КРІ нижестоящих подразделений составляются исходя из целевых показателей, поставленных руководством компании. Целевые показатели определяются на основе экономических прогнозов, а также стратегии компании. При итеративном подходе процедура установления КРІ включает ряд условных этапов. Информация сначала распространяется от высшего руководства к нижестоящим звеньям управления, а затем обобщается снизу вверх по иерархической структуре управления несколько раз в зависимости от ситуации.

Необходимо отметить, что между руководством и нижестоящими подразделениями при определении целевых значений КРІ возникает конфликт интересов.

Руководство стремится установить подразделениям целевое значения КРІ на максимально напряженном уровне. Такой подход позволяет руководству максимизировать прогноз в целом по системе. При этом руководство вводит мотивирующие факторы для подразделений, которые штрафуют подразделения при не достижении целевых значений КРІ и дополнительно премирует те подразделения, которые перевыполнили свои целевые значения.

Подразделения при планировании целевого значения КРІ стремятся гарантировать в будущем исполнение (а еще лучше – перевыполнение) своих плановых значений. Для решения данной задачи подразделения неохотно включают в целевое значение КРІ проекты, вероятность которых, по мнению

подразделения, недостаточно высока. Могут возникать и более негативные ситуации, когда подразделения (обладая наиболее полной информацией о высокой вероятности реализации того или иного проекта, которой нет у руководства) не включают в прогноз исполнения КРІ такие проекты, тем самым заведомо занижая целевое значение КРІ.

2. Обзор литературы

Теория управления организационными системами (ОС) исследует проблемы синтеза механизмов управления в социальных и экономических системах с учетом специфики человека как объекта управления, в том числе с учетом его активности. Механизмы принятия управленческих решений в ОС руководством (центром) на основании информации, поступающей от подчиненных (синонимы – агенты, игроки, эксперты), называются механизмами планирования.

При разработке механизмов планирования для обеспечения эффективности принимаемых решений необходимо, в том числе, учитывать возможность манипулирования – целенаправленного искажения агентами сообщаемой информации с целью обеспечения принятия более благоприятных для них решений. Механизмы планирования, устойчивые к подобному поведению со стороны агентов (т.е. механизмы, в которых агентам выгодно сообщать достоверную информацию), называются неманипулируемыми.

В [3] рассматриваются механизмы функционирования двухуровневых активных ОС, которые обеспечивают сообщение элементами нижнего уровня в центр достоверной информации о своих предпочтениях и возможностях. Получены необходимые и достаточные условия сообщения достоверных данных и доказана оптимальность принципа открытого управления, описанного в работах [2, 4].

В [6] для модели активной системы, состоящей из центра и активного элемента, в условиях неполной информированности центра о функции затрат активного элемента разработан оптимальный механизм функционирования. Оптимальный механизм включает в себя процедуру планирования и систему

стимулирования, включающую две составляющие: поощрение за выполнение плана и поощрение за «напряженность» плана.

В общем виде задача создания центром неманипулируемого механизма для активного элемента, решаемая методом встречных планов, описана в [8]. В работах [12, 13, 14] представлены результаты зарубежных исследований по проблеме встречных планов. В работе [11] приводится практическое применение метода встречных планов для решения задачи стимулирования планов работы исправительных учреждений уголовно-исполнительной системы, обеспечивающих требуемое значение комплексной оценки деятельности при заданной надежности (риске невыполнения).

3. Описание задачи

Опишем модель определения плановых значений одного КРП для нескольких подразделений, образующих ОС. Для этого будем использовать систему обозначений из [8]. В данной работе указанная задача называется «задача стимулирования»

Пусть есть n подразделений – агентов с номерами $i = 1, \dots, n$ объединение которых с центром образуют двухуровневую ОС. Первый уровень ОС образуют указанные агенты. Второй уровень – центр, который делает в игре первый ход, определяя для агентов их мотивационные факторы.

Каждое подразделение-агент i знает собственный реальный прогноз КРП, который равен R_i и неизвестен другим подразделениям, а также неизвестен центру ОС. При этом у каждого агента i есть множество различных потенциальных вариантов прогноза A_i . В нашей постановке задачи, когда все агенты устанавливают прогноз по одному и тому же КРП, можно считать, что множества вариантов прогноза совпадают для различных агентов. Другими словами, $A_i = A_j = A$.

Приведем пример множества A . Показатель КРП ПРПА «доля сокращения просроченной задолженности NPL90+ за квартал» рассчитывается как отношение суммы урегулированной просроченной задолженности ПРПА за квартал к объему просроченной задолженности на начало

квартала. Таким образом, для всех подразделений множество A данного КРІ будет одинаковым отрезком $A = [0\%; 100\%]$

Рассмотрим ситуацию, когда у каждого агента есть единственный вариант прогноза. Будем полагать, что фактическое значение КРІ, которое будет достигнуто агентом i , совпадает со значением прогноза R_i . Другими словами, это означает, что каждый агент адекватно оценивает то значение КРІ, которое может быть им достигнуто и может осуществить все действия для его достижения.

Обозначим $y_i \in A$ – действие агента i по выбору прогноза КРІ, или проще: y_i – значение прогноза КРІ, которое заявляет центру агент i . Тогда вектор действий агентов $y = (y_1, y_2, \dots, y_n) \in \prod_{j \in N} A = A'$.

Будем называть

$$y_{-i} = (y_1, y_2, \dots, y_{i-1}, y_{i+1}, \dots, y_n) \in \prod_{\substack{j \in N \\ j \neq i}} A = A_{-i}$$

обстановкой игры для агента i .

Сделаем два предположения относительно рассматриваемой ОС.

Во-первых, будем считать, что затраты агентов не зависят от того, какой прогноз они заявят и какой будет получен фактический результат, т.е. какое значение КРІ получит агент в конце отчетного периода. Данное допущение в случае планирования КРІ обосновано в том случае, когда все агенты-подразделения имеют повременную оплату труда, фиксированный график работы, и даже в том случае, когда возникает необходимость в привлечении какого-либо внешнего ресурса для более активной работы по выполнению КРІ, затраты на такое привлечение компенсируются центром ОС в полном объеме. Тогда по сравнению с общим случаем, описанным в работе [8], можно считать, что целевая функция $f_i(\sigma_i, y)$ агента i , зависящая от механизма стимулирования агента σ_i и вектора действий агентов, фактически состоит только из стимулирования $\sigma_i(y)$. Или $f_i(\sigma_i, y) = \sigma_i(y)$, $i \in N$.

Во-вторых, будем считать в рассматриваемой задаче, что единственная роль центра заключается в осуществлении

управления. То есть у центра отсутствует собственный (не опосредованный агентом) результат деятельности. Такое предположение обосновано в том случае, когда затраты центра на стимулирование агента i пренебрежимо малы по сравнению с тем результатом, который агент приносит центру при достижении своего прогноза $R_i \in A$.

При стандартном механизме планирования «сверху-вниз» агенту выгодно получить от центра как можно менее напряженный план. Заведомо перевыполняя такой план, агент гарантирует себе дополнительное вознаграждение. Для получения менее напряженного плана агенту не выгодно раскрывать центру всю имеющуюся информацию относительно перспектив выполнения КРІ. Более того, агент может заведомо исказить (в свою пользу) имеющуюся у него информацию относительно перспектив значения КРІ, которые могут быть им достигнуты. Явление сообщения агентами недостоверной информации называется манипулированием информацией, а механизмы, в которых агентам выгодно сообщение достоверной информации, называются неманипулируемыми.

Задача центра ОС состоит в том, чтобы предложить такую конструкцию установки неманипулируемого механизма стимулирования $\sigma(y) = (\sigma_1(y), \sigma_2(y), \dots, \sigma_n(y))$ целевых значений КРІ для всех агентов, чтобы в результате каждый агент при проведении планирования с подходом «снизу вверх» установил себе значение $y_i = R_i$.

Примем порядок функционирования ОС, аналогичный порядку в [8]. Центру и агентам на момент принятия решения о выбираемых стратегиях (для центра – это функции стимулирования, для агентов – выбираемые прогнозы) известны целевые функции и допустимые множества всех участников ОС. Напомним, что каждый агент i также знает значение своего реального прогноза R_i , который неизвестен остальным участникам ОС. Центр, обладая правом первого хода, выбирает параметры функций стимулирования (в дальнейшем мы увидим, что таких параметров три) и сообщает их агентам. После этого агенты при известных функциях стимулирования выбирают прогнозы, максимизирующие их целевые функции.

Изложенная постановка задачи соответствует механизму стимулирования встречных планов [8]. Для конкретизации применения данного механизма в приведенном случае в следующем разделе будут приведены вариант построения функций стимулирования $\sigma(y) = (\sigma_1(y), \sigma_2(y), \dots, \sigma_n(y))$ агентов центром ОС. Будет показано, что предложенный вариант стимулирования дает для агента максимальный эффект только в том случае, если агент сообщит в качестве прогноза значение $y_i = R_i$.

4. Определение целевой функции для метода встречного плана

Пусть заданы некоторая обстановка игры $y_{-i} = (y_1, y_2, \dots, y_{i-1}, y_{i+1}, \dots, y_n) \in \prod_{\substack{j \in N \\ j \neq i}} A = A_{-i}$ для агента i , плановое

значение КРП y_i , устанавливаемое агентом, а также для всех агентов центром ОС задан коэффициент $\alpha > 0$. Определим коэффициент K_i напряженности плана для агента i в соответствии с формулой

$$(1) \quad K_i = \left(\frac{(n-1)y_i}{\sum_{j=1}^n y_j - y_i} \right)^\alpha.$$

Фактически данный коэффициент возводит в степень $\alpha > 0$ отношение плана агента i к среднему плану, заявленному остальными агентами ОС.

Примеры зависимости коэффициента напряженности плана для степени $\alpha = 1/2$, $\alpha = 1$, $\alpha = 2$ от значений заявляемого плана агента i при среднем плане, заявленном остальными агентами ОС и равном 90,91, приведены на рис. 1.

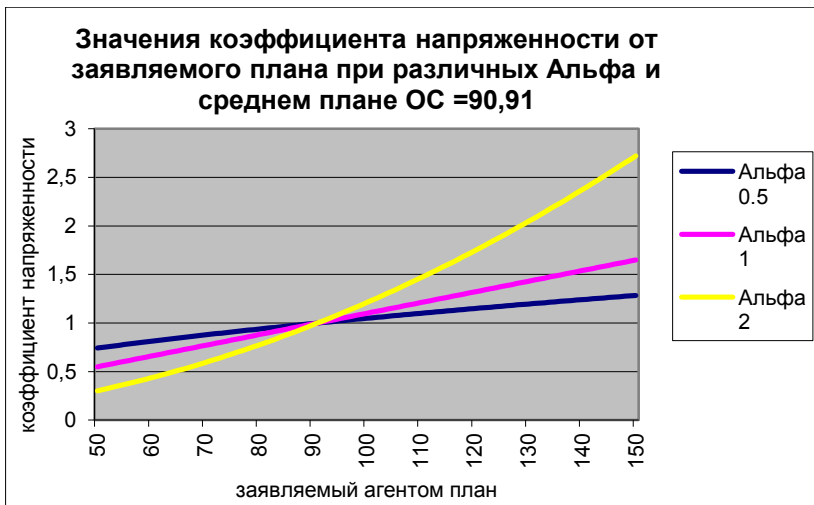


Рис. 1.

Определим коэффициент отклонения O_i фактического исполнения показателя агентом i от плановых показателей, заявленных агентом i . Обозначим r_i фактическое значение показателя КРІ, полученное агентом i в конце отчетного периода. Пусть для всех агентов центром ОС заданы коэффициенты β и γ . Тогда

$$(2) \quad O_i = \begin{cases} \gamma \left(\frac{r_i}{y_i} - 1 \right) + 1, & \text{если } \frac{r_i}{y_i} \geq 1, \\ \beta \left(\frac{r_i}{y_i} - 1 \right) + 1, & \text{если } \frac{r_i}{y_i} \leq 1. \end{cases}$$

С точки зрения центра ОС, перевыполнение агентом заявленного плана должно поощряться, однако не так сильно, как наказываться невыполнение плана. Другими словами, если агент i подстраховался и не заявил свой реальный прогноз R_i , то перевыполнение такого заниженного прогноза не должно давать агенту существенного выигрыша в значении коэффициента O_i . Для достижения такой логики стимулирования агента центр должен выбирать $\beta > 1$ и $0 < \gamma < 1$. В общем случае коэффициенты β и γ могут выбираться центром

независимо друг от друга. Одним из возможных вариантов их взаимозависимости является случай, когда $\beta > 1$ и $\gamma = 1/\beta$. Примеры зависимости коэффициента отклонения для значений $\beta = 2$ и $\gamma = 1/2$, $\beta = 3$ и $\gamma = 1/3$, $\beta = 5$ и $\gamma = 1/5$ от отношения факта исполнения $50 < r_i < 150$ агентом i к заявленному агентом плану $y_i = 100$ приведены на рис. 2.

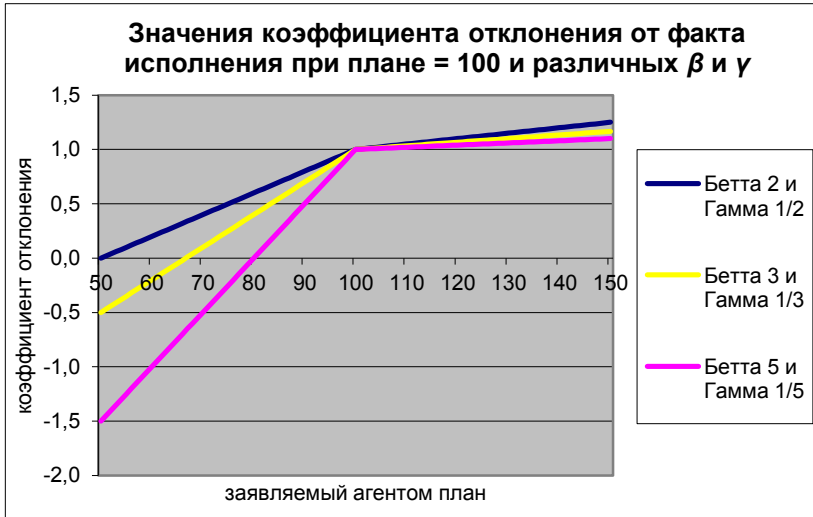


Рис. 2.

Определим для агента $i \in N$ функцию $\sigma_i(y)$ как произведение коэффициента напряженности плана и коэффициента отклонения: $\sigma_i(y) = O_i K_i$.

Из определения коэффициента напряженности плана следует, что выбор агентом i прогнозного значения y_i не влияет на значение среднего прогноза по остальным $n - 1$ агентам.

Обозначим $\hat{y}_{-i} = \frac{\sum_{j=1, j \neq i}^n y_j - y_i}{n-1}$.

Пусть для вычисления K_i задано $\alpha = 2$. Предположим, что известны три варианта отношения $\frac{y_i}{\hat{y}_{-i}} = 1$, $\frac{y_i}{\hat{y}_{-i}} = 0,9$, $\frac{y_i}{\hat{y}_{-i}} = 1,1$.

Тогда получим три варианта значений $K_i = 1$, $K_i = 0,81$ и $K_i = 1,21$. Влияние коэффициента напряженности плана для трех перечисленных вариантов значения на целевую функцию $\sigma_i(y) = O_i K_i$ с коэффициентом отклонения O_i для значений $\beta = 3$ и $\gamma = 1/3$ показано на рис. 3.

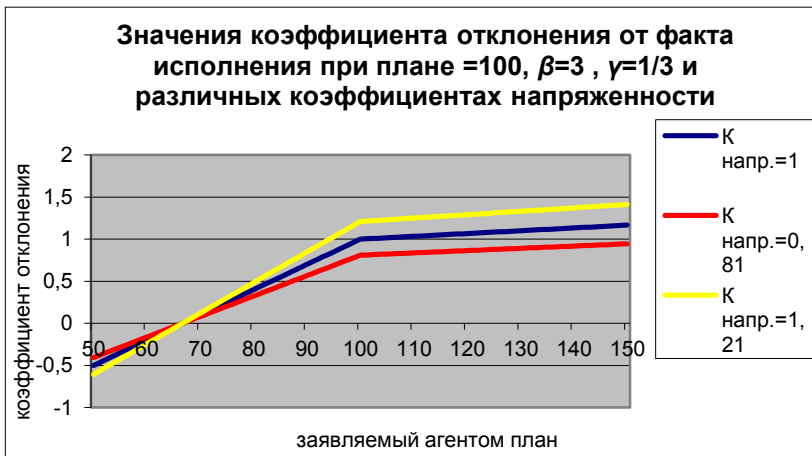


Рис. 3.

Отметим, что в данном случае, как и на рис. 2, рассматривается зависимость $\sigma_i(y)$ от фактического исполнения $50 < r_i < 150$ агентом I рассматриваемого КРП при заявленном агентом плане $y_i = 100$.

Сделаем еще одно замечание относительно функции $\sigma_i(y)$. В описанном примере построения $\sigma_i(y)$ одним из параметров, который отличает функцию $\sigma_i(y)$ от функции $\sigma_j(y)$ для агентов с номерами i и j , являются значения реального прогноза R_i и R_j , которые известны только самим агентам и неизвестны остальным участникам ОС. При этом на значение $\sigma_j(y)$ также оказывает влияние \hat{y}_{-i} – средняя величина прогноза, заявленного остальными $n - 1$ агентами ОС.

Таким образом, мы получили функцию, зависящую от трех скалярных величин: R_i – реального прогноза КРП, известного агенту i , y_i – заявляемого агентом i прогноза КРП, и \hat{y}_{-i} – средней

величины прогноза, заявленного остальными $n - 1$ агентами ОС. Для того чтобы подчеркнуть данные зависимости, будем в дальнейшем использовать обозначение $\sigma_i(y) \equiv \sigma(R_i, y_i, \hat{y}_{-i})$.

В момент осуществления процедуры планирования агенту i не известно фактическое значение показателя КРП r_i , которое он получит в конце отчетного периода. Агент i обладает только значением реального прогноза R_i и решает задачу, какое значение плана y_i заявить центру. Однако поскольку мы исходим из предположения, что агент i реализует свой прогноз, то в конце отчетного периода будет выполнено равенство $r_i = R_i$.

С ростом значения r_i величина O_i убывает. При значениях $y_i \leq R_i$ коэффициент $1 \leq O_i$. При значениях $R_i \leq y_i$ коэффициент $O_i \leq 1$ и при достаточно больших значениях y_i коэффициент O_i становится отрицательным. Примеры зависимости коэффициента отклонения O_i для значений $\beta = 2$ и $\gamma = 1/2$, $\beta = 3$ и $\gamma = 1/3$ и $\beta = 5$ и $\gamma = 1/5$ при условии $R_i = 100$ и диапазоне изменения $50 < y_i < 150$ приведены на рис. 4.

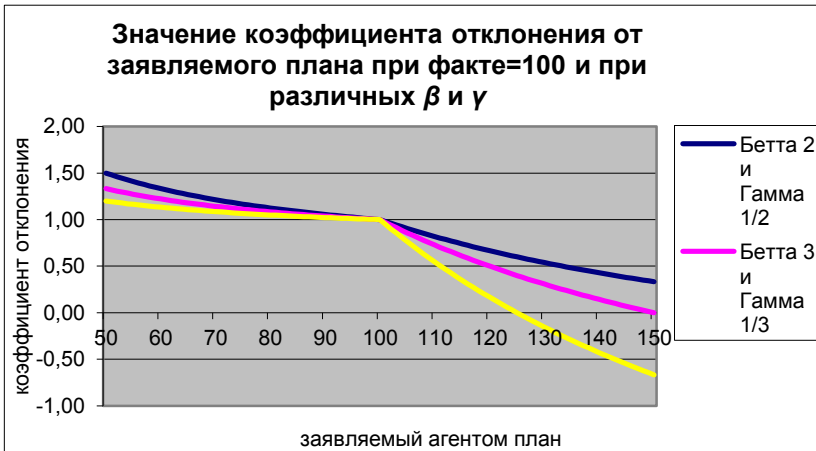


Рис. 4.

При принятии решения о выборе значения y_i агент i будет рассматривать не только значения коэффициента отклонения O_i , но и значение коэффициента напряженности плана K_i , а точнее

их произведения $\sigma(R_i, y_i, \hat{y}_{-i}) = O_i K_i$. Будем считать, что при расчете коэффициента напряженности использован параметр $\alpha = 2$.

Зависимость $\sigma(R_i, y_i, \hat{y}_{-i})$ от величины заявляемого прогноза $50 < y_i < 150$ для реального прогноза $R_i = 100$ и $\hat{y}_{-i} = 90,91$ (т.е. $100/90,91 = 110\%$), $\hat{y}_{-i} = 100$ (т.е. $100/100 = 100\%$) и $\hat{y}_{-i} = 111,11$ (т.е. $100/111,11 = 90\%$) приведена на рис. 5.

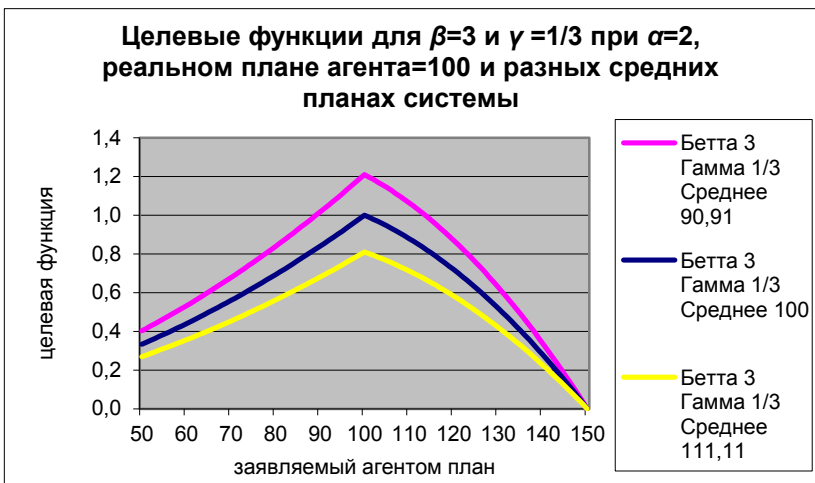


Рис. 5.

Как видно из графиков, максимум целевой функции $\sigma(R_i, y_i, \hat{y}_{-i})$ достигается при выполнении условия $y_i = R_i = 100$. Другими словами, агент i для максимизации целевой функции должен заявить план y_i , совпадающий с реальным прогнозом $R_i = 100$. Докажем данный вывод в общем случае.

Лемма. Пусть для одноуровневой ОС с n агентами заданы параметры $\beta > \alpha > \gamma > 0$ и вектор значений реальных прогнозов (R_1, R_2, \dots, R_n) . Тогда для целевой функции агента $\sigma(R_i, y_i, \hat{y}_{-i}) = O_i K_i$, $1 \leq i \leq n$, где

$$O_i = \begin{cases} \gamma \left(\frac{R_i}{y_i} - 1 \right) + 1, & \text{если } \frac{R_i}{y_i} \geq 1, \\ \beta \left(\frac{R_i}{y_i} - 1 \right) + 1, & \text{если } \frac{R_i}{y_i} \leq 1. \end{cases}$$

$$K_i = \left(\frac{(n-1)y_i}{\sum_{j=1}^n y_j - y_i} \right)^\alpha, \quad \hat{y}_{-i} = \frac{\sum_{j=1}^n y_j - y_i}{n-1},$$

существует равновесие в доминантных стратегиях (РДС) равное $y_i^* = R_i, 1 \leq i \leq n$.

Доказательство. В [8] доказано следующее утверждение:

Если в игре n лиц $y_i \in [a_i, b_i]$, функции выигрыша непрерывны по совокупности стратегий для каждого игрока, частная производная $\frac{\partial \sigma(R_i, y_i, \hat{y}_{-i})}{\partial y_i}$ существует и знакопостоянна, то существует РДС. При этом доминантной стратегией y_i^* игрока i будет стратегия

$$(3) \quad y_i^* = \begin{cases} a_i, & \frac{\partial \sigma(R_i, y_i, \hat{y}_{-i})}{\partial y_i} < 0, \\ b_i, & \frac{\partial \sigma(R_i, y_i, \hat{y}_{-i})}{\partial y_i} > 0. \end{cases} \quad 1 \leq i \leq n.$$

Используем данное утверждение для доказательства Леммы. Рассмотрим два интервала для функции $\sigma(R_i, y_i, \hat{y}_{-i}) = O_i K_i$. Интервал $y_i \in [0, R_i]$ и интервал $y_i \in [R_i, +\infty)$.

Вариант 1. Пусть $y_i \in [0, R_i]$. В этом случае получаем $\frac{R_i}{y_i} \geq 1$.

Тогда, согласно (2) получаем

$$\begin{aligned} \sigma(R_i, y_i, \hat{y}_{-i}) &= O_i K_i = \\ &= \left(\gamma \left(\frac{R_i}{y_i} - 1 \right) + 1 \right) \left(\frac{y_i}{\hat{y}_{-i}} \right)^\alpha = y_i^{\alpha-1} \frac{\gamma R_i}{\hat{y}_{-i}^\alpha} - y_i^\alpha \left(\frac{\gamma-1}{\hat{y}_{-i}^\alpha} \right). \end{aligned}$$

Тогда

$$\begin{aligned} \frac{\partial \sigma(R_i, y_i, \hat{y}_{-i})}{\partial y_i} &= y_i^{\alpha-2} \frac{\gamma R_i (\alpha-1)}{\hat{y}_{-i}^\alpha} - y_i^{\alpha-1} \alpha \left(\frac{\gamma-1}{\hat{y}_{-i}^\alpha} \right) = \\ &= \frac{y_i^{\alpha-2}}{\hat{y}_{-i}^\alpha} (\gamma R_i (\alpha-1) - y_i (\alpha \gamma - \alpha)). \end{aligned}$$

Докажем, что $\frac{\partial \sigma(R_i, y_i, \hat{y}_{-i})}{\partial y_i} > 0$, т.е. будем доказывать, что

функция $\sigma(R_i, y_i, \hat{y}_{-i})$ возрастает на интервале $y_i \in [0, R_i]$.

Величина $\frac{y_i^{\alpha-2}}{\hat{y}_{-i}^\alpha} > 0$ при условии $y_i \in [0, R_i]$. Тогда для

положительности производной должно выполняться условие $\gamma R_i (\alpha-1) - y_i (\alpha \gamma - \alpha) > 0$.

Это эквивалентно

$$\gamma R_i (\alpha-1) > y_i (\alpha \gamma - \alpha),$$

$$\frac{R_i}{y_i} > \frac{\alpha (\gamma-1)}{\gamma (\alpha-1)}.$$

С учетом условия $\frac{R_i}{y_i} \geq 1$ выберем минимальное значение

$\frac{R_i}{y_i} = 1$. В результате получаем

$$(4) \quad 1 > \frac{(\gamma-1)\alpha}{\gamma(\alpha-1)}.$$

Определим соотношение между положительными величинами α и γ , при котором неравенство (4) будет исполняться.

$$\frac{(\alpha-1)}{\alpha} > \frac{(\gamma-1)}{\gamma} \Rightarrow 1 - \alpha^{-1} > 1 - \gamma^{-1} \Rightarrow \alpha^{-1} < \gamma^{-1} \Rightarrow$$

$$\Rightarrow \frac{\gamma}{\alpha\gamma} < \frac{\alpha}{\alpha\gamma} \Rightarrow \gamma < \alpha.$$

Вариант 2. Пусть $y_i \in [R_i, \infty]$. В этом случае получаем $\frac{R_i}{y_i} \leq 1$. Тогда, согласно (2), получаем

$$\begin{aligned} \sigma(R_i, y_i, \hat{y}_{-i}) &= O_i K_i = \\ &= \left(\beta \left(\frac{R_i}{y_i} - 1 \right) + 1 \right) \left(\frac{y_i}{\hat{y}_{-i}} \right)^\alpha = y_i^{\alpha-1} \frac{\beta R_i}{\hat{y}_{-i}^\alpha} - y_i^\alpha \left(\frac{\beta-1}{\hat{y}_{-i}^\alpha} \right). \end{aligned}$$

Тогда

$$\begin{aligned} \frac{\partial \sigma(R_i, y_i, \hat{y}_{-i})}{\partial y_i} &= y_i^{\alpha-2} \frac{\beta R_i (\alpha-1)}{\hat{y}_{-i}^\alpha} - y_i^{\alpha-1} \alpha \left(\frac{\beta-1}{\hat{y}_{-i}^\alpha} \right) \\ &= \frac{y_i^{\alpha-2}}{\hat{y}_{-i}^\alpha} (\beta R_i (\alpha-1) - y_i (\alpha \beta - \alpha)). \end{aligned}$$

Докажем, что $\frac{\partial \sigma(R_i, y_i, \hat{y}_{-i})}{\partial y_i} < 0$, т.е. будем доказывать, что

функция $\sigma(R_i, y_i, \hat{y}_{-i})$ убывает на интервале $y_i \in [R_i, \infty]$.

Величина $\frac{y_i^{\alpha-2}}{\hat{y}_{-i}^\alpha} > 0$ при условии $y_i \in [R_i, \infty]$. Тогда для отрицательности производной должно выполняться условие $\beta R_i (\alpha-1) - y_i (\alpha \beta - \alpha) > 0$.

Это эквивалентно

$$\beta R_i (\alpha-1) < y_i \alpha (\beta-1),$$

$$\frac{R_i}{y_i} < \frac{\alpha (\beta-1)}{\beta (\alpha-1)}.$$

С учетом условия $\frac{R_i}{y_i} \leq 1$ выберем максимальное значение

$\frac{R_i}{y_i} = 1$. В результате получаем

$$(5) \quad 1 < \frac{(\beta-1)\alpha}{\beta(\alpha-1)}.$$

Определим соотношение между положительными величинами α и β , при котором неравенство (5) будет исполняться.

$$\begin{aligned} \frac{(\alpha-1)}{\alpha} < \frac{(\beta-1)}{\beta} &\Rightarrow 1-a^{-1} < 1-\beta^{-1} \Rightarrow a^{-1} > \beta^{-1} \Rightarrow \\ &\Rightarrow \frac{\beta}{\alpha\beta} > \frac{\alpha}{\alpha\beta} \Rightarrow \beta > \alpha. \end{aligned}$$

Таким образом, при выполнении условий $\beta > \alpha > \gamma > 0$ получаем, что на интервале $y_i \in [0, R_i]$ функция $\sigma(R_i, y_i, \hat{y}_{-i})$ монотонно возрастает и, согласно утверждению (3), для точки РДС выполняется условие $y_i^* = R_i$. Соответственно, на интервале $y_i \in [R_i, \infty]$ функция $\sigma(R_i, y_i, \hat{y}_{-i})$ монотонно убывает и, согласно утверждению (3), для точки РДС выполняется условие $y_i^* = R_i$.

Следовательно, на интервале $y_i \in [0, \infty]$ существует единственная точка $y_i^* = R_i$, которая является выбором агента, являющимся РДС. •

Рассмотрим пример игры агента с центром, в котором будут фигурировать приведенные выше графики.

Исходное состояние. Всем агентам, а также центру перед началом процедуры планирования известны формулы расчета K_i и O_i . Центр устанавливает параметры β , α , γ исходя из своей стратегии мотивации. Каждый агент знает свой реальный прогноз, который неизвестен остальным участникам ОС. Пусть для агента с номером i такой прогноз $R_i = 100$.

Шаг 1. Центр, исходя из своих предпочтений, выбирает значения параметров $\beta = 3$, $\alpha = 2$, $\gamma = 1/3$ и сообщает их всем агентам, в том числе и агенту с номером i . Заметим, что для выбранных значений выполняется условие Леммы, а именно $3 > 2 > 1/3 > 0$. В результате все участники ОС понимают, как меняется коэффициент напряженности плана (желтый график на рис. 1). Также всем участникам ОС становится известно, как изменяется коэффициент отклонения (желтый график на рис. 2).

Шаг 2. Агент с номером i , исходя из своего предположения о том, что в конце отчетного периода он со 100% вероятностью получит фактическое значение КРП $r_i = 100$, строит целевую

функцию на рис. 5. Анализируя данную функцию для любых значений \hat{y}_{-i} , агент с номером i приходит к выводу о том, что максимальное значение целевой функции он получит (при условии выполнения $r_i = R_i$), если заявит прогнозное значение $y_i = R_i$, причем независимо от того, каково будет значение \hat{y}_{-i} . В результате проведенного анализа агент с номером i заявляет центру свой прогноз $y_i = 100$.

Шаг 3. Центр, получив прогнозы от всех агентов, вычисляет значение \hat{y}_{-i} и сообщает его агенту i . Предположим, $\hat{y}_{-i} = 90,91$. Это означает, что агент i дал более напряженный прогноз плана $y_i = 100$, чем в среднем все остальные агенты ОС. Отношение $\frac{y_i}{\hat{y}_{-i}} = \frac{100}{90,91} = 1,1$. На основании полученного отношения и известного параметра $a = 2$ центр и агент i вычисляют значение $K_i = (1,1)^2 = 1,21$ и строят итоговую целевую функцию для агента i (желтый график на рис. 3).

В итоге если агент i в конце отчетного периода реализует свои планы и получит значение КРІ $r_i = 100$, то значение его целевой функции составит $O_i K_i = 1,21$.

Сделаем ряд заключительных замечаний по предложенному механизму определения оптимального значения КРІ.

Замечание 1. На основании реальных прогнозов R_i и R_j агенты должны (если система мотивации работает правильно) выбрать значения плановых показателей $y_i = R_i$ и $y_j = R_j$. При таком выборе коэффициенты напряженности планов K_i и K_j будут адекватно отражать напряженность планов агентов только в том случае, если агенты работают в одинаковых условиях.

Приведем пример. Предположим, что у первого агента портфель проблемных активов недостаточно обеспечен залогами, а у второго агента портфель переобеспечен. В данном случае первому агенту будет сложно соревноваться со вторым по такому показателю КРІ, как процент возврата денежными средствами на объем проблемного портфеля. При прочих равных самый напряженный прогноз первого агента будет ниже консервативного прогноза второго агента. Для того чтобы

сделать значения коэффициента напряженности плана адекватными в данной ситуации, необходимо дополнительно учитывать показатель обеспеченности портфеля, с которым работают агенты. Такой учет возможен, например, при использовании дополнительного поправочного коэффициента, отражающего обеспеченность портфеля агента по LGD-модели, на который будет умножаться целевая функция $\sigma(R_i, y_i, \hat{y}_{-i})$.

Замечание 2. Центр может модифицировать метод установления значения \hat{y}_{-i} . Это необходимо в том случае, например, если агенты ОС могут договориться между собой и в результате совместно дать заниженные прогнозы. В таких случаях значение \hat{y}_{-i} может определяться по формуле

$$\hat{y}_{-i} = \max \left\{ \begin{array}{l} \frac{\sum_{j=i}^n y_j - y_i}{n-1}, \\ y_{-i}^{\min} \end{array} \right.,$$

где y_{-i}^{\min} – некоторое минимальное значение KPI, устанавливаемое центром для агента i . В качестве ориентира для величины y_{-i}^{\min} могут рассматриваться данные LGD-модели, которые рассчитываются центром отдельно для каждого агента. Центр их сообщает агентам на шаге 1 вместе с параметрами β, α, γ . В этом случае агенты будут продолжать соревноваться друг с другом, устанавливая максимально напряженные планы, но центр дает им ориентир, который является минимальным планом. Для того чтобы агент мог получить результирующее значение функции $\sigma(R_i, y_i, \hat{y}_{-i})$ не ниже единицы (при условии исполнения агентом заявленного плана), агент должен заявить план не ниже y_{-i}^{\min} .

Рассмотрим отличия предложенной целевой функции $\sigma(R_i, y_i, \hat{y}_{-i})$ от целевой функции для механизма встречных планов в работе [11].

Во-первых, в [11] итоговая функция стимулирования определяется в виде суммы коэффициента отклонения и

коэффициента напряженности плана. В предложенной конструкции рассматривается произведение указанных коэффициентов.

Во-вторых, в [11] не определено в явном виде, какой план, заявляемый агентом, дает коэффициент напряженности плана, являющийся для центра ОС некой базовой величиной. В предложенном варианте такая базовая величина определяется, как среднее значение плана по остальным агентам ОС, но не ниже, чем y_{-i}^{\min} – некоторое минимальное значение КРІ, устанавливаемое центром для агента. В результате система мотивации с одной стороны стимулирует соревнование между агентами по установлению максимально напряженного плана, а с другой – не позволяет агентам вступить в сговор и заявить центру ОС минимально возможные планы.

5. Реализация метода встречных планов на примере Управления по работе с проблемной задолженностью Северо-Западного банка Сбербанка РФ

Метод встречных планов, изложенный в предыдущем разделе, реализован в качестве пилота по новой модели бизнес-планирования при процедуре планирования показателей КРІ для ПРПА Северо-Западного банка (СЗБ) Сбербанка РФ во втором квартале 2016 года. Пилот осуществлялся без выплаты фактического вознаграждения по его результатам при одновременной процедуре мотивации по традиционной модели. В рамках пилота сотрудники каждого ПРПА имели возможность дать свои прогнозы по каждому КРІ, которые отличались от плановых заданий, установленных для них исходя из данных LGD-модели. Такие прогнозные значения сравнивались с фактически полученными значениями в конце квартала.

Процедура планирования для ПРПА проходит ежеквартально и состоит в определении аппаратом СЗБ (который в данном случае является центром ОС) девяти установленных Центральным аппаратом (ЦА) показателей для каждого из семи головных отделений сберегательного банка

(ГОСБ), входящих в состав СЗБ. В данном случае ГОСБ могут рассматриваться в качестве агентов ОС. Перечень показателей, методика их расчетов не могут быть изменены аппаратом СЗБ, поскольку определяются ЦА, но плановые значения данных показателей аппарат СЗБ может устанавливать самостоятельно.

В состав СЗБ входят семь отдельных ГОСБ (ГОСБ по Санкт-Петербургу, ГОСБ по Ленинградской области, ГОСБ по Калининграду, ГОСБ по Мурманску, ГОСБ по Пскову, ГОСБ по Карелии, ГОСБ по Мурманску).

Дадим комментарии по девяти КРІ, плановые показатели по которым устанавливаются аппаратом СЗБ для каждого из семи ГОСБ. Портфель проблемной ссудной задолженности СЗБ в целом и каждого ГОСБ в частности состоит из проблемных активов (ПА) черной зоны (как правило, это кредиты со сроками просрочки более 90 дней) и остальных ПА, относящихся к красной зоне. Каждому ПА соответствует резерв, также относящийся к черной или красной зонам.

Показатель сокращения ПА (по красной, черной или в целом по портфелю ПА) рассчитывается как отношение объема сокращения величины ПА за квартал (по красной, черной или в целом по портфелю ПА) к объему ПА (по красной, черной или в целом по портфелю ПА) на начало квартала. Аналогично рассчитывается показатель восстановления РВПС в разрезе резерва красной зоны, черной зоны и резерва по всему портфелю ПА.

NPL90+ «После» означает отношение объема сокращения задолженности за квартал по ПА, которые имеют просроченную задолженность свыше 90 дней к объему такой задолженности на начало квартала. Аналогично NPL1095+ означает отношение объема сокращения задолженности за квартал по ПА, которые имеют просроченную задолженность свыше 1095 дней (т.е. более трех лет) к объему такой задолженности на начало квартала.

NPL90+ «До» означает отношение объема сокращения задолженности за квартал по ПА, которые имеют просроченную задолженность свыше 90 дней без учета списаний и дисконта по цессиям к объему такой задолженности на начало квартала. В размере объема списанных кредитов и дисконта по

заключенным цессиям заключается отличие между показателем NPL90+ «До» (в котором убытки банка по списанным кредитам и дисконтам по цессиям не учитываются) и NPL90+ «После» (в котором на снижение просроченной задолженности влияют дисконты и списания)

Все перечисленные девять КРІ являются относительными (т.е. показывают динамику изменения какого-либо показателя проблемного портфеля за квартал) и измеряются в процентах. Выгода аппарата СЗБ (как центра ОС) тем больше, чем больше фактическое значение любого из перечисленных КРІ (т.е. чем быстрее ГОСБ урегулирует тот или иной проблемный показатель) у каждого ГОСБ.

