ISSN 1819-3161

ПРОБЛЕМЫ УПРАВЛЕНИЯ

6/2013



РЕДКОЛЛЕГИЯ

Д-ра техн. наук Ф.Т. Алескеров, В.Н. Афанасьев, Р.Р. Бабаян, канд. техн. наук Л.П. Боровских, д-р техн. наук В.Н. Бурков, акад. РАН С.Н. Васильев, д-ра техн. наук В.А. Виттих, Б.Г. Волик, А.А. Дорофеюк, акад. РАН С.В. Емельянов, д-р техн. наук Э.Л. Ицкович, чл.-корр. РАН И.А. Каляев, д-ра техн. наук В.Ю. Кнеллер, О.С. Колосов, О.П. Кузнецов, В.В. Кульба, В.Г. Лебедев, акад. РАН В.А. Левин, д-р физ.-мат. наук Г.Г. Малинецкий, д-ра техн. наук В.Д. Малюгин, А.С. Мандель, чл.-корр. РАН Н.А. Махутов, акад. РАН Е.А. Микрин, д-р экон. наук Р.М. Нижегородцев, чл.-корр. РАН Д.А. Новиков, д-ра техн. наук В.Н. Новосельцев, К.Б. Норкин, Б.В. Павлов, чл.-корр. РАН П.П. Пархоменко, д-ра техн. наук Ф.Ф. Пащенко, С.А. Редкозубов, чл.-корр. РАН А.Ф. Резчиков, д-ра техн. наук Л.И. Розоноер (США), Е.Я. Рубинович, В.Ю. Рутковский, акад. ГАН М. Салуквадзе (Грузия), д-р техн. наук А.Э. Софиев, акад. РАН Е.А. Федосов, д-р техн. наук А.Д. Цвиркун

РУКОВОДИТЕЛИ РЕГИОНАЛЬНЫХ РЕДСОВЕТОВ

ВЛАДИВОСТОК – д-р техн. наук О.В. Абрамов, \$\alpha\$ (4232) 31-02-02 ВОЛГОГРАД – д-р физ.-мат. наук А.А. Воронин, \$\alpha\$ (8442) 40-55-06 ВОРОНЕЖ – д-р техн. наук С.А. Баркалов, \$\alpha\$ (4732) 76-40-07 КРАСНОЯРСК – д-р техн. наук В.В. Огурцов, \$\alpha\$ (3912) 27-63-89 КУРСК – д-р техн. наук С.Г. Емельянов, \$\alpha\$ (4712) 50-48-00 ЛИПЕЦК – д-р техн. наук А.К. Погодаев, \$\alpha\$ (4742) 32-80-44 МИНСК – д-р физ.-мат. наук А.В. Тузиков, \$\alpha\$ (10-37517) 284-21-40 ПЕРМЬ – д-р техн. наук В.Ю. Столбов, \$\alpha\$ (3422) 39-12-97 САМАРА – д-р техн. наук В.Г. Засканов, \$\alpha\$ (8463)35-64-07 САРАТОВ – д-р техн. наук В.А. Твердохлебов, \$\alpha\$ (8452) 22-10-42 ТАГАНРОГ – д-р. техн. наук В.М. Курейчик, \$\alpha\$ (8634) 39-32-60 ТИРАСПОЛЬ – д-р физ.-мат. наук С.И. Берилл, \$\alpha\$ (10-373553) 9-44-87 УФА – д-р техн. наук Б.Г. Ильясов, \$\alpha\$ (3472) 23-78-35



CONTROL SCIENCES

Научно-технический журнал

6 номеров в год ISSN 1819-3161

УЧРЕДИТЕЛЬ

Федеральное государственное бюджетное учреждение науки Институт проблем управления им. В.А. Трапезникова РАН

Главный редактор чл.-корр. РАН Д.А. Новиков

Заместители главного редактора

Л.П. Боровских, Ф.Ф. Пащенко

Редактор Т.А. Гладкова

Выпускающий редактор Л.В. Петракова

Издатель ООО «Сенсидат-Плюс»

Адрес редакции 117997, ГСП-7, Москва, ул. Профсоюзная, д. 65, к. 410. Тел./факс (495) 334-92-00

E-mail: pu@ipu.ru

Интернет: http://pu.mtas.ru

Оригинал-макет и электронная версия подготовлены ИП Прохоров О. В.