Существующая процедура планирования ПРПА состоит в установлении плановых показателей ГОСБ на основе LGD-модели, которые рассчитываются ЦА. В результате плановые показатели КРІ для ГОСБ назначаются директивно и не учитывают мнение самого ГОСБ. Расхождение в данных LGD-модели с мнением ГОСБ, как правило, обусловлено тем, что LGD-модель качественно оценивает большой портфель с большим числом активов. Объем портфеля ГОСБ относительно небольшой и может качественно отличаться от среднестатистических данных, что дает отклонения как в сторону увеличения возможности ГОСБ урегулировать тот или иной показатель, так и в сторону уменьшения такой возможности. План по урегулированию ПА для каждого из девяти КРІ, утвержденный на основании существующей модели премирования, приведен в Приложении 1.

Альтернативным методом установления плановых значений КРІ для ГОСБ является метод встречных планов, рассмотренный в предыдущем разделе. Для получения встречных планов по каждому из девяти КРІ, до ГОСБ были доведены параметры функции $\beta = 1$, $\alpha = 0,3$, $\gamma = 0,2$, а также вид целевой функции $\sigma(R_i, y_i, \hat{y}_{-i}) = O_i K_i$ в соответствии с Леммой. Для выбранных значений выполняется условие Леммы, а именно $1 > 0,3 > 0,2 > 0$.

На основании полученных данных, исходя из своего портфеля ПА, каждый ГОСБ сообщил аппарату СЗБ свой

встречный план по каждому из девяти КРІ. Данные планы по урегулированию ПА в рамках пилота по новой модели бизнес-планирования приведены в Приложении 2.

Фактический процент урегулирования по каждому из девяти КРІ за второй квартал 2016 года в разрезе каждого ГОСБ приведен в Приложении 3.

Сравним результаты планирования для ГОСБ по LGD-модели и результаты, полученные путем применения механизма встречных планов. В качестве критерия точности предлагаемого метода планирования для каждого КРІ будем использовать среднеквадратичное отклонение фактического значения данного КРІ от планового значения по семи ГОСБ. Таким образом, получим среднеквадратичные отклонения для планирования на основе LGD-модели и планирования на основе встречных планов по каждому из девяти КРІ. Итоговые данные по полученным среднеквадратичским отклонениям приведены в Приложении 4.

Сделаем выводы на основе полученных данных по результатам пилота.

Вывод 1. Как видно из Приложения 4, среднеквадратичное отклонение плановых от фактических значений по методу встречных планов дает меньший разброс, чем среднеквадратичное отклонение при планировании на основе LGD-модели. Данный вывод подтверждается по каждому из девяти КРІ, которые являются, по сути, девятью экспериментами сравнения двух методов планирования.

Вывод 2. Суммарные плановые значения семи ГОСБ по методу встречных планов в рассмотренном эксперименте отличаются от суммарных плановых значений по LGD-модели как в меньшую сторону по шести КРІ, так и в большую сторону (по сокращению ПА красной зоны, восстановлению резерва по черной зоне и NPL1095+). Это подтверждает, что при использовании предложенной модели установления планов, ГОСБ будут (если у них есть соответствующая информация) устанавливать более напряженные планы по ряду КРІ, чем планы по LGD-модели.

6. Заключение

Рассмотрен ряд известных систем КРІ, в том числе для подразделений по работе с проблемными активами в банке. Отмечается, что многими авторами [1, 7, 9] рассматривалась задача, при которой КРІ-система управления результатами бизнеса не только отслеживала, но и стимулировала достижение компанией (или подразделением по работе с проблемными активами) своих целей. Для решения такой задачи необходимо наладить собственно механизм установки этих целей.

Дается формулировка задачи определения целевых значений одного КРІ для нескольких подразделений образующих систему.

Изложенная постановка задачи соответствует механизму стимулирования встречных планов [8]. Для случая единственного варианта реального прогноза R_i доказывается Лемма, задающая ограничения на параметры целевой функции агента, позволяющие Центру решить поставленную задачу, причем такое решение будет иметь свойства равновесия в доминантной стратегии (РДС).

Предложенная методика определения КРІ по методу встречных планов была применена в пилотном проекте на базе Северо-Западного банка Сбербанка РФ. В качестве параметра, показывающего качество планирования каждого из девяти КРІ, предложена величина среднеквадратичного отклонения фактически полученных агентами значений КРІ от устанавливаемых плановых значений. Все перечисленные девять КРІ являются относительными (т.е. показывают динамику изменения какого-либо показателя проблемного портфеля за квартал) и измеряются в процентах.

На основе полученных данных по результатам Пилота были рассчитаны среднеквадратичные отклонения плановых значений КРІ от фактических значений, получаемых Подразделениями-агентами по результатам квартала. По методу встречных планов данный параметр дает меньший разброс, чем при планировании на основе LGD-модели. Указанная зависимость отмечается по каждому из девяти КРІ, назначаемых для ПРПА.

Необходимо отметить, что по трем из девяти КРІ заявляемый агентами прогноз стал более напряженным, чем прогноз, рассчитанный Центром на основании LGD-модели. Таким образом, прогноз значения КРІ по предложенному методу встречных планов дает не только более точные прогнозы, но и (в том случае, когда подразделение-агент это может) устанавливает более напряженные плановые значения КРІ.

Литература

1. БУГРОВ Д. *Метрика эффективности* // Вестник McKinsey. – 2003. – №1.
2. БУРКОВ В.Н. *Основы математической теории активных систем.* – М.: Наука, 1977.
3. БУРКОВ В.Н., ЕНАЛЕЕВ А.К. *Оптимальность принципа открытого управления. Необходимые и достаточные условия достоверности информации в активных системах* // Автоматика и телемеханика. – 1985. – №3. – С. 73–80.
4. БУРКОВ В.Н., ЛЕРНЕР А.Я. *Принцип открытого управления.* – М.: ИАТ, 1974.
5. ГУБКО М.В., НОВИКОВ Д.А. *Теория игр в управлении организационными системами.* – М.: Институт проблем управления им. В.А. Трапезникова РАН, 2005.
6. ЕНАЛЕЕВ А.К. *Оптимальный механизм функционирования в активной системе с обменом информацией* // Управление большими системами. – 2010. – Вып. 29. – С. 108–127.
7. КАПЛАН Р.С., НОРТОН Д.П. *Организация, ориентированная на стратегию* / Пер. с англ. – М.: ЗАО «Олимп-Бизнес», 2003. – 416 с.
8. *Механизмы управления: Учебное пособие* / Под ред. Д.А. Новикова. – М.: УРСС, 2011.

9. ОЛЬВЕ Н.Г., РОЙ Ж., ВЕТТЕР М. *Оценка эффективности деятельности компании. Практическое руководство по использованию сбалансированной системы показателей / Пер. с англ.* – М.: «Вильямс», 2003. – 304 с.
10. ПАНОВ М.М. *Оценка деятельности и система управления компанией на основе KPI.* – М.: Инфра-М, 2013. – 255 с.
11. РОССИХИНА Л.В., АГАФОНОВ С.М., БАРКАЛОВ С.А. *Задача стимулирования оптимальных планов работы исправительных учреждений // Моделирование, оптимизация и информационные технологии.* – 2014. – №4(7).
12. DOWD J., MCGONIGLE W., DJATEJ A. *The optimal level of budgetary goal difficulty: an experimental study // Int. Journal of Managerial and Financial Accounting.* – 2010. – Vol. 2(2). – P. 153–160.
13. HARRIS M., RAVIV A. *Optimal incentive contracts with imperfect information // Journal of Economic Theory.* – 1979. – Vol. 20, No. 2. – P. 231–259.
14. LEWIS T.R., SAPPINGTON D.E.M. *Information Management in Incentive Problems // Journal of Political Economy.* – 1997. – Vol. 105(4). – P. 796–821.

Приложение 1.

ГОСБ	Сокращение ПА			Восстановление РВПС			NPL1095+	Остаток NPL90+	
	Общее	Красная	Черная	Общее	Красная	Черная		NPL90+ «До»	NPL90+ «После»
ГО ЛО	4,8%	1,8%	13,9%	8,7%	2,5%	14,4%	35,0%	60,7%	70,4%
ГО СПб	8,6%	12,6%	3,7%	7,8%	45,7%	4,2%	35,0%	0,0%	0,0%
Калининград	10,7%	71,7%	7,2%	8,9%	562,0%	6,6%	35,0%	10,5%	70,8%
Карелия	3,4%	4,2%	2,6%	3,8%	27,2%	1,9%		1,4%	5,0%
Мурманск	57,5%	62,7%	5,7%	10,0%	22,9%	5,7%	35,0%	0,0%	0,0%
Новгород	5,8%	9,4%	1,9%	6,3%	12,7%	2,7%	35,0%	0,2%	0,0%
Псков	2,9%	3,5%	1,3%	5,1%	9,6%	1,0%		3,1%	0,0%

Приложение 2.

ГОСБ	Сокращение ПА			Восстановление РВПС			NPL1095+	Остаток NPL90+	
	Общее	Красная	Черная	Общее	Красная	Черная		NPL90+ «До»	NPL90+ «После»
ГО ЛО	7,1%	4,7%	14,3%	20,7%	2,3%	37,3%	70,6%	61,4%	73,6%
ГО СПб	8,0%	13,1%	1,7%	3,0%	13,9%	1,9%	40,2%	1,6%	7,8%
Калининград	14,2%	5,6%	14,6%	13,3%	1,5%	13,4%	54,3%	15,9%	76,2%
Карелия	13,3%	14,8%	11,6%	13,6%	19,2%	13,1%		21,5%	30,7%
Мурманск	57,7%	62,8%	7,1%	14,0%	18,9%	12,3%	93,3%	30,3%	44,0%
Новгород	7,8%	9,9%	5,6%	2,6%	1,5%	3,2%	43,1%	42,4%	45,9%
Псков	2,9%	3,1%	2,2%	2,4%	2,7%	2,2%		6,8%	3,4%

Приложение 3.

ГОСБ	Сокращение ПА			Восстановление РВПС			NPL1095+	Остаток NPL90+	
	Общее	Красная	Черная	Общее	Красная	Черная		NPL90+ «До»	NPL90+ «После»
ГО ЛО	4,8%	2,7%	11,1%	40,3%	46,8%	34,4%	72,6%	60,8%	70,6%
ГО СПб	7,8%	11,9%	2,7%	4,9%	23,6%	3,1%	26,1%	5,1%	22,6%
Калининград	9,4%	19,6%	8,8%	8,3%	44,3%	8,2%	2,2%	13,1%	70,6%
Карелия	13,1%	14,4%	11,5%	13,9%	26,7%	12,8%		22,5%	32,7%
Мурманск	0,1%	0,0%	1,4%	5,1%	9,3%	3,7%	90,8%	26,0%	40,0%
Новгород	7,7%	9,9%	5,3%	2,8%	2,4%	3,1%	0,0%	42,9%	45,8%
Псков	2,4%	3,1%	0,3%	16,0%	33,4%	0,4%		6,1%	3,3%

Приложение 4.

	Сокращение ПА			Восстановление РВПС			NPL 1095+	Остаток NPL90+	
	Общее	Красная	Черная	Общее	Красная	Черная		NPL90+ «До»	NPL90+ «После»
Планирование на основе LGD-модели	58,2%	82,1%	11,1%	35,5%	520,9%	22,9%	83,1%	54,6%	70,6%
Планирование с использованием встречных планов	57,8%	64,4%	9,0%	26,0%	70,7%	10,7%	69,1%	6,4%	16,7%
Отклонение	0,4%	17,7%	2,1%	9,5%	450,2%	12,2%	14,0%	48,2%	53,9%

COUNTER-PLAN METHOD FOR DETERMINING KPI'S TARGET SIGNIFICANCE IN THE TWO-LEVEL ORGANIZATIONAL SYSTEM

Viacheslav Davydov, Moscow Institute of Electronics and Mathematics HSE, Moscow (v.davydov@hse.ru).

Abstract: We consider a three-parameter model for stimulating the agent of the organizational system (OS), which uses the modification of the counter-plan method. For the proposed model, restrictions are introduced on the parameters under which an agent's choice equals to his own forecast and is the dominant strategy (DS) of the agent. It is proved that such a choice for all organizational system agents is an equilibrium in dominant strategy. An example of the application of the model for stimulating units for dealing with distressed assets (PRPA) in the North-West Bank of the Sberbank of the Russian Federation is given. There is a more accurate determination of the PPRA plans for the proposed model in comparison with the traditional "top-down" planning method.

Keywords: method of counter plans, incentives, KPI, problem assets of banks.

Статья представлена к публикации членом редакционной коллегии Э.Ю. Калимулиной.

Поступила в редакцию 25.07.2017.

Опубликована 31.05.2018.

УДК 021.8 + 025.1

ББК 78.34

МНОГОИНДЕКСНАЯ ЗАДАЧА ПОСТРОЕНИЯ РАСПИСАНИЯ СЕМИНАРА¹

Афраймович Л. Г.², Ильин С. Ю.³

(Нижегородский государственный университет
им. Н.И. Лобачевского, Нижний Новгород)

Целью работы является расширение области применимости потоковых методов анализа многоиндексных задач при решении одной задачи теории расписаний. В работе рассматривается многоиндексная постановка задачи построения расписания семинара. Рассмотрено два частных случая: задача с декомпозиционной матрицей эффективности расписания и задача с матрицей эффективности, отражающей соответствие аудиторий лекциям. Для первого частного случая показана возможность сведения задачи построения расписания семинара к задаче поиска потока в сети. Для второго предложен эвристический алгоритм решения, приведены результаты вычислительного эксперимента.

Ключевые слова: построение расписания, многоиндексная задача, потоковый алгоритм, эвристика.

1. Введение

Задачи планирования возникают во многих сферах жизнедеятельности [6]. В данной работе мы рассматриваем одну из задач планирования – задачу построения расписания при

¹ Работа выполнена при поддержке РФФ, проект № 15-11-30022 «Глобальная оптимизация, суперкомпьютерные вычисления и приложения».

² Лев Григорьевич Афраймович, доктор физико-математических наук, профессор (levafraymovich@gmail.com).

³ Сергей Юрьевич Ильин, студент (ilin_nmov_pf@mail.ru).

проведении семинара, которая заключается в следующем. Компания организует выездной образовательный семинар – определяет перечень лекций, возможных лекторов, возможные аудитории для проведения лекций. Известны временные слоты доступности аудиторий и лекторов. Заданы коэффициенты, отражающие условную «эффективность» проведения лекции каждым из лекторов в каждой из аудиторий в каждый из временных слотов. Требуется построить расписание проведения семинара, удовлетворяющее заданным ограничениям и оптимизирующий общий критерий эффективности построенного расписания.

Задачи планирования в большом числе постановок оказываются NP-трудными, что, согласно современной гипотезе, означает отсутствие эффективных алгоритмов решения [5, 10]. В литературе рассматриваются подходы к решению задач планирования, основанные на алгоритмах дискретной оптимизации, таких как метод ветвей и границ, метод динамического программирования [7], рассматриваются эвристические алгоритмы планирования [11], исследуются частные подклассы задач планирования, для решения которых применимы ε -приближенные алгоритмы [8].

Одним из перспективных направлений при разработке эффективных алгоритмов исследования задач дискретной оптимизации является нахождение подклассов задач, для решения которых применимы потоковые методы. Важной мотивацией данных исследований являются активно развивающиеся эффективные потоковые алгоритмы [9]. В [1–4] были найдены необходимые и (или) достаточные условия сводимости многоиндексных задач к задачам поиска потока в сети. На основе результатов сводимости в [1–4] были построены эффективные алгоритмы решения многоиндексных задач транспортного типа, удовлетворяющих условиям сводимости.

Данная работа является продолжением исследований, связанных с анализом сводимости оптимизационных задач к задаче поиска потока в сети при анализе задач теории планирования. В работе дается многоиндексная постановка задачи построения расписания семинара и рассматривается два частных подкласса. Для первого подкласса показано, что задача

имеет декомпозиционную структуру, что согласно [3] гарантирует сводимость к задаче поиска потока в сети. Для второго подкласса предложен эвристический алгоритм, идея которого состоит в построении специальной потоковой модели, поиске оптимального потока, его декомпозиции на простые циклы и построении решения исходной задачи через соответствие простых циклов сетевой модели и переменных исходной задачи планирования.

2. Постановка задачи

2.1. СОДЕРЖАТЕЛЬНАЯ ПОСТАНОВКА

Рассмотрим содержательную постановку задачи построения расписания семинара. Известны доступные аудитории, лекторы и множество лекций, которые необходимо провести в рамках семинара. Для каждой аудитории известно время ее доступности. Для каждого лектора известно максимальное число лекций и множество лекций, которые он может провести, а также доступное время. Для семинара определено максимальное число лекций, которые могут идти параллельно. Нужно составить расписание семинара: определить время проведения лекций, назначить лекторов и аудитории таким образом, чтобы «обобщенная эффективность» расписания принимала экстремальное значение.

При решении задачи будем предполагать, что все лекции имеют равную длительность, а временные слоты модельного времени разбито на такты, каждый из которых соответствует интервалу времени, в который может быть проведена лекция.

2.2. ИСХОДНЫЕ ПАРАМЕТРЫ

Введём следующие исходные параметры: I – множество аудиторий, J – множество лекторов, K – множество лекций, T – множество тактов времени. Далее пусть

$$S_{it} = \begin{cases} 1, & \text{если аудитория } i \text{ свободна в такт } t, \\ 0 & \text{иначе;} \end{cases}$$

$$L_{jk} = \begin{cases} 1, & \text{если лектор } j \text{ может провести лекцию } k, \\ 0 & \text{иначе;} \end{cases}$$

$$H_{jt} = \begin{cases} 1, & \text{если лектор } j \text{ свободен в такт } t, \\ 0 & \text{иначе;} \end{cases}$$

G_j – максимальное количество семинаров, которое готов провести лектор j ; B – максимальное количество лекций, которые могут проводиться параллельно; C_{ijkt} – коэффициент, отражающий эффективность проведения лектором j лекции k в аудитории i в такт t , $i \in I$, $j \in J$, $k \in K$, $t \in T$.

2.3 ВАРЬИРУЕМЫЕ ПАРАМЕТРЫ

Введём следующие варьируемые параметры:

$$x_{ijkt} = \begin{cases} 1, & \text{если лектор } j \text{ пров. лекцию } k \text{ в ауд. } i \text{ в такт } t, \\ 0 & \text{иначе;} \end{cases}$$

$i \in I, j \in J, k \in K, t \in T$.

2.4 СИСТЕМА ОГРАНИЧЕНИЙ

Допустимым расписанием будем называть набор варьируемых параметров x_{ijkt} , $i \in I$, $j \in J$, $k \in K$, $t \in T$, удовлетворяющих следующей системе ограничений:

$$(1) \quad \sum_{j \in J} \sum_{k \in K} x_{ijkt} \leq S_{it}, \quad i \in I, t \in T;$$

(аудиторию можно занимать только когда она свободна);

$$(2) \quad \sum_{i \in I} \sum_{t \in T} x_{ijkt} \leq L_{jk}, \quad j \in J, k \in K;$$

(лектора можно назначить только на лекцию, которую он может провести);

$$(3) \quad \sum_{i \in I} \sum_{k \in K} x_{ijkt} \leq H_{jt}, \quad j \in J, t \in T;$$

(лектора можно назначить только когда он свободен);

$$(4) \quad \sum_{i \in I} \sum_{k \in K} \sum_{t \in T} x_{ijkt} \leq G_j, \quad j \in J;$$

(количество лекций, проводимых лектором, не может превышать максимально возможного значения);

$$(5) \sum_{i \in I} \sum_{j \in J} \sum_{t \in T} x_{ijkt} = 1, k \in K;$$

(каждая лекция должна быть проведена одним лектором, в одной аудитории, в один такт времени);

$$(6) \sum_{j \in J} \sum_{k \in K} x_{ijkt} \leq 1, i \in I, t \in T;$$

(в каждый такт времени в аудитории проводится не более одной лекции);

$$(7) \sum_{i \in I} \sum_{k \in K} x_{ijkt} \leq 1, j \in J, t \in T;$$

(в каждый такт времени лектор проводит не более одной лекции);

$$(8) \sum_{i \in I} \sum_{j \in J} \sum_{k \in K} x_{ijkt} \leq B, t \in T;$$

(в каждый такт времени параллельно проводятся не больше B лекций);

$$(9) x_{ijkt} \in \{0, 1\}, i \in I, j \in J, k \in K, t \in T;$$

(естественные ограничения на переменные).

2.5 КРИТЕРИЙ ОПТИМИЗАЦИИ

Задача заключается в поиске допустимого расписания, при котором принимает минимальное значение критерий

$$(10) \sum_{i \in I} \sum_{j \in J} \sum_{k \in K} \sum_{t \in T} C_{ijkt} x_{ijkt} \rightarrow \min,$$

характеризующий суммарную эффективность построенного расписания.

2.6 ВЫЧИСЛИТЕЛЬНАЯ СЛОЖНОСТЬ

Несложно увидеть, что задача построения расписания семинара (1)–(10) является NP-трудной, так как к ней за полиномиальное время сводится классическая NP-трудная аксиальная задача о назначениях [5], что обуславливает поиск частных подклассов задачи построения расписания семинара, для исследования которых применимы эффективные методы.

3. Матрица эффективности с декомпозиционной структурой

Известны классы многоиндексных задач, допускающие сведение к задачам поиска потока в сети [1–4]. Покажем, что поставленная задача построения расписания семинара формализуется как многоиндексная задача транспортного типа. При этом, когда многоиндексная матрица C_{ijkt} , определяющая критерий эффективности расписания, имеет декомпозиционную структуру (см. [3]), задача построения расписания семинара сводима к задаче поиска потока в сети.

Несложно увидеть, что система ограничений (1)–(9) эквивалентна следующей системе:

$$(11) \sum_{j \in J} \sum_{k \in K} x_{ijkt} \leq S_{it}, i \in I, t \in T;$$

$$(12) \sum_{i \in I} \sum_{t \in T} x_{ijkt} \leq L_{jk}, j \in J, k \in K;$$

$$(13) \sum_{i \in I} \sum_{k \in K} x_{ijkt} \leq H_{jt}, j \in J, t \in T;$$

$$(14) \sum_{i \in I} \sum_{k \in K} \sum_{t \in T} x_{ijkt} \leq G_j, j \in J;$$

$$(15) \sum_{i \in I} \sum_{j \in J} \sum_{t \in T} x_{ijkt} = 1, k \in K;$$

$$(16) \sum_{i \in I} \sum_{j \in J} \sum_{k \in K} x_{ijkt} \leq B, t \in T;$$

$$(17) x_{ijkt} \in Z_+, i \in I, j \in J, k \in K, t \in T.$$

Согласно [1], задача (11)–(17), (10) относится к классу $W(M)$, $M = \{\{j, k\}, \{i, t\}, \{i, k\}, \{i, k, t\}, \{i, j, t\}, \{i, j, k\}\}$. Здесь множество M является множеством подмножеств индексов, по которым происходит суммирование многоиндексной матрицы неизвестных в системе ограничений (11)–(16). Так, например, в ограничении (11) происходит суммирование по индексам j, k при фиксированных индексах i, t . Таким образом, ограничение (11) определяет элемент $\{j, k\}$ множества M и т.д.

Далее рассмотрим частный случай задачи (1)–(10), когда матрица C_{ijkt} обладает следующей декомпозиционной структурой. Пусть заданы матрицы U_{ij} , V_{jk} , W_{jt} , $i \in I$, $j \in J$, $k \in K$, $t \in T$. Тогда рассмотрим частный случай, когда

$$(18) C_{ijkt} = U_{it} + V_{jt} + W_{jk}, i \in I, j \in J, k \in K, t \in T.$$

Критерий оптимальности в данном случае будет иметь вид

$$(19) \sum_{i \in I, j \in J, k \in K, t \in T} (U_{it} + V_{jt} + W_{jk}) x_{ijkt} \rightarrow \min.$$

Согласно [1], задача (11)–(16), (19) относится к классу $W(M, H)$, где $M = \{\{j, k\}, \{i, t\}, \{i, k\}, \{i, k, t\}, \{i, j, t\}, \{i, j, k\}\}$, $H = \{\{i, t\}, \{j, t\}, \{k, j\}\}$. Здесь множество M , как уже отмечено выше, определяет систему ограничений (11)–(17), H – множество подмножеств индексов, задающих декомпозиционное представление (18) матрицы критерия оптимальности (19).

Введем ряд вспомогательных определений, предложенных в [1]. Пусть F – множество индексов многоиндексной задачи и $f \subseteq F$, тогда обозначим $\bar{f} = F \setminus f$. Далее, пусть M – некоторое множество подмножеств индексов, $M \subseteq 2^F$. Тогда обозначим $\tilde{M} = \{\bar{f}, f \in M\}$.

Пусть f_1, f_2, \dots, f_k – некоторое разбиение множества F . Тогда будем говорить, что множество M обладает f_1, f_2, \dots, f_k -декомпозиционной структурой, если $M \subseteq \{f_i / i = 1, 2, \dots, k\} \cup \cup \{f_i \cup f_{i+1} \mid i = 1, 2, \dots, k-1\}$. В работах [1, 3] было показано, что если множества \tilde{M} и H одновременно являются f_1, f_2, \dots, f_k -декомпозиционными, то задача $W(M, H)$ сводима к задаче поиска потока в сети.

Согласно введенным выше обозначениям, $F = \{i, j, k, t\}$ для задачи (11)–(17), (19). Тогда рассмотрим разбиение $f_1 = \{k\}$, $f_2 = \{j\}$, $f_3 = \{t\}$, $f_4 = \{i\}$ множества F .

Несложно увидеть, что множества \tilde{M} и H одновременно являются f_1, f_2, f_3, f_4 -декомпозиционными и задача (11)–(17), (19) сводима к задаче поиска потока в сети. Тогда согласно [1] существует алгоритм решения задачи (11)–(17), (19), требующий $O(n^2 \log^2 n)$ вычислительных операций, где $n = |I \times J \times K \times T|$.

4. Матрица эффективности, отражающая соответствие аудиторий лекциям

Рассмотрим частный случай задачи (1)–(10), когда матрица эффективности отражает соответствие аудиторий лекциям. Подобная ситуация возникает, например, когда аудитория обладает некоторой вместительностью, оборудована специальной аппаратурой и т.д., а при этом для лекций задано пожелание о вместимости аудиторий, наличии аппаратуры и т.д. Тогда обозначим через O_{ik} признак соответствия аудитории i лекции k : O_{ik} равен 0, если лекция и аудитория соответствуют друг другу; и равен 1 в противном случае, $i \in I$, $k \in K$.

Критерий оптимальности при этом будет иметь вид

$$(20) \sum_{i \in I, j \in J, k \in K, t \in T} O_{ik} x_{ijkt} \rightarrow \min.$$

Критерий (20) соответствует задаче построения расписания семинара с минимизацией количества лекций, проводимых в «несоответствующих» аудиториях.

Для решения задачи (1)–(9), (20) предлагается эвристический алгоритм. Ниже приводится описание данного алгоритма и результаты вычислительного эксперимента.

4.1. АЛГОРИТМ ПОИСКА РЕШЕНИЙ ЗАДАЧ СООТВЕТСТВИЯ

Шаг 1. Построим сетевую потоковую модель. Пусть $V = K \cup K' \cup J \cup J' \cup T \cup T' \cup I$ – множество вершин, где $K' = \{k' | k \in K\}$, $J' = \{j' | j \in J\}$, $T' = \{t' | t \in T\}$, $A = K \times K' \cup K' \times J \cup J \times J' \cup J' \times T \cup T \times T' \cup T' \times I \cup I \times K$ – множество дуг сети. Если a – дуга сети, то через l_a , u_a , и c_a обозначим верхнюю пропускную способность, нижнюю пропускную способность и стоимость дуги a соответственно, $a \in A$. Тогда определим:

$$\begin{aligned} l_{(k,k')} &= 1, u_{(k,k')} = 1, c_{(k,k')} = 0, (k, k') \in K \times K', \\ l_{(k',j)} &= 0, u_{(k',j)} = L_{jk}, c_{(k',j)} = 0, (k', j) \in K' \times J, \\ l_{(j,j')} &= 0, u_{(j,j')} = G_j, c_{(j,j')} = 0, (j, j') \in J \times J', \\ l_{(j',t)} &= 0, u_{(j',t)} = H_{jt}, c_{(j',t)} = 0, (j', t) \in J' \times T, \end{aligned}$$

$$l_{(t,t')} = 0, u_{(t,t')} = B, c_{(t,t')} = 0, (t, t') \in T \times T',$$

$$l_{(t',i)} = 0, u_{(t',i)} = S_{it'}, c_{(t',i)} = 0, (t', i) \in T' \times I,$$

$$l_{(i,k)} = 0, u_{(i,k)} = 1, c_{(i,k)} = O_{ik}, (i, k) \in I \times K.$$

Шаг 2. Найдем циркуляцию минимальной стоимости в данной сети и декомпозируем поток на простые циклы [3].

Шаг 3. Рассмотрим все циклы декомпозиции потока. По построению циклы сети имеют длину, кратную 7. Если длина цикла равна 7, т.е. имеет вид

$$k, k', j, j', t, t', i, k,$$

то определим

$$x_{ijkt} = 1.$$

Если длина цикла равна $7p$, т.е. цикл имеет вид

$k_1, k'_1, j_1, j'_1, t_1, t'_1, i_1, k_2, k'_2, j_2, j'_2, t_2, t'_2, i_2, \dots, k_p, k'_p, j_p, j'_p, t_p, t'_p, i_p, k_1$, то декомпозируем данный цикл на подциклы длины 7. Среди возможных четырёх способов декомпозиции выберем наилучший, исходя из стоимостных коэффициентов O_{ik} . Далее определим переменные задачи построения расписания семинара, соответствующие подциклам длины 7, равными 1.

По построению полученное решение будет являться допустимым решением задачи (1)–(9), (20).

4.2. РЕЗУЛЬТАТЫ ВЫЧИСЛИТЕЛЬНОГО ЭКСПЕРИМЕНТА

Для того чтобы оценить качество эвристического алгоритма будем сравнивать значение критерия с его нижней оценкой, полученной при помощи линейной релаксации задачи (1)–(9), (20). Линейной релаксацией задачи (1)–(9), (20) является задача (1)–(8), (21), (20), в которой условие целочисленности переменных (9) заменено на линейное ограничение (21):

$$(21) \quad 0 \leq x_{ijkt} \leq 1, \quad i \in I, j \in J, k \in K, t \in T.$$

Оптимум задачи (1)–(8), (21), (20) будет являться нижней оценкой для исходной задачи (1)–(9), (20) в связи с тем, что множество допустимых решений системы (1)–(8), (21) включает в себя множество допустимых решений системы (1)–(9).

Таблица 1. Результаты вычислительного эксперимента

Размерность ¹	Число проведённых экспериментов	Оценка работы эвристического алгоритма ²
8	10	10%
9	10	6%
10	10	10%
13	10	10%

¹ За размерность n примем $I = J = K = T = n$.

² $(F_3 - F^*)/|K|$, где F_3 – значение, полученное эвристикой; F^* – нижняя оценка; $|K|$ – число семинаров.

5. Заключение

Рассмотрена NP-трудная многоиндексная задача построения расписания семинара. Рассмотрено два частных случая: задача с декомпозиционной матрицей эффективности расписания и задача с матрицей эффективности, отражающей соответствие аудиторий лекциям. Для первого частного случая показана возможность сведения задачи построения расписания семинара к задаче поиска потока в сети. Тем самым предложен полиномиальный алгоритм решения, требующий $O(n^2 \log^2 n)$ вычислительных операций, где $n = |I \times J \times K \times T|$. Для второго случая предложен полиномиальный эвристический алгоритм решения, основанный на поиске потока в специальной сетевой модели и декомпозиции найденного потока. Приведены результаты вычислительного эксперимента, показывающие отклонение значения критерия на решении, которое было найдено эвристическим алгоритмом, от нижней оценки задачи. Полученные результаты демонстрируют перспективность предложенного подхода – на тестовой выборке значение критерия отклонялось от нижней оценки не более чем на 10%.

Дальнейшее направление исследования связано с исследованием иных частных подклассов задачи построения расписания семинара и дальнейшей проработкой эвристического подхода, основанного на декомпозиции циркуляции соответствующей потоковой модели.

Литература

1. АФРАЙМОВИЧ Л.Г. *Эвристический метод решения целочисленных декомпозиционных многоиндексных задач* // Автоматика и телемеханика. – 2014. – №8. – С. 3–18.
2. АФРАЙМОВИЧ Л.Г. *Многоиндексные транспортные задачи с 2-вложенной структурой* // Автоматика и телемеханика. – 2013. – №1. – С. 116–134.
3. АФРАЙМОВИЧ Л.Г. *Многоиндексные транспортные задачи с декомпозиционной структурой* // Автоматика и телемеханика. – 2012. – №1. – С. 130–147.
4. АФРАЙМОВИЧ Л.Г. *Циклическая сводимость многоиндексных систем линейных неравенств транспортного типа* // Известия Российской академии наук. Теория и системы управления. – 2010. – №4. – С. 83–90.
5. ГЭРИ М., ДЖОНСОН Д. *Вычислительные машины и труднорешаемые задачи.* – М.: Мир, 1982.
6. ЛАЗАРЕВ А.А., МУСАТОВА Е.Г., КВАРАЦХЕЛИЯ А.Г., ГАФАРОВ Е.Р. *Теория расписаний. Задачи железнодорожного планирования.* – М.: ИПУ РАН, 2012.
7. PINEDO M.L. *Scheduling Theory, Algorithms, and Systems.* – Springer, 2016.
8. JANSEN K. PRÄDEL L., SCHWARZ U.M., SVENSSON O. *Faster Approximation Algorithms for Scheduling with Fixed Jobs* // Seventeenth Computing: The Australasian Theory Symposium, CATS-2011, Perth, Australia, January, 2011. – Vol. 119. – P. 3–10.
9. AHUJA R.K., MAGNATI T.L., ORLIN J.B. *Network flows: theory, algorithms, and applications.* – N.J. Prentice Hall, 1993.
10. BRUCKER P. *Scheduling algorithms.* – Springer-Verlag Berlin Heidelberg, 2007.
11. SAMPSON S., WEISS E. *Increasing service levels in conference and educational scheduling: A heuristic approach* // Management Science. – 1995. – No. 41(11). – P. 1816–1825.

MULTINEDEX PROBLEM OF SEMINAR SCHEDULING

Lev Afraymovich, Nizhny Novgorod State University, Nizhny Novgorod, Dr.Sc., professor (levafraymovich@gmail.ru).

Sergey Ilyin, Nizhny Novgorod State University, Nizhny Novgorod, Dr.Sc., student (ilin_nnov_pf@mail.ru).

Abstract: The goal of the paper is to expand the field of applicability of flow algorithms for analyzing multi-index problems in solving one problem of scheduling theory. The paper considers a multi-index statement of seminar scheduling problem. Two particular cases are considered: the problem with the decomposition matrix of schedule effectiveness and the problem with the efficiency matrix, which reflects the correspondence of audiences to lectures. For the first special case, the possibility of reduction the seminar scheduling problem to network flow problem is shown. For the second one, a heuristic solution algorithm is proposed, and the results of the computational experiment are presented.

Keywords: scheduling, multi-index problems, flow algorithm, heuristics.

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

*Поступила в редакцию 15.10.2017.
Опубликована 31.05.2018.*

УДК 004.93
ББК 32.972.1

ОБЗОР АЛГОРИТМОВ ФОНЕТИЧЕСКОГО КОДИРОВАНИЯ

Выхованец В. С.¹,

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

Ду Ц.², Сакулин С. А.³

(МГТУ им. Н.Э. Баумана, Москва)

Приведен обзор алгоритмов фонетического кодирования, предназначенных для определения схожести слов по звучанию (произношению). Алгоритмы фонетического кодирования разделены на алгоритмы для сравнения слов и алгоритмы определения расстояния между словами. Описаны алгоритмы сравнения слов SoundEx, NYSIIS, Daitch-Mokotoff, Metaphone, Polyphone и алгоритмы вычисления расстояния между словами Левенштейна, Джаро, на основе N-грамм. Для каждого алгоритма указаны его достоинства и недостатки, приводится аналог алгоритма для русского языка. Для устранения общих недостатков алгоритмов фонетического кодирования предложено использовать не последовательности букв слов, а последовательности их элементарных звуков.

Ключевые слова: алгоритм фонетического кодирования, фонетическое расстояние, индексирование слов по звучанию.

¹ Валерий Святославович Выхованец, доктор технических наук, доцент (valery@vykhovanets.ru).

² Цзяньмин Ду, аспирант (forsola@qq.com).

³ Сергей Александрович Сакулин, кандидат технических наук, доцент (ss141291@yandex.ru).

1. Введение

Алгоритмы фонетического кодирования представляют собой алгоритмы индексирования слов по их звучанию, которые на основе последовательности букв слова и правил произношения преобразуют их в кодирующий текст (код, индекс, ключ). При совпадении или близости кодирующего текста для двух различных слов делается вывод о близости этих слов по звучанию.

Первым представителем алгоритмов фонетического кодирования для английского языка стал алгоритм SoundEx [69], который использовался в тридцатые годы для кодирования фамилий при переписи населения. В алгоритме используется метод кодирования, призванный устранить орфографические и типографические ошибки в именах. Например, алгоритм SoundEx вычислит одинаковый кодирующий текст «S530» для таких слов как «Smith», «Smithe» и «Smyth», т.е. отождествит эти слова как одинаковые по звучанию.

С появлением и развитием компьютерных технологий появилось множество других алгоритмов фонетического кодирования, в том числе и для различных естественных языков. В качестве примеров такого рода алгоритмов следует назвать алгоритмы NYSIS [77], улучшенный SoundEx [75], Metaphone [40] и др.

Алгоритмы фонетического кодирования включают в себя не только алгоритмы для сравнения слов, но и алгоритмы определения расстояния между словами при поиске по звучанию. На практике наибольшее распространение получили алгоритмы вычисления расстояния Левенштейна [2] и Джаро [30], а также расстояния на основе N -грамм [33].

Алгоритмы фонетического кодирования широко используются в областях где требуется сопоставление акустических данных с текстовыми образцами. Например, в области распознавания речи, для проверки написания и исправления слов, для поиска в базах данных, при добыче полезных ископаемых, для идентификации пользователей, для кросс-языкового кодирования речи, при поиске в сети интернет и т.д.

Несмотря на появление и развитие новых моделей и методов распознавания речи, таких как вероятностные и статистические алгоритмы, скрытые марковские модели, нейронные сети, машинное обучение и др., алгоритмы фонетического кодирования не утратили своей актуальности, так как являются базовыми для применения этих моделей и методов на практике [14, 22, 44].

Настоящая статья посвящена обзору алгоритмов фонетического кодирования, главная цель которого – показать те решения, которые используются для сравнения строк по звучанию, а также выявить тенденции их развития.

2. Базовые алгоритмы

2.1. АЛГОРИТМ SOUNDEX

Алгоритм SoundEx запатентован почти сто лет назад [69]. Алгоритм SoundEx для английского языка состоит из пяти шагов.

Шаг 1. Сохранить первую букву слова.

Шаг 2. Удалить из слова буквы A, E, I, O, U, Y, W и H.

Шаг 3. Заменить оставшиеся буквы на цифры согласно таблице 1.

Таблица 1. Кодирование букв в алгоритме SoundEx

Цифра	Буквы
1	B, P, F, V
2	C, S, K, G, J, Q, X, Z
3	D, T
4	L
5	M, N
6	R

Шаг 4. Если в коде имеются группа из одинаковых цифр, то заменить эту группу первой цифрой, за исключением цифр, разделяющих буквы W и H в исходном слове.

Шаг 5. Сформировать результирующий код из первой буквы слова и трех первых цифр, полученный на предыдущих шагах. Если в коде менее трех цифр, то дополнить код нулями.