Фото на четвертой странице обложки В.М. Бабикова

Отпечатано в ИПУ РАН

Заказ № 115

Подписано в печать 25.11.2013 г.

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

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

Журнал входит в Перечень ведущих рецензируемых журналов и изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени доктора и кандидата наук

Подписные индексы:

80508 и 81708 в каталоге Роспечати; 38006 в объединенном каталоге «Пресса России»

Цена свободная

© Федеральное государственное бюджетное учреждение науки Институт проблем управления им. В.А. Трапезникова РАН

ПРОБЛЕМЫ УПРАВЛЕНИЯ

СОДЕРЖАНИЕ

6.2013

Математические проблемы управления
Непейвода Н.Н. Алгебраический подход к управлению
Информационные технологии в управлении
Новиков Д.А. Большие данные: от Браге — к Ньютону
Болодурина И.П., Парфенов Д.И. Управление потоками запросов при доступе к широкополосным мультимедийным образовательным ресурсам системы дистанционного обучения
Управление в социально-экономических системах
Макаров В.Л., Бахтизин А.Р., Сушко Е.Д. Компьютерное моделирование взаимодействия между муниципалитетами, регионами, органами государственного управления
Горелик В.А., Золотова Т.В. О некоторых оценках устойчивости фондового рынка и влиянии на них информированности инвесторов
Сараев П.В., Сяглова Ю.Е. Повышение эффективности валютного хеджирования на основе результатов нейроструктурного прогнозирования 48
Управление сложными технологическими процессами и производствами
Волкова И.О., Губко М.В., Сальникова Е.А. Активный потребитель: задача оптимизации потребления электроэнергии и собственной генерации 53
Зак Ю.А. Построение расписаний выполнения заданий для параллельно работающих систем машин и поточных линий
Мещеряков В.Н., Мещерякова О.В., Сараев П.В. Математическое
моделирование и управление процессами в системах автоматизации промышленных установок с помощью многозначных нейронных сетей
Хроника
Симпозиум «Меры сложности», посвященный 75-летию А.Я. Червоненкиса 76
* * *
Указатель статей, опубликованных в 2013 г
Contants & abstracts 80

УДК 519.681.2:512.579 + 004.942

АЛГЕБРАИЧЕСКИЙ ПОДХОД К УПРАВЛЕНИЮ

Н.Н. Непейвода

Рассмотрены алгебраические структуры, порождаемые программами моделирования. Намечены их связи с численным моделированием, информатикой, суперкомпьютерами, качественной теорией систем. Сформулированы задачи дальнейших исследований.

Ключевые слова: моделирование, алгебраические системы, обратимые вычисления, группоид.

ВВЕДЕНИЕ

При математическом моделировании сначала строится модель рассматриваемой системы на языке математики, а затем алгоритм и программа, реализующие ее. Полученные результаты анализируются и интерпретируются в рамках как математической модели, так и исходной системы. Традиционно под математической моделью понимается численная модель (дифференциальное уравнение или дифференциально-разностный полудискретный процесс). Если удается, то система обрабатывается точными методами (например, методами качественной теории дифференциальных уравнений). Но чаще всего алгоритмическая модель является приближением к математической (дискретизация непрерывной системы), а программа приближением к алгоритму. В итоге численные результаты требуют серьезной неформальной обработки, чтобы получить содержательно требуемые качественные выводы о поведении системы.

Дальнейшее развитие средств моделирования путем грубой силы (увеличения количества и быстродействия процессоров и памяти), который был магистральным направлением в последние 40 лет, подошло к трем барьерам. Мощность компьютеров объективно ограничивается скоростью света и тепловыделением, сложность программ — пределом Чейтина. Пока что не видно никаких средств попытаться обойти предел, который ставит скорость света. А вот два других предела можно попытаться преодолеть. Но чтобы это сделать, надо сначала разобраться в их природе.

Алгебраический подход направлен на преодоление двух барьеров: теплового и сложностного. Но развитие в данном направлении потребовало, прежде всего, пересмотра фундаментальных концепций алгебраической теории моделирования.

Алгебры программных систем начали развиваться с 1960-х гг.: Глушков [1], дальнейшее развитие этого подхода см. в работах [2, 3]; Маурер [4], далее в работах [5, 6]; алгебры функциональных программ — прекрасное и полное изложение важнейших результатов дано в работе [7]. Все эти подходы носят на себе печать одного и того же «первородного греха»: в алгебры в том или ином виде непосредственно закладываются операторы языков программирования. Тем самым вместо того, чтобы на базе теории выдать рекомендации для практики, может быть, в новом и неожиданном направлении, ставится задача изучения существующей практики.

В настоящей работе намечается направление, ортогональное традиционному: алгебраические модели на базе теории алгебраических систем [8], и уже затем интерпретация результатов в рамках программных конструкций. Заодно рассматриваются взаимосвязи этого подхода с качественной теорией систем и возможности, которые открываются для преодоления второго и третьего пределов.

1. ТЕПЛО И ЭНЕРГИЯ: ПРЕДЕЛ ЛАНДАУЭРА

Тепловой предел связан с фундаментальным термодинамическом результатом фон Неймана и Ландауэра. В соответствии с принципом Ландау-



эра минимальное выделение тепла на выработку одного бита информации

$$E_{diss} = k_R T \ln 2$$
, Дж, (1)

где k_B — постоянная Больцмана, T — абсолютная температура в кельвинах [9, 10].

Доказательство принципа Ландауэра не зависит от физической структуры вычислителя. Также не влияют на общую потерю энергии от нереверсивности и сверхнизкие температуры: значение предела Ландауэра (1) формально снижается, но на охлаждение требуется тот же порядок энергии [11]. Эта формула показывает предел производительности традиционных процессоров, т. е. фундаментальный предел производительности идеального классического вычислителя — $2,5 \times 10^{26}$ операций/с [12, 13]. А вычислителя, использующего операции с плавающей точкой (принципиально неточные) — 10^{22} операций/с (там же). Соображения надежности вычислений ука-

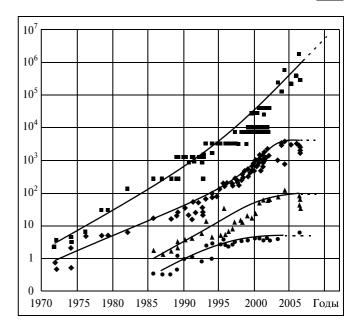
Соображения надежности вычислений указывают на еще более низкие верхние границы [12, 14—16]. Разберем это подробнее. Вероятность ошибки при мощности сигнала E_{sig} равна $e^{-E_{sig}/E_{diss}}$. Для случая электрического сигнала $E_{sig} = CV^2/2$, где C — емкость элемента, V — разность напряжений при переключении. Для надежности нужно $E_{diss} = 100k_BT$. Этот предел уже достигнут.

Ситуацию иллюстрирует диаграмма (см. рисунок), любезно предоставленная Л.К. Эйсымонтом.

Верхняя кривая на рисунке — рост числа процессоров, вторая сверху — тактовая частота, МГц, третья — выделяемое тепло, Дж, нижняя — производительность одного процессора, операций/с. Видно, что рост идет лишь благодаря грубой силе и технологическим ухищрениям, позволяющим размещать больше процессоров на ограниченном объеме и отводить генерируемое ими тепло.