Пример 1. Рассмотрим следующие два слова: «Lee» и «Shaw». Применяв алгоритм, описанный выше, получаем кодирующий текст «L000» для слова «Lee» и «S000» – для слова «Shaw».

Пример 2. Для пояснения шага 4 рассмотрим слово «Ashcraft». Из-за наличия шага 4 кодирующий текст для этого слова будет «A226», а не «A261».

Как видно из таблицы 1, основной принцип кодирования букв в алгоритме состоит в том, что близкие по звучанию буквы кодируются одной и той же цифрой, а непронизносимые буквы удаляются. Однако алгоритм не свободен от недостатков.

Первый недостаток состоит в том, что существуют близкие по звучанию слова, которые имеют разный кодирующий текст.

Пример 3. Для слов «Lee» и «Leigh», имеющих одно и то же звучание, получаем различные коды: «L000» и «L200» соответственно.

Второй недостаток обратный первому: существуют различные по звучанию слова, которые имеют одинаковый кодирующий текст.

Пример 4. Слова «Gauss» и «Ghosh» с разным звучанием имеют код «G200».

Для частичного устранения проблем, выявленных в процессе эксплуатации алгоритма, в настоящее время для английского языка используется модифицированный алгоритм SoundEx со следующей таблицей кодирования.

Таблица 2. Кодирование букв в модифицированном алгоритме

Цифра	Буквы
1	B, P
2	F, V
3	C, S, K
4	G, J
5	Q, X, Z

6	D, T
7	L
8	M, N
9	R

Из таблицы 2 видно, что из групп 1 и 2 исходного алгоритма образованы новые группы со своим цифровым кодом. Дополнительно к этому длина кодирующего текста в модифицированном алгоритме не ограничивается четырьмя знаками.

Как показывают эксперименты, на одно значение кода SoundEx приходится до 21 фамилии. У модифицированного алгоритма таких фамилий две или три.

Алгоритм SoundEx сильно зависит от языка. Поэтому разработано множество модификаций этого алгоритма для различных языков: китайского [20], испанского [7], персидского [36], арабского [56], малайского [53] и т.д. Для русского языка используется кодирование букв в соответствии с таблицей 3, а удаляемыми буквами являются буквы У, Е, Ё, Ы, А, О, Э, Я, И, Ю, Ъ, Ь.

Таблица 3. Кодирование русских букв в алгоритме SoundEx

Цифра	Буквы
1	Б, П
2	Ф, В
3	Ж, З, С, Х
4	К, Г,
5	Ц, Ч, Ш, Щ
6	Д, Т
7	Л, Й
8	М, Н
9	Р

Несмотря на свои недостатки, алгоритм SoundEx имеет положительные отзывы [37]. В настоящее время алгоритм широко используется для сравнения слов по звучанию и значительно повышает вероятность идентификации слов. Благодаря своей

простоте и низкой вычислительной сложности алгоритм SoundEx стал стандартом и встроен в механизм поиска почти всех известных систем управления базами данных [27].

2.2. АЛГОРИТМ NYSIIS

Алгоритм NYSIIS разработан в 1970 году для использования в одноименной информационной системе «New York State Identification and Intelligence System» [77]. Этот алгоритм дает несколько лучшие результаты по сравнению с алгоритмом SoundEx, используя более сложные правила преобразования исходного слова в результирующий код.

Алгоритм учитывает произношение слов английского языка и состоит из шести шагов.

Шаг 1. Преобразование префикса (начала слова) путем следующих подстановок: MAC → MCC; KN → N; K → C; PH, PF → FF; SCH → SSS.

Шаг 2. Преобразование суффикса (конца слова) путем следующих подстановок: EE → Y; IE → Y; DT, RT, RD, NT, ND → D.

Шаг 3. Преобразование слова в целом путем следующих подстановок: EV → AF; A, E, I, O, U → A; Q → G; Z → S; M → N; KN → N; K → C; SCH → SSS; PH → FF, W → A.

Шаг 4. Удаление H после гласных и S, A в конце слова.

Шаг 5. Преобразование суффикса слова путем подстановки AY → Y.

Шаг 6. Ограничить полученный код 6 знаками.

Как видно из описания алгоритма, в NYSIIS используется большое число правил, связывающих написание с произношением, а также учитывается функционирование гласных звуков при произношении слов (все гласные заменяются гласной A).

При сравнительном исследовании алгоритмов фонетическом кодирования установлено, что алгоритм NYSIIS более всего подходит для кодирования фамилий, где дает отличные результаты.

Пример 5. Фамилии «Brain», «Brown» и «Brun» имеют код «Bran», фамилии «Carr», «Core», «Copp» и «Kipp» – код «Car»,

фамилии «Dane», «Dean», «Dent» и «Dionne» – код «Dan», фамилии «Smith», «Schmit», «Schmidt» – код «Snat», фамилии «Trueman», «Truman» – код «Tranan» [27].

2.3. АЛГОРИТМ DAITCH-MOKOTOFF SOUNDEX

В процессе использования фонетических алгоритмов установлено, что алгоритмы SoundEx и NYSIIS плохо работают со словами других языков. Для учета специфики произношения слов в других языках разработан алгоритм Daitch-Mokotoff SoundEx, названный по фамилиям авторов, который имеет такие усовершенствования как использование большей длины кода и учет различных произношений слов, выражающийся во множественном кодировании одного и того же слова.

Алгоритм имеет более сложные правила преобразования исходного слова в его код. Как и у алгоритма NYSIIS, в формировании результирующего кода участвуют не только одиночные буквы, но и их последовательности. Преобразования слова в числовой код осуществляются по следующей таблице [75]

Таблица 4. Кодирование букв в улучшенном алгоритме SoundEx

Последовательность букв	Н	Г	О
AI, AJ, AY, EI, EY, EJ, OI, OJ, OY, UI, UJ, UY	0	1	
AU	0	7	
IA, IE, IO, IU	1		
EU	1	1	
A, UE, E, I, O, U, Y	0		
J	1	1	1
SCHTSCH, SCHTSH, SCHTCH, SHTCH, SHCH, SHTSH, STCH, STSCH, STRZ, STRS, STSH, SZCZ, SZCS	2	4	4
SHT, SCHT, SCHD, ST, SZT, SHD, SZD, SD	2	43	43
CSZ, CZS, CS, CZ, DRZ, DRS, DSH, DS, DZH, DZS, DZ, TRZ, TRS, TRCH, TSH, TTSZ, TTZ, TZS, TSZ, SZ, TTCH, TCH, TTSCCH, ZSCH, ZHSH, SCH, SH, TTS, TC, TS, TZ, ZH, ZS	4	4	4
SC	2	4	4

DT, D, TH, T	3	3	3
CHS, KS, X	5	54	54
S, Z	4	4	4
CH, CK, C, G, KH, K, Q	5	5	5
MN, NM	66	66	66
M, N	6	6	6
FB, B, PH, PF, F, P, V, W	7	7	7
H	5	5	
L	8	8	8
R	9	9	9

Порядок преобразований соответствует порядку последовательностей букв в таблице 4. Колонки H, Г и O задают цифровые коды для последовательности букв из первой колонки: H – в начала слова, Г – перед гласной, O – в остальных случаях.

Альтернативные коды слова, учитывающие различные произношения, поучаются для слов, образованных из исходного путем следующих подстановок: CH → KH, TCH; CK → K, TSK; C → K, TZ; J → Y, DZH; RS → RTZ, ZH.

Пример 6. Имя «Peters» с SoundEx кодом «P362» при использовании алгоритма Daitch-Mokotoff SoundEx получает два кода: 739400 для формы «Peters» и 734000 для формы «Petertz», а имя «Jackson» с SoundEx кодом «J250» – четыре кода: 154600, 454600, 145460, 445460 для форм «Jackson», «Yakson», «Jakson» и «Jatskon» соответственно.

2.4. АЛГОРИТМ METAPHONE

Metaphone – еще один алгоритм фонетического кодирования слов с учётом основных правил английского языка, разработанный в 1990 г. [40]. Он отличается от предыдущих алгоритмов тем, что реализует более сложные правила преобразования. Другое отличие состоит в том, что буквы не разбиваются на группы и не кодируются цифрами. На выходе алгоритм даёт код переменной длины, состоящий из букв.

Алгоритм включает в себя 16 шагов [10].

Шаг 1. Удаление повторяющихся соседних букв кроме буквы С.

Шаг 2. Преобразования префикса слова путем следующих подстановок: KN → N; GN → N; PN → N; AE → E; WR → R.

Шаг 3. Удаление суффикса MB.

Шаг 4. Преобразования последовательности букв с буквой С путем следующих подстановок: CIA → XIA; SCH → SKH; CH → XH; CI → SI; CE → SE; CY → SY; C → K.

Шаг 5. Преобразования последовательности букв с буквой D путем следующих подстановок: DGE → JGE; DGY → JGY; DGI → JGY; D → T.

Шаг 6. Подстановка GH → H, если GH стоит не в конце слова и не перед гласной.

Шаг 7. Подстановка суффиксов GN → N и GNED → NED.

Шаг 8. Преобразования последовательности букв с буквой G путем следующих подстановок: GI → JI; GE → JE; GY → JY; G → K.

Шаг 9. Удаление H после гласных, но не перед гласными.

Шаг 10. Преобразования слова в целом путем следующих подстановок: CK → K; PH → F; Q → K; V → F; Z → S.

Шаг 11. Преобразования последовательности букв с буквой S путем следующих подстановок: SH → XH; SIO → XIO; SIA → XIA.

Шаг 12. Преобразования последовательности букв с буквой T путем следующих подстановок: TIA → XIA; TIO → XIO; TH → O; TCH → CH.

Шаг 13. Подстановка префикса WH → W. Если после W нет гласной, то удаление W.

Шаг 14. Подстановка префикса X → S, в середине слова – X → KS.

Шаг 15. Удаление Y, которые не находятся перед гласными.

Шаг 16. Удаление всех гласных, кроме начальной.

Пример 7. Фамилии «Brain», «Brown» и «Brun» имеют код «BRN», фамилии «Carr», «Core», «Copp» и «Kipp» – код «KP», фамилии «Dane», «Dean» и «Dionne» – код «TN», но фамилия «Dent» – код «TNT» (см. пример 5), фамилия «Smith» – код «SM0», фамилия «Schmit» – код «SXMT», фамилия «Schmidt» – код «SXMTT», фамилии «Trueman», «Truman» – код «TRMN».

Позднее в 2000 г. была разработана вторая версия алгоритма, которая получила название Double Metaphone [41], где, в отличие от первой версии, применимой только к английскому языку, учитываются особенности произношения слов, заимствованных из других языков [41]. Для этих слов результатом работы алгоритма являются два кода – по одному для каждого варианта произношения.

Хотя Double Metaphone имеет преимущества перед алгоритмом Metaphone, он по-прежнему имеет некоторые ограничения [45]. В частности, все так же встречаются слова с разным произношением и одинаковым кодом, например, слова «Alice» и «Elsa» и «Ullo» кодируются как «ALS».

В 2009 г. появилась третья коммерческая версия алгоритма под названием Metaphone 3. Утверждается, что новый алгоритм увеличивает точность отождествления слов с 89% – Double Metaphone, до 98% – Metaphone 3 [58]. Также Metaphone 3 начал поддерживать заимствованные слова из большего числа языков. Алгоритм Metaphone 3 достаточно сложный и включает большое число правил. Описание этого алгоритма на языке Java занимает более семи тысяч строк [42].

Алгоритм Metaphone был адаптирован к русскому языку [1]. Для русского языка алгоритм Metaphone состоит из пяти шагов.

Шаг 1. Преобразование гласных путем следующих подстановок: О, Ы, Я → А; Ю → У; Е, Ё, Э, ЙО, ЙЕ → И.

Шаг 2. Оглушение согласных букв, за которыми следует любая согласная, кроме Л, М, Н или Р, либо согласных на конце слова путем следующих подстановок: Б → П; З → С; Д → Т; В → Ф; Г → К.

Шаг 3. Удаление повторяющихся букв.

Шаг 4. Преобразование суффикса слова путем следующих подстановок: УК, ЮК → 0; ИНА → 1; ИК, ЕК → 2; НКО → 3; ОВ, ЕВ, ИЕВ, ЕЕВ → 4; ЫХ, ИХ → 5; АЯ → 6; ЫЙ, ИЙ → 7; ИН → 8; ОВА, ЕВА, ИЕВА, ЕЕВА → 9; ОВСКИЙ → @; ЕВСКИЙ → #; ОВСКАЯ → \$; ЕВСКАЯ → %.

Шаг 5. Удаление букв Ъ, Ь и дефиса.

Однако из-за небольшого числа правил Metaphone для русского языка не отождествляет некоторые фонетические схожие слова.

2.5. АЛГОРИТМ POLYPHONE

Специально для русского языка разработан алгоритм фонетического кодирования Polyphone [59]. При разработке алгоритма учтены морфологические, фонологические, фонетические и исторические аспекты произношения слов в русском языке.

Алгоритм состоит из следующих шагов.

Шаг 1. Подстановка вместо латинских букв схожих букв русского алфавита: А → А; Е → Е; О → О; С → С; Х → Х; В → В; М → М; Н → Н. Последние три подстановки применяются только для прописных букв.

Шаг 2. Удаление всех букв, не принадлежащих алфавиту русского языка.

Шаг 3. Удаление букв Ъ, Ы.

Шаг 4. Замена двух одинаковых букв одной.

Шаг 5. Замена одиночных букв: А, Е, Ё, И, О, Ы, Э, Я → А; Б → П; В → Ф; Г → К; Д → Т; З → С; Щ → Ш; Ж → Ш; М → Н; Ю → У.

Шаг 6. Преобразование слова путем выполнения следующих подстановок: АКА → АФА; АН → Н; ЗЧ → Ш; ЛНЦ → НЦ; ЛФСТФ → ЛСТФ; НАТ → Н; НТЦ → НЦ; НТ → Н; НТА → НА; НТК → НК; НТС → НС; НТСК → НСК; НТШ → НШ; ОКО → ОФО; ПАЛ → ПЛ; РТЧ → РЧ; РТЦ → РЦ; СП → СФ; ТСЯ → Ц; СТЛ → СЛ; СТН → СН; СЧ → Ш; СШ → Ш; ТАТ → Т; ТСА → Ц; ТАФ → ТФ; ТС → ТЦ; ТЦ → Ц; ТЧ → Ч; ФАК → ФК; ФСТФ → СТФ; ШЧ → Ш.

Пример 8. В соответствии с описанным выше алгоритмов слова «ТЕЛЕГРАММА» и «АППАРАТ» преобразуются в «ТАЛАКРАМА» и «АПАРАТ».

Для нечеткого сравнения слов предлагается использовать сумму простых чисел, закрепленных за каждой буквой преобразованного слова: А → 2; П → 3; К → 5; Л → 7; М → 11; Н → 13;

Р → 17; С → 19; Т → 23; У → 29; Ф → 31; Х → 37; Ц → 41;
Ч → 43; Щ → 47; Э → 53; Я → 59.

Пример 9. Слова из предыдущего примера получать следующие числовые коды: слово «ТАЛАКРАМА» – код 71, «АПАРАТ» – код 49.

В работе приводятся результаты сравнения алгоритма Polyphone с другими алгоритмами. В частности, процент правильно отождествленных слов у алгоритма Polyphone оценен как 95,12 %, в то время как наилучший результат после транслитерации русских слов по ГОСТ Р 52535.1-2006 у алгоритма NYSIS составляет 75,97 %, у алгоритма SoundEx – 90,24 %, у алгоритма Metaphone – 90,29 %, у алгоритма Double Metaphone – 96,15 %, у алгоритма Daitch-Mokotoff Soundex – 96,84 %.

В то же время утверждается, что точность нечеткого сравнения слов достигает 98,8 %.

3. Фонетическое расстояние

Основным методом, реализуемым в рассмотренных ранее алгоритмах фонетического кодирования, является метод эквивалентных преобразований слова по звучанию, при котором часть слова, принадлежащая некоторому множеству (классу эквивалентности) заменяются кодом этого множества или его типичным представителем.

При этом заметно, что части слов из одного множества, близких по звучанию, также близки по написанию. При введении соответствующей метрики на словах можно поставить задачу определения схожести слов по звучанию путем подсчета расстояния между словами по написанию.

Этот подход используется в другом классе алгоритмов фонетического кодирования, в котором вычисление фонетического кода слова заменяется попарным сравнением слов путем вычисления расстояния между ними в некотором метрическом пространстве. Предполагается, что речевой и письменный строй естественного языка сильно коррелирует между собой. Осталось только найти соответствующую метрику.

3.1. РАССТОЯНИЕ ЛЕВЕНШТЕЙНА

Расстояние Левенштейна – это мера различия двух слов относительно минимального числа операций вставки, удаления и замены, необходимых для преобразования одного слова в другое [2].

Расстояние Левенштейна $L(i, j)$ между двумя словами a и b длиной i и j при $\min(i, j) = 0$ по определению равно $\max(i, j)$, а при $\min(i, j) > 0$ находится из следующего рекуррентного уравнения:

$$L(i, j) = \min \left(\begin{array}{l} L(i, j-1) + 1, \\ L(i-1, j) + 1, \\ L(i-1, j-1) + m(i, j) \end{array} \right),$$

где $m(i, j)$ равно нулю, если i -я буква слова a равна j -й букве слова b , и единице – в противном случае.

Установлено [18], что более 80 % ошибок написания слов составляют ошибки перестановки букв. Поэтому расстояние Левенштейна в настоящее время определяется не на трех, а на четырех операциях: вставка, удаление, замена и транспозиция.

Алгоритм вычисления расстояния Левенштейна $L(a, b)$ между словами a и b имеет вычислительную сложность $\Theta(|a| \cdot |b|)^1$ при объеме памяти $\Theta(\min(|a|, |b|))$, где $|x|$ – длина слова x [80].

Одной из проблем использования расстояния Левенштейна для определения фонетической близости двух слов является необходимость задания минимального расстояния между фонетически схожими словами ε . Чтобы минимизировать число ошибок используют функцию ε , зависящую от максимальной длины сравниваемых слов n . Обычно принимают $\varepsilon(n < 5) = 0$, $\varepsilon(4 < n < 9) = 1$, $\varepsilon(8 < n) = 3$.

Функции, вычисляющие расстояние Левенштейна, реализованы на многих языках программирования [11].

¹ Знак Θ указывает на асимптотическое ограничение сверху и снизу, $f(n) \in \Theta(g(n)) \leftrightarrow \exists(C, D, N > 0) \forall(n > N) |C \cdot g(n)| \leq f(n) \leq |D \cdot g(n)|$.

3.2. РАССТОЯНИЕ НА ОСНОВЕ N -ГРАММ

N -граммой называется последовательность из N элементов (букв). Для определения фонетической близости двух слов подсчитывается число общих N -грамм. Обычно N принимается равным трем.

Пример 7. Рассмотрим два слова, которые имеют одинаковое звучание: «Thomson» и «Thompson». Разобьём эти слова на триграммы. В результате получим, что слово «Thomson» включает триграммы «ТНО», «НОМ», «ОМС», «МСО» и «СОН», а слово «Thompson» – «ТНО», «НОМ», «ОМР», «МРС», «РСО» и «СОН». Общими триграммами этих слов являются триграммы «ТНО», «НОМ» и «СОН». Доля общих триграмм составляет $3/6$, что определяет расстояние между словами, равное трем (подсчет велся относительно слова с большим числом триграмм).

Пример 8. Рассмотрим другие два слова: «Dane» и «Dean». Слово «Dane» имеет триграммы «DAN» и «ANE», слово «Dean» – «DEA» и «EAN». Хотя эти слова имеют одинаковое звучание, общие триграммы у них отсутствуют.

Как видно из примеров, различие между расстоянием, определенным с помощью триграмм, и расстоянием Левенштейна заключается в том, что различающаяся буква имеет значительное влияние на первое расстояние и слабое влияние на второе. Очевидно также, что вычислении расстояния между словами на основе N -грамм дает лучший результат для более длинным слов, чем для коротких.

Обычно N -граммы используются для нечеткого сравнения слов, которое не затрагивает фонетических аспектов. Например, для идентификации языка, так как установлено, что на достаточно длинных текстах каждый язык имеет свое распределение N -грамм, для сжатия текстов, для «угадывания» следующих букв и т.п. Другое применение N -грамм – индексирование данных в поисковых системах [33].

3.3. РАССТОЯНИЕ ДЖАРО

Неформальное определение расстояния Джаро между двумя словами – это минимальное число однобуквенных изменений,

которое необходимо выполнить для преобразования одного слова в другое [30, 31]. Чем меньше расстояние Джаро, тем более схожи сравниваемые слова.

Расстояние Джаро $D(a, b)$ между двумя словами a и b при $m > 0$ определяется так:

$$D(a, b) = w_1 \frac{m}{|a|} + w_2 \frac{m}{|b|} + w_3 \frac{m-t}{m},$$

где w_1 , w_2 , и w_3 – весовые коэффициенты, $w_1 + w_2 + w_3 = 1$; m – число совпадающих букв (число букв, разнесенных не более чем на половину длины самого короткого слова); t – половина числа транспозиций (половина числа совпадающих букв, отличающихся порядковыми номерами). При $m = 0$ по определению $D(a, b) = 0$.

Пример 9. Найдём расстояние Джаро между словами «Dape» и «Dean». Число совпадающих букв m равно четырём, а половина транспозиций $t = 3/2$. Приняв коэффициенты, равные $1/3$, и округляя t до целого числа, получаем $d = 11/12$. После округления имеем $d = 1$. Действительно, слово «Dean» преобразуется в слово «Dape» путем транспозиции буквы E в конец слова.

В работе [81] предложено улучшение формулы вычисления расстояния Джаро исходя из предположения о том, что начало слова более значимо по сравнению с оставшейся его частью:

$$d'(a, b) = d(a, b) + s[1 - d(a, b)]/10,$$

где $d'(a, b)$ – улучшенное расстояние Джаро, s – число начальных общих букв слова, не превышающее четырех.

Следует заметить, что вычислительная сложность определения расстояния Джаро самая высокая из всех ранее рассмотренных алгоритмов [46]. Самая низкая вычислительную сложность оказалась у алгоритма вычисления расстояния на основе N -грамм.

4. Применение фонетического кодирования

Алгоритмы фонетического кодирования получили широкое распространение в современных информационных технологиях как в неизменном, так и в адаптированном виде, например:

- для нормализации данных в социальных сетях [9, 52, 68];
- для устранения дублирования данных [6, 49, 54, 57];
- для фонетического поиска [48, 79];
- при переводе с одного языка на другой [17];
- для поиска в медицинских базах данных [50, 74];
- для строковых метрик в распределенных базах данных [23];
- в поисковых сервисах сети интернет [3, 69];
- для извлечения семантических данных из онтологий [29];
- для отбора слов морфологических генераторов [37];
- для отождествления слов при восстановлении данных [82];
- для первичной обработки голосовых запросов [72];
- для распознавания чрезвычайных ситуаций [77];
- при анализе и синтезе речи [25, 61];
- при анализе данных в интеллектуальных системах [35];
- при обработке и хранении персональных данных [24, 39];
- при проверке правописания на различных языках [5, 43, 71];
- для создания хэш-тегов в технологии Big Data [13];
- для распознавания этнической принадлежности [0];
- при статистической идентификации объектов [19];
- для поиска речевых документов [66, 67];
- для генерации реалистичных персональных данных [15];
- для генерации хэш-ключей при анализе данных [16];
- при предсказании поведения пользователей [61];
- при интеграции данных на основе топонимов [73];
- для индексирования музыкальных произведений [26, 51];
- при обработке исторических документов [20];
- для исправления орфографических ошибок [32, 64];
- для ускоренного ввода текста с клавиатуры [55];
- для анализа настроения пользователей [76];
- для кросс-языкового извлечения данных [12];
- для определения сходства коротких сообщений [63];
- для межязыковой транслитерации [4, 8, 0];

– для распознавания слов при обработке текстов [0].

5. Заключение

Хотя основные алгоритмы фонетического кодирования предложены в прошлом веке, исследование и разработка других алгоритмов никогда не прекращались. Однако новых результатов в этой области получено не было, и основным итогом всех исследований явилось только улучшение базовых алгоритмов [34, 65].

Несмотря на все улучшения, общим недостатком алгоритмов фонетического кодирования все еще остается наличие в получаемых результатах ложноположительных и ложноотрицательных ошибок. Это объясняется тем, что эти алгоритмы работают не на последовательности элементарных звуков, из которых состоят слова, а на их текстовом представлении, которое, к тому же, искажено правилам письма того или иного естественного языка.

Очевидно, имеющиеся различия между языком-речью и языком-письмом делают алгоритмы фонетического кодирования не всегда корректными. Видится перспективным фонетическое преобразование сравниваемых слов в последовательность элементарных звуков (аллофонов, фонем, дифонов, трифонов) перед использованием базовых алгоритмов фонетического кодирования.

Теоретической основой такого подхода могут служить результаты, появившиеся в последнее время в области синтеза и распознавания речи. В частности, разработаны и опробованы алгоритмы, позволяющие как «озвучивать» текст, так и выделять из речевого потока последовательности элементарных звуков [33].

Литература

1. КАНЬКОВСКИ П. «Как ваша фамилия?» или русский *MetaPhone* // Программист. – 2002. – Вып. 8. – С. 36–39.
2. ЛЕВЕНШТЕЙН В.И. Двоичные коды с исправлением выпадений, вставок и замещений символов // Докл. АН СССР. – 1965. – Вып. 163 (4). – С. 845–848.
3. ABDULHAYOGLU M.A., THIJS B. *Use of ResearchGate and Google CSE for author name disambiguation* // *Scientometrics*. – Akademiai Kiado, Budapest, Hungary, 2017 – Vol. 111. – P. 1965–1985.
4. ADEEL Z.M., IQBAL R.N., MASOOD S.A. *English to Urdu transliteration: An application of Soundex algorithm* // IEEE Int. Conference on Information and Emerging Technologies, Karachi, Pakistan, June 14–16, 2010. – P. 1–5.
5. ALMEIDA G., AVANCO L., DURAN M.S., et al. *Evaluating Phonetic Spellers for User-Generated Content in Brazilian Portuguese* // Int. Conference on Computational Processing of the Portuguese Language, Tomar, Portugal, June 13–15, 2016. – Springer International Publishing, 2016. – P. 361–373.
6. ANGELES M.P., PEREZ-FRANKO L.F. *Analysis of string encoding functions during de-duplication process* // Int. Conference on Informatics, Electronics & Vision, Fukuaka, Japan, June 15–18, 2015. – P. 1–6.
7. ANGELES M.P., ESPINO-GAMEZ A., GIL-MONCADA J. *Comparison of a Modified Spanish Phonetic, Soundex, and Phonex coding functions during data matching process* // Int. Conference on Informatics, Electronics & Vision, Fukuaka, Japan, June 15–18, 2015. – P. 1–5.
8. ANONTHANASAP O., KETNA M., LEELANUPAB T. *Automated English mnemonic keyword suggestion for learning Japanese vocabulary* // IEEE Int. Conference on Information Technology and Electrical Engineering (ICITEE), Chiang Mai, Thailand, October 29–30, 2015. – P. 638–643.
9. BILAL A. *Lexical normalization of Twitter Data* // Science and Information Conference, London, UK, July 28–30, 2015. – P. 326–328.

10. BINSTOCK A., REX J. *Practical Algorithms for Programmers*. – Addison-Wesley, 1995. – 577 p.
11. *Calculate Levenshtein distance between two strings*. – URL: <http://php.net/manual/en/function levenshtein.php> (дата обращения: 08.09.2017).
12. SHHEDA P., FARUQUI M., MITRA P. *Handling OOV Words in Indian-language – English CLIR* // European Conference on Information, Barcelona, Spain, April 1–5, 2012. – Retrieval. – Berlin Heidelberg: Springer-Verlag, 2012. – P. 476–479.
13. CHERICHI S., FAIZ R. *Upgrading Event and Pattern Detection to Big Data* // Int. Conference on Computational Collective Intelligence, Halkidiki, Greece, September 28–30, 2016. – Springer International Publishing, 2016. – P. 377–386.
14. CHOUDHURY M., SARAF R., JAIN V., MUKHERJEE A., SARKAR S, BASU A. *Investigation and modeling of the structure of texting language* // Int. Journal of Document Analysis and Recognition. – Springer-Verlag, 2007. – Vol. 10. – P. 157–174.
15. CHRISTEN P., PUDJIJONO A. *Accurate Synthetic Generation of Realistic Personal Information* // Pacific-Asia Conference on Knowledge Discovery and Data Mining, Bangkok, Thailand, April 27–30, 2009. – Berlin Heidelberg: Springer-Verlag, 2009. – P. 507–514.
16. CHRISTEN P. *Geocode Matching and Privacy Preservation* // Privacy, Security and Trust in KDD, Las Vegas, USA, August 24, 2008. – Berlin Heidelberg: Springer-Verlag, 2009. – P. 7–24.
17. CHUNG J.M., LU C.Y., LEE H.M., HO J.M. *Automatic English-Chinese name translation by using Web-Mining and phonetic similarity* // IEEE Int. Conference on Information Reuse & Integration. Las Vegas, USA, August 3–5, 2011. – P. 283–287.
18. DAMERAU F.J. *A technique for computer detection and correction of spelling errors* // ACM. – 1964. – No. 7(3). – P. 171–176.
19. DENK M. *Framework for Statistical Entity Identification in R* // Data Analysis, Machine Learning and Applications. – Berlin Heidelberg: Springer-Verlag, 2008. – P. 335–342.

20. DONGHUI L., DEWEI P. *Spelling Correction for Chinese Language Based on Pinyin-Soundex Algorithm* // Int. Conference on Internet Technology and Applications, Wuhan, China, August 16–18, 2011. – P. 1–3.
21. ERNST-GERLACH A., FUHR N. *Advanced Training Set Construction for Retrieval in Historic Documents* // Asia Information Retrieval Symposium, Taipei, Taiwan, December 1–3, 2010. – Berlin Heidelberg: Springer-Verlag, 2010. – P. 131–140.
22. GAONA M.A., GELBUKH A., BANDYOPADHYAY S. *Recognizing Textual Entailment Using a Machine Learning Approach* // Mexican Int. Conference on Artificial Intelligence, Pa-chuca, Mexico, November 8–13, 2010. – Berlin Heidelberg: Springer-Verlag, 2010. – P. 177–185.
23. GIRAUD-CARRIER C., GOODLIFFE J., JONES B.M. CUEVA S. *Effective record linkage for mining campaign contribution data* // *Knowledge and Information Systems* – London: Springer-Verlag, 2015. – Vol. 45. – P.389–416.
24. GRZEBALA P., CHEATHAM M. *Private Record Linkage: Comparison of Selected Techniques for Name Matching* // Int. Semantic Web Conference, Heraklion, Crete, Greece, May 29 – June 2, 2016. – Springer International Publishing, 2016. – P. 593–606.
25. JIAN H.-L. *Speech Driven by Artificial Larynx: Potential Advancement Using Synthetic Pitch Contours* // Int. Conference on Universal Access in Human-Computer Interaction, Los Angeles, USA, August 2–7, 2015. – Springer International Publishing, 2015. – P. 312–321.
26. HAN Y., MIN L., ZOU Y. et al. *LRC Sousou: A Lyrics Retrieval System* // Int. Conference of Young Computer Scientists, Engineers and Educators, Harbin, China, January 10–12, 2015. – Berlin Heidelberg: Springer-Verlag, 2015. – P. 464–467.
27. HERZOG T.N., SCHEUREN F.J., WINKLER W.E. *Data Quality and Record Linkage Techniques*. – New York: Springer, 2007. – P. 115–121.

28. HONG S.G., JANG S., CHUNG Y.H. et al. *News Media Analysis Using Focused Crawl and Natural Language Processing: Case of Lithuanian News Websites* // Int. Conference on Information and Software Technologies, Kaunas, Lithuania, September 13–14, 2012. – Berlin Heidelberg: Springer-Verlag, 2012. – P. 48–61.
29. HU B., HU B. *On Capturing Semantics in Ontology Mapping* // LLC: Springer Science + Business 2008. – Vol. 11. – P. 361–385.
30. JARO M.A. *UNIMATCH – a computer system for generalized record linkage under conditions of uncertainty* // Spring Joint Computer Conference, Anaheim, USA December 5–17, 1972. – P. 523–530.
31. JARO M.A. *Advances in record-linkage methodology as applied to matching the 1985 census of Tampa, Florida* // Journal of the American Statistical Association. – 1989. –No. 84(406). – P. 414–420.
32. JORDAO C.C., ROSA J.L. *Metaphone-pt_BR: The Phonetic Importance on Search and Correction of Textual Information* // Int. Conference on Intelligent Text Processing and Computational Linguistics, New Delhi, India, March 11–17, 2012. – Berlin Heidelberg: Springer-Verlag, 2012. – P. 297–305.
33. JURAFSKY D., MARTIN J.H. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. – Pearson Prentice Hall, 2009. – 988 p.
34. KARAKASIDIS A., VERYKIOS V.S. *Privacy Preserving Record Linkage Using Phonetic Codes* // Balkan Conference in Informatics, Thessaloniki, Greece, September 17– 19, 2009. – P. 101–106.
35. KARAKASIDIS A., KOLONIARI G., VERYKIOS V.S. *Privacy Preserving Blocking and Meta-Blocking* // Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Porto, Portugal, September 7–11, 2015. – Springer International Publishing, 2015. – P. 232–236.

36. KAVEH-YAZDY F., ZAREH-BIDOKI A.M. *Aleph or Aleph-Maddah, that is the question! Spelling correction for search engine autocomplete service* // Int. Conference on Computer and Knowledge, Mashhad, Iran, October 29–30, 2014. – P. 273–278.
37. KNUTH D.E. *The Art of Computer Programming. Vol. 3.* / Second Edition. – Addison-Wesley, 1998.
38. KOCOŃ J., PIASECKI M. *Named Entity Matching Method Based on the Context-Free Morphological Generator* // Int. Conference on Natural Language Processing, Warsaw, Poland, September 17–19, 2014. – Springer International Publishing, 2014. – P. 34–44.
39. KROLL M., STEINMETZER S. *Who Is 101101111... 1110110010? Automated Cryptanalysis of Bloom Filter Encryptions of Databases with Several Personal Identifiers* // Int. Joint Conference on Biomedical Engineering Systems and Technologies, Lisbon, Portugal, January 12–15, 2015. – Springer International Publishing, 2015 – P. 341–356.
40. LAWRENCE P. *Hanging on the Metaphone* // Computer Language. – 1990. – Vol. 7, No. 12. – P. 39–44.
41. LAWRENCE P. *The Double Metaphone Search Algorithm* // C/C++ Users Journal. – 2000. – No. 18(6). – P. 38–43.
42. LAWRENCE P. *Metaphone 3: Version 2.1.3.* – URL: <https://searchcode.com/codesearch/view/2366000/> (дата обращения: 18.02.2012).
43. LI J., OUAZZANE K., JING Y. et al. *Evolutionary Ranking on Multiple Word Correction Algorithms Using Neural Network Approach* // Int. Conference on engineering Applications of Neural Networks, London, UK, August 27–29, 2009. – Berlin Heidelberg: Springer-Verlag, 2009. – P. 409–418.
44. LI L., QUAZZANE K., KAZEMIAN H., JING Y., BOYD R. *A neural network based solution for automatic typing errors correction* // Neural Computing and Applications. – 2011. – Vol. 20. – P. 889–896.
45. LISBACH B., MEYER M. *Linguistic Identity Matching.* – Springer Fachmedien Wiesbaden, 2013. – P. 118–120.

46. LOUPPE G., AL-NATSHEH H.T., SUSIK M., MAGUIRE E.J. *Ethnicity Sensitive Author Disambiguation Using Semi-Supervised Learning* // Int. Conference on Knowledge Engineering and the Semantic Web, Prague, Czech Republic, September 21–23, 2016. – Springer International Publishing, 2016. – P. 272–287.
47. MAARIF H.A., AKMELIAWATI R., HTIKE Z.Z., GUNAWAN T.S. *Complexity Algorithm Analysis for Edit Distance* // Int. Conference on Computer and Communication Engineering, Kuala Lumpur, Malaysia September 23–24, 2014. – P. 135–137.
48. MANDAL A.K., HOSSAN D., NADIM. *Developing an efficient search suggestion generator, ignoring spelling error for high speed data retrieval using Double Metaphone Algorithm* // Int. Conference on Computer and Information Technology, Dhaka, Bangladesh, December 23–25, 2010. – P. 317–320.
49. MARTINS B. *A Supervised Machine Learning Approach for Duplicate Detection over Gazetteer Records* // Int. Conference on GeoSpatial Semantics, Brest, France, May 12–13, 2011. – Berlin Heidelberg: Springer-Verlag, 2011. – P. 34–51.
50. MASON-BLAKLEY F., LU L., PRICE M., ROUDSARI A. *An RCT Simulation Study on Performance and Accuracy of Inexact Matching Algorithms for Patient Identity in Ambulatory Care Settings* // Int. Conference on Healthcare Informatics, Dallas, USA, October 21–23, 2015. – P. 8–17.
51. MEINTANIS K., SHIPMAN F.M. *Visual Expression for Organizing and Accessing Music Collections in MusicWiz* // Int. Conference on Theory and Practice of Digital Libraries, Glasgow, UK, September 6–10, 2010. – Berlin Heidelberg: Springer-Verlag, 2010. – P. 80–91.
52. MOSQUERA A., MOREDA P. *The Study of Informality as a Framework for Evaluating the Normalisation of Web 2.0 Texts* // Int. Conference on Application of Natural Language to Information Systems, Groningen, The Netherlands, June 26–28, 2012. – Berlin Heidelberg: Springer-Verlag, 2012. – P. 241–246.

53. MUTALIB A., NOAH S.A. *Phonetic coding methods for Malay names retrieval* // IEEE Int. Conference on Semantic Technology and Information, Putrajaya, Malaysia June 28–29, 2011. – P. 125–129.
54. MUTHMANN K., LOSER A. *Detecting Near-Duplicate Relations in User Generated Forum Content* // OTM Confederated International Conferences «On the Move to Meaningful Internet Systems», Hersonissos, Crete, Greece, October 25–29, 2010. – Berlin Heidelberg: Springer-Verlag, 2010. – P. 698–707.
55. OUAZZANE K., LI J., KAZEMIAN H.B. *An Intelligent Keyboard Framework for Improving Disabled People Computer Accessibility* // Engineering Applications of Neural Networks, Corfu, Greece, September 15–18, 2011. – International Federation for Information Processing, 2011. – P. 382–391.
56. OUSIDHOUM N.D., BENSOU N. *Towards the Refinement of the Arabic Soundex* // Int. Conference on Application of Natural Language to Information Systems, Salford, UK, June 19–21, 2013. – Berlin Heidelberg: Springer-Verlag, 2013. – P. 309–314.
57. OWONIBI M., KOENIG-RIES B. *A Quality Management Workflow Proposal for a Biodiversity Data Repository* // Int. Conference on Conceptual Modeling, Atlanta, USA, October 27–29, 2014. – Springer International Publishing, 2014. – P. 157–167.
58. PANDE B.P., DHAMI H.S. *Application of Natural Language Processing Tools in Stemming* // Int. Journal of Computer Applications. – 2011. – No. 27(6). – P. 14–19.
59. PARAMONOV V.V., SHIGAROV A.O., RUZHNIKOV G.M. et al. *Polyphon: An Algorithm for Phonetic String Matching in Russian Language* // Int. Conference on Information and Software Technologies, Druskininkai, Lithuania, October 13–15, 2016. – Springer International Publishing, 2016. – P. 568–579.

60. PARMAR V.P., PANDYA A.K., KUMBHARANA C.K. *Determining the character replacement rules and implementing them for phonetic identification of given words to identify similar pronunciation words* // IEEE Int. Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE), New Delhi, India, February 25–27, 2015. – P. 272–277.
61. PARVATHY A.G., VASUDEVAN G.B., KUMAR A., BALAKRISHNAN R. *Leveraging Call Center Logs for Customer Behavior Prediction* // Int. Symposium on Intelligent Data Analysis, New Delhi, India, February 25–27, 2015. – Berlin Heidelberg: Springer-Verlag, 2009. – P. 143–154.
62. PFEIFFER A., KURAEVA A., FOULONNEAU M. et al. *Automatically Generated Metrics for Multilingual Inline Choice Questions on Reading Comprehension* // Int. Computer Assisted Assessment Conference, Zeist, The Netherlands, June 22–23, 2015. – Springer International Publishing, 2015. – P. 80–95.
63. PINTO D., VILARINO D., ALEMAN Y. et al. *The Soundex Phonetic Algorithm Revisited for SMS Text Representation* // Int. Conference on Text, Speech and Dialogue, Brno, Czech Republic, September 3–7, 2012. – Berlin Heidelberg: Springer-Verlag, 2012. – P. 47–55.
64. PURAO S., STOREY V.C., SUGUMARAN V. et al. *Repurposing Social Tagging Data for Extraction of Domain-Level Concepts* // Int. Conference on Application of Natural Language to Information Systems, Alicante, Spain, June 28–30, 2011. – Berlin Heidelberg: Springer-Verlag, 2011. – P. 185–192.
65. RAKIBUL H., REAZ A. *FM-Chord: Fault-tolerant Chord supporting misspelled queries* // Int. Conference on Computers and Information Technology, Dhaka, Bangladesh, December 21–23, 2009. – P. 651–656.
66. REYES-BARRAGAN M., VILLASENOR-PINEDA L., MONTES-Y-GOMEZ M. *A Soundex-Based Approach for Spoken Document Retrieval* // Mexican Int. Conference on Artificial Intelligence, Atizapan de Zaragoza, Mexico, October 27–31, 2008. – Berlin Heidelberg: Springer-Verlag, 2008. – P. 204–211.

67. REYES-BARRAGAN A., MONTES-Y-GOMEZ M., VILLASENOR-PINEDA L. *Combining Word and Phonetic-Code Representations for Spoken Document Retrieval* // Int. Conference on Intelligent Text Processing and Computational Linguistics, Tokyo, Japan, February 20–26, 2011. – Berlin Heidelberg: Springer-Verlag, 2011. – P. 458–466.
68. ROEDLER R., KERGL D., RODOSEK G.D. *Profile Matching Across Online Social Networks Based on Geo-Tags* // Advances in Nature and Biologically Inspired Computing. – Springer International Publishing, 2016. – P. 417–428.
69. RUSSELL R.C, MARGARET K.O. *US Patent 1262167, 1435663*. – 1918, 1922.
70. SANCHEZ-VILAS F., LAMA M., VIDAL J.C. et al. *Combining Uncorrelated Similarity Measures for Service Discovery* // Int. Workshop on Resource Discovery, Paris, France, November 5, 2010. – Berlin Heidelberg: Springer-Verlag, 2012. – P. 160–180.
71. SCHIERLE M., SCHULZ S., ACKERMANN M. *From Spelling Correction to Text Cleaning – Using Context Information* // Data Analysis, Machine Learning and Applications. – Berlin Heidelberg: Springer-Verlag. – 2008. – P. 397–404
72. SCHNEIDER J.M., FERNANDEZ J., MARTINEZ P. *A Proof-of-Concept for Orthographic Named Entity Correction in Spanish Voice Queries* // Int. Workshop on Adaptive Multimedia Retrieval, Copenhagen, Denmark, October 24–25, 2012. – Springer International Publishing, 2014. – P. 181–190.
73. SMART P.D., JONES C.B., TWAROCH F.A. *Multi-source Toponymal Data Integration and Mediation for a Meta-Gazetteer Service* // Int. Conference on Geographic Information Science. Zurich, Switzerland, April 30, 2010. – Berlin Heidelberg: Springer-Verlag, 2010. – P. 234–248.
74. SOMAN S., SRIVASTAVA P, MURTHY B.K. *Unique Health Identifier for India: An algorithm an feasibility analysis on patient data* // Int. Conference on E-health Networking, Application & Services. Boston, USA, October 13–17, 2015. – P. 250–255.

75. *Soundexing and Genealogy by Gary Mokotff.* – URL: <http://www.avotaynu.com/soundex.htm> (дата обращения: 08.09.2017).
76. SOUZA M., VIEIRA R. *Sentiment Analysis on Twitter Data for Portuguese Language // Sentiment Analysis on Twitter Data for Portuguese Language.* – Berlin Heidelberg: Springer-Verlag, 2012. – P. 241–247.
77. TAFT R.L. *Name Search Techniques // New York State Identification and Intelligence system: Special Report No. 1.* – New York: Albany, 1970.
78. SOUZA J., BOTEGA L.C., SEGUNDO S. et al. *Conceptual Framework to Enrich Situation Awareness of Emergency Dispatchers // Int. Conference on Human Interface and the Management of Information, Los Angeles, USA, August 2–7, 2015.* – Springer International Publishing, 2015. – P. 33–44.
79. TISSOT H., PESCHL G., FABRO M.D. *Fast Phonetic Similarity Search over Large Repositories // Int. Conference on Database and Expert Systems Applications. Munich, Germany, September 1–4, 2014.* – Springer International Publishing, 2014. – P. 74–81.
80. WAGNER R.A., FISCHER M.J. *The string-to-string correction problem // ACM.* – 1974. – No. 21(1). – P. 168–173.
81. WINKLER W.E. *String comparator metrics and enhanced decision rules in the Fellegi-Sunter model of record // Proc. of the Section on Survey Research Methods.* – American Statistical Association, 1990. – P. 354–359.
82. ZAMPIERI M., AMORIM R.C. *Between Sound and Spelling: Combining Phonetics and Clustering Algorithms to Improve Target Word Recovery // Int. Conference on Natural Language Processing, Warsaw, Poland, September 17–19, 2014.* – Springer International Publishing, 2014. – P. 438–449.

OVERVIEW OF PHONETIC ENCODING ALGORITHMS

Valeriy Vykhovanets, Trapeznikov Institute of Control Sciences of RAS, Moscow, Dr. Sc., associative professor (valery@vykhovanets.ru).

Jianming Du, Bauman Moscow State University, Moscow, postgraduate student (forsola@qq.com)

Sergey Sakulin, Bauman Moscow State University, Moscow, PhD, associative professor (ss141291@yandex.ru)

Abstract: This paper gives an overview of the phonetic encoding algorithms, designed to determine the similarity of words in sound (pronunciation). Phonetic encoding algorithms are divided into algorithms for comparing words and algorithms for determining the distance between words. Word comparison algorithms such as SoundEx, NYSIIS, Daitch-Mokotoff, Metaphone, Polyphone and algorithms for determining the distance between words such as Levenshtein, Jaro, N-grams are described. For each algorithm, its advantages and disadvantages are indicated, an analogue of the algorithm for the Russian language is given. To eliminate the common shortcomings of phonetic encoding algorithms, it is proposed to use not the sequence of letters of words, but the sequence of their elementary sounds. In this case, word recognition, record linkage, indexing words by sounds are expected to improve.

Keywords: phonetic encoding algorithms, phonetic distance, record linkage, indexing words by sound.

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

*Поступила в редакцию 12.09.2017.
Опубликована 31.05.2018.*

УДК 519.179.2
ББК 22.176 + 65.23

АЛГОРИТМЫ УПРАВЛЕНИЯ РЕСУРСАМИ В ОБОБЩЕННЫХ СТОХАСТИЧЕСКИХ СЕТЕВЫХ ГРАФИКАХ

Иванов Н. Н.¹

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

Для обобщенного стохастического сетевого графика рассматриваются три метода управления свободными ресурсами, направленных на ускорение выполнения в реальном времени моделируемого процесса управления. Оценка целесообразности и проверка эффективности предложенных методов проводится на основе дерева состояний сетевого графика, моделирующего процесс управления. Обсуждаются средства имитационного моделирования, позволяющие по дереву состояний получить сравнительные данные по каждому из предлагаемых методов.

Ключевые слова: обобщенный стохастический сетевой график, дерево состояний сетевого графика, время выполнения сетевого графика, имитационное моделирование.

1. Введение

Сетевые графики, созданные первоначально в предположении о том, что времена прохождения отдельных дуг являются детерминированными величинами, в настоящее время рассматриваются в предположениях, связанных со случайными временами прохождения дуг с известными распределениями [1]. Такие графы получили название стохастических сетевых графиков (ССГ).

В работах [2, 3] рассматривается обобщенный вариант ССГ, названный обобщенным стохастическим сетевым графиком

¹ Николай Николаевич Иванов, доктор технических наук, доцент (ivanov.nni@yandex.ru).

(ОССГ) и имеющий своим отличием допущение в качестве дисциплины, помимо стандартной, при которой событие наступает после прохождения всех дуг, входящих в соответствующую данному событию вершину (дисциплина «И»), альтернативной дисциплины «(исключенное) ИЛИ», при которой свершение события происходит в результате прохождения первой (по времени) из дуг, входящих в соответствующую ему вершину.

Применительно к процессам в управляющих параллельных вычислительных системах (ВС), структурированным с помощью ОССГ, последние могут рассматриваться как модели, на которых осуществляется контроль и обеспечение временной надежности управляющих программных комплексов. Это предполагает, что для заданных распределений времени выполнения отдельных работ в каждом таком комплексе предварительно на основе поведения моделирующего ОССГ может быть сделан вывод о возможности выполнения управляющей программы не более чем за заданное директивное время [4].

Для выполнения требования временной надежности ВС могут быть привлечены различные алгоритмы использования свободных вычислительных ресурсов для ускорения выполнения управляющей программы в случае, если эта совокупность выбрана исходя из понятия степени параллелизма ОССГ. Напомним, степень параллелизма определяется как верхняя оценка такого числа исполнителей, при котором отсутствие очередей сочетается с минимизацией простоев исполнителей [3].

Необходимость использования свободных вычислительных ресурсов определяется пользователем, а их целесообразность и эффективность может быть подтверждена обсуждаемыми в работе способами и соответствующими этим способам инструментами имитационного моделирования.

2. Алгоритмы управления временем выполнения ОССГ, основанные на использовании свободных ресурсов

В процессе выполнения ОССГ могут возникать моменты, когда исполнительные устройства (например, каналы ВС), выбранные исходя из степени параллелизма ОССГ [3] или по ка-

ким-либо иным критериям, не будут задействованы полностью. При этом возникает возможность использования свободных ресурсов в целях ускорения процесса выполнения ОССГ. Ресурс, использованный при прохождении некоторой дуги, считается свободным в момент окончания ее прохождения.

Алгоритмы управления, нацеленные на ускорение выполнения ОССГ, могут быть построены с учетом того, могут ли свободные ресурсы использоваться с разделением на части или их использование не предполагает такой возможности, а в последнем случае включение в работу свободного ресурса допускается только в момент начала прохождения дуги и не предусмотрена возможность его изъятия до окончания ее прохождения, либо неделимый свободный ресурс может быть включен в работу независимо от того, какая доля времени уже затрачена на прохождение дуги.

2.1. АЛГОРИТМЫ С ДЕЛЕНИЕМ РЕСУРСОВ В МОМЕНТЫ ОКОНЧАНИЯ ПРОХОЖДЕНИЯ ДУГ

Эти алгоритмы предусматривают неизменное число исполнителей для каждой дуги в любой момент выполнения ОССГ до момента самого раннего окончания прохождения какой-либо дуги. При этом для прохождения любой дуги используется не менее одного исполнителя, а общее их число определяется либо по степени параллелизма ОССГ [3], либо выбор числа исполнителей проведен в результате предварительного имитационного моделирования дерева состояний [3] при заданных распределениях времени прохождения дуг. Ниже для этого числа принято обозначение N .

Алгоритм I, распределяющий ресурсы в равных долях по всем одновременно проходимым дугам. Алгоритм предполагает, что для каждой дуги при ее прохождении может быть использовано разное (возможно, нецелое) число (но не менее одного) исполнителей, при этом в любой момент времени сумма этих чисел, одинаковых для любой дуги в любой момент времени, по всем одновременно проходимым дугам строго равна степени параллелизма N . После завершения прохождения некоторой дуги производится пересчет чисел исполнителей для всех проходимых дуг в соответствии с равенством $k = N/n$, где

k – число исполнителей, назначаемых для каждой дуги; n – число одновременно проходимых в рассматриваемый момент времени дуг.

Алгоритм II, распределяющий ресурсы в равных целых долях по всем одновременно проходимым дугам. Алгоритм предполагает, что для каждой дуги при ее прохождении может быть использовано разное целое число исполнителей, но при этом в любой момент времени сумма этих чисел, одинаковых для любой дуги в любой момент времени, по всем одновременно проходимым дугам не превосходит N . После завершения прохождения некоторой дуги производится пересчет чисел исполнителей для всех проходимых дуг в соответствии с равенством $k = \lfloor N/n \rfloor$, где k – целое число исполнителей, назначаемых для каждой дуги; n – число одновременно проходимых в рассматриваемый момент времени дуг; N – степень параллелизма ОССГ.

2.2. АЛГОРИТМ III С ДЕЛЕНИЕМ РЕСУРСОВ В МОМЕНТЫ СВЕРШЕНИЯ СОБЫТИЙ

При использовании этого алгоритма пересчет числа подключаемых резервных ресурсов происходит в более редкие моменты времени, а именно, только в моменты свершения событий в ОССГ. При этом дуги, которые еще не закончили свое прохождение в момент свершения события, продолжают прохождение с тем числом исполнителей, которое было на момент свершения события. После свершения события накапливается некоторое количество свободных ресурсов, которые распределяются в равных долях (но не менее одного исполнителя для каждой дуги) между теми дугами, которые инициализировались в результате свершения события. При таком распределении ресурсов каждая дуга с момента своей инициализации и до момента окончания прохождения проходит с фиксированным числом исполнителей.

В рассматриваемом алгоритме должны соблюдаться следующие правила: 1) число назначаемых исполнителей для каждой дуги должно быть не менее единицы; 2) все свободные исполнители распределяются по инициализируемым дугам поровну в целых числах, в максимальном количестве в соответствии с

накопленным ресурсом; 3) минимальное общее число исполнителей ОССГ выбирается так, чтобы ни для какого набора случайных времен прохождения дуг не существовало событий, для которых недостаточно числа свободных исполнителей для назначения по одному исполнителю для каждой инициализируемой дуги.

В этом алгоритме максимальное число исполнителей может превзойти степень параллелизма. Каково будет это превышение, может показать описываемое ниже имитационное моделирование этого алгоритма.

Приведенными алгоритмами не исчерпываются все возможные варианты методов управления ресурсами в ОССГ. Разработчик может использовать другие алгоритмы, более подходящие для исследуемой системы, исходя из ее возможностей и особенностей. Однако при этом он может воспользоваться теми способами моделирования алгоритмов управления ресурсами, которые, по мнению автора, являются достаточно универсальными.

3. Имитационное моделирование алгоритмов управления ресурсами

Предпочтение, которое может отдать пользователь тому или иному алгоритму управления процессом выполнения ОССГ с помощью привлечения свободных ресурсов, может быть основано на их имитационном моделировании. Для этих целей наиболее эффективным является, по мнению автора, построения траекторий в дереве состояний ОССГ [3] для каждого случайного набора времен прохождения дуг с включением алгоритмов управления ресурсами в механизм прохождения траекторий в дереве состояний.

Напомним, что состоянием ОССГ считается вектор, составленный из номеров всех дуг, проходимых в текущий момент времени. Каждая дуга дерева состояний помечена номером дуги сетевого графика, исходящей из состояния дерева, в котором данная дуга объявлена в качестве составляющей вектора состояния. Начальным состоянием дерева состояний назначается век-

тор из номеров дуг, исходящих из начальной вершины сетевого графика. Конечному состоянию соответствует пустой вектор.

Алгоритм построения дерева состояний состоит в рекурсивном пополнении состояний дерева, начиная с начального. Если предположить, что построено состояние s k -го уровня этого дерева, то для построения состояний $(k + 1)$ -го уровня, непосредственно связанных с состоянием s , нужно из s вывести дуги числом, соответствующим числу дуг сетевого графика, входящих в s . Каждая такая дуга u , помеченная номером выбранной дуги сетевого графика, ведет в некоторое состояние q , которое содержит дуги, входящие в s , за исключением u , и пополненное дугами, активируемыми в случае, если прохождение дуги u вызвало свершение некоторого события. Построение дерева состояний в лексикографическом порядке может вестись так, что если состояние, соответствующее рассматриваемой вершине, в процессе построения дерева уже было достигнуто ранее, то оно подчеркивается, и построение ее продолжения обрывается. В этом случае результатом такого построения будет укороченное дерево состояний. Если же процедура подчеркивания не используется, то результатом построения будет полное дерево состояний.

В качестве примера для проверки предложенных выше алгоритмов управления методом имитационного моделирования был выбран ОССГ, представленный на рис. 1. Нормальные независимые распределения времен прохождения дуг и их обозначения принимались в соответствии с данными, приведенными ниже.

$$m_{01} = m_{12} = m_{23} = 8, m_{02} = 5,5\sqrt{2}, m_{13} = 6, m_{14} = 6\sqrt{2},$$

$$m_{24} = 10\sqrt{2}, m_{34} = 7, m_{35} = 3, m_{45} = 9;$$

$$d_{01} = d_{14} = d_{02} = d_{12} = d_{13} = d_{23} = d_{24} = d_{34} = d_{35} = d_{45} = 1.$$

Было принято следующее предположение: скорость прохождения каждой дуги пропорциональна используемому ресурсу: $v_k = kv_1$, где k – число исполнителей, привлеченных для прохождения дуги; v_1 – скорость прохождения при одном исполнителе. Таким образом, при $k > 1$ время прохождения дуги, определенное для $k = 1$, уменьшается в k раз. Возможны и другие

зависимости времени прохождения дуги от числа исполнителей, определяемые пользователем.

Рассматриваемый ОССГ, как это показано в [3], имеет степень параллелизма, равную 4. Там же показано, что при выбранных нами распределениях времен прохождения дуг эта степень параллелизма достижима.

По этой причине в статистических экспериментах выбиралось 4 исполнителя. Однако при экспериментах с алгоритмом III было выявлено, что для отсутствия очередей необходимо наличие 5 исполнителей.

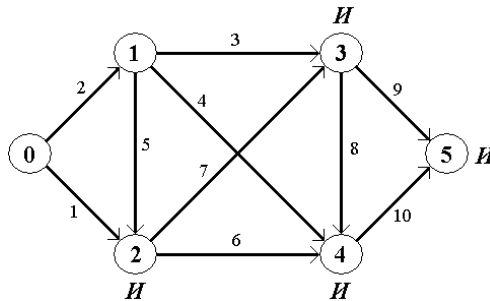


Рис. 1. Пример ОССГ

Моделирующие программы строились на основании двух принципов. Первоначально рассматривалась простая возможность построения программ, основанная на определении траекторий в укороченном дереве состояний при заданном случайном векторе времен прохождения дуг. Этот путь требовал предварительного построения укороченного дерева состояний, что может осуществляться машинным методом. Для стандартного режима без управления и алгоритмов I–II возможно было ограничиться построением укороченного дерева состояний, представленного на рис. 2.

Заметим, что дерево состояний ОССГ, в котором все вершины типа «И», имеет высоту, равную числу дуг сетевого графика. Такую же высоту будет иметь дерево состояний при наличии вершин типа «ИЛИ» без блокирования входных для них дуг [3]. При наличии вершин типа «ИЛИ» с блокированием дуг

(с остановкой прохождения всех входящих в эту вершину дуг после прихода первой по времени дуги) высота дерева состояний станет меньше на величину, равную суммарному числу входных дуг в вершины типа «ИЛИ» за вычетом числа таких вершин. Что касается ширины (полного) дерева состояний, то в общем случае ее можно оценить с помощью общего числа путей.

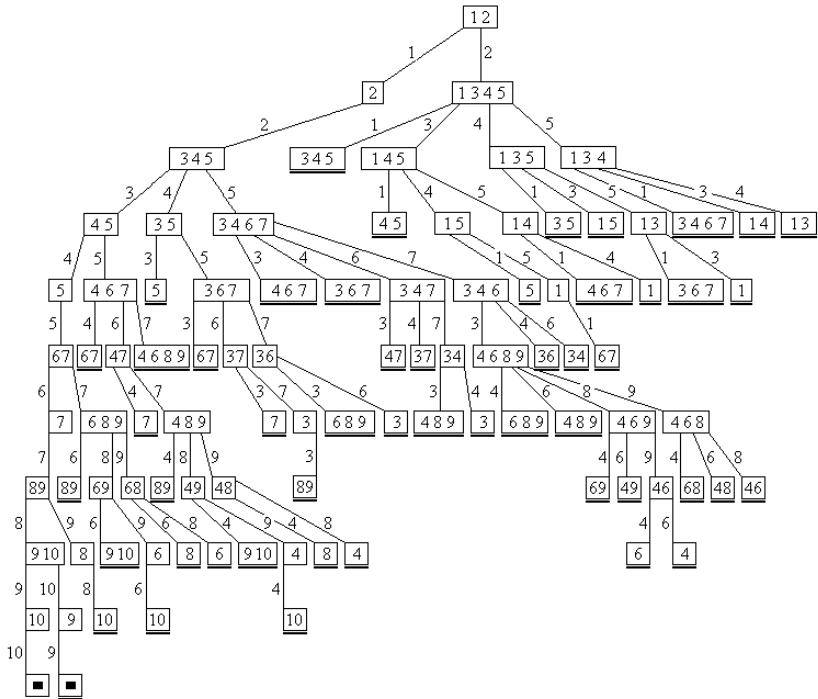


Рис. 2. Укороченное дерево состояний ОССГ по рис. 1

Однако для моделирования алгоритма III, как показано ниже, требовалось построение полного дерева состояний, что существенно увеличивало объем программы. Поэтому для алгоритма III этот метод моделирования не использовался.

По этой причине для моделирования всех алгоритмов был рассмотрен способ программирования, основанный на построе-

нии траекторий в полном дереве состояний без построения самого дерева. Такой метод предполагал построение траектории в дереве состояний для заданного набора случайных времен прохождения дуг непосредственно по графу ОССГ.

На рис. 3, *a* показан фрагмент сетевого графика, в котором события, соответствующие верхнему ряду вершин, свершились и проходятся дуги *a*, *b*, *c*, *d*. Степень параллелизма этого ОССГ равна 5. Вершины 1 и 2 имеют тип «И».

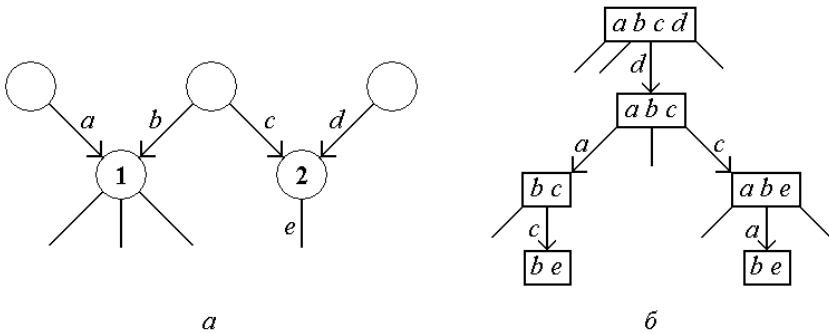


Рис. 3. Фрагмент ОССГ (а) и соответствующий ему фрагмент дерева состояний (б)

Если первой пройдена дуга *d*, то из состояния *abcd* на дереве состояний, фрагмент которого показан на рис. 3, б, произойдет переход в состояние *abc*. Если дуга *a* будет пройдена раньше дуг *b* и *c*, то произойдет переход в состояние *bc*, если же дуга *c* будет пройдена раньше дуг *a* и *b*, то произойдет переход в состояние *abe* и свершится событие 2. Далее из состояния *bc* по дуге *c* произойдет переход в состояние *be*, а из состояния *abe* по дуге *a* также в состояние *be*. Однако эти два состояния на дереве состояний могут обладать разными числами исполнителей для дуг *b* и *e*. Это окажется возможным, например, если дуги *a*, *b*, *c* и *d* проходились с помощью, соответственно, 1, 2, 1 и 1 исполнителей. Тогда, как нетрудно проследить, в левом состоянии *be* дуги *b* и *e* проходятся с помощью 2 и 3 исполнителей соответственно, а в правом состоянии *be* эти же дуги проходятся с по-

мощью 2 исполнителей каждая, что говорит о невозможности слияния этих двух одноименных состояний.

Второй способ моделирования имел преимущество перед способом, основанном на построении укороченного дерева состояний, определяемое практически независимым от размерности ОССГ объемом моделирующей программы.

Для верификации обоих методов было произведено сравнение результатов счета, получаемых для алгоритмов I–II обоими способами, показавшее их полное совпадение.

Результаты статистических испытаний четырех режимов выполнения ОССГ, в число которых входил стандартный режим без управления и три рассмотренных в п. 2 алгоритма I–III, приведены в таблице 1.

В таблице приняты следующие обозначения: m – оценка среднего времени выполнения ОССГ; σ – среднеквадратическое отклонение этого времени; Δ – ширина интервала доверия при уровне значимости $\alpha = 0,05$ и числе испытаний 10^5 .

Таблица 1.

Параметр	Стандарт	Алгоритмы управления		
		I	II	III
m	40,34	19,85	22,17	26,91
σ	2,04	0,78	1,02	2,44
Δ	$2,5 \cdot 10^{-2}$	10^{-2}	$1,3 \cdot 10^{-2}$	$3 \cdot 10^{-2}$

Как видно из таблицы 1, алгоритм I обеспечивает почти двукратное уменьшение среднего времени выполнения ОССГ. Однако дробление исполнительных ресурсов не всегда возможно. Менее эффективными в порядке убывания являются алгоритмы II и III, в которых дробление не используется. Однако относительная недостаточная эффективность компенсируется простотой подключения и использования ресурсов во время прохождения дуг.

4. Заключение

Предложен ряд алгоритмов привлечения резервных исполнительных ресурсов с целью ускорения прохождения отдельных дуг и как следствие ускорения процесса выполнения ОССГ. Для этих алгоритмов в статье предложен инструмент их анализа с помощью дерева состояний ОССГ. Основываясь на рассмотренных в статье результатах, пользователь может разрабатывать иные алгоритмы управления и проводить проверку их эффективности имитационными средствами предложенными способами.

Литература

1. ГОЛЕНКО-ГИНЗБУРГ Д.И. *Стохастические сетевые модели планирования и управления разработками*. – Воронеж: Научная мысль, 2010. – 283 с.
2. ИВАНОВ Н.Н. *Аналитико-имитационное моделирование обобщенных стохастических сетевых графиков // Управление большими системами*. – 2015. – Вып. 53. – С. 27–44.
3. ИВАНОВ Н.Н. *Степень параллелизма обобщенных стохастических сетевых графиков // Управление большими системами*. – 2017. – Вып. 65. – С. 6–23.
4. ИГНАТУЩЕНКО В.В., ИСАЕВА Н.А. *Резервирование взаимосвязанных программных модулей для управляющих параллельных вычислительных систем: организация, оценка отказоустойчивости, формализованное описание // Автоматика и телемеханика*. – 2008. – №10. – С. 142–161.

ALGORITHMS OF RESOURCE MANAGEMENT IN GENERALIZED STOCHASTIC NETWORKS

Nikolay Ivanov, Institute of Control Sciences of RAS, Moscow,
Doctor of Science (Moscow, Profsoyuznaya st., 65,
ivanov.nni@yandex.ru).

Abstract: Generalized stochastic network contains vertices corresponding to events, the fulfillment of which can occur in the discipline "AND" and "OR". These networks can be used for simulation of real-time control processes using parallel computing systems. The weight of edges in these networks are considered random variables distributed according to given laws. The main purpose of modeling of such networks is monitoring and ensuring of the specified directive execution time of the management process. The time management of the network schedule can be managed by utilization of free resources (processors, channels of parallel computing systems, etc.). For a generalized stochastic network, three methods of managing free resources are considered, aimed at accelerating the execution of a simulated control process in real time. Evaluation of the feasibility and verification of the effectiveness of the proposed methods is carried out on the basis of the state tree of the network modeling the management process. The state tree defines those states of the network process that require fewer resources than what is called the degree of parallelism. The degree of parallelism determines the maximum number of resources that ensures the absence of queues. Simulation tools are discussed that allow you to obtain comparative data on each of the proposed methods in the tree of states.

Keywords: the generalized stochastic network, the state tree of the network, the execution time of the network, simulation.

*Статья представлена к публикации
членом редакционной коллегии А.А. Лазаревым.*

Поступила в редакцию 15.06.2017.

Опубликована 31.05.2018.

УДК 519.854.2 + 519.83
ББК 22.18

ДИНАМИЧЕСКАЯ АДАПТАЦИЯ ГЕНЕТИЧЕСКОГО АЛГОРИТМА МАРШРУТИЗАЦИИ ТРАНСПОРТА НА БОЛЬШИХ СЕТЯХ

Захаров В. В.¹, Мугайских А. В.²

(Санкт-Петербургский государственный университет,
Санкт-Петербург)

Описывается процедура динамической адаптации генетического алгоритма для тестовых задач коммивояжера на больших сетях, позволяющая получать более экономичные маршруты за то же время вычислений. Эффективность предложенной процедуры подтверждается результатами вычислительных экспериментов получения численного решения набора тестовых задач из библиотеки TSPLib и устойчивого уменьшения средней длины генерируемых решений по сравнению с решениями, предоставляемыми исходной эвристикой. Тем самым демонстрируется методика использования свойства временной несостоятельности эвристических алгоритмов для целей повышения уровня их эффективности. Оценка временной состоятельности эвристического алгоритма проводится с помощью нового критерия – экспериментального уровня временной состоятельности.

Ключевые слова: временная состоятельность, генетический алгоритм, задачи маршрутизации.

Введение

Одной из важнейших задач транспортной логистики является оптимизация затрат, которая позволяет при том же объё-

¹ Виктор Васильевич Захаров, доктор физико-математических наук, профессор (v.zaharov@spbu.ru).

² Александр Всеволодович Мугайских, аспирант (alexander.mugaiskih@gmail.com).

ме используемых ресурсов увеличить прибыль. Для транспортных компаний существенная минимизация затрат на транспортировку может быть достигнута за счёт построения эффективных маршрутных планов транспортных средств. Особенно существенный эффект снижения затрат может быть достигнут на больших транспортных сетях. Именно по этой причине эффективным алгоритмам решения задачи маршрутизации транспорта, позволяющим генерировать менее затратные маршруты, уделяется пристальное внимание со стороны исследователей. Одним из важных объектов применения математических методов в транспортной логистике является задача маршрутизации транспортных средств. Классическим примером задачи маршрутизации является задача коммивояжёра (англ. *travelling salesman problem*), которую мы будем подробно рассматривать.

Основной мотивацией написания данной статьи являлась проверка гипотезы о временной несостоятельности (динамической неустойчивости) эвристических алгоритмов маршрутизации и разработка метода повышения уровня динамической устойчивости и эффективности используемых алгоритмов на примере генетического алгоритма. Статья структурирована следующим образом. Кроме введения, она включает семь разделов. Первый посвящен описанию задачи коммивояжёра и обзору методов ее решения. Во втором разделе приводятся основные понятия и термины генетического алгоритма, используемые в тексте. Третий раздел включает в себя описание базовой схемы генетического алгоритма и одного из наиболее эффективных инструментов построения решения – кроссовера Греффенстетта. В четвертом разделе обсуждается проблема динамической неустойчивости (временной несостоятельности) эвристических алгоритмов маршрутизации и новый критерий – экспериментальный уровень временной состоятельности алгоритма. Пятый раздел посвящен описанию процедуры динамической адаптации эвристических алгоритмов транспортной маршрутизации применительно к генетическому алгоритму. В разделе 6 приводятся результаты численных экспериментов по оценке уровня временной состоятельности ге-

нетического алгоритма и сравнительный анализ результатов применения генетического алгоритма с кроссовером Грегфенстетта и алгоритма его динамической адаптации при решении ряда тестовых задач из широко известной среди специалистов библиотеки TSPLib. Седьмой раздел содержит краткие выводы и планы дальнейших исследований.

1. Задача коммивояжёра (TSP) и методы её решения

1.1. Постановка задачи коммивояжёра

В классической постановке задача коммивояжёра заключается в поиске самого выгодного маршрута, проходящего через все указанные в условиях задачи вершины хотя бы по одному разу с возвратом в исходную вершину. Задача коммивояжёра может быть представлена в виде математической модели – графа $G = (V, E)$. Множество вершин $V = \{v_1, v_2, \dots, v_n\}$ в графе соответствуют городам в задаче коммивояжёра, множество рёбер $E = \{(v_i, v_j) : v_i, v_j \in V, i \neq j\}$ – коммуникациям, соединяющим пары городов. Каждому ребру из множества E можно сопоставить критерий выгодности маршрута.

В зависимости от постановки задачи вес ребра (v_i, v_j) графа G определяет время перехода, стоимость поездки или расстояние между вершинами v_i и v_j соответственно. Задача коммивояжёра может быть сформулирована как задача поиска на графе G гамильтонова контура с наименьшим общим весом.

Пусть $I = \{1, \dots, n\}$ – множество индексов вершин из тестовой задачи. Определим целевую функцию f как суммарную длину маршрута, включающего в себя все вершины рассматриваемой задачи.

Параметрами задачи являются элементы матрицы расстояний $C = \{c_{ij}\}$, $i, j \in I$. В случае асимметричной матрицы C задача коммивояжера моделируется ориентированным графом. В симметричном случае количество возможных маршрутов вдвое меньше асимметричного случая и верно равенство $c_{ij} = c_{ji}$. Далее будем рассматривать задачи только с симметричной матрицей расстояний, в которых длина маршрута между двумя вер-

шинами v_i и v_j определяется как евклидово расстояние: $c_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$, где $(x_i, y_i), (x_j, y_j)$ – координаты вершин v_i, v_j соответственно.

Переменными задачи являются элементы бинарной матрицы переходов между вершинами $Z = \{z_{ij}\}$, $i, j \in I$, которые равны 1, если в построенном маршруте для тестовой задачи присутствует ребро (v_i, v_j) , 0 – иначе.

Таким образом, задача коммивояжера может быть сформулирована и решена в рамках линейного целочисленного программирования [22, 40, 41]. Требуется найти

$$f = \sum_{i \in I} \sum_{j \neq i, j \in J} c_{ij} z_{ij} \rightarrow \min$$

при следующем наборе ограничений:

$$\begin{aligned} & 0 \leq z_{ij} \leq 1, z_{ij} \in \mathbb{Z} \\ (1) \quad & \sum_{j \in I, j \neq i} z_{ij} = 1, \forall i \in I, \\ (2) \quad & \sum_{i \in I, i \neq j} z_{ij} = 1, \forall j \in I, \\ (3) \quad & \sum_{i \in S, j \in S} z_{ij} = |S| - 1 \quad (S \subset V, |S| > 1). \end{aligned}$$

Ограничения (1) гарантируют, что в каждую вершину можно попасть по одному ребру, исходящему из другой вершины. Аналогично (2) показывают, что из каждой вершины существует только один путь, ведущий к другой вершине. Группа неравенств (3) обеспечивает связность маршрута обхода городов в допустимом решении, т.е. он не может состоять из двух и более несвязных частей.

Задача коммивояжера занимает центральное место среди задач комбинаторной оптимизации [6, 10, 12]. Имеет широкое практическое применение в различных областях: транспортная логистика [15], робототехника [19, 21], теория расписаний. Примерами данной задачи могут служить построение маршрутных планов для парка машин логистической компании; наблюдение группы роботов за объектами [44]; размещение датчиков в сенсорной

сети [48]; поставка запчастей на производстве [39]; планирование авиаперелётов [25]. Несмотря на кажущуюся простоту постановки задачи коммивояжёра, ей по-прежнему уделяется пристальное внимание со стороны специалистов при рассмотрении задачи в новых формулировках и при различных ограничениях [2, 4, 13, 14].

1.2. Алгоритмы решения задачи коммивояжёра

Все методы решения задачи TSP, как правило, относят к одной из двух категорий: точные и эвристические алгоритмы [7]. Среди точных алгоритмов решения данной задачи выделим метод полного перебора, метод ветвей и границ, метод ветвей и отсечений и метод динамического программирования.

Метод ветвей и границ и метод ветвей и отсечений являются развитием алгоритма полного перебора. Основная идея заключается в проверке критерия ограничивающей функции, по которому можно приостановить построение ветви дерева перестановок на определенном уровне. Известные модификации данных подходов позволяют решать задачу коммивояжёра с несколькими сотнями вершин, однако они требуют высокой вычислительной мощности компьютера.

Идея динамического программирования, предложенная Р. Беллманом [17], М. Хелдом и Р. Карпом [32], применима к задаче TSP и заключается в многошаговом процессе принятия решений, на каждой шаге которого необходимо определить функцию Беллмана и найти оптимальный маршрут обхода оставшихся вершин. Число операций, требуемых для вычисления итогового маршрута, растет экспоненциально по мере увеличения количества вершин тестовой задачи.

Ввиду NP-сложности задачи TSP точные методы не всегда могут эффективно применяться для задач большой размерности. По этой причине используют эвристические методы, которые генерируют решения, близкие к оптимальному, но за приемлемое по сравнению с точными алгоритмами время. Экспериментальный анализ эвристических алгоритмов решения задачи TSP и её подклассов проведен в работе [30].

По способу формирования итогового маршрута эвристические алгоритмы можно разделить на два класса: конструктивные (tour construction heuristics) и итерационные эвристики (tour improvement heuristics).

Конструктивные эвристики генерируют один уникальный маршрут для каждой тестовой задачи без последующего его улучшения [37]. К данному классу относят алгоритм ближайшего соседа [36], жадный алгоритм [31], эвристики вставок [27], алгоритм Кристофидеса [20]. Перечисленные алгоритмы последовательно строят допустимое решение, добавляя в него вершины, пока полный маршрут не будет сформирован. Часть решения, построенная к текущему моменту, останется неизменной до конца работы алгоритма. Как правило, с помощью данных эвристик можно получить решения, уступающие в длине оптимальному около 10–15%.

Другой класс эвристических алгоритмов решения задачи TSP – итерационные эвристики – начинает работать с уже готовым маршрутом, построенным одним из конструктивных методов, последовательно улучшая его на каждой итерации. Существует несколько способов сделать это: широко используемыми являются операции 2-замена и 3-замена [3], интегрированные в алгоритмы локального поиска и алгоритм Лина–Кернигана [33, 34].

Среди итерационных методов можно выделить отдельный класс алгоритмов, который носит название метаэвристик (metaheuristics) [18, 26]. Это достаточно общие итерационные процедуры, использующие рандомизацию и элементы самообучения, интенсификацию и диверсификацию поиска, адаптивные механизмы управления, а также конструктивные эвристики и методы локального поиска. Метаэвристики принято делить на траекторные методы, когда каждой итерации соответствует одно допустимое решение, и на алгоритмы, которые работают сразу с популяцией решений. К первой группе относят методы имитации отжига (simulated annealing) [5, 11, 28], поиск с запретами (tabu search) [49], поиск с чередующимися окрестностями

ми (variable neighbourhood search). Ко второй – генетические алгоритмы (genetic algorithms), муравьиные колонии (ant colony optimization) [23], вероятностные жадные алгоритмы (greedy randomized adaptive search procedure) и др. [47].

2. Термины и понятия

Далее в статье используются следующие термины и понятия.

Эвристический алгоритм – алгоритм решения задачи, правильность которого для всех возможных случаев не доказана, но про который известно, что он даёт достаточно хорошее решение в большинстве случаев.

Генетический алгоритм – эвристический алгоритм для решения задач оптимизации путём случайного подбора, комбинирования и вариации искомых параметров с использованием механизмов, аналогичных естественному отбору в природе, генерирующий различные решения при каждом запуске в виду механизмов рандомизации на основных этапах метода.

Особь – допустимое решение задачи в генетическом алгоритме.

Популяция особей – конечный набор допустимых решений задачи, с которым впоследствии будет работать генетический алгоритм.

Хромосома – совокупность параметров-генов, описывающих особь. Применительно к задаче коммивояжёра в хромосоме особи записан порядок обхода вершин тестовой задачи, соответствующий одному допустимому решению; гены в хромосоме соответствуют вершинам тестовой задачи.

Ген – это атомарный элемент хромосомы.