В работе Ландауэра [10] показан и теоретический метод обхода термодинамического барьера: вычисления должны быть обратимыми, тогда значительная часть выделенного тепла поглощается на месте. Беннет [17] показал, что любая машина Тьюринга может быть смоделирована обратимой машиной Тьюринга, и отметил еще одно важное свойство вычислений, которые могут преодолевать предел Ландауэра: обратимы должны быть не только действия, но и программа.

Чтобы пояснить, что такое обратимость, воспользуемся наиболее распространенной моделью действий. Система представляется как множество состояний, а действия — как функции на этом множестве. Эта модель не является полностью адекватной в нашем случае и ни в коем случае не



Диаграмма, иллюстрирующая положение дел на переднем крае суперкомпьютеров

может рассматриваться как основная, но для демонстрации некоторых примеров на привычном языке она годится. Тогда из анализа доказательств фон Неймана и Ландауэра выявляются следующие свойства обратимости.

- 1. Никакие два состояния не должны переходить в одно и то же. Для дискретных систем с конечным числом состояний это означает биективность отображений.
- 2. Имеется возможность после получения результата запустить систему таким образом, что она пройдет в обратном направлении путь от результата к исходным данным через те же промежуточные состояния. Это в совокупности со свойством 1 влечет биективность отображений и для бесконечных систем.
- 3. Обратное вычисление может быть выполнено до прямого, и в результате их композиции получается тождественное преобразование. На самом деле это следствие свойств 1 и 2, но слишком часто не замечаемое на практике.

Беннет не обратил внимания на то, что выделение памяти и ее освобождение — существенно необратимые действия, а его последователи увлеклись позитивной стороной результатов Беннета. И совершенно было проигнорировано предупреждение Беннета насчет программы. Оно показывает, почему стандартная модель действий неадекватна в данном случае: одним из действий должно быть обращение последовательности действий.



Не только из соображений тепловыделения возникает обратимость. Общеизвестно, что квантовые вычисления по своей природе обратимы. Поэтому, если когда-нибудь удастся создать квантовый процессор, он будет способен лишь на обратимые вычисления. Далее, вычисления на молекулярном уровне тоже обратимы по своей природе. Вычисления на уровне нанокристаллов (все равно, электрические или световые) тоже часто обратимы. Нанокристалл вычисляет конечную группу. Группы вычислений кристаллов более широкий класс, чем шубниковские группы [18—21], поскольку групповыми являются также вычисления в кристаллах с упорядоченными дислокациями. Вычисления на сверхпроводниках тоже должны быть обратимыми, если мы не желаем вскипятить жидкий азот.

Стандартным представлением обратимых вычислений в математике являются группы. Поэтому обратимость требует алгебраических, а не численных, моделей.

2. ПАРАДОКС ИЗОБРЕТАТЕЛЯ: ЭФФЕКТ ОРЕВКОВА И ПРЕДЕЛ ЧЕЙТИНА

Рассмотрим теперь вторую трудность: накопление громадных количеств кода, зачастую плохо сделанного. Такой код становится необозримым и неулучшаемым. Попытки улучшить перешедшую некоторый барьер программу приводят к тому, что на каждую исправленную ошибку появляется две новых, что даже рассматривается как некий «объективный закон». Подкрепляет такие убеждения то, что даже в сообществе свободного софта, где каждая программа, а особенно системные программы, подвергается беспощадному анализу на понятность кода, в последние годы начался кризис. Старые программы становятся все менее перестраиваемыми, а приток молодых и энергичных волонтеров, на которых 20 лет развивалось сообщество, резко сократился в связи с почти запредельной сложностью вхождения в существующие проекты [22]. Тем самым надежность существующего программного обеспечения оказалась практически уникально низкой среди всех современных высокотехнологичных систем. Каждая программа с гарантией содержит ошибки.

Это неразрывно связано еще с одним, более «мягким», но тоже объективным барьером: пределом Чейтина [23]. В этой статье установлено существование принципиально непознаваемых объектов, причем предел непознаваемости конечен и зависит от «сложности» познающего субъекта.