Скращивание, кроссовер – основной этап генетического алгоритма, в котором принимают участие две особи одной популяции – родительские особи. Путём комбинирования и наследования особенностей обоих родительских особей (далее – родителей) генерируются потомственные особи (далее – потомки), происходит улучшение качественного признака одной потомственной особи в частности и всей популяции в целом.

Мутация – один из этапов генетического алгоритма, в котором принимает участие одна особь из популяции, проводится с целью восстановления особей, выпавших из популяции, и формирования генов, которые не были представлены в исходной популяции.

Селекция – заключительный этап генетического алгоритма при формировании новой популяции особей, заключается в отборе наиболее приспособленных особей для их включения в дальнейший процесс эволюции.

3. Генетический алгоритм и кроссовер Греффенстетта

Генетический алгоритм – метаэвристический алгоритм, в основе которого лежат операции, свойственные естественному отбору, происходящему в природе. Впервые был представлен в 1975 г. Джоном Холландом как одно из направлений эволюционных алгоритмов [35]. Не детализируя описание генетического алгоритма, коротко приведем схему классического генетического алгоритма, которая состоит из следующих этапов:

1. Инициализация начальной популяции, s особей.
2. **Пока** не выполнен критерий останова, **выполнять**:
3. Производится $s/2$ операций скрещивания между 2 случайными особями.
4. Случайно выбранные особи из популяции мутируют.
5. Осуществляется оценка функции приспособленности и селекция.

Решением задачи коммивояжера является маршрут, соответствующий хромосоме самой приспособленной особи популяции последнего поколения. Генетический алгоритм является универсальной эвристической процедурой, которая может быть имплементирована в программный комплекс для решения задач транспортной маршрутизации в различных постановках таких как TSPTW (travelling salesman problem with time windows), IRP (inventory routing problem), VRP (vehicle routing problem), MDVRP (multi-depot vehicle routing problem), PDP (pickup and

delivery problem) и др. Кроме того, генетическая эвристика способна решать как саму задачу, так и её подзадачи, что является необходимым условием для проведения процедуры динамической адаптации алгоритма.

Как известно, основным этапом генетического алгоритма, о котором речь пойдёт в разделе 5, является операция скрещивания. Предложенные в литературе методы скрещивания являются допустимыми для задачи коммивояжёра и дают в целом приемлемые результаты, однако они не используют информацию о расстоянии между вершинами тестовых задач TSP. Разрешил эту проблему Джон Греффенстетт, предложив в своей работе [29] новый вид скрещивания, который использует информацию о расстоянии между вершинами.

В применении к задачам TSP с симметричной матрицей переходов с 70 и 100 вершинами тестовой библиотеки TSPLib [43] использование эвристики Греффенстетта на этапе скрещивания генетического алгоритма позволило получить более короткие маршруты в сравнении с частичным скрещиванием (partially mapped crossover), циклическим скрещиванием (cycle crossover), кроссовером перестановки рёбер (edge recombination crossover), двухточечным скрещиванием (two-point crossover), равномерным скрещиванием (uniform order-based crossover), модифицированным одноточечным скрещиванием (shuffle crossover) и кроссовером Ямамура (sub-tour exchange crossover). Подробнее о перечисленных видах скрещивания можно прочитать в работе [38]. Для тестовой задачи с 100 вершинами среднее отклонение длины маршрута, сгенерированного с помощью кроссовера Греффенстетта, от длины точного оптимального решения является наименьшим по сравнению перечисленными кроссоверами и составляет 11,9% [38].

В работе [42] также было произведено сравнение 8 видов кроссовера для решения задачи VRP, которая является обобщением задачи TSP на случай нескольких транспортных средств. Были описаны следующие виды скрещивания: упорядоченное скрещивание (order crossover), частичное скрещивание, кроссовер пере-

становки рёбер, циклическое скрещивание, кроссовер альтернативных рёбер (*alternating edges crossover*), эвристический кроссовер Греффенстетта и две его модификации [42]. Лучшие результаты среди рассмотренных алгоритмов скрещивания на тестовых задачах VRP стандартной библиотеки Кристофидеса [24] показал эвристический кроссовер Греффенстетта и две его модификации.

Таким образом, рассмотренный оператор скрещивания генетического алгоритма превосходит другие альтернативы в генерации маршрутов меньшей длины, и он был имплементирован в программную реализацию генетического алгоритма решения задачи коммивояжёра для дальнейших исследований, результаты которых представлены в следующих разделах статьи.

Ниже приведен классический алгоритм кроссовера Греффенстетта:

1. Случайную вершину v необходимо записать в первый ген потомка
2. Пока маршрут полностью не сформирован, **выполнять**:
3. Сравнить два ребра, выходящие из вершины v у родителей, выбрать самое короткое. Вторую вершину данного ребра обозначить через v_{min} .
4. **Если** вершина v_{min} уже присутствует в хромосоме потомка, **то** за v_{min} принять случайную вершину, не принадлежащую потомку.
5. Записать v_{min} в хромосому потомственной особи, $v = v_{min}$.

4. Динамическая неустойчивость эвристических алгоритмов маршрутизации

В работе [1] авторы отмечают, что процесс поиска оптимального решения в генетических алгоритмах направляется исключительно полученными значениями целевой функции в предыдущих точках пространства решений. При этом никак не используются предположения о таких свойствах целевой функции и ограничений как выпуклость, дифференцируемость, выполнение соотношения оптимальности Беллмана.

Сформулированный в 1957 г. принцип оптимальности Беллмана применительно к задачам маршрутизации утверждает, что любая часть оптимального маршрутного плана сохраняет свойство оптимальности в каждой подзадаче, рассматриваемой вдоль сгенерированной траектории движения, в процессе реализации начального маршрута [16]. Однако данное свойство не выполняется для решений задач маршрутизации, полученных с использованием большинства итерационных эвристик. Конструктивные эвристики почти всегда генерируют динамически устойчивые маршруты, однако не обеспечивают достаточный уровень эффективности по сравнению с итерационными.

Используя принципы динамической устойчивости решений, описанные в [9], введём следующие обозначения:

Пусть P – множество тестовых задач, $S(p)$ – конечное множество решений, сгенерированных эвристическим алгоритмом для тестовой задачи $p \in P$. Каждому решению $s(p)$ соответствует маршрут, представляющий собой последовательность обхода вершин тестовой задачи. Рассмотрим последовательность обхода вершин одного маршрута и разобьём её на T частей так, что количество вершин, которые были пройдены за первые t частей маршрута, вычисляется по формуле $n(t, s(p)) = \lfloor n_0 t / T \rfloor$, где n_0 – количество вершин в тестовой задаче p , а T – параметр алгоритма, определяемый в начале эксперимента. Таким образом, каждая часть маршрута (кроме последней) содержит одинаковое число вершин. Под периодом t , где $t = 0, 1, \dots, T - 1$, будем понимать промежуток времени в пути, которому соответствует часть t маршрута исходного решения $s(p)$.

Пусть $s^+(t, p)$ – часть маршрута $s(p)$, соответствующая порядку обхода городов после периода t . Часть $s^-(t, p)$ исходного решения включает города маршрута $s(p)$, посещённые за периоды $\tau = 0, 1, \dots, t$. Тогда каждое решение можно представить в виде объединения двух частей $s(p) = s^-(t, p) \cup s^+(t, p)$. При этом верно следующее: $s^+(0, p) = s(p)$.

Рассмотрим подзадачу $p(s^-(t, p))$, которая отличается от начальной тестовой задачи p тем, что множество городов сокра-

щено за счёт исключения тех из них, которые входят в часть маршрута $s^-(t, p)$. Точка отправления в подзадаче $p(s^-(t, p))$ находится в последней вершине маршрута $s^-(t, p)$. Обозначим через $s(p(s^-(t, p)))$ некоторое решение подзадачи $p(s^-(t, p))$, генерируемое эвристическим алгоритмом.

Определение 1. Маршрут (решение) $s(p)$, сгенерированный определенной эвристикой, будем называть состоятельным во времени (динамически устойчивым) относительно этой эвристики, если для каждого $t = 1, \dots, T - 1$ и любого $s(p(s^-(t, p)))$, сгенерированного той же эвристикой, верно неравенство

$$(4) \quad f(s^+(t, p)) \leq f(s(p(s^-(t, p)))) ,$$

где f – минимизируемая целевая функция для данной задачи маршрутизации транспорта.

Определение 2. Маршрут (решение) $s(p)$, сгенерированный некоторой эвристикой, будем называть несостоятельным во времени (динамически неустойчивым) относительно этой эвристики, если хотя бы для одного t' найдется сгенерированное этой же эвристикой решение $s(p(s^-(t', p)))$, для которого выполняется неравенство

$$(5) \quad f(s^+(t, p)) > f(s(p(s^-(t', p)))) .$$

Свойство временной состоятельности решения в применении к задачам транспортной маршрутизации означает, что при движении согласно начальному маршрутному плану в подзадачах, генерируемых по методу описанному выше, не будет найдено решение выгоднее текущего. И наоборот, временная несостоятельность решения означает, что хотя бы в одном из периодов будет сгенерировано решение выгоднее текущего.

Опишем процедуру оценки уровня временной состоятельности эвристического алгоритма H для конкретного класса задач транспортной маршрутизации.

1. Рассмотрим множество P тестовых задач из данного класса задач транспортной маршрутизации.

2. Для каждой тестовой задачи $p \in P$ сформируем множество N различных решений, полученных эвристическим алгорит-

мом H . Решения будут различны, так как операторы скрещивания, мутации используют рандомизацию в своих алгоритмах.

3. Для каждого решения проведём M экспериментов его проверки на временную состоятельность. Параметр M служит для получения средней оценки показателя временной состоятельности решения. Обычно полагаем $M = 5$. Начиная с $t = 1$ сформулируем текущую задачу $p(s^-(t, p))$, найдём её решение $s(p(s^-(t, p)))$ с помощью алгоритма H . Далее проверим начальное решение $s(p)$ на состоятельность после первого периода, т.е. справедливость неравенства (4). Если неравенство выполнено, то переходим к следующему периоду. Иначе фиксируем номер периода, на котором свойство временной состоятельности нарушилось и запускаем следующий тест.

Количество экспериментов, в которых начальное решение $s(p) \in N$ перестало быть состоятельным во времени после периода t обозначим через $b(s(p), t)$. Из принципа оптимальности Беллмана следует, что для оптимального решения тестовой задачи, т.е. полученного точными методами, выполнено условие (5).

$$(6) \quad \sum_{t=1}^{T-1} b(s(p), t) = 0.$$

Определение 3. Экспериментальным уровнем временной состоятельности эвристического алгоритма H будем называть величину, определяемую по следующей формуле:

$$(7) \quad conH = 1 - \frac{1}{M|P|} \sum_{p \in P} \frac{1}{|N|} \sum_{s(p) \in N} \sum_t^{T-1} b(s(p), t).$$

Уровень временной состоятельности для эвристического алгоритма может меняться в зависимости от тестовых задач, на которых он был рассчитан, однако можно заметить, что $0 \leq conH \leq 1$. Чем ближе значение величины $conH$ к 0, тем чаще решение полученное данной эвристикой теряет оптимальность в ходе своей реализации. Для маршрутов, сгенерированных точными методами, значение данной оценки равно 1.

5. Динамическая адаптация генетического алгоритма

Понятие динамической устойчивости решений было широко исследовано в последние годы для задач теории игр. Используя основную идею, изложенную в статьях [50, 51, 52], мы предложим процедуру (алгоритм) динамической адаптации эвристических алгоритмов в применении к различным классам задач транспортной маршрутизации на примере задачи TSP и генетического алгоритма.

На начальном этапе сгенерируем множество N различных решений тестовой задачи генетическим алгоритмом. Из множества N выберем лучшее решение, с минимальным значением целевой функции. Для него проведём M экспериментов проверки на временную состоятельность. Пусть выбранный маршрут является несостоятельным во времени, начиная с периода t , т.е. после периода t будет найдено лучшее решение в текущей подзадаче $p(s^-(t, p))$. Для уменьшения общей длины обхода всех вершин текущий маршрутный план следует изменить согласно новому решению, полученному в рассмотренной подзадаче. Общая схема работы динамической адаптации состоит в следующем [8].

Алгоритм 1 (Динамическая адаптация ГА для TSP).

1. Генерируем множество N решений эвристическим алгоритмом H для $p \in P$.
2. Определим $s_1(p) = \arg \min_{s(p) \in N} f(s(p))$.
3. От $t = 1$ до $T - 1$ **выполнять**
4. Сформулируем текущую задачу $p(s_t^-(t, p))$, получим множество N решений.
5. Определим $s_{t+1}^*(p) = \arg \min_{s(p(s_t^-(t, p))) \in N} f(s(p(s_t^-(t, p))))$.
6. Проверим свойство временной состоятельности для $s_t(p)$.
7. Если $f(s_t^+(t, p)) > f(s_{t+1}^*(p))$, тогда маршрут следует изменить на $s_{t+1}(p) = s_t^-(t, p) \cup s_{t+1}^*(p)$.

6. Численные эксперименты

6.1. Расчёт экспериментального уровня временной

состоятельности генетического алгоритма с эвристическим кроссовером Грэффенстетта

Для проведения эксперимента были рассмотрены пять тестовых задач из стандартной библиотеки TSPLib и применён генетический алгоритм с эвристическим кроссовером Грэффенстетта их решения.

В вычислительном эксперименте использовался следующий набор параметров: $|P| = 5$, $M = 5$, $T = 5$, $|N| = 20$. Для каждой из тестовых задач было сгенерировано по 20 различных решений. Количество тестов для одного решения $s(p)$ равнялось 5. Исходный маршрут разбивался на пять периодов. Фиксировался номер периода t , после которого решение $s(p)$ теряло свойство временной состоятельности. Данным значениям соответствуют столбцы $b(s, t)$ в таблице 1.

Таблица 1. Оценка временной состоятельности генетического алгоритма для TSP

Тест. задача	Кол-во тестов	$b(s, t)$				Кол-во сост. во врем.
		$t = 1$	$t = 2$	$t = 3$	$t = 4$	
att48	100	31	16	26	8	19
eil51	100	64	14	11	3	8
berlin52	100	19	27	4	4	46
st70	100	16	22	4	1	57
eil101	100	15	28	3	2	52
Сумма	500	145	107	48	18	182

Среднее значение экспериментального уровня временной состоятельности генетического алгоритма для решения задачи TSP: $conGA = 0,364$.

Данное значение является довольно низким: только треть сгенерированных на начальном этапе решений сохраняет свойство оптимальности в процессе своей реализации. Это означает,

что существуют другие маршруты, которые можно получить динамически с помощью того же эвристического алгоритма, значения целевой функции которых будет меньше, чем у текущих решений.

Вычислим значение уровня временной состоятельности для решений, полученных с помощью процедуры динамической адаптации. Проведём эксперимент по той же схеме и с параметрами, определёнными ранее. Результаты экспериментов приведены в таблице 2.

Таблица 2. Оценка временной состоятельности динамической адаптации генетического алгоритма для решения задачи TSP

Тест. задача	Кол-во тестов	$b(s, t)$				Кол-во сост. во врем.
		$t = 1$	$t = 2$	$t = 3$	$t = 4$	
att48	100	24	5	11	1	59
eil51	100	47	11	7	5	30
berlin52	100	2	12	5	3	78
st70	100	1	7	5	5	82
eil101	100	1	6	4	4	85
Сумма	500	75	41	32	18	334

Среднее значение экспериментального уровня временной состоятельности динамически адаптированного генетического алгоритма для решения задачи TSP равно 0,668, что почти в два раза выше, чем для генетического алгоритма до адаптации.

Представим зависимость количества состоятельных во времени решений, полученных в процессе эксперимента в первые t периодов, от t . Общее количество проведенных прогонов (тестов) равно $M|N||P|$. Рассмотрим функцию (8), значения которой равно количеству состоятельных во времени решений после завершения каждого периода:

$$(8) \quad c(s, t) = M|N||P| - \sum_{k=0}^t b(s, k).$$

На рис. 1 непрерывная линия соответствует значениям функции $c(s, t)$ при расчёте экспериментального уровня временной

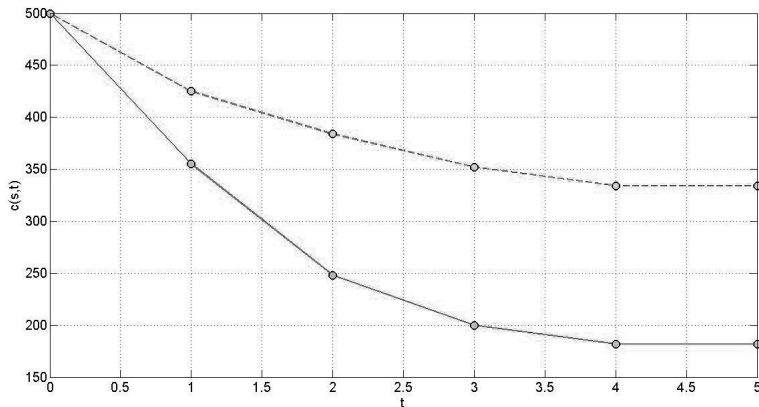


Рис. 1. Зависимость количества состоятельных во времени решений от периода для генетического алгоритма и процедуры динамической адаптации ГА

состоятельности генетического алгоритма с эвристическим кроссовером Грэффенстетта, а пунктирная линия – значениям этой функции для динамической адаптации этого алгоритма.

Сравнив значения таблиц 1 и 2 отметим, что после процедуры динамической адаптации значение экспериментального уровня временной состоятельности решений увеличилось как для каждой тестовой задачи, так и в среднем для всех.

6.2. Сравнение генетического алгоритма с эвристическим кроссовером Грэффенстетта и его динамической адаптации

На примере задачи TSP и генетического алгоритма было показано, что динамический подход позволяет получать решения с более высоким уровнем временной состоятельности.

Проверим, как изменится длина маршрутных планов при данной процедуре. В таблице 3 представлено сравнение длин маршрутов (решений), сгенерированных генетическим алгоритмом и динамической адаптацией генетического алгоритма. Средние значения целевой функции приведены для 500 прогонов алгоритма.

ритмов. Критерием останова для каждого запуска генетического алгоритма с эвристическим кроссовером Греффенстетта выступало заданное число поколений. Оно равнялось 100. Процедура динамической адаптации проводилась по схеме, приведённой в алгоритме 3, при следующем наборе параметров: $T = 5$, $|N| = 20$.

Введём следующие обозначения: значение $l_{GA}(p)$ равно длине маршрута, полученного генетическим алгоритмом для тестовой задачи p , $l_{DAGA}(p)$ – длина маршрута, полученного при использовании динамически адаптированного генетического алгоритма. В последних двух столбцах таблицы 3 приведены значения целевой функции, соответствующие эффективному запуску. Эффективным запуском процедуры динамической адаптации для тестовой задачи p будем называть запуск, для которого значение величины $l_{GA}(p) - l_{DAGA}(p)$ максимально в проведённом эксперименте.

Таблица 3. Сравнение длины маршрутных планов, сгенерированных генетической эвристикой с эвристическим кроссовером Греффенстетта и её динамической адаптацией

Тестовая задача	Среднее значение f		Эффективный запуск	
	Генетич. алгоритм	Динами. адаптация	Генетич. алгоритм	Динам. адаптация
att48	12270,72	11723,79	11657	10928
eil51	493,72	464,64	493	448
berlin52	8836,31	8257,36	8570	7893
st70	803,25	758,61	765	708
eil101	779,30	720,338	725	665

Из таблицы 3 видно, что для каждой из рассмотренных тестовых задач среднее значение сгенерированных решений алгоритмом динамической адаптации меньше, чем классическим генетическим алгоритмом с эвристическим кроссовером Греффенстетта. Статистику улучшения проведём по формуле

$$(9) \quad k = \frac{l_{GA}(p) - l_{DAGA}(p)}{l_{GA}(p)} 100\%.$$

Средний процент улучшения равен 6,01%. Максимальное значение улучшения получено для тестовой задачи eil51 и составляет 9,12%.

6.3. Время работы динамической адаптации генетического алгоритма

Процедура динамической адаптации генетического алгоритма с эвристическим кроссовером Греффенстетта требует дополнительных запусков эвристики на каждом периоде, тем самым увеличивая время получения одного маршрутного плана по сравнению с классическим подходом. Проведём следующий эксперимент. Пусть t_{dyn} – средняя продолжительность работы (в секундах) алгоритма динамической адаптации при генерировании маршрута. Ограничим время работы классической генетической эвристики величиной t_{dyn} . Для каждой тестовой задачи проведём по 100 запусков генетического алгоритма и его динамической адаптации и сравним результаты двух методов при фиксированной продолжительности их работы, равной t_{dyn} .

Таблица 4. Сравнение среднего значения длины генерируемых маршрутов при работе алгоритмов при различных фиксированных значениях t_{dyn}

Тестовая задача	Среднее значение f		t_{dyn}
	Генетический алгоритм	Динамическая адаптация	
att48	12273,51	11723,79	120,1
eil51	492,44	464,64	165,5
berlin52	8820, 25	8257,36	206,35
st70	807,37	758,61	1300,4
eil101	786,54	720,338	3403,7

Как видно из таблицы 4, при фиксированном времени работы метод динамической адаптации генетической эвристики способен генерировать лучшие решения по сравнению с генетическим алгоритмом с эвристическим кроссовером Греффенстетта. Данное замечание стоит учитывать в задачах большой размер-

ности, когда время генерации одного маршрута может занимать десятки минут. Как показал проведенный эксперимент, в случае заданного в условии задачи промежутка времени для генерации маршрутного плана выгоднее использовать процедуру динамической адаптации вместо классической эвристики, распределив время между соответствующими периодами работы алгоритма.

7. Заключение

В ходе исследования был подробно описан и реализован на языке программирования Java генетический алгоритм с эвристическим кроссовером Греффенстетта решений задачи TSP. Произведена оценка уровня временной состоятельности решений, генерируемых данной эвристикой. Проведенные расчеты для набора тестовых задач из библиотеки TSPLib продемонстрировали, что генетический алгоритм с эвристическим кроссовером Греффенстетта генерирует маршруты с невысоким уровнем временной состоятельности.

Для генерации решений с более высоким уровнем временной состоятельности предложена процедура динамической адаптации генетического алгоритма с эвристическим кроссовером Греффенстетта, в результате применения которой уровень временной состоятельности решений, генерируемых полученной эвристикой, увеличен в два раза. При этом среднее значение длины маршрутов уменьшилось на 6,01%. Таким образом, использование на практике предложенного динамического подхода, как показал численный эксперимент, позволяет генерировать маршруты меньшей длины.

Было показано, что при фиксированном ограничении на время работы процедура динамической адаптации генетического алгоритма с кроссовером Греффенстетта позволяет получать маршруты меньшей длины.

Стоит отметить, что предложенный метод динамической адаптации может быть применён и к другим эвристическим алгоритмам решения задачи TSP. Использование данного метода при решении задач транспортной маршрутизации таких как TSPTW,

IRP, MDVRP, PDP возможно, если эвристический алгоритм способен решать как саму задачу, так и её подзадачи [45, 46].

Литература

1. БОРИСОВСКИЙ П.А., ЕРЕМЕЕВ А.В. *О сравнении некоторых эволюционных алгоритмов* // Автоматика и телемеханика. – 2004. – №3. – С. 3–9.
2. БРОНШТЕЙН Е.М., ЗАЙКО Т.А. *Детерминированные оптимизационные задачи транспортной логистики* // Автоматика и телемеханика. – 2010. – №10. – С. 133–147.
3. ЕРЗИН А.И. *Задачи маршрутизации*. – Новосибирск: РИЦ НГУ, 2014. – 95 с.
4. ИВАНКО Е.Е. *Метод масштабирования в приближенном решении задачи коммивояжёра* // Автоматика и телемеханика. – 2011. – №12. – С. 115–129.
5. ИПАТОВ А.В. *Модифицированный метод отжига в задаче маршрутизации транспорта* // Труды института математики и механики УрО РАН. – 2011. – Т. 17. – №4. – С. 121–125.
6. МЕЛАМЕД И.И., СЕРГЕЕВ С.И., СИГАЛ И.Х. *Задача коммивояжёра. Вопросы теории* // Автоматика и телемеханика. – 1989. – №9. – С. 3–33.
7. МЕЛАМЕД И.И., СЕРГЕЕВ С.И., СИГАЛ И.Х. *Задача коммивояжёра. Приближенные алгоритмы* // Автоматика и телемеханика. – 1989. – №11. – С. 3–26.
8. МУГАЙСКИХ А.В. *Динамическая адаптация генетического алгоритма для задачи коммивояжёра* // Процессы управления и устойчивость. – 2015. – Т. 2. – №1. – С. 665–670.
9. ПЕТРОСЯН Л.А., ЗЕНКЕВИЧ Н.А. *Принципы устойчивой кооперации* // Управление большими системами. – 2009. – Вып. 3. – С. 100–120.
10. САЗОНОВ В.В., СКОБЕЛЕВ П.О., ЛАДА А.Н. И ДР. *Применение мультиагентных технологий в транспортной задаче с временными окнами и несколькими пунктами погрузки* // Управление большими системами. – 2016. – Вып. 64. – С. 65–80.

11. ТОВСТИК Т.М, ЖУКОВА Е.В. *Алгоритм приближенного решения задачи коммивояжера* // Вестник С.-Петерб. ун-та. – 2013. – Сер. 1, вып. 1. – С. 101–109.
12. УРАКОВ А.Р., ТИМЕРЯЕВ Т.В. *Алгоритм решения динамической задачи поиска кратчайших расстояний в графе* // Управление большими системами. – 2017. – Вып. 65. – С. 60–86.
13. ЧЕНЦОВ А.Г. *Задача последовательного обхода мегаполисов с условиями предшествования* // Автоматика и телемеханика. – 2014. – №4. – С. 170–190.
14. ЧЕНЦОВ А.Г. *Одна параллельная процедура построения функции Беллмана в обобщённой задаче курьера с внутренними работами* // Автоматика и телемеханика. – 2012. – №3. – С. 134–139.
15. ANBUUDAYASANKAR S.P., GANESH K., MOHAPATRA S. *Survey of Methodologies for TSP and VRP* // Models for Practical Routing Problems in Logistics. – 2014. – P. 11–42.
16. BELLMAN R. *Dynamic Programming*. – Princeton: Princeton University Press, 1957. – 392 p.
17. BELLMAN R. *Dynamic programming treatment of the travelling salesman problem* // Journal of the ACM. – 1962. – Vol. 9. – P. 61–63.
18. BLUM C., ROLI A. *Metaheuristics in combinatorial optimization: Overview and conceptual comparison* // ACM Comput. Surv. – 2003. – Vol. 35, No. 3. – P. 268–308.
19. CHIU K.-M., LIU J.-S. *Robot routing using clustering-based parallel genetic algorithm with migration* // Proc. of Merging Fields Of Computational Intelligence And Sensor Technology. – 2011. – №9. – P. 42–49.
20. CHRISTOFIDES N. *Worst-case analysis of a new heuristic for the travelling salesman problem* // Technical Report 388. – Graduate School of Industrial Administration, Carnegie Mellon University, 1976. – 11 p.

21. COMARELA G., GONCALVES K., PAPPA G.L., ALMEIDA J., ALMEIDA V. *Robot routing in sparse wireless sensor networks with continuous ant colony optimization* // Proc. of the 13th Annual Conference Companion on Genetic and Evolutionary Computation (GECCO'11), New York, NY, USA, ACM. – 2011. – P. 599–606.
22. DANTZIG G.B. *Linear programming and extensions*. – Princeton: Princeton Univ. Press, 1963. – 219 p.
23. DORIGO M., GAMBARDELLA L.M. *Ant colonies for the travelling salesman problem* // Biosystems. – Vol. 43, No. 2. – 1997. – P. 73–81.
24. DORRONSORO B. *The VRP Web*. – Languages and Computation Sciences Department, University of Malaga. – URL: <http://www.bernabe.dorronsoro.es/vrp/>.
25. FURINI F., PERSIANI C.A., TOTH P. *The Time Dependent Traveling Salesman Planning Problem in Controlled Airspace* // Transportation Research Part B: Methodological. – 2016. – Vol. 90. – P. 38–55.
26. GENDREAU M., POTVIN J.-Y. *Handbook of Metaheuristics*. – Springer Publishing Company, 2010. – 649 p.
27. GENDREAU M., YEHERTZO A., LAPORTE G., STAN M. *A Generalized Insertion Heuristic for the Traveling Salesman Problem with Time Windows* // Oper. Res. – 1998. – Vol. 46, No. 3. – P. 330–335.
28. GRANVILLE V., KRIVANEK M., RASSON J.-P. *Simulated annealing: A proof of convergence* // IEEE Trans. Pattern Anal. Mach. Intell., 1994. – Vol. 16, No. 6. – P. 652–656.
29. GREFENSTETTE J.J., GOPAL R., ROSMAITA B.J., VAN GUCHT D. *Genetic algorithms for the traveling salesman problem* // Proc. of the 1st Int. Conference on Genetic Algorithms. – Hillsdale: Lawrence Erlbaum Associates, 1985. – P. 160–168.
30. GUTIN G., PUNNEN A.P. *The traveling salesman problem and its variations*. – Kluwer Academi, 2002. – 837 p.

31. GUTIN G., YEO A., ZVEROVICH A. *Traveling salesman should not be greedy: domination analysis of greedy-type heuristics for the tsp* // Discrete Applied Mathematics. – 2002. – Vol. 117, No. 1–3. – P. 81–86.
32. HELD M., KARP R.M. *A dynamic programming approach to sequencing problems* // Proc. of the 1961 16th ACM National Meeting, ACM. – 1961. – P. 201–204.
33. HELSGAUN K. *An effective implementation of the lin-kernighan traveling salesman heuristic* // European Journal of Operational Research, 2000. – Vol. 126. – P. 106–130.
34. HELSGAUN K. *General k-opt submoves for the lin-kernighan tsp heuristic* // Math. Program. Comput. – 2009. – Vol. 1, No. 2–3. – P. 119–163.
35. HOLLAND J.H. *Adaptation in natural and artificial systems*. – Ann Arbor: The University of Michigan Press, 1975. – 183 p.
36. HURKENS COR A.J., WOEGINGER G.J. *On the nearest neighbor rule for the traveling salesman problem* // Oper. Res. Lett., 2004. – Vol. 32, No. 1. – P. 1–4.
37. JOHNSON D.S., MCGEOCH L.A. *Experimental analysis of heuristics for the stsp*. In *Local Search in Combinatorial Optimization*. – Wiley and Sons, 2001. – 80 p.
38. KHAN I.H. *Assessing Different Crossover Operators for Travelling Salesman Problem* // I.J. Intelligent Systems and Applications. – 2015. – Vol. 1. – P. 19–25.
39. KNEPPER R.A., LAYTON T., ROMANISHIN J., RUS D. *Ikeabot: An autonomous multi-robot coordinated furniture assembly system* // Robotics and Automation. – 2013. – P. 855–862.
40. MILLER C.E., TUCKER A.W., ZEMLIN R.A. *Integer programming formulation of traveling salesman problems* // Journal of the ACM. – 1960. – Vol. 7, No. 4. – P. 326–329.
41. PAPADIMITRIOU C.H., STEIGLITZ K. *Combinatorial Optimization: Algorithms and Complexity*. – Upper Saddle River: Prentice-Hall, Inc., 1982. – 496 p.

42. PULJIC K., MANGER C.R. *Comparison of eight evolutionary crossover operators for the vehicle routing problem* // Mathematical Communications. – 2013. – Vol. 18. – P. 359–375.
43. REINELT G. *Travelling Salesman Problem Library*. – URL: <http://comopt.ifi.uni-heidelberg.de/software/TSPLIB95/>.
44. SIVASOUNDARI A., KALAIMANI S. *Wireless surveillance robot with motion detection and live video transmission* // Int. Journal of Emerging Science and Engineering. – 2013. – No. 1. – P. 147–165.
45. SHIROKIKH V.A., ZAKHAROV V.V. *Dynamic Adaptive Large Neighbourhood Search for Inventory Routing Problem* // Advances in Intelligent Systems and Computing. – 2015. – Vol. 359. – P. 231–241.
46. SHIROKIKH V.A., ZAKHAROV V.V. *Heuristic evaluation of the characteristic function in the Cooperative Inventory Routing Game* // Journal on Vehicle Routing Algorithms. – 2017. – P. 1–14.
47. TALBI E.-G. *Metaheuristics for Bi-level Optimization*. – Springer Publishing Company, Inc., 2013. – 288 p.
48. WANG Y., WU C.H. *Robot-assisted sensor network deployment and data collection* // Proc. of the 2007 IEEE Int. Symposium on Computational Intelligence in Robotics and Automation. – 2007. – P. 467–472.
49. ZACHARIASEN M, DAM M. *Tabu Search on the Geometric Traveling Salesman Problem* // Proc. from Metaheuristics Int. Conference, Colorado, 1996. – P. 571–587.
50. ZAKHAROV V.V., DEMENTIEVA M. *Multistage cooperative games and problem of time consistency* // Int. Game Theory Review. – 2004. – Vol. 6. – P. 157–170.
51. ZAKHAROV V.V., SHCHEGRYAEV A.N. *Multi-period cooperative vehicle routing games* // Contributions to Game Theory and Management. – 2014. – Vol. 7, No. 2. – P. 349–359.

52. ZAKHAROV V.V., SHCHEGRYAEV A.N. *Stable Cooperation in Dynamic Vehicle Routing Problems* // Automation and Remote Control. – 2015. – Vol. 76, No. 5. – P. 935–943.

DYNAMIC ADAPTATION OF GENETIC ALGORITHM FOR THE LARGE-SCALE ROUTING PROBLEMS

Victor Zakharov, Saint-Petersburg State University, Saint-Petersburg, Doctor of Science, professor (v.zaharov@spbu.ru).
Alexander Mugayskikh, Saint-Petersburg State University, Saint-Petersburg, student (alexander.mugaiskih@gmail.com).

Abstract: This paper is devoted to implementation of the dynamic adaptation procedure for genetic algorithm used for the traveling salesman problem on large-scale networks. It is shown that solutions obtained by genetic algorithm can be improved during its dynamic adaptation and allow finding the more effective routes for the equal time. To evaluate effectiveness of new approach computational experiments were performed on well-known benchmark instances from TSPLib. As a result, the experimental level of time consistency of improved solution considerably increases compare to basic one. Dynamically adapted genetic algorithm has demonstrated possibility to produce solutions with higher level of time consistency. At the same time proposed procedure reduces length of the one solution in certain experiment as well as average length of all routes in it.

Keywords: time consistency, genetic algorithm, routing problems.

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

Поступила в редакцию 17.04.2017.

Дата опубликования 31.05.2018.

УДК 330.4
ББК 65.27

ДИНАМИЧЕСКАЯ МОДЕЛЬ ОЦЕНКИ БАЛАНСА СОЛИДАРНО-РАСПРЕДЕЛИТЕЛЬНОЙ ПЕНСИОННОЙ СИСТЕМЫ

Калашников П. В.¹

*(Дальневосточный федеральный университет,
Владивосток)*

Описывается построение динамической актуарной модели оценки баланса солидарно-распределительной пенсионной системы. На основе построения актуарной модели и анализа статистических данных делается долгосрочный прогноз уровня сбалансированности взносов и выплат, связанных с формированием страховой части трудовой пенсии по старости на общих основаниях. В ходе исследования проводится анализ управляющих воздействий на параметры актуарного базиса, позволяющих сократить уровень возникающего дефицита бюджета Пенсионного фонда РФ в долгосрочном периоде.

Ключевые слова: управление пенсионной системой, актуарный базис, прогноз уровня сбалансированности бюджета ПФР.

1. Введение

Обеспечение достойного уровня жизни старшего поколения является приоритетной целью социально-экономической политики большинства современных стран. В настоящее время пенсионная реформа является предметом широкого обсуждения. Ее огромная социальная значимость очевидна и является фундаментом для эффективного поступательного развития общества в целом, активного вовлечения граждан в экономическую

¹ Павел Викторович Калашников, аспирант 3-го года обучения (pkalash_89@mail.ru).

жизнь страны на протяжении всего периода трудовой деятельности.

Объектом исследования является пенсионная система Российской Федерации, рассматриваемая в контексте моделирования динамики величины страховых взносов и обязательств ПФР по выплате трудовой пенсии по старости на общих основаниях.

Предметом исследования является оценка основных показателей бюджета Пенсионного фонда в части величины взносов и обязательств по выплате страховой части трудовой пенсии по старости на общих основаниях.

Задачами исследования являются: построение модели оценки уровня сбалансированности бюджета Пенсионного фонда Российской Федерации (ПФР), расчет величины дефицита ПФР при существующей структуре населения и законодательной базе, а также анализ инструментов социально-экономической политики, позволяющих обеспечить балансировку страховых взносов и обязательств ПФР по старости на общих основаниях в долгосрочном периоде.

На состояние пенсионной системы в настоящее время оказывает значительное влияние неблагоприятная демографическая ситуация, которая выражается в увеличении числа людей старших возрастов и росту их относительной доли в общем количестве населения. По данным статистического учета трудовую пенсию по старости получают в настоящее время 35 555 тыс. человек, что составляет порядка 24% от общей численности населения, а общая численность пенсионеров, получающих различные виды пенсий, превышает 42 млн. человек [16]. Данный факт обуславливает постоянный рост уровня демографической нагрузки на трудоспособное население со стороны пожилых людей [11].

Прогноз уровня демографической нагрузки на трудоспособное население по данным Федеральной службы государственной статистики [7] представлен на рис. 1.

Из анализа данных прогноза можно сделать вывод о росте рассматриваемого показателя в долгосрочном периоде, обусловленном увеличением числа лиц пожилого возраста, что в свою очередь подтверждается исследованием половозрастной структуры населения представленной на рис. 2 [15].

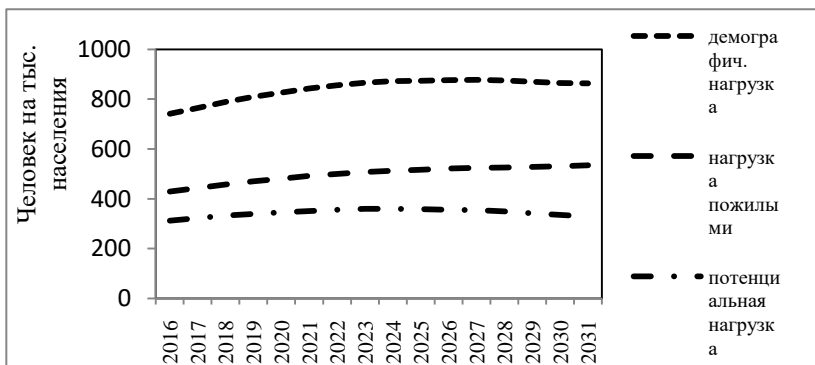


Рис. 1. Прогноз уровня демографической нагрузки на трудоспособное население

Половозрастная структура населения отражает распределение население по возрастным группам, а также соотношение мужчин и женщин в каждом возрастном интервале.

На рис. 2 приведен регрессивный тип половозрастной структуры, что отражает процесс старения населения Российской Федерации в долгосрочном периоде и обуславливает рост численности лиц старших возрастных групп при относительно низком уровне рождаемости.

Задача актуарного оценивания солидарно-распределительной системы пенсионного страхования включает в себя анализ демографических, социально-экономических, а также институциональных параметров (пенсионного законодательства). Каждая из указанных групп параметров в свою очередь подразделяется на широкий спектр подзадач, рассмотрение которых в совокупности позволяет дать прогноз состояния пенсионной системы в целом, а также разработать механизмы для ее эффективного функционирования в краткосрочном и долгосрочном периоде. Основные подходы к решению рассматриваемой задачи представлены в работах [1, 4, 8, 10, 12, 13].

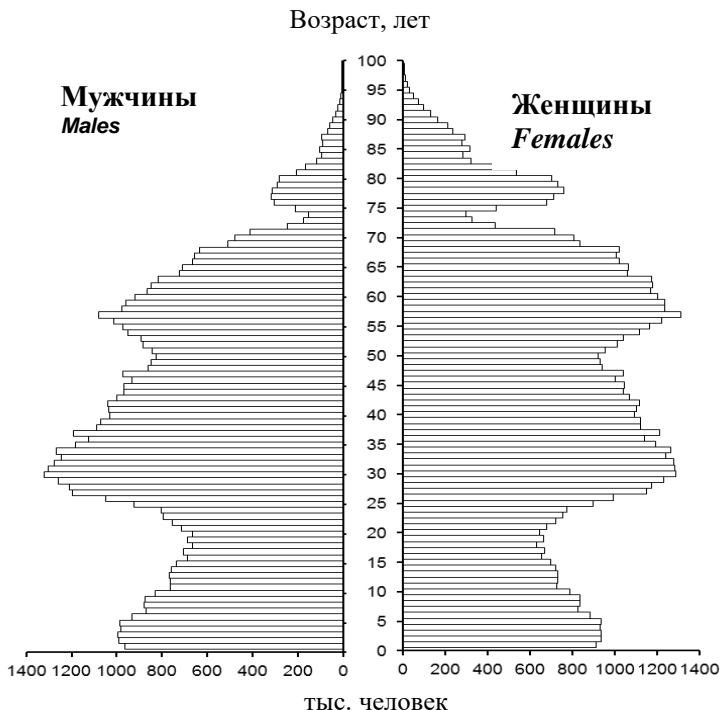


Рис. 2. Половозрастная структура населения РФ на начало 2017 г.

Мировой опыт построения моделей оценки эффективности функционирования пенсионных систем на уровне государства описан в работах [6, 9, 14, 17, 18, 19, 20].

Модель Всемирного банка PROST (Pension Reform Option Simulation Toolkit) является одной из базовых актуарных моделей, применяемых для оценки качественных изменений государственной пенсионной системы, построенной по солидарно-распределительному принципу [14].

Рассматриваемая модель представляет собой ряд взаимосвязанных между собой компонентов для проведения расчетов по оценке уровня баланса пенсионной схемы в долгосрочном периоде. Демографический блок модели содержит в себе описание числа плательщиков взносов и получателей пенсии для

каждой из возрастных групп населения, входящих в пенсионную схему. Параметрический и пенсионный блоки модели учитывают такие показатели, как размер ставки страховых взносов, средней заработной платы для населения, занятого в экономике, средней величины выплачиваемых пенсионных пособий новым участникам пенсионной схемы, а также уровень индексации выплат для пенсионеров. При проведении расчетов получают оценки объема взносов в государственный пенсионный фонд и величины обязательств по выплате пенсионных пособий. Сопоставление этих показателей позволяет сделать вывод о сбалансированности солидарно-распределительной пенсионной системы, реализуемой на государственном уровне, в долгосрочном периоде. Рассматриваемая модель предназначена для получения ответов на следующие основные вопросы: продолжительность функционирования пенсионной системы в рамках действующего законодательства, изменение обязательств государства по выплате пенсионных пособий с течением времени, оценка уровня возможного дефицита средств, возникающего в ходе функционирования изучаемой схемы пенсионного обеспечения.

Пенсионная модель Международной организации труда (ILO Pension model) [18], так же как и модель Всемирного банка, ориентирована на оценку сбалансированности в долгосрочной перспективе пенсионных систем на уровне государства.

Модель MOT позволяет решить следующие задачи: определение соотношения доходов и обязательств государственного пенсионного фонда в долгосрочном периоде, а также расчет ставки взносов, соответствующей достижению баланса пенсионной системы.

В рассматриваемой модели выделяют следующие основные компоненты: демографический блок, макроэкономический блок, блок расходов и доходов пенсионной системы.

Начальные параметры демографического блока задаются экзогенно и содержат в себе информацию о половозрастной структуре населения, вероятностях дожития до соответствующих возрастов, определяемых на основе таблиц смертности, коэффициентах рождаемости и миграционного прироста. Модель позволяет осуществить прогноз изменения численности

населения по возрастно-половым группам в каждый год расчетного периода.

Макроэкономический блок содержит в себе экзогенно задаваемые данные по основным параметрам социально-экономического развития государства, таким как уровень инфляции, изменение среднего уровня заработной платы, уровень занятости и безработицы среди населения.

Расчет взносов в государственную пенсионную систему и выплат по пенсионным обязательствам осуществляется исходя из значений параметров, содержащихся в демографическом и макроэкономическом блоке модели, а также данных о среднем уровне пенсии по старости и величине ее индексации. Прогноз состояния пенсионной системы делается на основе сопоставления данных расчетов по обязательствам и взносам, поступающим в нее в долгосрочном периоде.

Основным недостатком в рассматриваемых моделях является универсальный характер описания пенсионной схемы, не позволяющий учитывать особенности российского законодательства в области обязательного пенсионного страхования. Также описанные выше модели позволяют получать информацию о качественных изменениях в состоянии пенсионной схемы в долгосрочном периоде. Однако получение точных количественных оценок уровня сбалансированности пенсионной системы в долгосрочном периоде при их помощи практически невозможно по причине огрубленного характера расчетов.

В мировой практике также следует отметить использование вычисляемых балансовых моделей общего равновесия (ВМОР) для моделирования пенсионных систем на уровне государства. Основные подходы по построению подобного сорта моделей для конкретных стран описаны в работах [17, 19, 20].

Модели общего балансового равновесия представляют собой совокупность взаимосвязанных между собой блоков, каждый из которых содержит ряд уравнений, переменных состояния, а также параметров, описывающих закон и критерии качества его функционирования. Наряду с этим в моделях присутствуют также балансовые соотношения и функции реакции блоков модели на изменения значений параметров актуарного базиса, откалиброванные на основе анализа статистических

данных. Пенсионная система рассматривается как часть целостной взаимосвязанной экономической системы государства, подверженная изменению макроэкономических показателей и сама влияющая на установление равновесия. Достижения баланса пенсионной системы рассматривается в парадигме равновесия Эрроу–Дебре и закона Вальраса.

К преимуществам описанного подхода можно отнести балансовую сходимость переменных модели, обусловленную представлением данных в форме матрицы социальных счетов, описывающих основные макроэкономические показатели рассматриваемого государства. Доходы и расходы всех основных моделируемых агентов экономики, таких как государство, фирмы, домохозяйства, в рассматриваемой модели совпадают. Балансовые соотношения должны выполняться как для всей экономики в целом, так и отдельных ее компонентов, представляющих соответствующе блоки модели. Рассматриваемая модель позволяет оценить структурные сдвиги в экономике, обусловленные изменениями в пенсионной системе.

Основной недостаток описанного подхода заключается в том, что модель ВМОР технически не позволяет оценивать уровень дефицита пенсионной системы в среднесрочном периоде, так как ввиду балансовой сходимости переменных модели они прогнозируются как сбалансированные, однако в реальности рассматриваемой схема пенсионного обеспечения может иметь недостаток средств, необходимых для исполнения обязательств.

Российская аналитическая пенсионная модель, разработанная Независимым актуарным аналитическим центром, описана в работе [3]. Аналогично модели Всемирного банка PROST рассматриваемая модель содержит демографический, макроэкономический, а также пенсионный блоки, применяемые для оценки уровня сбалансированности страховых взносов и обязательств ПФР по выплате пенсий населению в долгосрочном периоде. Данная модель реализована в форме взаимосвязанных макросов MS Excel и предназначена для анализа эффективности функционирования пенсионной системы РФ. Применяемая модель построена на основе модели Всемирного банка и адаптирована к использованию в рамках действующего российского пенсион-

ного законодательства. Однако в рамках построения рассматриваемой модели не определен комплекс управляющих воздействий на параметры актуарного базиса, позволяющий минимизировать величину дефицита бюджета ПФР в долгосрочном периоде, а также отсутствует описание мер социально-экономической политики, направленных на реализацию данной задачи на практике.

Все вышеприведенные недостатки основных моделей, применяемых в международной и отечественной практике, свидетельствуют о необходимости создания актуарной модели, адекватным образом описывающей действующую систему пенсионного страхования и позволяющей строить достоверные оценки уровня сбалансированности бюджета Пенсионного фонда РФ в долгосрочной перспективе. Также разработанная модель должна позволять проводить анализ параметров актуарного базиса с целью уменьшения возможного актуарного дефицита средств Пенсионного фонда РФ, необходимых для выплат трудовой пенсии по старости на общих основаниях. Основная научная новизна реализуемого в данной работе подхода заключается в построении актуарной модели оценки баланса взносов и обязательств ПФР по выплате страховой части трудовой пенсии по старости в долгосрочном периоде, позволяющей определить комплекс управляющих воздействий на параметры актуарного базиса, сводящих к минимуму возникающий дефицит средств, необходимых для исполнения рассматриваемых обязательств. Также построенная модель позволяет произвести анализ дифференцированного влияния управляющих воздействий на параметры актуарного базиса и сравнить эффективность проведения различных мер социально-экономической политики, направленных на обеспечение баланса пенсионной системы РФ.

2. Модель оценки баланса солидарно-распределительной пенсионной системы

Проведем последовательное построение модели баланса солидарно-распределительной пенсионной системы.

Блок демографических параметров модели включает в себя анализ половозрастной структуры населения, прогноз численно-

сти пенсионеров и занятых в экономике в каждый год расчетного периода, а также прогноз численности новорожденных и величины миграционного прироста.

Базовым понятием, применяемым при анализе демографической ситуации и построении долгосрочных прогнозов, является таблица смертности населения [5]. Вероятность дожития для представителя начальной совокупности в возрасте x лет до возраста $x + 1$ лет задается соотношением

$$(1) \quad p_x = \frac{l_{x+1}}{l_x},$$

где l_{x+1} , l_x – численность представителей начальной совокупности, доживающих до возраста $x + 1$ и x лет соответственно.

Обозначим через $inv_x^{(m)}$, $inv_x^{(f)}$ – вероятности события «не получить инвалидность» для лица мужского и женского пола соответственно, находящегося в возрасте x лет и продолжающего трудовую деятельность в возрасте $x + 1$ год. Указанные показатели рассчитываются на основе таблиц инвалидности населения РФ, разработанных Независимым актуарным информационно-аналитическим центром [2]. По предположению модели, лица, получившие инвалидность в процессе трудовой деятельности, считаются выбывшими из пенсионной схемы обязательного пенсионного страхования по старости на общих основаниях. Случаи возможного возвращения данных лиц в активную трудовую деятельность после снятия инвалидности не рассматриваются. Дифференциация групп инвалидности не производится.

Используя выше приведенные обозначения численность мужчин и женщин данного возраста, доживающих до следующего года расчетного периода с номером $s + 1$, можно оценить исходя из следующего рекуррентного соотношения:

$$(2) \quad men_{x+1}^{(s+1)} = men_x^{(s)} p_x^{(m)} inv_x^{(m)},$$

$$(3) \quad fem_{x+1}^{(s+1)} = fem_x^{(s)} p_x^{(f)} inv_x^{(f)},$$

где

$men_x^{(s)}$ – численность представителей начальной совокупности мужского населения в возрасте x лет в год расчетного периода с номером s ;

$fem_x^{(s)}$ – численность представителей начальной совокупности женского населения в возрасте x лет в год расчетного периода с номером s ;

$p_x^{(m)}$ – вероятность дожития до возраста $x + 1$ лет для представителя мужского населения в возрасте x лет;

$p_x^{(f)}$ – вероятность дожития до возраста $x + 1$ лет для представителя женского населения в возрасте x лет.

Численность новорожденных мужского и женского пола определяется в соответствии с прогнозом Федеральной службы государственной статистики и задается в модели экзогенно.

Моделирование изменения численности населения за счет лиц прибывающих и отбывающих на постоянное место жительства осуществляется исходя из величины сальдо внешней миграции, на основе соотношений

$$(4) \quad \begin{aligned} migrmen_{x+1}^{(s+1)} &= (imen_x^{(s)} - emen_x^{(s)})p_x^{(m)}inv_x^{(m)}, \\ migrfem_{x+1}^{(s+1)} &= (ifem_x^{(s)} - efem_x^{(s)})p_x^{(f)}inv_x^{(f)}, \end{aligned}$$

где

$imen_x^{(s)}$, $emen_x^{(s)}$ – численность иммигрантов и эмигрантов среди мужского населения в возрасте x лет в год, предшествующий году расчетного периода с номером $s + 1$, прибывающих или отбывающих на постоянное место жительства;

$ifem_x^{(s)}$, $efem_x^{(s)}$ – численность иммигрантов и эмигрантов среди женского населения в возрасте x лет в год, предшествующий году расчетного периода с номером $s + 1$ прибывающих или отбывающих на постоянное место жительства.

Численность временных трудовых мигрантов мужского $imew_x^{(s)}$ и женского $ifemw_x^{(s)}$ пола в возрасте x лет в год расчетного периода с номером s описывается с помощью экзогенных переменных модели, в соответствии с данными Федеральной службы государственной статистики и имеющимися прогнозами по величине рабочей силы, привлекаемой из-за рубежа. По предположению модели значения указанных переменных для возрастных групп населения, приобретающих право на получение трудовой пенсии по старости на общих основаниях, полагаются равными нулю.

Общее количество мужского и женского населения с учетом рождаемости и миграции рассчитывается исходя из соотношений

$$(5) \quad \begin{aligned} ment_x^{(s)} &= migrmen_x^{(s)} + men_x^{(s)} + imenw_x^{(s)}, \\ femt_x^{(s)} &= migrfem_x^{(s)} + fem_x^{(s)} + ifemw_x^{(s)}, \end{aligned}$$

где $ment_x^{(s)}$ – общая численность мужского населения в возрасте x лет в год расчетного периода с номером s ; $femt_x^{(s)}$ – общая численность женского населения в возрасте x лет в год расчетного периода с номером s .

Общее количество $N_x^{(s)}$ пенсионеров по старости в данный год расчетного периода с номером s определяется по формуле

$$(6) \quad N^{(s)} = \sum_{x=60}^{100} ment_x^{(s)} + \sum_{x=55}^{100} femt_x^{(s)}.$$

В настоящее время идет широкое обсуждение варианта реформы пенсионной системы РФ, предусматривающего увеличение, а также уравнивание возраста выхода на пенсию для мужского и женского населения. Прогноз численности пенсионеров по старости на общих основаниях в зависимости от изменения момента наступления пенсионного возраста, полученный на основе построенной актуарной модели, представлен на рис. 3, 4.

Для моделирования величины доходов бюджета Пенсионного фонда РФ введем следующие обозначения:

α_1 – ставка тарифа по страховым взносам в ПФР для граждан старше 1967 г. рождения и 1967 г. рождения и моложе, не производящих выплат на накопительную часть трудовой пенсии по старости;

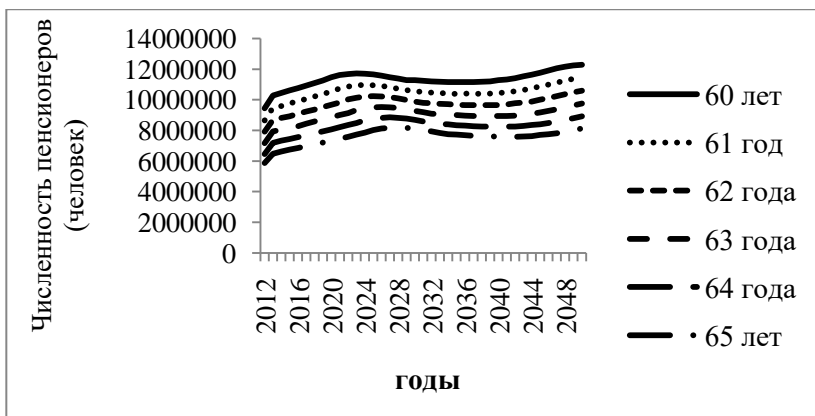


Рис. 3. Прогноз численности пенсионеров по старости на общих основаниях в составе мужского населения в зависимости от возраста выхода на пенсию

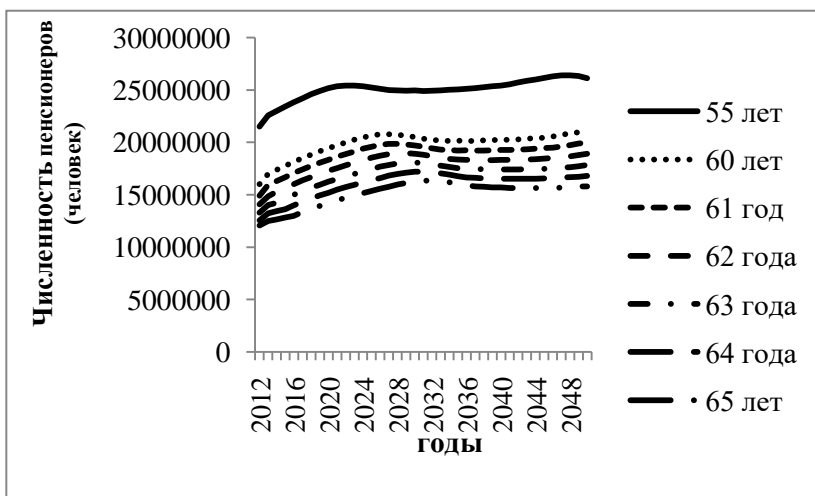


Рис. 4. Прогноз численности пенсионеров по старости на общих основаниях в составе женского населения в зависимости от возраста выхода на пенсию

α_2 – ставка тарифа по страховым взносам в ПФР для граждан 1967 г. рождения и моложе, производящих выплаты на накопительную часть трудовой пенсии по старости;

${}_m W_x^{(s)}$ – средний уровень номинальной заработной платы мужского населения в возрасте x лет в год расчетного периода с номером s ;

${}_f W_x^{(s)}$ – средний уровень номинальной заработной платы женского населения в возрасте x лет в год расчетного периода с номером s ;

$\delta_x^{(s)}$ – уровень занятости мужского населения в возрасте x лет в год расчетного периода с номером s

$\gamma_x^{(s)}$ – уровень занятости женского населения в год расчетного периода с номером s ;

${}_m \beta_x^{(s)}$ – отношение численности мужского населения 1967 г. рождения и моложе в возрасте x лет занятого в экономике, у которого отсутствует накопительная часть трудовой пенсии по старости, к общей численности мужского населения данной возрастной группы и социальной принадлежности в год расчетного периода с номером s ;

${}_m \beta_x^{(s)}$ – отношение численности мужского населения 1967 г. рождения и моложе в возрасте x лет занятого в экономике, у которого присутствует накопительная часть трудовой пенсии по старости, к общей численности мужского населения данной возрастной группы и социальной принадлежности в год расчетного периода с номером s ;

${}_f \beta_x^{(s)}$ – отношение численности женского населения 1967 г. рождения и моложе в возрасте x лет занятого в экономике, у которого отсутствует накопительная часть трудовой пенсии по старости, к общей численности женского населения данной возрастной группы и социальной принадлежности в год расчетного периода с номером s

${}_f \beta_x^{(s)}$ – отношение численности женского населения 1967 г. рождения и моложе в возрасте x лет занятого в экономике, у которого присутствует накопительная часть трудовой пенсии, к общей численности женского населения данной возрастной группы и социальной принадлежности в год расчетного периода с номером s .

Если в качестве базового года расчетного периода принять 2012 г, то величина страховых взносов, поступающих в ПФР от мужского населения в год расчетного периода с номером $s = 0, \dots, 13$, задается в виде следующих соотношений:

$$(7) \quad \begin{aligned} {}_1V_m^{(s)} &= \alpha_1 \sum_{x=18}^{45+s} m w_x^{(s)} m_1 \beta_x^{(s)} \delta_x^{(s)} ment_x^{(s)}, \\ {}_2V_m^{(s)} &= \alpha_2 \sum_{x=18}^{45+s} m w_x^{(s)} m_2 \beta_x^{(s)} \delta_x^{(s)} ment_x^{(s)}, \\ {}_{old}V_m^{(s)} &= \alpha_1 \sum_{x=46+s}^{59} m w_x^{(s)} \delta_x^{(s)} ment_x^{(s)}, \\ {}_{mon}V_m^{(s)} &= {}_1V_m^{(s)} + {}_2V_m^{(s)} + {}_{old}V_m^{(s)}, \end{aligned}$$

где ${}_1V_m^{(s)}$ – величина страховых взносов в Пенсионный фонд РФ от мужского населения младше и 1967 г. рождения, не имеющего накопительной части трудовой пенсии по старости; ${}_2V_m^{(s)}$ – величина страховых взносов в Пенсионный фонд РФ от мужского населения младше и 1967 г. рождения, имеющего накопительную часть трудовой пенсии по старости; ${}_{old}V_m^{(s)}$ – величина страховых взносов в Пенсионный фонд РФ от мужского населения старше 1967 г. рождения; ${}_{mon}V_m^{(s)}$ – величина страховых взносов в Пенсионный фонд РФ от мужского населения в каждый месяц года расчетного периода с номером s .

Для каждого года расчетного периода с номером $s = 14, \dots, 39$ расчет величины взносов в ПФР от мужского населения производится по формулам

$$(8) \quad \begin{aligned} {}_1V_m^{(s)} &= \alpha_1 \sum_{x=18}^{59} m w_x^{(s)} m_1 \beta_x^{(s)} \delta_x^{(s)} ment_x^{(s)}, \\ {}_2V_m^{(s)} &= \alpha_2 \sum_{x=18}^{59} m w_x^{(s)} m_2 \beta_x^{(s)} \delta_x^{(s)} ment_x^{(s)}, \\ {}_{mon}V_m^{(s)} &= {}_1V_m^{(s)} + {}_2V_m^{(s)}. \end{aligned}$$

Если в качестве базового года расчетного периода принять 2012 г, то величина страховых взносов, поступающих в ПФР от женского населения в год расчетного периода с номером $s = 0, \dots, 8$, задается в виде следующих соотношений:

$$\begin{aligned}
 {}_1V_f^{(s)} &= \alpha_1 \sum_{x=18}^{45+s} {}_f w_x^{(s)} {}_{f1} \beta_x^{(s)} \gamma_x^{(s)} femt_x^{(s)}, \\
 (9) \quad {}_2V_f^{(s)} &= \alpha_2 \sum_{x=18}^{45+s} {}_f w_x^{(s)} {}_{f2} \beta_x^{(s)} \gamma_x^{(s)} femt_x^{(s)}, \\
 {}_{old}V_f^{(s)} &= \alpha_1 \sum_{x=46+s}^{54} {}_f w_x^{(s)} \gamma_x^{(s)} femt_x^{(s)}, \\
 {}_{mon}V_f^{(s)} &= {}_1V_f^{(s)} + {}_2V_f^{(s)} + {}_{old}V_f^{(s)},
 \end{aligned}$$

где ${}_1V_f^{(s)}$ – величина страховых взносов в Пенсионный фонд РФ от женского населения младше и 1967 г. рождения, не имеющего накопительной части трудовой пенсии по старости; ${}_2V_f^{(s)}$ – величина страховых взносов в Пенсионный фонд РФ от женского населения младше и 1967 г. рождения, имеющего накопительную часть трудовой пенсии по старости; ${}_{old}V_f^{(s)}$ – величина страховых взносов в Пенсионный фонд РФ от женского населения старше 1967 г. рождения; ${}_{mon}V_f^{(s)}$ – величина взносов в Пенсионный фонд РФ от женского населения за каждый месяц года расчетного периода с номером s .

Для каждого года расчетного периода с номером $s = 9, \dots, 39$ расчет величины взносов в ПФР от женского населения производится по формуле

$$\begin{aligned}
 {}_1V_f^{(s)} &= \alpha_1 \sum_{x=18}^{54} {}_f w_x^{(s)} {}_{f1} \beta_x^{(s)} \gamma_x^{(s)} femt_x^{(s)}, \\
 (10) \quad {}_2V_f^{(s)} &= \alpha_2 \sum_{x=18}^{54} {}_f w_x^{(s)} {}_{f2} \beta_x^{(s)} \gamma_x^{(s)} femt_x^{(s)}, \\
 {}_{mon}V_f^{(s)} &= {}_1V_f^{(s)} + {}_2V_f^{(s)}.
 \end{aligned}$$

Общая величина взносов в Пенсионный фонд РФ на финансирование страховой части трудовой пенсии V_s определяется на основе соотношения

$$(11) \quad V_s = 12({}_{mon}V_m^{(s)} + {}_{mon}V_f^{(s)}).$$

Величина накопленных обязательств AL_s в год расчетного периода с номером s описывается соотношением

$$(12) \quad AL_s = 12P_0 \prod_{t=1}^s (1 + \theta_t) \left(\sum_{x=60}^{100} ment_x^{(s)} + \sum_{x=55}^{100} femt_x^{(s)} \right),$$

где P_0 – средняя величина пенсии в базовый год; θ_t – размер индексации средней пенсии в каждый последующий год расчет-

ного периода относительно базового в соответствии с уровнем инфляции предыдущего года.

Уровень баланса взносов и обязательств Пенсионного фонда РФ по выплате страховой части трудовой пенсии по старости на общих основаниях рассчитывается для каждого года расчетного периода с номером s на основе соотношения

$$(13) B_s = V_s - AL_s.$$

Исходя из значений величины B_s , делается вывод о дефиците ($B_s < 0$), профиците ($B_s > 0$) или сбалансированности ($B_s = 0$) средств, необходимых для выполнения обязательств ПФР по выплате трудовой пенсии по старости на общих основаниях, и поступивших в пенсионную систему страховых взносов.

3. Анализ результата влияния управляющих воздействий на параметры актуарного базиса

Рассмотрим применение построенной модели для оценки баланса солидарно-распределительной пенсионной системы, а также определения комплекса управляющих воздействий на параметры актуарного базиса, позволяющего минимизировать возможный дефицит средств ПФР по исполнению обязательств, связанных с выплатой страховой части трудовой пенсии по старости на общих основаниях в долгосрочном периоде.

Значения основных параметров модели для базового года расчетного периода представлены в таблице 1.

Результаты расчетов по построенной модели оценки баланса солидарно-распределительной пенсионной системы представлены на рис. 5, 6, 7.

Прогноз уровня сбалансированности взносов и обязательств ПФР по выплате страховой части трудовой пенсии по старости на общих основаниях представлен на рис. 5.

Таблица 1. Значение основных параметров модели для базового года расчетного периода

Название параметра	Обозначение	Значение
Уровень инфляции	θ_0	0,07
Ставка страховых взносов в пенсионный фонд для лиц, не имеющих накопительной части трудовой пенсии (общий режим налогообложения)	α_1	0,22
Ставка страховых взносов в пенсионный фонд для лиц, имеющих накопительную часть трудовой пенсии (общий режим)	α_2	0,16
Размер средней пенсии по старости на общих основаниях в базовый год расчетного периода	P_0	9790
Возраст выхода на пенсию по старости на общих основаниях для мужского населения		60 лет
Возраст выхода на пенсию для женского населения		55 лет
Численность новорожденных в составе мужского населения	$men_x^{(0)}$	970063
Численность новорожденных в составе женского населения	$fem_x^{(0)}$	932021

При изучении результатов расчетов, представленных на графике, можно сделать вывод о дефиците средств ПФР, необходимых для исполнения обязательств по выплате основного вида пенсионных пособий, а именно пенсии по старости на общих основаниях. Величина дефицита средств значительно увеличивается в долгосрочном периоде.



Рис. 5. Прогноз уровня сбалансированности страховых взносов и выплат ПФР трудовой пенсии по старости в долгосрочном периоде

При анализе возможных управляющих воздействий на параметры актуарного базиса, направленных на сокращение дефицита бюджета ПФР в долгосрочном периоде, рассматриваются следующие: увеличение рождаемости на 100 000 человек ежегодно, увеличение величины миграционного прироста до величины 500 тыс. человек в год и численности легально работающих на территории временных трудовых мигрантов до 2,5 млн человек ежегодно, повышение пенсионного возраста для мужского и женского населения до 65 лет, а также полную отмену накопительной части пенсии по старости для населения младше 1967 г. рождения и внесение взносов только по страховому компоненту рассматриваемого пенсионного пособия. Подобный характер изменения параметров актуарного базиса основан на анализе динамики рассматриваемых величин в период, предшествующий базовому году проведения расчетов, исследовании долгосрочных прогнозов социально-экономического развития РФ, подготовленных Федеральной службой государственной статистики, а также изучении мирового опыта проведения реформ пенсионной системы на уровне государства.

Расчет уровня сбалансированности бюджета ПФР для каждого из рассмотренных сценариев изменения параметров актуарного базиса представлен на рис. 6. Среди предполагаемых мер социально-экономической политики наибольшее сокращение дефицита средств ПФР обеспечивает увеличение пенсионного возраста, а также отмена накопительной части трудовой пенсии по старости.

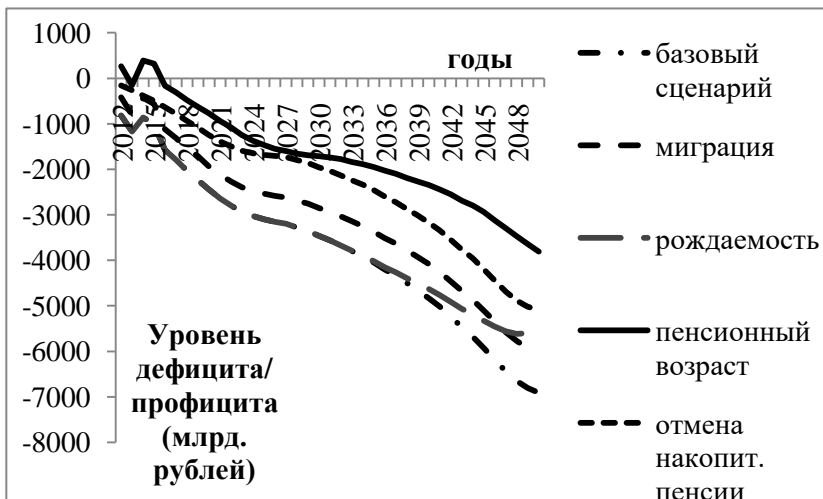


Рис. 6. Прогноз уровня сбалансированности страховых взносов и выплат ПФР трудовой пенсии по старости в долгосрочном периоде для различных вариантов управления параметрами актуарного базиса

Уровень сбалансированности страховых взносов и выплат ПФР трудовой пенсии по старости на общих основаниях в случае одновременной реализации комплекса управляющих воздействий на параметры актуарного базиса представлен на рис. 7.

Параметрическим триггером, переводящим уровень сбалансированности бюджета ПФР к новому равновесному состоянию, является увеличение пенсионного возраста для мужского и женского населения до 65 лет в 2025 г. Переход к новому равновесному состоянию сбалансированности ПФР в более поздние

сроки возможен только путем большего увеличения возраста выхода на пенсию.

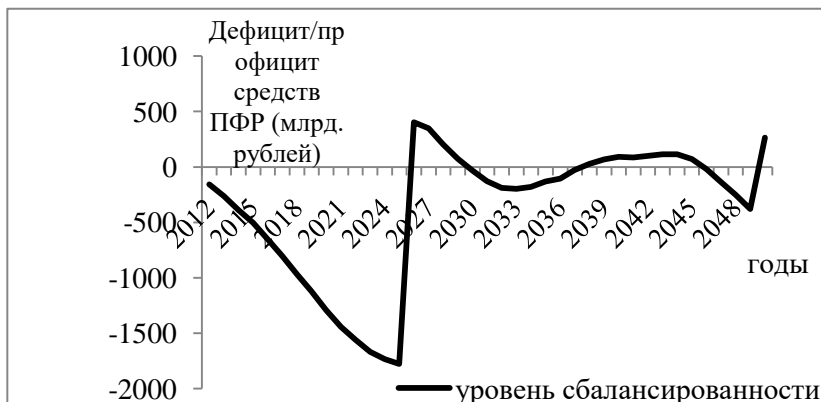


Рис. 7. Прогноз уровня сбалансированности страховых взносов и выплат ПФР по трудовой пенсии по старости на общих основаниях при одновременном изменении параметров актуарного базиса

Проведенный анализ позволяет сделать вывод о неэффективности действующей системы обязательного пенсионного страхования РФ в долгосрочном периоде. Наибольшее сокращение дефицита бюджета ПФР обеспечивается такими мерами социально-экономической политики, как увеличение возраста выхода на пенсию для мужского и женского населения до 65 лет, а также отмена накопительной части трудовой пенсии по старости.

Исследование управляющих воздействий на параметры актуарного базиса позволило определить комплекс мер социально-экономической политики, которые минимизируют величину недостатка средств ПФР, необходимых для исполнения обязательств по выплате трудовой пенсии по старости на общих основаниях населению РФ в долгосрочном периоде.

Литература

1. АГАПЦЕВ С.А. *Пенсионная реформа в России: цели и результаты*. – М.: «Тиссо-Полиграф», 2008. – 239 с.
2. БАСКАКОВ В.Н. *Страхование от несчастных случаев на производстве: актуарные основы*. – М.: Academia, 2001. – 192 с.
3. БАСКАКОВ В.Н. *Пенсионная система: модель для России и зарубежный опыт* // Серия «Научные доклады: независимый экономический анализ». – №140. – М.: МОНФ, 2003. – 54 с.
4. БАСКАКОВ В.Н. *Моделирование пенсионной системы: возмещение утраченного заработка* // *Актуарий*. – 2007. – №1. – С. 15–24.
5. БАУЭРС Н., ГЕРБЕР Х. и др. *Актуарная математика*. – М.: Янус-К, 2001. – 656 с.
6. ДЕГТЯРЬ Л.С. *Мировой опыт пенсионных реформ и реформирование пенсионной системы в России* // Аналитический вестник Совета Федерации ФС РФ. – 2002. – №33(189). – С. 40–71
7. *Демографический прогноз численности населения Российской Федерации до 2035 г.* – [Электронный ресурс]: данные Федеральной службы государственной статистики. – URL: http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/population/demography/# (дата обращения 1.09.2017).
8. ЗАХАРОВ И.Н. *Современная практика актуарного оценивания пенсионной системы Российской Федерации* // *Российское предпринимательство*. – 2011. – №3. – С. 18–24.
9. РОИК В.Д. *Эволюция пенсионных систем: мировые тенденции и опыт России* // *Человек и труд*. – 2008. – №8. – С. 17–25.
10. СИМОНЕНКО В.Н. *Сценарий моделирования пенсионной системы в разрезе современного состояния экономики* // *Вестник ТОГУ*. – 2010. – №4(19). – С. 145–152.
11. СМIRHOV И.В. *Демография: учебное пособие*. – Калуга: филиал СЗАГС в г. Калуга, 2004. – 138 с.

12. СОЛОВЬЕВ А.К. *Пенсионные системы в контексте страховых принципов* // Журнал новой экономической ассоциации. – 2012. – №3(15). – С. 141–156.
13. СОЛОВЬЕВ А.К. *Актuarные расчеты в системе пенсионного страхования*. – М.: «Финансы и статистика», 2005. – 239 с.
14. ХОЛЫЦМАН Р. *Обеспеченная старость в XXI веке: пенсионные системы и реформы в международной перспективе*. – Вашингтон: Всемирный банк, 2001. – 28 с.
15. *Численность населения Российской Федерации по полу и возрасту на 1 января 2017 г.* – [Электронный ресурс]: данные Федеральной службы государственной статистики. – URL: http://www.gks.ru/bgd/regl/b17_111/Main.htm (дата обращения: 1.09.2017).
16. *Численность пенсионеров и средний размер назначенных пенсий по видам пенсионного обеспечения и категориям пенсионеров*. – [Электронный ресурс]: данные Федеральная служба государственной статистики РФ. – URL: http://www.gks.ru/free_doc/new_site/population/urov/urov_p2.htm (дата обращения: 1.09.2017).
17. SIN Y. *China social security reform. Technical Analysis of Strategic Options*. – Washington DC, World Bank, 2000.
18. *The ILO Pension Model, a Technical Guide*. – URL: <http://www.ilo.org> (дата обращения 12.12.2016).
19. WANG Y. *Options and impact of China's pension reform: a complete general equilibrium analysis* // Journal of Comparative Economic. – 2004. – No. 32. – P. 105–127.
20. ZHUNG C. *Linking pension and state enterprise reform in China* // Thesis of the China Development Forum Conference on China Social Security System, Beijing, 2000.

THE DYNAMIC MODEL OF ASSESSMENT OF THE BALANCE OF THE SOLIDARY-DISTRIBUTION PENSION SYSTEM

Pavel Kalashnikov, Far East Federal University, Vladivostok, the third year postgraduate student (pkalash_89@mail.ru).

Abstract: The author describes how to build a dynamic actuarial balance model of the distribution component of the budget of the Pension Fund. The author offers a long-term forecast of the level of balance of contributions and benefits (associated with the formation of the insurance part of the labour old-age pension) which is based on the construction of actuarial models and analysis of statistical data. In the course of the study, the control actions on the parameters of the actuarial basis are analyzed, which allow reducing the level of the emerging budget deficit of the Russian Federation Pension Fund in the long-term period.

Key words: control of the pension system, actuarial basis, forecast of the level of the budget balance of the Pension Fund of the Russian Federation.

Статья представлена к публикации членом редакционной коллегии Р.М. Нижегородцевым.

*Поступила в редакцию 19.09.2017.
Опубликована 31.05.2018.*

УДК 004.942 + 004.272.43

ББК 32.973.202

**ПОДХОД К ОРГАНИЗАЦИИ ОТКАЗОУСТОЙЧИВЫХ
ЦИФРОВЫХ СИСТЕМ
СО СТРУКТУРОЙ МИНИМАЛЬНОГО
КВАЗИПОЛНОГО ГРАФА С ДВУМЯ ПУТЯМИ
МЕЖДУ ДВУМЯ АБОНЕНТАМИ
(НА ПРИМЕРЕ ГРАФА РАЗМЕРА 11×11)**

Ведешенков В. А.¹,

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

Разработанный подход к организации отказоустойчивых анализируемых цифровых систем (ЦС) заключается в поочередном устранении влияния неисправного компонента различных фрагментов ЦС. Анализируемый фрагмент ЦС состоит из проверяемого абонента, коммутатора и двух линий связи коммутатора с проверяющим и проверяемым абонентами. Для каждого фрагмента строятся программные базовые модули трех видов. Устранение влияния неисправного компонента проводится замещением функций неисправного компонента функциями соответствующих исправных компонентов. Представленный подход гарантирует устранение влияния одного неисправного компонента любого типа ЦС анализируемой структуры. Необходимым условием реализации предлагаемого подхода является наличие двух сопряженных путей между каждой парой абонентов ЦС анализируемой структуры.

¹ **Виктор Алексеевич Ведешенков**, доктор технических наук, старший научный сотрудник (vva@ipu.ru, Москва, ул. Профсоюзная, д. 65, тел. (495) 334-75-90).

Ключевые слова: цифровая система, минимальный квазиполный граф, отказоустойчивость, фрагмент, абонент, базовый модуль.

1. Введение

В настоящей статье под цифровыми системами (ЦС) понимаются модели многомашинных или многопроцессорных вычислительных систем, отражающие необходимые диагностические свойства и параметры анализируемых вычислительных систем.

Минимальный квазиполный граф образуется на основе однородного двудольного графа, одну долю которого составляют коммутаторы $m \times m$, а другую – m -портовые абоненты. Значение m выбирается минимальным, при котором любые два узла в одной доле связаны σ путями длины 2 через разные узлы в другой доле. В одной доле имеется n коммутаторов, а в другой – n абонентов. Каждый такой путь проходит через один коммутатор, и разные пути проходят через разные коммутаторы. Для рассматриваемых топологий параметры n и m связаны соотношением $n = m(m - 1)/\sigma + 1$ и не могут быть взяты произвольно [6, 7].

В числе возможных областей применения графов с подобной новой структурой называют отказоустойчивые многомашинные вычислительные системы (МВС) реального времени, где, например, подмножество вершин одной доли представляет совокупность процессорных элементов или вычислительных машин, а подмножество вершин другой доли – коммутаторы [8, 9].

В задачах диагностирования неисправных компонентов цифровых систем со структурой минимального квазиполного графа предполагалось, что последующее восстановление работоспособности ЦС проводится с участием обслуживающего персонала [1–4]. Между тем для обеспечения отказоустойчивости ЦС должна иметь программно-аппаратные средства автоматического восстановления работоспособности после отказа компонентов системы. Поскольку физический ремонт неисправных компонентов в ЦС анализируемой

структуры не может производиться, то в дальнейшем будем говорить об устранении влияния неисправных компонентов с помощью аппаратно-программной реконфигурации ЦС.

Методы построения отказоустойчивых систем существенно зависят от структуры и параметров объекта разработки. В связи с этим появление новых структур вычислительных систем требует безусловной коррекции или доработки известных методов. Трудности построения отказоустойчивых систем существенно возрастают с увеличением числа устраняемых отказавших компонентов. В связи с этим для построения отказоустойчивых ЦС представляется полезным фрагментный подход, в котором необходимые программные процедуры выполняются над частью (фрагментом) всей ЦС [4, 5]. Вследствие одинаковой структуры и состава фрагментов методы и средства обеспечения отказоустойчивости, разработанные для одного фрагмента, можно «настроить» для устранения влияния неисправных компонентов в других фрагментах.

2. Постановка задачи

Исходная ЦС со структурой минимального квазиполного графа $S_{n,m,\sigma}$ представлена таблицей связей, в которой n вершин одной доли представляют абоненты (процессоры), n вершин другой доли – коммутаторы, причем каждый абонент связан с каждым другим абонентом двумя линиями связи, проходящими через два разных коммутатора.

Будем считать, что допускаются устойчивые отказы компонентов: абонентов, коммутаторов и линий связи, причем неисправности компонентов таковы, что прекращают работу тех компонентов, в которых они возникли, и не влияют на работоспособность других компонентов, не имеющих связей с данным неисправным компонентом.

Для обеспечения отказоустойчивости в ЦС выделены резервные компоненты.

Требуется разработать подход к обеспечению отказоустойчивости ЦС без изменения исходной структуры минимального квазиполного графа с двумя путями между двумя

абонентами (на примере графа с параметрами $n = 11$, $m = 5$, $\sigma = 2$).

3. Подход к обеспечению отказоустойчивости ЦС со структурой минимального квазиполного графа с двумя путями между двумя абонентами (на примере графа размера 11×11)

Традиционный способ обеспечения отказоустойчивости вычислительных систем включает следующие этапы:

- контроль работоспособности системы и обнаружение факта ее отказа;
- диагностирование технического состояния компонентов с целью идентификации неисправных компонентов;
- замещение (ремонт) неисправных компонентов однотипными исправными компонентами;
- возобновление работы восстановленной системы.

Предлагаемый подход к обеспечению отказоустойчивости ЦС анализируемой структуры реализует эти этапы с учетом структурных особенностей ЦС и заключается в следующем.

С помощью контролирующего теста поочередно тестируется техническое состояние компонентов одного фрагмента ЦС: работоспособен – не работоспособен. В зависимости от результата тестирования проверяющий абонент данного фрагмента выбирает направление последующих действий:

- продолжить контроль ЦС – при успешном результате тестирования или
- перейти к восстановлению работоспособности данного фрагмента – при неудачном результате тестирования. Для этого проверяющий абонент выполняет диагностическую проверку и идентифицирует компонент или группу компонентов, подозреваемых в неисправности. Затем выполняются подпрограммы реконфигурации, замещающие исполнение функций неисправного компонента их исполнением соответствующими исправными компонентами. Такое замещение позволяет восстановить работоспособность ЦС и перейти к тестированию следующего фрагмента.

Реализация такого подхода основана на следующих положениях.

1. Для сохранения исходной структуры ЦС при разработке отказоустойчивой ЦС необходимы специализированные программные средства.

2. Для упрощения процедур обеспечения отказоустойчивости будем считать, что отказы компонентов являются достаточно редкими, чтобы интервалы между моментами возникновения двух соседних (по времени) отказов позволяли устранить влияние более ранних отказов.

3. Предположение об одном неисправном компоненте (в каждый момент времени) позволяет считать, что исправны все компоненты ЦС, кроме одного из тех, которые входят в состав фрагмента ЦС, отказ которого обнаружен контролирующим тестом. Это означает, что действия (операции), выполняемые этими компонентами ЦС, являются правильными, а их результаты достоверны. При разработке процедур устранения влияния последующих отказов компонентов нужно учитывать наличие ранее отказавших компонентов.

4. Неисправный компонент K_x не влияет на работоспособность других компонентов ЦС, а потому техническое состояние компонента K_x (работоспособен – не работоспособен) можно определить только по результату той проверки, в выполнении которой участвует компонент K_x .

5. Будем считать, что анализируемый фрагмент F_j включает абонент $A_{j,2}$, коммутатор $C_{j,1}$ и линии связи $l(a_{j,1}-c_{j,1})$, $l(c_{j,1}-a_{j,2})$. Назовем проверкой элементарную тестовую операцию над компонентами фрагмента F_j , выполняемую следующим образом.

Исправный проверяющий абонент $A_{j,1}$ передает абоненту $A_{j,2}$ через коммутатор $C_{j,1}$ и соответствующие линии связи запрос (проверку) $P_{j,1}$. Абонент $A_{j,2}$, получив такое сообщение, подтверждает его получение абоненту $A_{j,1}$, который запоминает такой результат как оценку $r_{j,1} = 0$. Если в течение интервала Δt такого подтверждения от абонента $A_{j,2}$ не приходит, то абонент $A_{j,1}$ запоминает этот результат как оценку $r_{j,1} = 1$. Эти оценки используются абонентом $A_{j,1}$ для выбора последующих тестовых процедур. (В последующем тексте будут использоваться

проверки четырех типов разного назначения, но обязательной функцией этих проверок является исполнение проверки в указанном выше смысле.)

6. Каждый абонент рассматриваемой ЦС связан с каждым другим абонентом двумя линиями связи, проходящими через два разных коммутатора.

Будем называть сопряженными пути, проходящие между абонентами A_1 и A_2 в одном и том же направлении, но через разные коммутаторы C_1 и C_2 . Коммутаторы C_1 и C_2 также будем называть сопряженными относительно абонентов A_1 и A_2 .

7. Из-за невозможности физического ремонта неисправных компонентов для восстановления работоспособности ЦС, нарушенной отказом компонента K_x , исполнение его функций нужно передать другим исправным компонентам и таким образом *устранить влияние неисправного компонента K_x на работоспособность ЦС.*

8. Для тестирования и устранения влияния неисправного компонента фрагмента F_j будем использовать базовые программные модули $B_{j,1}$, $B_{j,2}$, $B_{j,3}$ (в дальнейшем, просто базовые модули $B_{j,1}$, $B_{j,2}$, $B_{j,3}$), состоящие из проверок $P_{j,1}$, $P_{j,2}$, $P_{j,3}$, $P_{j,4}$, $P_{j,5}$ следующего функционального назначения:

- первая проверка $P_{j,1}$ обнаруживает неисправный компонент в составе фрагмента F_j ;

- вторая проверка $P_{j,2}$ диагностирует техническое состояние фрагмента F_j и идентифицирует один или несколько компонентов, подозреваемых в неисправности;

- третья проверка $P_{j,3}$ – вход в (и выход из) программы восстановления функций, нарушенных первым отказом абонента $A_{j,2}$;

- четвертая проверка $P_{j,4}$ восстанавливает функции, нарушенные отказом второго компонента фрагмента F_j ;

- пятая проверка $P_{j,5}$ является сопряженной для проверки $P_{j,3}$.

9. Будем считать входом базового модуля B_j ту проверку P_j , с которой начинается исполнение проверок этого модуля.

Совокупность базовых модулей, построенных для тестирования технического состояния всех фрагментов ЦС, позволяет устранить влияние нескольких неисправных

компонентов ЦС. Точное число компонентов, чьи неисправности можно устранить с использованием данного подхода, зависит от ряда факторов, и прежде всего от взаимного расположения неисправных компонентов.

Рассмотрим подробнее процедуры, выполняемые при устранении влияния неисправных компонентов в ЦС с параметрами $n = 11$, $m = 5$, $\sigma = 2$.

4. Состав контролирующего теста для ЦС со структурой минимального квазиполного графа с двумя путями между двумя абонентами (с параметрами графа $n = 11$, $m = 5$, $\sigma = 2$)

Структура связей анализируемой ЦС показана в таблице 1. Строки и столбцы таблицы 1 озаглавлены номерами абонентов $A_{1,j}$, $A_{2,k}$, $j, k = 1, \dots, 11$. В клетке на пересечении j -й строки и k -го столбца записаны номера коммутаторов $C_{1,s}$ и $C_{2,t}$ ($s, t = 12, \dots, 22$), через которые проходят сопряженные пути для данной пары абонентов. Так как связи между абонентами двусторонние, то номера коммутаторов $C_{1,j}$ и $C_{2,k}$ связывающих абонента j -й строки с абонентом k -го столбца, совпадают с номерами $C_{1,k}$ и $C_{2,j}$, связывающих абонента k -й строки с абонентом j -го столбца. Например, в клетках 3, 7 и 7, 3 записаны номера коммутаторов 14 и 18, являющихся сопряженными для абонентов 3 и 7. Поскольку каждую пару абонентов связывает «индивидуальная» пара путей, то такая пара коммутаторов записана только в этих двух клетках. Знаки \emptyset в диагональных клетках отражают тот факт, что абонент не взаимодействует сам с собой.

Все процедуры по устранению влияния неисправных компонентов в отказоустойчивой ЦС основаны на ее контролирующем тесте. Назначение контролирующего теста – протестировать техническое состояние каждого компонента системы хотя бы один раз.

В соответствии с записями таблицы 1 в состав рассматриваемой ЦС входят 11 абонентов с номерами 1, ..., 11, включая резервный абонент 6, 11 коммутаторов с номерами

12, ..., 22 и 110 линий связи между указанными абонентами и коммутаторами, причем половина из них являются резервными.

В таблице 2 показан вариант контролирующего теста для данной ЦС, включающий проверки $P_{j,1}$, также приведены проверки $P_{j,2}$, сопряженные с проверками $P_{j,1}$. Все проверки таблицы 2 условно разделены на 11 групп, в каждой из которых проверяющий абонент является общим (имеет один и тот же номер). Количество проверок в одной группе равно значению m – числу портов у абонента и коммутатора.

Таблица 2 построена для $m = 5$. Таким образом, таблица контролирующего теста (таблица 2) содержит $R = n \times m = 55$ проверок типа $P_{j,1}$ и 55 проверок типа $P_{j,2}$.

В таблице 2 приняты следующие обозначения:

– проверки $P_{j,1}$ входят в состав контролирующего теста для начальной конфигурации ЦС, включающей исправные компоненты;

– проверки $P_{j,2}$ являются сопряженными для соответствующих проверок $P_{j,1}$ (выделены курсивом, не входят в контролирующий тест для начальной конфигурации);

– полужирным шрифтом выделены проверки, исполняемые резервным абонентом 6 (они также не входят в контролирующий тест для начальной конфигурации). Таким образом, начальная (рабочая) конфигурация включает 10 абонентов, 11 коммутаторов и 45 линий связи.

При построении конкретных проверок $P_{j,1}$ для каждого проверяющего абонента $A_{1,j}$ в таблице связей (таблица 1) нужно найти номера коммутаторов $C_{j,i}$, к которым присоединен абонент $A_{1,j}$ например, абонент 3 присоединен к коммутаторам 14, 15, 16, 18, 21, абонент 5 – к коммутаторам 12, 16, 17, 18, 20 и т.д. Таким образом, для любой проверки проверяющий абонент, коммутатор и линия связи между ними определяются однозначно с помощью записей таблицы 1.

Второй (проверяемый) абонент $A_{2,k}$ для каждой проверки также выбирается из таблицы связей: он должен быть присоединен к коммутатору этой же проверки. Ограничения на подбор таковы: все вторые абоненты, тестируемые одним и тем же проверяющим абонентом, должны быть разными, а их общее количество должно быть равно $m = 5$. Подбор вторых абонентов

для проверок контролирующего теста имеет переборный характер, но позволяет получить приемлемое решение.

Таблица 1. Таблица связей ЦС размера 11×11

$A_{1,j} \backslash A_{2,k}$	1	2	3	4	5	6	7	8	9	10	11
1	∅	13, 14	14, 16	16, 19	12, 16	13, 19	14, 19	12, 19	13, 16	12, 14	12, 13
2	13, 14	∅	14, 15	15, 17	17, 20	13, 17	14, 20	15, 20	13, 20	14, 17	13, 15
3	14, 16	14, 15	∅	15, 16	16, 18	18, 21	14, 18	15, 21	16, 21	14, 21	15, 18
4	16, 19	15, 17	15, 16	∅	16, 17	17, 19	19, 22	15, 19	16, 22	17, 22	15, 22
5	12, 16	17, 20	16, 18	16, 17	∅	17, 18	18, 20	12, 20	16, 20	12, 17	12, 18
6	13, 19	13, 17	18, 21	17, 19	17, 18	∅	18, 19	19, 21	13, 21	17, 21	13, 18
7	14, 19	14, 20	14, 18	19, 22	18, 20	18, 19	∅	19, 20	20, 22	14, 22	18, 22
8	12, 19	15, 20	15, 21	15, 19	12, 20	19, 21	19, 20	∅	20, 21	12, 21	12, 15
9	13, 16	13, 20	16, 21	16, 22	16, 20	13, 21	20, 22	20, 21	∅	21, 22	13, 22
10	12, 14	14, 17	14, 21	17, 22	12, 17	17, 21	14, 22	12, 21	21, 22	∅	12, 22
11	12, 13	13, 15	15, 18	15, 22	12, 18	13, 18	18, 22	12, 15	13, 22	12, 22	∅

Одно из ограничений, которое следует учитывать при выполнении этого перебора: для любой пары абонентов в клетке таблицы 1 записано два номера коммутаторов. Один из этих коммутаторов должен входить в состав проверки $P_{j,1}$, а второй коммутатор – в состав проверки $P_{j,2}$, используемой для замещения соответствующих неисправных линий связи и коммутаторов. Следствием такого ограничения является то, что каждая линия связи будет тестироваться только в одной проверке типа $P_{j,1}$ или типа $P_{j,2}$, при этом и все проверки будут различными.

Таблица 2. Состав контролирующего теста для ЦС размера 11×11

B_j	Проверки $P_{j,1} = A_j \rightarrow C_{i,1} \rightarrow A_k$ $P_{j,2} = A_j \rightarrow C_{i,2} \rightarrow A_k$	B_j	Проверки $P_{j,1} = A_j \rightarrow C_{i,1} \rightarrow A_k$ $P_{j,2} = A_j \rightarrow C_{i,2} \rightarrow A_k$
1	$P_{1,1} = 3 \rightarrow 14 \rightarrow 7$	31	$P_{31,1} = 7 \rightarrow 14 \rightarrow 1$
2	$P_{1,2} = 3 \rightarrow 18 \rightarrow 7$	32	$P_{31,2} = 7 \rightarrow 19 \rightarrow 1$
3	$P_{2,1} = 3 \rightarrow 15 \rightarrow 2$	33	$P_{32,1} = 7 \rightarrow \mathbf{18} \rightarrow \mathbf{6}$
4	$P_{2,2} = 3 \rightarrow 14 \rightarrow 2$	34	$P_{32,2} = 7 \rightarrow \mathbf{19} \rightarrow \mathbf{6}$
5	$P_{3,1} = 3 \rightarrow 16 \rightarrow 1$	35	$P_{33,1} = 7 \rightarrow 19 \rightarrow 4$
6	$P_{3,2} = 3 \rightarrow 14 \rightarrow 1$		$P_{33,2} = 7 \rightarrow 22 \rightarrow 4$
7	$P_{4,1} = 3 \rightarrow 18 \rightarrow 5$		$P_{34,1} = 7 \rightarrow 20 \rightarrow 9$
8	$P_{4,2} = 3 \rightarrow 16 \rightarrow 5$		$P_{34,2} = 7 \rightarrow 22 \rightarrow 9$
9	$P_{5,1} = \mathbf{3} \rightarrow \mathbf{21} \rightarrow \mathbf{6}$		$P_{35,1} = 7 \rightarrow 22 \rightarrow 10$
10	$P_{5,2} = \mathbf{3} \rightarrow \mathbf{18} \rightarrow \mathbf{6}$		$P_{35,2} = 7 \rightarrow 14 \rightarrow 10$
6	$P_{6,1} = 5 \rightarrow 12 \rightarrow 1$	36	$P_{36,1} = 8 \rightarrow 12 \rightarrow 10$
7	$P_{6,2} = 5 \rightarrow 16 \rightarrow 1$	37	$P_{36,2} = 8 \rightarrow 21 \rightarrow 10$
8	$P_{7,1} = 5 \rightarrow 16 \rightarrow 3$	38	$P_{37,1} = 8 \rightarrow 15 \rightarrow 11$
9	$P_{7,2} = 5 \rightarrow 18 \rightarrow 3$	39	$P_{37,2} = 8 \rightarrow 12 \rightarrow 11$
10	$P_{8,1} = 5 \rightarrow 17 \rightarrow 2$	40	$P_{38,1} = 8 \rightarrow 19 \rightarrow 7$
11	$P_{8,2} = 5 \rightarrow 20 \rightarrow 2$		$P_{38,2} = 8 \rightarrow 20 \rightarrow 7$
12	$P_{9,1} = 5 \rightarrow 18 \rightarrow 11$		$P_{39,1} = 8 \rightarrow 20 \rightarrow 2$
13	$P_{9,2} = 5 \rightarrow 12 \rightarrow 11$		$P_{39,2} = 8 \rightarrow 15 \rightarrow 2$
14	$P_{10,1} = 5 \rightarrow 20 \rightarrow 8$		$P_{40,1} = 8 \rightarrow 21 \rightarrow 9$
15	$P_{10,2} = 5 \rightarrow 12 \rightarrow 8$		$P_{40,2} = 8 \rightarrow 20 \rightarrow 9$
11	$P_{11,1} = 1 \rightarrow 12 \rightarrow 11$	41	$P_{41,1} = 9 \rightarrow 13 \rightarrow 1$
12	$P_{11,2} = 1 \rightarrow 13 \rightarrow 11$	42	$P_{41,2} = 9 \rightarrow 16 \rightarrow 1$
13	$P_{12,1} = \mathbf{1} \rightarrow \mathbf{13} \rightarrow \mathbf{6}$	43	$P_{42,1} = 9 \rightarrow 16 \rightarrow 5$
14	$P_{12,2} = \mathbf{1} \rightarrow \mathbf{19} \rightarrow \mathbf{6}$	44	$P_{42,2} = 9 \rightarrow 20 \rightarrow 5$
15	$P_{13,1} = 1 \rightarrow 14 \rightarrow 10$	45	$P_{43,1} = 9 \rightarrow 20 \rightarrow 2$
16	$P_{13,2} = 1 \rightarrow 12 \rightarrow 10$		$P_{43,2} = 9 \rightarrow 13 \rightarrow 2$
17	$P_{14,1} = 1 \rightarrow 16 \rightarrow 4$	46	$P_{44,1} = 9 \rightarrow 21 \rightarrow 10$
18	$P_{14,2} = 1 \rightarrow 19 \rightarrow 4$	47	$P_{44,2} = 9 \rightarrow 22 \rightarrow 10$
19	$P_{15,1} = 1 \rightarrow 19 \rightarrow 8$	48	$P_{45,1} = 9 \rightarrow 22 \rightarrow 4$
20	$P_{15,2} = 1 \rightarrow 12 \rightarrow 8$	49	$P_{45,2} = 9 \rightarrow 16 \rightarrow 4$
16	$P_{16,1} = 2 \rightarrow 13 \rightarrow 9$	50	$P_{46,1} = 10 \rightarrow 12 \rightarrow 11$
17	$P_{16,2} = 2 \rightarrow 20 \rightarrow 9$		$P_{46,2} = 10 \rightarrow 22 \rightarrow 11$
18	$P_{17,1} = 2 \rightarrow 14 \rightarrow 3$		$P_{47,1} = 10 \rightarrow 14 \rightarrow 7$
19	$P_{17,2} = 2 \rightarrow 15 \rightarrow 3$		$P_{47,2} = 10 \rightarrow 22 \rightarrow 7$
20	$P_{18,1} = 2 \rightarrow 15 \rightarrow 4$		$P_{48,1} = \mathbf{10} \rightarrow \mathbf{17} \rightarrow \mathbf{6}$
	$P_{18,2} = 2 \rightarrow 17 \rightarrow 4$		$P_{48,2} = \mathbf{10} \rightarrow \mathbf{21} \rightarrow \mathbf{6}$
	$P_{19,1} = 2 \rightarrow 17 \rightarrow 10$		$P_{49,1} = 10 \rightarrow 21 \rightarrow 3$
	$P_{19,2} = 2 \rightarrow 14 \rightarrow 10$		$P_{49,2} = 10 \rightarrow 14 \rightarrow 3$
	$P_{20,1} = 2 \rightarrow 20 \rightarrow 5$		$P_{50,1} = 10 \rightarrow 22 \rightarrow 9$
	$P_{20,2} = 2 \rightarrow 17 \rightarrow 5$		$P_{50,2} = 10 \rightarrow 21 \rightarrow 9$

Таблица 2 (продолжение).

21	$P_{21,1} = 4 \rightarrow 15 \rightarrow 8$ $P_{21,2} = 4 \rightarrow 19 \rightarrow 8$	51	$P_{51,1} = 11 \rightarrow 12 \rightarrow 8$ $P_{51,2} = 11 \rightarrow 15 \rightarrow 8$
22	$P_{22,1} = 4 \rightarrow 16 \rightarrow 9$ $P_{22,2} = 4 \rightarrow 22 \rightarrow 9$	52	$P_{52,1} = 11 \rightarrow 13 \rightarrow 2$ $P_{52,2} = 11 \rightarrow 15 \rightarrow 2$
23	$P_{23,1} = 4 \rightarrow 17 \rightarrow 5$ $P_{23,2} = 4 \rightarrow 16 \rightarrow 5$	53	$P_{53,1} = 11 \rightarrow 15 \rightarrow 3$ $P_{53,2} = 11 \rightarrow 16 \rightarrow 3$
24	$P_{24,1} = 4 \rightarrow 19 \rightarrow 6$ $P_{24,2} = 4 \rightarrow 17 \rightarrow 6$	54	$P_{54,1} = 11 \rightarrow 18 \rightarrow 7$ $P_{54,2} = 11 \rightarrow 22 \rightarrow 7$
25	$P_{25,1} = 4 \rightarrow 22 \rightarrow 7$ $P_{25,2} = 4 \rightarrow 19 \rightarrow 7$	55	$P_{55,1} = 11 \rightarrow 22 \rightarrow 4$ $P_{55,2} = 11 \rightarrow 15 \rightarrow 4$
26	$P_{26,1} = 6 \rightarrow 13 \rightarrow 11$ $P_{26,2} = 6 \rightarrow 18 \rightarrow 11$		
27	$P_{27,1} = 6 \rightarrow 17 \rightarrow 5$ $P_{27,2} = 6 \rightarrow 18 \rightarrow 5$		
28	$P_{28,1} = 6 \rightarrow 18 \rightarrow 3$ $P_{28,2} = 6 \rightarrow 21 \rightarrow 3$		
29	$P_{29,1} = 6 \rightarrow 19 \rightarrow 1$ $P_{29,2} = 6 \rightarrow 13 \rightarrow 1$		
30	$P_{30,1} = 6 \rightarrow 21 \rightarrow 8$ $P_{30,2} = 6 \rightarrow 19 \rightarrow 8$		

Достоверность результатов выполнения проверок гарантируется исправностью проверяющих абонентов $A_{j,1}$. Чтобы удовлетворить этому требованию, будем назначать проверку $P_{j,1}$ к выполнению только в том случае, если исправность проверяющего абонента $A_{j,1}$ подтверждена в ранее выполненных проверках.

Исключение составляет проверка $P_{1,1}$, для которой проверяющим является абонент 3: при его неисправности система контроля остановится. Чтобы обойти влияние его неисправности, нужно запустить контролирующий тест с другого начального абонента, например, с абонента 5. В предположении, что в ЦС есть только один неисправный компонент, такая смена начального абонента позволит гарантированно запустить контролирующий тест и довести процесс до идентификации неисправного компонента, в том числе, и абонента 3.

5. Состав базового модуля $B_{1,1}$ для фрагмента исправной ЦС с параметрами графа $n = 11$, $t = 5$, $\sigma = 2$

В таблице 3 показаны компоненты, входящие в состав базового модуля $B_{1,1}$ (одного из 55 базовых модулей, необходимых для анализа исправной ЦС). В таблице 3 абоненты и коммутаторы записаны их номерами (без символов компонентов), а линии связи показаны парой номеров (через черточку) в скобках.

При описании действий компонентов базового модуля $B_{1,1}$ приняты следующие обозначения:

– компоненты с номерами 3, 7, 6 обозначают абоненты A_3 , A_7 , A_6 соответственно;

– компоненты с номерами 14, 18, 21 обозначают коммутаторы C_{14} , C_{18} , C_{21} соответственно;

– компоненты с номерами (3–14) и (14–7) обозначают линии связи $l(a_3-c_{14})$, $l(c_{14}-a_7)$ соответственно.

Таким образом, запись $P_{1,1} = 3 \rightarrow 14 \rightarrow 7$ означает следующее.

При выполнении проверки $P_{1,1}$ абонент 3 через линию связи (3–14), коммутатор 14 и линию связи (14–7) передает сообщение абоненту 7. Исправный абонент 7 отсылает ответ на запрос абоненту 3, который фиксирует в клетке столбца «Оценки» оценку $r_{1,1} = 0$, если же в течение интервала Δt ответ не придет, то оценку $r_{1,1} = 1$.

Для записи формул диагноза использован алгебрологический метод из [8], являющийся формализацией хорошо известного в технической диагностике метода пересечений. В алгебрологическом методе для отображения технического состояния неисправного компонента используется его номер в прямой форме, а исправного компонента – номер с отрицанием. Так, для оценки $r_{1,1} = 0$ в клетке столбца «Диагноз» записана конъюнкция номеров с отрицанием, означающая, что все компоненты, участвовавшие в проверке $P_{1,1}$, исправны. Для оценки $r_{1,1} = 1$ в клетке столбца «Диагноз» записана дизъюнкция номеров в прямой форме, означающая, что неисправен хотя бы один из компонентов, участвовавших в проверке $P_{1,1}$.

Таблица 3. Состав базового модуля $V_{1,1}$ для исправной ЦС размера 11×11

$V_{j,1}$	Проверки $P_{j,1}$ $P_{j,2}$ $P_{j,3}$	Проверя- емые компо- ненты	Оценки	Диагноз	Действия	Следующие проверки
1,1	$P_{1,1} =$ $= 3 \rightarrow 14 \rightarrow 7$	(3-14), 14, (14-7), 7	$r_{1,1} = 0$	$\overline{(3-14)} \cdot 14$ $\cdot \overline{(14-7)} \cdot \overline{7}$	\rightarrow	$P_{2,1}$
			$r_{1,1} = 1$	$(3-14) \vee 14$ $\vee (14-7) \vee 7$	\rightarrow	$P_{1, \text{нач } 2}$
	$P_{1,2} =$ $= 3 \rightarrow 18 \rightarrow 7$	7	$r_{1,1} = 1$ $r_{1,2} = 0$	$(3-14) \vee 14 \vee$ $(14-7)$	$Rec_{1,2,0}$	$P_{2,1}$
			$r_{1,1} = 1$ $r_{1,2} = 1$	7	\rightarrow	$P_{1,3}$
			$P_{1,3} =$ $= 3 \rightarrow 21 \rightarrow 6$	K_x	$r_{1,3} = 0$	$\overline{7 K_x}$
	$r_{1,3} = 1$	K_x			$Отказ$ $ЦС$	-

Стрелка (\rightarrow) в клетке столбца «Действия» означает переход к выполнению проверки, показанной в клетке последнего столбца. Таким образом, при оценке $r_{1,1} = 0$ выполняется переход к проверке $P_{2,1}$, а при оценке $r_{1,1} = 1$ выполняется переход к диагностической проверке $P_{1,2}$.

В проверке $P_{1,2}$ участвует коммутатор 18, сопряженный с коммутатором 14 из проверки $P_{1,1}$ для абонентов 3 и 7. Поэтому в клетке столбца «Проверяемые компоненты» для проверки $P_{1,2}$ записан только компонент 7 из числа компонентов, которые тестировались проверкой $P_{1,1}$.

Покажем подробности получения формул диагноза, показанных в строках для проверки $P_{1,2}$. Функции состояния компонентов, тестируемых на проверке $P_{1,2}$, определяются оценкой ее выполнения и имеют вид:

$$F_{1,2} (\text{при } r_{1,2} = 0) = \overline{7},$$

$$F_{1,2} (\text{при } r_{1,2} = 1) = 7.$$

Функции состояния компонентов по результатам тестирования на проверках $P_{1,1}$ и $P_{1,2}$, являются конъюнкцией функций состояния, полученных для каждой из этих проверок.

В результате получим формулы, показанные в клетках «Диагноз» для проверки $P_{1,2}$:

$$(1) \quad F_{1,1} \text{ (при } r_{1,1} = 1) \wedge F_{1,2} \text{ (при } r_{1,2} = 0) = [(3-14) \vee 14 \vee \\ \vee (14-7) \vee 7] \wedge \bar{7} = [(3-14) \vee 14 \vee (14-7)],$$

$$F_{1,1} \text{ (при } r_{1,1} = 1) \wedge F_{1,2} \text{ (при } r_{1,2} = 1) = [(3-14) \vee 14 \vee \\ \vee (14-7) \vee 7] \wedge 7 = 7.$$

Поочередное выполнение проверок $P_{1,1}$ и $P_{1,2}$ позволяет различить неисправность абонента 7 от неисправностей коммутатора 14 и прилегающих линий связи (3–14) и (14–7). Кроме того, подключение проверки $P_{1,2}$ вместо проверки $P_{1,1}$ устраняет влияние неисправного компонента (3–14), 14 или (14–7), обнаруженного проверкой $P_{1,1}$. Таким образом, необходимая реконфигурация ЦС, обозначенная сокращением $Rec_{1,2,0}$, («реконфигурация после проверки $P_{1,2}$ с оценкой $r_{1,2} = 0$ ») состоит в замене $P_{1,1}$ на $P_{1,2}$. После получения оценки $r_{1,2} = 0$ можно продолжать контролирующий тест ЦС с проверки $P_{2,1}$, что показано в последней клетке этой строки.

Как видно из формул (1), для пары оценок $r_{1,1} = 1$ и $r_{1,2} = 1$ формируется диагноз: подозревается в неисправности абонент 7. Для устранения влияния неисправного абонента 7 нужно заменить проверки с участием абонента 7 на соответствующие проверки, исполняемые резервным абонентом 6, показанные в таблице 4 раздела 5.3. Действия, проводимые для таких замещений, обозначены в клетке таблицы 3 сокращением $Rec_{1,2,1}$ («реконфигурация после проверки $P_{1,2}$ с оценкой $r_{1,2} = 1$ »). Выходом из программы этих действий является проверка $P_{1,3} = 3 \rightarrow 21 \rightarrow 6$. После ее выполнения с оценкой $r_{1,3} = 0$ можно продолжить контролирующий тест ЦС с проверки $P_{2,1}$ базового модуля B_2 , что показано в последней клетке этой строки.

В двух строках проверки $P_{1,3}$ показан символ компонента K_x , который в неисправном состоянии может нарушить выполнение проверки $P_{1,3}$ и привести к оценке $r_{1,3} = 1$. Так как анализ функционирования ЦС при одновременном отказе двух компонентов не входил в постановку задачи, то такой диагноз отмечен словом «Отказ ЦС». При отсутствии второго неисправного компонента K_x получим оценку $r_{1,3} = 0$ и переход

к реконфигурации ЦС для устранения влияния неисправного абонента 7 (отмечен символом $Rec_{1,3,0}$) с последующим возвратом на проверку $P_{2,1}$ контролирующего теста.

Базовый модуль $B_{1,1}$ правильно выполнит проверки $P_{1,1}$, $P_{1,2}$, $P_{1,3}$ при условии, что абонент 3 исправен. Для того чтобы устранить влияние его неисправности на оценку $r_{1,1} = 1$ (проверка $P_{1,1}$ не «прошла»), нужно запустить контролирующий тест с другого начального абонента $A_{2,нач}$, что отмечено символом $P_{1,нач} 2$ в клетке столбца «Следующие проверки». Следует отметить, что в последующих базовых модулях $B_{j,1}$, $j = 2, \dots, 55$, начальный абонент $A_{1,j}$, включая абонент 3, исправен, диагноз будет правильным, и в клетке «Следующие проверки» будет вписана проверка $P_{j+1,1}$.

Для построения состава базового модуля B_j нужно знать состав проверки $P_{j,1}$ из контролирующего теста и таблицу связей для нахождения коммутатора $C_{i,2}$, сопряженного с коммутатором $C_{i,1}$ из проверки $P_{j,1}$. Заменяя номера компонентов из проверки $P_{1,1}$ номерами компонентов из проверки $P_{j,1}$ в записях для базового модуля B_1 , получим содержимое строк для базового модуля B_j .

6. Способы устранения влияния первого отказавшего компонента фрагмента анализируемой ЦС

Для устранения влияния на работоспособность ЦС компонента, подозреваемого в неисправности, необходимо передать исполнение его функций другим исправным компонентам. Для реализации такого подхода в анализируемой ЦС предусмотрены следующие аппаратные резервы:

- абонент 6;
- 55 линий связи, сопряженных с 55 линиями связи из рабочей конфигурации и включенных в состав проверок типа $P_{j,2}$ (показаны курсивом в таблице 2);
- 10 линий связи, присоединенных к резервному абоненту 6, они будут использованы в проверках типа $P_{j,3}$ (показаны далее в таблице 5 полужирным шрифтом),

в зависимости от номера неисправного абонента в качестве резервных будут использованы те или иные линии связи.

Как показано в таблице 3, последовательное исполнение проверки $P_{j,1}$ с оценкой $r_{j,1} = 1$ и проверки $P_{j,2}$ с оценкой $r_{j,2} = 0$ диагностирует в числе подозреваемых в неисправности коммутатор 14 и линии связи (3–14), (14–7) и устраняет их влияние. Если же проверка $P_{j,2}$ исполнена с оценкой $r_{j,2} = 1$, то абонент 7 диагностируется в числе подозреваемых в неисправности. Для устранения влияния неисправного абонента 7 исполняется проверка $P_{1,3}$. Если проверка $P_{1,3}$ исполнена с оценкой $r_{1,3} = 0$, то это подтверждает правильность диагноза: неисправен абонент 7. Для устранения влияния неисправного абонента запускается подпрограмма, обозначенная символом $Rec_{1,3,0}$ (подробно рассматривается ниже), после исполнения которой управление передается входной проверке $P_{2,1}$ базового модуля $B_{2,1}$.

Для восстановления работоспособности ЦС, нарушенной отказом одного из ее компонентов, предлагается использовать следующие способы устранения влияния неисправных компонентов, учитывающие функциональные особенности компонентов ЦС.

1. Неисправные коммутатор и линия связи.

Из приведенного описания видно, что влияние неисправного коммутатора и прилегающих линий связи устраняется единообразно – исполнением проверки $P_{j,2}$ с оценкой $r_{j,2} = 0$ и последующим переходом на продолжение контролирующего теста с входной проверки $P_{j+1,1}$ следующего базового модуля $B_{j+1,1}$.

При этом следует учесть, что каждый коммутатор участвует в пяти проверках. Поэтому действия по замене «отказавших» проверок $P_{j,1}$ «исправными» проверками $P_{j,2}$ нужно выполнить в пяти базовых модулях, в состав которых входит неисправный коммутатор.

Для примера, в левой части таблицы 4 показаны необходимые проверки $P_{j,2}$ для устранения влияния неисправного коммутатора 14. Поскольку проверки $P_{j,2}$ являются сопряженными с соответствующими проверками $P_{j,1}$, то они также показаны в таблице 4. (В правой части таблицы 4

приведены проверки $P_{j,4}$, о которых подробнее будет рассказано в разделе 7.)

Таблица 4. Таблица резервных проверок для неисправного коммутатора 14

$B_{j,1}$	Проверки $P_{j,1} = A_{j,1} \rightarrow C_{i,1} \rightarrow A_{j,2}$ $P_{j,2} = A_{j,1} \rightarrow C_{i,2} \rightarrow A_{j,2}$	Проверки $P_{j,2} = A_{j,1} \rightarrow C_{i,2} \rightarrow A_{j,2}$ $P_{j,4} = A_{j,1} \rightarrow C_{i,3} \rightarrow$ $A_{i,3} \rightarrow C_{i,4} \rightarrow A_{j,2}$
$B_{1,1}$	$P_{1,1} = 3 \rightarrow 14 \rightarrow 7$ $P_{1,2} = 3 \rightarrow 18 \rightarrow 7$	$P_{1,2} = 3 \rightarrow 18 \rightarrow 7$ $P_{1,4} = 3 \rightarrow 15 \rightarrow$ $4 \rightarrow 22 \rightarrow 7$
$B_{13,1}$	$P_{13,1} = 1 \rightarrow 14 \rightarrow 10$ $P_{13,2} = 1 \rightarrow 12 \rightarrow 10$	$P_{13,2} = 1 \rightarrow 12 \rightarrow 10$ $P_{13,4} = 1 \rightarrow 13 \rightarrow$ $9 \rightarrow 21 \rightarrow 10$
$B_{17,1}$	$P_{17,1} = 2 \rightarrow 14 \rightarrow 3$ $P_{17,2} = 2 \rightarrow 15 \rightarrow 3$	$P_{17,2} = 2 \rightarrow 15 \rightarrow 3$ $P_{1,4} = 2 \rightarrow 17 \rightarrow$ $5 \rightarrow 16 \rightarrow 3$
$B_{31,1}$	$P_{31,1} = 7 \rightarrow 14 \rightarrow 1$ $P_{31,2} = 7 \rightarrow 19 \rightarrow 1$	$P_{31,2} = 7 \rightarrow 19 \rightarrow 1$ $P_{1,4} = 7 \rightarrow 20 \rightarrow$ $8 \rightarrow 12 \rightarrow 1$
$B_{47,1}$	$P_{47,1} = 10 \rightarrow 14 \rightarrow 7$ $P_{47,2} = 10 \rightarrow 22 \rightarrow 7$	$P_{47,2} = 10 \rightarrow 22 \rightarrow 7$ $P_{1,4} = 10 \rightarrow 17 \rightarrow$ $4 \rightarrow 19 \rightarrow 7$

Может показаться, что таблицы, подобные таблице 4, нужно строить для каждого коммутатора. Но в состав этих таблиц будут входить проверки $P_{j,2}$, уже «вписанные» в состав соответствующих базовых модулей. Поэтому модули $P_{j,2}$ выполняют необходимые замены автоматически, так как они проработаны заранее при построении базовых модулей $B_{j,1}$. Поэтому последовательная (в порядке выполнения контролирующего теста) замена пяти проверок $P_{j,1}$ на $P_{j,2}$ позволит устранить влияние неисправного коммутатора (в таблице 4 – коммутатора 14).

2. Неисправный абонент

Для устранения влияния неисправного абонента в ЦС зарезервирован абонент 6. Влияние неисправного абонента устраняется с помощью резервных линий связи, подключающих в работу резервный абонент 6 вместо неисправного абонента.

Физически эти линии связи существуют в структуре ЦС с момента ее построения. Задача состоит в организации подключения этих линий связи для устранения влияния того или иного неисправного абонента.

Как видно из записей таблицы 2, в полном контролирующем тесте каждый абонент, включая резервный абонент 6, входит в состав пяти проверок в качестве проверяющего (назовем «активного») абонента и пяти проверок в качестве проверяемого (назовем «пассивного») абонента.

В таблице 5 приведены индексы резервных проверок для неисправного абонента 7. (Для описания проверок в таблице 5 использованы те же обозначения, что и в таблице 3.) В левом столбце таблицы 5 показаны проверки для ситуаций, когда неисправный абонент является пассивным (проверяемым), в правом столбце – для ситуаций, когда неисправный абонент является активным (проверяющим). Для каждой ситуации приведены две проверки: $P_{j,1}$, которая взята из контролирующего теста – таблицы 2, и $P_{j,3}$ (показана полужирным шрифтом), построенная следующим образом. В проверке $P_{j,1}$ номер неисправного абонента A_n ($A_{j,1}$ или $A_{j,2}$) заменяется номером резервного абонента A_p , и для пары: исправный абонент $A_{j,1}$ и резервный абонент A_p (в левом столбце), резервный абонент A_p и исправный абонент $A_{j,2}$ (в правом столбце) по таблице связей (таблица 1) находятся номера сопряженных коммутаторов C_1 и C_2 .

В проверке $P_{j,3}$ подставляется один из этих номеров (C_1 или C_2), что автоматически определяет номера необходимых линий связи.

Аналогичные таблицы нужно построить и для каждого из остальных абонентов ЦС.

Каждая из частей таблицы 5, построенная для неисправного абонента 7, определяет состав проверок $P_{j,3}$ конфигурации ЦС, устраняющей влияние неисправного абонента 7. Переход от исходной конфигурации исправной ЦС к этой конфигурации ЦС может быть организован разными способами. Функциональная роль любого способа заключается в замене номера неисправного абонента 7 номером резервного абонента 6 во всех сообщениях (проверках), в которых ранее участвовал, будучи исправным,

абонент 7. Возможность реализации выбранного способа не вызывает сомнений, так как реконфигурация выполняется исправными компонентами ЦС.

Таблица 5. Таблица резервных проверок для неисправного абонента 7

$V_{j,1}$	Проверки $P_{j,1} = A_{j,1} \rightarrow C_{i,1} \rightarrow A_n$ $P_{j,3} = A_{j,1} \rightarrow C_{i,2} \rightarrow A_p$ для пассивного абонента 7	V_j	Проверки $P_{j,1} = A_n \rightarrow C_{i,1} \rightarrow A_{j,2}$ $P_{j,3} = A_p \rightarrow C_{i,2} \rightarrow A_{j,2}$ для активного абонента 7
$B_{1,1}$	$P_{1,1} = 3 \rightarrow 14 \rightarrow 7$ $P_{1,3} = 3 \rightarrow 21 \rightarrow 6$	B_{31}	$P_{31,1} = 7 \rightarrow 14 \rightarrow 1$ $P_{31,3} = 6 \rightarrow 13 \rightarrow 1$
$B_{25,1}$	$P_{25,1} = 4 \rightarrow 22 \rightarrow 7$ $P_{25,3} = 4 \rightarrow 17 \rightarrow 6$	$B_{33,1}$	$P_{33,1} = 7 \rightarrow 19 \rightarrow 4$ $P_{33,3} = 6 \rightarrow 17 \rightarrow 4$
$B_{38,1}$	$P_{38,1} = 8 \rightarrow 19 \rightarrow 7$ $P_{38,3} = 8 \rightarrow 19 \rightarrow 6$	$B_{34,1}$	$P_{34,1} = 7 \rightarrow 20 \rightarrow 9$ $P_{34,3} = 6 \rightarrow 21 \rightarrow 9$
$B_{47,1}$	$P_{47,1} = 10 \rightarrow 14 \rightarrow 7$ $P_{47,3} = 10 \rightarrow 17 \rightarrow 6$	$B_{35,1}$	$P_{35,1} = 7 \rightarrow 22 \rightarrow 10$ $P_{35,3} = 6 \rightarrow 17 \rightarrow 10$
$B_{54,1}$	$P_{54,1} = 11 \rightarrow 18 \rightarrow 7$ $P_{54,3} = 11 \rightarrow 13 \rightarrow 6$		

Для устранения влияния каждого неисправного абонента в памяти резервного абонента 6 необходимо иметь таблицу, построенную аналогично таблице в правом столбце таблицы 5. Таким образом, для рассматриваемого примера в памяти резервного абонента 6 нужно иметь 10 таблиц замен для абонентов 1 – 5, 7 – 11.

7. Способы устранения влияния второго отказавшего компонента анализируемой ЦС

Для устранения влияния второго отказавшего компонента построим базовые модули типа $V_{j,2}$ и $V_{j,3}$, с которыми будут работать проверяющие абоненты $A_{j,1}$ после того как в предыдущем цикле контроля и восстановления при исполнении проверки $P_{j,1}$ был зафиксирован отказ.

Базовый модуль типа $B_{j,2}$ имеет входную проверку $P_{j,2}$ и запускается в том случае, если первым был отказ коммутатора или линии связи.

Базовый модуль типа $B_{j,3}$ имеет входную проверку $P_{j,3}$ и запускается в том случае, если первым был отказ абонента.

Для реализации функционального назначения базовых модулей типа $B_{j,2}$ и $B_{j,3}$ введем проверки типа $P_{j,4}$ и $P_{j,5}$.

Проверки типа $P_{j,4}$ тестируют абонент $A_{j,2}$ через исправный абонент $A_{j,3}$ и коммутаторы $C_{j,3}$, $C_{j,4}$, отличные от коммутаторов $C_{j,1}$, $C_{j,2}$ в проверках $P_{j,1}$, $P_{j,2}$. Поэтому назовем обходящей проверкой типа $P_{j,4}$, которая позволяет устранить влияние неисправного коммутатора $C_{j,2}$ и прилегающих линий связи во входной проверке $P_{j,2}$.

Проверки типа $P_{j,5}$ являются сопряженными для проверок типа $P_{j,3}$ и выполняют функции, аналогичные функциям проверок типа $P_{j,2}$ для проверок типа $P_{j,1}$.

Действия ЦС, выполняемые после обнаружения отказа второго компонента, однозначно определяются тем, проверкой какого типа обнаружен второй отказ. А именно:

– если отказ обнаружен проверкой $P_{j,1}$, то это означает, что второй отказавший компонент никак не связан с компонентами, неисправность которых обнаружена в первом цикле контроля, а потому действия по устранению влияния второго неисправного компонента нужно проводить как для первого отказавшего компонента в соответствии с правилами, определенными для базового модуля типа $B_{j,1}$;

– если же отказ обнаружен проверками $P_{j,2}$ или $P_{j,3}$, то это означает, что второй отказавший компонент связан с компонентами, неисправность которых обнаружена в первом цикле контроля; а потому действия по устранению влияния второго неисправного компонента нужно проводить в соответствии с правилами, определенными для базовых модулей $B_{j,2}$, $B_{j,3}$.

Рассмотрим подробнее эти действия для неисправных компонентов разных типов на примере базовых модулей $B_{1,2}$ и $B_{1,3}$ (таблица 6), построенных на основе базового модуля $B_{1,1}$ (таблица 3); при этом будем рассматривать

отказовые ситуации, в которых второй неисправный компонент обнаружен проверками типа $P_{j,2}$ или $P_{j,3}$. (Отметим, что обозначения и комментарии в таблице 6 аналогичны ранее использованным в таблице 3.)

Таблица 6. Состав базовых модулей $B_{1,2}$ и $B_{1,3}$

$B_{1,2}$ $B_{1,3}$	Проверки $P_{1,2}$, $P_{1,4}$, $P_{1,3}$, $P_{1,5}$	Проверяемые компоненты	Оценки	Диагноз	Действия	Следующие проверки
1,2	$P_{1,2} =$ $= 3 \rightarrow 18 \rightarrow 7$	(3-18), 18, (18-7), 7	$r_{1,2} = 0$	$\overline{(3-18)} \cdot \overline{18} \cdot$ $\overline{(18-7)} \cdot \overline{7}$	\rightarrow	$P_{2,1}$
			$r_{1,2} = 1$	$(3-18) \vee 18 \vee$ $\vee (18-7) \vee 7$	\rightarrow	$P_{1,4}$ (отказ ЦС)
	$P_{1,4} =$ $= 3 \rightarrow 15 \rightarrow 4 \rightarrow 19 \rightarrow 7$	7	$r_{1,2} = 1$ $r_{1,4} = 0$	$(3-18) \vee 18 \vee$ $\vee (18-7)$	$Rec_{1,4,0}$	$P_{2,1}$
			$r_{1,2} = 1$ $r_{1,4} = 1$	7	\rightarrow	$P_{1,3}$
			$P_{1,3} =$ $3 \rightarrow 21 \rightarrow 6$	K_x	$r_{1,3} = 0$	7, $\overline{K_x}$
	$r_{1,3} = 1$	K_x			Отказ ЦС	-
1,3	$P_{1,3} =$ $= 3 \rightarrow 21 \rightarrow 6$	(3-21), 21, (21-6), 6	$r_{1,3} = 0$	$\overline{(3-21)}$ $\overline{21} \cdot \overline{(21-6)} \cdot \overline{6}$	\rightarrow	$P_{2,1}$
			$r_{1,3} = 1$	$(3-21) \vee 21 \vee$ $\vee (21-6) \vee 6$	\rightarrow	$P_{1,5}$
	$P_{1,5} =$ $= 3 \rightarrow 18 \rightarrow 6$	6	$r_{1,5} = 0$	$(3-21) \vee 21 \vee$ $\vee (21-6)$	\rightarrow	$P_{2,1}$ (отказ ЦС)
			$r_{1,5} = 1$	6	Отказ ЦС	-

1. Неисправная линия связи.

Предположим, что неисправна одна из линий связи (3-18) или (18-7). Тогда в процессе контроля на проверке $P_{1,2}$ из базового модуля $B_{1,2}$ будет получена оценка $r_{1,2} = 1$, для которой диагноз имеет вид:

$$Dz = (3-18) \vee 18 \vee (18-7) \vee 7,$$

и последующая рекомендация: перейти к проверке

$$P_{1,4} = 3 \rightarrow 15 \rightarrow 4 \rightarrow 19 \rightarrow 7.$$

Исполнение этой проверки с оценкой $r_{1,4} = 0$ в сочетании с оценкой предыдущей проверки $r_{1,2} = 1$ позволяет заключить, что компонент 7 исправен, а влияние одного из подозреваемых в неисправности компонентов – коммутатора 18, линии связи (3–18), линии связи (18–7) – устранено подключением проверки $P_{1,4}$ вместо проверки $P_{1,2}$.

В общем случае между абонентами 3 и 7 можно построить 8 проверок типа $P_{j,4}$, различающихся номерами промежуточного абонента $A_{j,3}$. Если наложить ограничение, чтобы проверки типа $P_{j,4}$ различались также номерами коммутаторов $C_{j,3}$, $C_{j,4}$, то останутся такие проверки:

$$P_{1,41} = 3 \rightarrow 15 \rightarrow 4 \rightarrow 19 \rightarrow 7,$$

$$P_{1,42} = 3 \rightarrow 21 \rightarrow 8 \rightarrow 20 \rightarrow 7,$$

$$P_{1,43} = 3 \rightarrow 16 \rightarrow 9 \rightarrow 22 \rightarrow 7.$$

В состав таблицы 6 нужно вставить любую из этого списка, так как по своим функциональным свойствам они равноценны.

2. Неисправный коммутатор.

Предположим, что первым неисправным компонентом был коммутатор 14. Как видно из левого столбца таблицы 4, в процессе исполнения контролирующего теста его неисправность будет обнаружена проверками типа $P_{j,1}$: $P_{1,1}$, $P_{13,1}$, $P_{17,1}$, $P_{31,1}$, $P_{47,1}$, и его влияние будет устранено проверками типа $P_{j,2}$: $P_{1,2}$, $P_{13,2}$, $P_{17,2}$, $P_{31,2}$, $P_{47,2}$, при этом абоненты $A_{j,1}$ переключат входы на соответствующие базовые модули $B_{j,2}$.

Предположим, что вторым неисправным компонентом оказался коммутатор 18. В процессе второго прогона контролирующего теста его неисправность будет обнаружена проверками типа $P_{j,1}$: $P_{4,1}$, $P_{9,1}$, $P_{28,1}$, $P_{32,1}$, $P_{54,1}$, а его влияние будет устранено проверками типа $P_{j,2}$: $P_{4,2}$, $P_{9,2}$, $P_{28,2}$, $P_{32,2}$, $P_{54,2}$, что нетрудно проверить с помощью записей таблицы 2. Кроме того, неисправность коммутатора 18 будет обнаружена проверкой $P_{1,2}$ из базового модуля $B_{1,2}$ (таблица 6). Как видно из таблицы 6, для неисправного коммутатора 18 проверка $P_{1,2}$ даст оценку $r_{1,2} = 1$, а для устранения влияния неисправного коммутатора 18 потребуется обходящая проверка $P_{1,4}$.

Действуя по аналогии с данным примером, нетрудно показать, что обходящая проверка типа $P_{j,4}$ нужна для устранения влияния кратной неисправности двух сопряженных коммутаторов, используемых в проверках типа $P_{j,1}$ и $P_{j,2}$.

Что касается несопряженных пар неисправных коммутаторов, то их влияние будет устранено двумя группами из пяти проверок, обнаруживающих каждый неисправный коммутатор как первый, построенными аналогично группе для коммутатора 14 из левого столбца таблицы 6.

3. Неисправный абонент.

В разделе 6 показано, что влияние первой неисправности компонента типа абонент устраняется замещением неисправного абонента резервным исправным абонентом и десяти проверок. Факт подобных действий фиксируется в соответствующих проверяющих абонентах переключением входа на базовый модуль $V_{j,3}$. Как видно из записей таблицы 6, последующая неисправность линии связи или коммутатора будет обнаружена проверкой $P_{1,3}$, а ее влияние будет устранено после исполнения проверки $P_{1,5}$ с оценкой $r_{1,5} = 0$. Так как для оценки $r_{1,5} = 1$ диагноз означает, что неисправен абонент 6, то в соседней клетке стоит запись *Отказ ЦС*. Такая запись объясняется тем, что для замещения неисправного абонента нужен исправный абонент; но второго резервного абонент в ЦС нет, а потому – отказ ЦС.

Из приведенного описания видно, что предложенный подход к организации устранения влияния неисправных компонентов правильно исполняет назначенные функции, если отказы компонентов обнаруживаются проверками типа $P_{j,1}$. Более сложная организация требуется для устранения влияния кратных отказов, возникающих в компонентах, исполняющих функции одного и того же базового модуля $V_{j,1}$. Поэтому в клетке «Следующие проверки» таблицы 6 имеются две записи: $P_{1,4}$ и (отказ ЦС). Объяснение таково:

– при первом входе в базовый модуль $V_{1,2}$ выполняется переход на обходящую проверку $P_{1,4}$ для устранения влияния обнаруженной неисправности;

– при втором входе в базовый модуль $V_{1,2}$ производится останов ЦС из-за сложности возникшей отказовой ситуации.

4. Обобщая результаты разделов 6 и 7, можно сделать следующие выводы.

– Представленный подход гарантирует устранение влияния одного неисправного компонента цифровых систем анализируемой структуры.

– Подобные гарантии распространяются на устранение влияния второго неисправного компонента, если он не входит в состав компонентов, исполняющих функции того же базового модуля типа $V_{j,1}$, на котором обнаружен первый неисправный компонент.

– Для устранения влияния парных отказов коммутаторов и линий связи, исполняющих функции одного и того же базового модуля $V_{j,1}$, используются обходящие проверки типа $P_{1,4}$. Обходящие проверки успешно устраняют влияние соответствующего неисправного компонента, но их исполнение требует вдвое больше времени, чем проверки остальных типов. Такое замедление может оказаться неприемлемым для организации синхронных систем.

– Как видно из приведенных результатов, необходимым условием реализации предлагаемого подхода к устранению влияния неисправных компонентов является наличие двух сопряженных путей между каждой парой абонентов ЦС анализируемой структуры, т.е. нужно иметь параметр $\sigma = 2$. В свою очередь, такое значение параметра σ определяет только одну резервную линию связи для каждой рабочей линии связи. Для устранения влияния одновременного отказа такой пары линии связи необходима обходящая проверка, вносящая замедление в процесс функционирования ЦС.

7. Заключение

Разработанный подход к организации отказоустойчивости цифровых систем со структурой минимального квазиполного графа с двумя путями между двумя абонентами заключается в последовательном устранении влияния неисправного компонента отдельных фрагментов ЦС. Анализируемый фрагмент ЦС состоит из проверяемого абонента, коммутатора

и двух линий связи коммутатора с проверяющим и проверяемым абонентами. Для каждого фрагмента строятся программные базовые модули трех видов. На основании полученного диагноза технического состояния компонентов фрагмента проводится замещение функций компонента, подозреваемого в неисправности, функциями соответствующих исправных компонентов. Такое замещение позволяет восстановить работоспособность ЦС и перейти к тестированию следующего фрагмента.

Представленный подход гарантирует устранение влияния одного неисправного компонента любого типа цифровых систем анализируемой структуры. Подобные гарантии распространяются на устранение влияния второго неисправного компонента, если он не входит в состав компонентов, исполняющих функции того же базового модуля типа $B_{j,1}$, на котором обнаружен первый неисправный компонент. Для устранения влияния парных отказов в компонентах, исполняющих функции одного и того же базового модуля $B_{j,1}$, введены обходящие проверки, чье исполнение занимает вдвое больше времени, чем проверки остальных типов. Такое замедление может оказаться неприемлемым для организации синхронных систем.

Необходимым условием реализации предлагаемого подхода к устранению влияния неисправных компонентов является наличие двух сопряженных путей между каждой парой абонентов ЦС анализируемой структуры, т.е. нужно иметь параметр $\sigma = 2$. Что же касается других параметров ЦС, то конкретные значения n и m определяют количество необходимых базовых модулей $R = 3(n \times m)$, но по существу не влияют на их организацию. Таким образом, представленный подход может быть использован для анализа ЦС со структурой минимального квазиполного графа с двумя путями между двумя абонентами с параметрами графа, отличными от размера 11×11 . Но параметры n и m должны удовлетворять соотношению $n = m(m - 1)/\sigma + 1$ [1, 2].

Литература

1. ВЕДЕШЕНКОВ В.А., КУРАКО Е.А., ЛЕБЕДЕВ В.Н. *О диагностируемости компонентов цифровых систем со структурой минимального квазиполного графа размера 7×7 с 2 путями между 2 абонентами* // Управление большими системами. – 2015. – Вып. 58. – С. 90–114.
2. ВЕДЕШЕНКОВ В.А. *Организация диагностирования одиночных неисправных компонентов цифровых систем со структурой минимального квазиполного графа размера 7×7 с 2 путями между 2 абонентами* // Проблемы управления. – 2016. – №1. – С. 65–72.
3. ВЕДЕШЕНКОВ В.А. *О диагностируемости цифровых систем со структурой минимального квазиполного графа размера 7×7* // Автоматика и телемеханика. – 2016. – №3. – С. 152–165.
4. ВЕДЕШЕНКОВ В.А. *Подход к фрагментному диагностированию компонентов цифровых систем со структурой минимального квазиполного графа (на примере графа размера 7×7)* // Проблемы управления. – 2016. – №6. – С. 53–58.
5. ВЕДЕШЕНКОВ В.А. *Фрагментарный подход к диагностированию компонентов цифровых систем со структурой минимального квазиполного графа (на примере графа размера 7×7)* // Управление большими системами. – 2017. – Вып. 66. – С. 171–197.
6. КАРАВАЙ М.Ф., ПОДЛАЗОВ В.С. *Распределенный полный коммутатор как «идеальная» системная сеть для многопроцессорных вычислительных систем* // Управление большими системами. – 2011. – Вып. 34. – С. 92–116.
7. КАРАВАЙ М.Ф., ПАРХОМЕНКО П.П., ПОДЛАЗОВ В.С. *Комбинаторные методы построения двудольных однородных минимальных квазиполных графов (симметричных блок-схем)* // Автоматика и телемеханика. – 2009. – №2. – С. 153–170.

8. ALVERSON B., FROESE E., KAPLAN L., ROWETH D. *Cray XCTM Series Network, WP-Aries01-1112*. 2012 Cray Inc. – 28 p. – URL: <http://www.cray.com/sites/default/files/resources/CrayXCNetwork.pdf> (дата обращения: 16.03.2017).
9. ALVERSON R., ROWETH D., KAPLAN L. *CRAY INC. The Gemini System Interconnect // 18th IEEE Symposium on High Performance Interconnects*. – 2009. – P. 83–87.

THE APPROACH TO AN ORGANIZATION FAULT – TOLERANT DIGITAL SYSTEMS WITH MINIMAL QUASICOMPLETE GRAPH STRUCTURES WITH TWO PATHS BETWEEN TWO ABONENTS (AN EXAMPLE OF GRAPH BY DIMENSION 11 × 11)

Victor Vedeshenkov, Institute of Control Sciences of RAS, Moscow, Doctor of Science, senior researcher associate (vva@ipu.ru).

Abstract: The developed approach to the organization of analyzable fault-tolerant digital systems consists of an alternate elimination of the effect of one faulty component of various fragments of digital systems. The analysed fragment of digital system consists of tested abonent, a commutator and two lines of commutator with tested abonent and testing abonent. The programming basic modules of three types are developed for each fragment. The elimination of influence of fault component holds the substitution of the faulty component function by non-faulty backup functions. The developed approach guarantees the elimination of influence of one fault component of any type for analysed structure of digital systems. Prerequisite for the realization of the proposed approach are two paths between each pair of abonents.

Keywords: digital systems, minimal quasicomplete graph, fault-tolerance, fragment, abonent, basic module.

Статья представлена к публикации членом редакционной коллегии В.М. Вишневым.

*Поступила в редакцию 25.06.2017.
Опубликована 31.05.2018.*

УДК 519.63
ББК 34.327

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ПРОЦЕССА НАГРЕВА СЛЯБА В МЕТОДИЧЕСКИХ ПЕЧАХ ТОЛКАТЕЛЬНОГО ТИПА

Алтыева М. Ч.¹

*(Липецкий государственный технический университет,
Липецк)*

Рассматривается процесс нагрева слэбов в методических печах. Данный процесс описывается математической моделью, которая представляет собой уравнение в частных производных параболического типа, дополняемое граничными условиями 2-го и 3-го рода. Для решения задачи определения распределения температур в слэбе в процессе нагрева используется метод конечно-разностной аппроксимации.

Ключевые слова: уравнение теплопроводности, метод конечных разностей, метод Гаусса, методическая печь, сварочная зона.

1. Введение

В современном мире металлы и сплавы имеют широкий круг использования, а расширенное производство заготовок и изделий из металлических материалов с термической обработкой позволяет повысить качество конечного продукта.

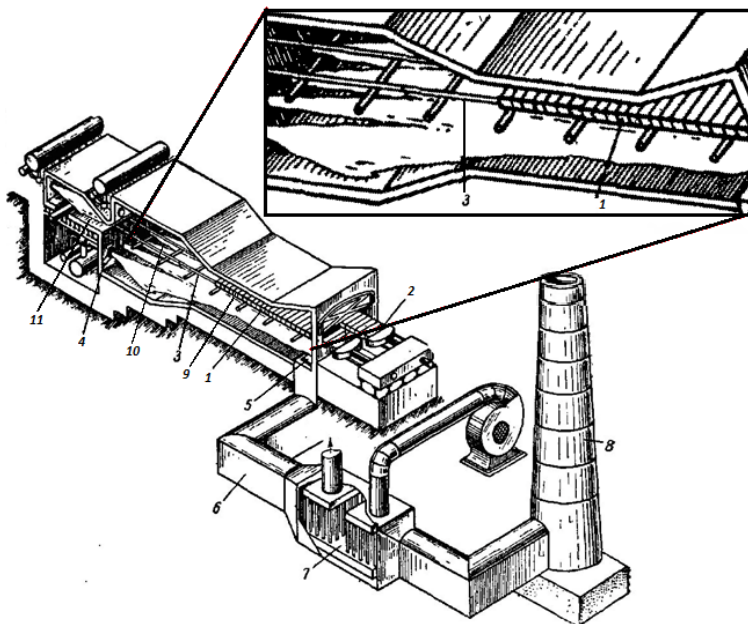
Построение математических моделей и моделирование технологических процессов играют большую роль в производстве какого-либо продукта, о чем свидетельствуют работы [1–2]. По словам авторов книги [8], в прокатном производстве широкое распространение для нагрева металла получили методические

¹ Марина Чарыевна Алтыева, магистрант ЛГТУ
(94marinka94@gmail.com).

печи. В данной работе рассматривается нагрев сляба в сварочной зоне методической печи толкательного типа.

2. Описание технологического процесса в методической печи

Из [8] возьмем описание методической печи (рис. 1). Топливо сжигается при помощи горелок, а продукты сгорания движутся вдоль рабочего пространства печи в направлении, противоположном движению металла. Металл, в свою очередь, толкателем перемещается по водоохлаждаемым (глиссажным) трубам и через окно выдачи нагретый металл попадает на рольганг и по нему к стану.



1 – металл, 2 – толкатель, 3 – водоохлаждаемая труба, 4 – горелка,
5 – дымовой канал, 6 – боров, 7 – рекуператор, 8 – дымовая труба,
9 – методическая зона, 10 – сварочная зона, 11 – томильная зона

Рис. 1. Описание методической печи

Методическая печь имеет три теплотехнические зоны, предназначенные для разделения стадий нагрева:

- **методическая зона** предназначена для утилизации тепла уходящих газов;
- **сварочная зона** предназначена для постепенного нагрева металла. Здесь металл имеет почти готовую температуру;
- **томильная зона** предназначена для выравнивания температуры по сечению металла [7].

В [3] изложена проблема методических печей, а именно то, что в сварочной зоне до высоких температур нагревается только поверхность металла; температура средних слоев меньше температуры поверхности, т.е. создается перепад температур по сечению металла, недопустимый по технологическим требованиям.

Так как длина сляба перпендикулярна глоссажным трубам (см. рис. 1), то распределение температуры в различных поперечных сечениях приблизительно одинаково, поэтому изменением температуры по ширине заготовки можно пренебречь.

Для того чтобы изготавливать качественную продукцию и выбирать оптимальные режимы достижения качества при минимальных затратах, необходимо смоделировать этот процесс.

3. Формализованная постановка задачи

3.1. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

Системы дифференциальных уравнений в частных производных лежат в основе математического описания многих физических процессов. Опираясь на работу [4], опишем математическую модель распределения температур в слябе в процессе нагрева в сварочной зоне методической печи с учетом двух пространственных измерений (x, y) . Таким образом, распределение температуры металла по длине и высоте можно определить, решая двумерное уравнение теплопроводности:

$$(1) \quad \rho C_p \frac{\partial T}{\partial t} = \frac{\partial}{\partial x} \left(\lambda \frac{\partial T}{\partial x} \right) + \frac{\partial}{\partial y} \left(\lambda \frac{\partial T}{\partial y} \right).$$

Данное уравнение дополняется начальным условием:

$$(2) \quad T(x, y, t_0) = \phi(x, y)$$

и граничными условиями:

– для верхней поверхности сляба:

$$(3) \quad \lambda \frac{\partial T}{\partial y} = \sigma \left[\left(\frac{T_s}{100} \right)^4 - \left(\frac{T_m}{100} \right)^4 \right];$$

– для боковых граней с расчетным участком:

$$(4) \quad \lambda \frac{\partial T}{\partial x} = \sigma_2 \left(\frac{T_s}{100} - \frac{T_m}{100} \right);$$

– для нижней поверхности сляба:

$$(5) \quad \lambda \frac{\partial T}{\partial y} = \frac{\alpha_1 (T_s - T_T)}{1 + \left(\frac{x}{R} \right)^2} - \alpha_2 (T_s - T_m).$$

Все параметры формул представлены в таблице 1.

Таблица 1. Параметры формул

Параметр	Пояснение
λ	коэффициент теплопроводности
x, y	координаты в пространстве
ρ	массовая плотность
C_p	удельная теплоемкость
t	время
σ, σ_2	приведенный коэффициент теплообмена излучением
T_m	температура поверхности металла
T_s	температура греющей среды
α_1, α_2	коэффициент теплопередачи
T_T	температура глиссажной трубы
R	радиус глиссажной трубы

3.2. АППРОКСИМИРОВАННАЯ МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

В [5–6] показано, что существует большое количество численных методов для решения уравнений. В данной работе используем метод сеток, который обладает довольно высокой эффективностью и имеет широкое практическое применение.

Чтобы построить расчетную схему, необходимо преобразовать уравнение (1) и граничные условия (3)–(5):

$$(6) \quad T_{j,k}^{i+1} \left(\frac{2\tau\lambda}{\rho C_p (\Delta x)^2} + \frac{2\tau\lambda}{\rho C_p (\Delta y)^2} + 1 \right) - \frac{\tau\lambda}{\rho C_p (\Delta x)^2} \cdot T_{j+1,k}^{i+1} - \\ - \frac{\tau\lambda}{\rho C_p (\Delta x)^2} T_{j-1,k}^{i+1} - \frac{\tau\lambda}{\rho C_p (\Delta y)^2} T_{j,k+1}^{i+1} - \frac{\tau\lambda}{\rho C_p (\Delta y)^2} T_{j,k-1}^{i+1} = T_{j,k}^i;$$

$$(7) \quad T_{j,k}^{i+1} - T_{j,k-1}^{i+1} = \frac{\Delta y \sigma \left[\left(\frac{T_s}{100} \right)^2 - \left(\frac{T_m}{100} \right)^2 \right]}{\lambda};$$

$$(8) \quad T_{j,k}^{i+1} - T_{j-1,k}^{i+1} = \frac{\Delta x \sigma_2 \left(\frac{T_s}{100} - \frac{T_m}{100} \right)}{\lambda};$$

$$(9) \quad T_{j,k+1}^{i+1} - T_{j,k}^{i+1} = \frac{\Delta y \alpha_1 (T_s - T_T)}{\lambda \left(1 + \left(\frac{x}{R} \right)^2 \right)} - \frac{\Delta y \alpha_2 (T_s - T_m)}{\lambda}.$$

Таким образом, из уравнений (6)–(9) получим систему линейных алгебраических уравнений (СЛАУ), которая решается методом Гаусса:

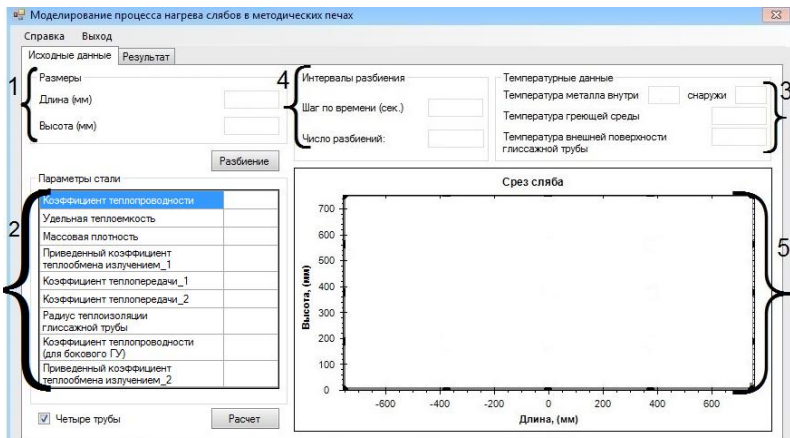
$$(10) \left\{ \begin{array}{l} T_{j,k}^{i+1} \left(\frac{2\tau\lambda}{\rho C_p (\Delta x)^2} + \frac{2\tau\lambda}{\rho C_p (\Delta y)^2} + 1 \right) - \frac{\tau\lambda}{\rho C_p (\Delta x)^2} \cdot T_{j+1,k}^{i+1} - \\ - \frac{\tau\lambda}{\rho C_p (\Delta x)^2} T_{j-1,k}^{i+1} - \frac{\tau\lambda}{\rho C_p (\Delta y)^2} T_{j,k+1}^{i+1} - \frac{\tau\lambda}{\rho C_p (\Delta y)^2} T_{j,k-1}^{i+1} = T_{j,k}^i, \\ T_{j+1,k}^{i+1} - T_{j,k}^{i+1} = \frac{\Delta x \sigma_2 \left(\frac{T_s}{100} - \frac{T_m}{100} \right)}{\lambda}, \\ T_{j,k}^{i+1} - T_{j-1,k}^{i+1} = \frac{\Delta x \sigma_2 \left(\frac{T_s}{100} - \frac{T_m}{100} \right)}{\lambda}, \\ T_{j,k+1}^{i+1} - T_{j,k}^{i+1} = \frac{\Delta y \alpha_1 (T_s - T_T)}{\lambda \left(1 + \left(\frac{x}{R} \right)^2 \right)} - \frac{\Delta y \alpha_2 (T_s - T_m)}{\lambda}, \\ T_{j,k}^{i+1} - T_{j,k-1}^{i+1} = \frac{\Delta x \sigma \left(\left(\frac{T_s}{100} \right)^4 - \left(\frac{T_m}{100} \right)^4 \right)}{\lambda}. \end{array} \right.$$

4. Исследование построенной модели нагрева сляба в методической печи

4.1. СВЕДЕНИЯ О ПРОГРАММЕ

Программа «Моделирование процесса нагрева слябов в методических печах» разработана в интегрированной среде Microsoft Visual Studio 2010 на языке программирования C#.

Данная программа предназначена для моделирования процесса нагрева слябов в методических печах толкательного типа и получения числовых и графических характеристик температурного поля среза сляба за весь период нагрева. На рис. 2 показан интерфейс программы.



1 – размерности сляба, 2 – входные параметры стали, 3 – температурные данные,
4 – данные о разбиении исходной области, 5 – дискретизованная область

Рис. 2. Интерфейс ПО

Перед началом работы заполняем области 1–4 и нажимаем кнопку «Расчет». Когда программа посчитает, то необходимо перейти во вкладку «Результат».

4.2. РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ

Для расчета введем исходные данные, представленные в таблице 2.

На рисунке 3 представлен результат работы программы, здесь показано температурное поле сляба при выходе из сварочной зоны.

По результатам вычислений ПО разница температур между нижней и верхней поверхностью составила 49.6°C , что обеспечивает необходимые условия технологического процесса, при которых перепад температур после нагрева в сварочной зоне печи должен находиться в интервале $40\text{--}50^{\circ}\text{C}$. Также здесь явно видно влияние четырех глссажных труб на перепад температур заготовки по длине.

Таблица 2. Исходные данные

Название	Ед. измерения	Значение
Длина сляба	мм	4000
Высота сляба	мм	250
Шаг по времени	с	1
Число разбиений	шт	20
Температура металла внутри	°C	980
Температура металла снаружи	°C	1000
Температура греющей среды	°C	1340
Температура внешней поверхности глиссажной трубы	°C	1270
Коэффициент теплопроводности	Вт/(м·°C)	27
Удельная теплоемкость	Дж/(кг·°C)	691
Массовая плотность	кг/м ³	7662
Приведенный коэффициент теплообмена излучением	Вт/(м ² ·K ⁴)	3,13
Коэффициент теплопередачи 1	Вт/(м ² ·K)	4,6
Коэффициент теплопередачи 2	Вт/(м ² ·K)	4,1
Радиус теплоизоляции глиссажной трубы	мм	150
Коэффициент теплопроводности (для бокового ГУ)	Вт/(м·°C)	21
Приведенный коэффициент теплообмена излучением 2	Вт/(м ² ·K ⁴)	2,73

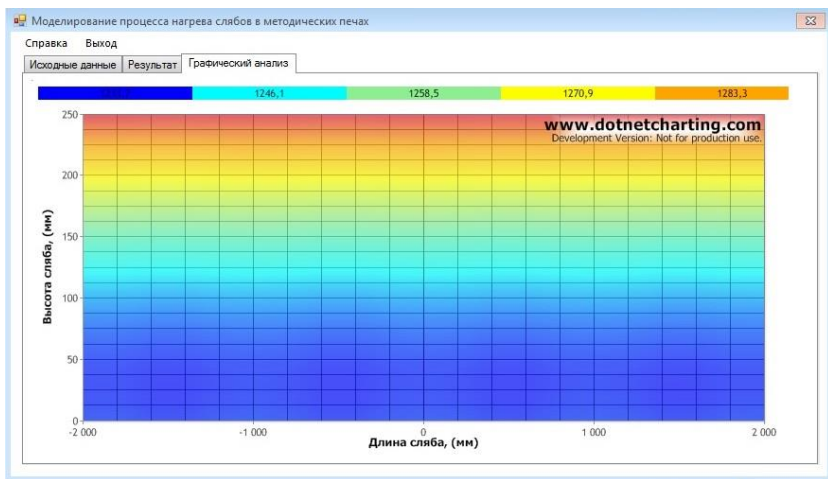


Рис. 3. Результат работы ПО

5. Выводы

На основе построенной математической модели, разработанного программного обеспечения и при известных параметрах стали имеется возможность моделировать процесс нагрева сляба в сварочной зоне методической печи.

Исходя из полученных результатов моделирования, существует возможность выбора режимов работы печи, обеспечивающих минимальный перепад температур по сечению металла.

Таким образом, с помощью предварительного расчета модель дает возможность подобрать оптимальное значение температур, позволяющее снизить термические напряжения в заготовке, тем самым повышая качество конечного продукта.

Литература

1. ГАЛКИН А.В., КУЗИНА Н.Е., ПИМЕНОВ В.А. *Математическое моделирование теплового режима процесса остывания и затвердевания слитка в УНРС* // Материалы XI Всерос. школы-конференции молодых ученых «Управление большими системами» (УБС'2014), 9-12 сентября 2014 г., Арзамас / Ин-т проблем упр. Им. В.А. Трапезникова; Арзамас. политехн. ин-т Нижегородск. гос. техн. ун-та; под общ. ред. Д.А. Новикова, П.В. Пакшина. – М.: ИПУ РАН, 2014. – С. 961–972.
2. ГАЛКИН А.В., ФИЛИППОВ Д.А., ПИМЕНОВ В.А. *Математическое моделирование напряженно-деформированного состояния сляба в УНРС* // Материалы XII Всерос. школы-конференции молодых ученых «Управление большими системами» (УБС'2015), 7-11 сентября 2015 г., Волгоград / Ин-т проблем упр. Им. В.А. Трапезникова; Волгоград. гос. ун-т; под общ. ред. Д.А. Новикова, А.А. Воронина. – М.: ИПУ РАН, 2015. – С. 600–611.
3. ГУБИНСКИЙ В.И. *Металлургические печи*. – Днепропетровск: НМетАУ, 2006. – 83 с.

4. ДРУЖИНИН Г.М., САМОЙЛОВИЧ Ю.А., ПОПОВ Е.В. *Способы снижения термической неоднородности массивных стальных слябов при нагреве в методических нагревательных печах* // Черная металлургия. – 2013. – №7. – С. 27–32.
5. МЕРКУЛОВА Н.Н., МИХАЙЛОВ М.Д. *Разностные схемы для обыкновенных дифференциальных уравнений: Учеб. пособие.* – Томск: ИДО ТГУ, 2014. – 122 с.
6. САМАРСКИЙ А.А., ГУЛИН А.В. *Численные методы: Учеб. пособие для вузов.* – М: Наука, 1989. – 432 с
7. СЕВЕРИН Т.С. *Обоснование параметров автоматизированного управления нагревом слитков в методической трехзонной печи: Автореф. магистра.* – Донецк: ДонНТУ, 2008. – 20 с.
8. ТАЙЦ Н.Ю., РОЗЕНГАРТ Ю.И. *Методические нагревательные печи.* – М: Металлургиздат, 1964. – 410 с.

MATHEMATICAL MODELING THE PROCESS OF SLAB HEATING AS COMPONENT OF PRODUCTION CONTROL

Marina Altyeva, Lipetsk State Technical University, Lipetsk, master student (94marinka94@gmail.com).

Abstract: This paper describes modeling the process of slab heating in continuous furnaces. The process is described using mathematical model consisting of parabolical type partial derivative equation with second and third type boundary conditions. The finite-difference approximation is used to solve the problem. Further, the system of linear algebraic equations is solved by the Gauss method. Because of the complexity of the solution, the software "Modeling of the process of slabs heating in methodical furnaces" was developed. Programming in the integrated environment of Microsoft Visual Studio 2010 in the programming language C#. She gives an opportunity to obtain a numerical and graphical result.

Thus, the paper presents a model for heating slabs in methodical seals, which yields an adequate result and can be applied by technologists in setting suitable operating conditions for the furnace.

Keywords: heat equation, finite-difference method, Gaussian elimination, continuous furnace, heating section.

*Статья представлена к публикации
членом редакционной коллегии Э.Ю. Калимулиной.*

Поступила в редакцию 11.05.2017.

Опубликована 31.05.2018.