Рассмотрим понятие алгоритмической сложности объекта¹. Это *минимальная длина программы и данных, порождающих искомый объект*. Объект может быть очень большим (например, двоичная

запись числа $33^{33^{33}}$), но воспроизводится короткой программой и данными, и поэтому прост.

Понятие алгоритмической сложности, первоначально введенное Колмогоровым [24] для стандартной концепции вычислимости, оказывается корректным для очень широкого класса концепций и уже, чем тьюрингова вычислимость, и шире ее [25]. Теорема Чейтина [23], как показано в работе [25], сохраняется для самых различных понятий сложности и при очень слабых предположениях о формальной системе. Поэтому ее можно сформулировать в качестве методологического закона:

«Для каждого субъекта или сообщества объекты, превосходящие некоторую сложность, являются принципиально непознаваемыми. Предел непознаваемости для каждого индивидуума свой, и область познаваемости может быть неравномерно развита в разных областях».

Таким образом, если системы становятся все сложнее, с некоторого момента они оказываются принципиально непонимаемыми. Длина и структурная сложность программ оказываются сейчас во многих случаях решающим фактором.

Возникает вопрос, является ли такая ситуация объективным законом развития больших формализованных систем? Сразу же виден контрпример: математика. В ней результаты, первоначально получаемые как сложные конструкции на грани понимания, в дальнейшем постепенно свертываются во все более и более короткие, и «разбухания кода» не происходит. Например, теорема полноты классической логики первоначально доказывалась на 200 страницах, а теперь доказательство умещается на двух, одна из которых занята необходимыми определениями. Проанализируем, благодаря чему это достигается.

Еще в 1920-е гг. Д. Гильберт [26] уверенно утверждал, что развитие математики практически невозможно без идеальных высокоуровневых по-

¹ Обратим внимание, что слово «сложность» употребляется в различных контекстах по отношению к совершенно разным характеристикам. Вычислительная сложность принимает во внимание количество операций для решения задачи, но игнорирует характеристики самой программы. Структурная сложность измеряет характеристики структуры объекта, почти игнорируя его величину. Логическая сложность измеряет уровень объекта или высказывания. Алгоритмическая в некотором смысле сочетает характеристики структурной сложности и сложности, измеряемой просто как количество знаков в коде



нятий. В конце концов на практике нас интересуют реальные ответы, выражающиеся конечными сущностями. Например, численная оценка или область фазового пространства, в которой находится система; последнее конечно ввиду того, что область представляется как элемент дискретного разбиения пространства или же как элемент конечной алгебраической структуры, качественно описывающей подмногообразия пространства состояний системы. Но получить нетривиальные реальные результаты с приемлемыми затратами труда получается, лишь пройдя через идеальные абстрактные понятия и их теоретические преобразования, которые можно рассматривать как интеллектуальные орудия.

Д. Пойа [27] заметил этот эффект на примерах из элементарной математики: чтобы доказать простое утверждение, часто требуются значительно более сложные леммы. Хао Ван [28] подтвердил этот факт для арифметики, построив последовательность примитивно-рекурсивных функций f_n , для каждой из которых доказательство утверждения $\forall x f_n(x) = 0$ может быть проведено лишь с помощью математической индукции по формулам не менее чем с n переменами кванторов. В.П. Оревков доказал [29], что аналогичный эффект имеет место уже в логике, где, в принципе, от лемм можно избавиться (теорема о нормализации вывода). А именно, он построил последовательность формул, доказательство каждой из которых с промежуточными леммами имеет длину 13n + 7, а без

помощи лемм (прямое) — как минимум $2^{(2^{(...2)})}$ (n раз). Идею конструкции Оревкова можно пояснить на двух функциональных примерах: числовом и абстрактном (в λ -исчислении).

Пример 1. Пусть у нас есть натуральные числа и одна лишь операция добавления 1. Рассмотрим рекурсивное определение

$$\varphi(x, 0) = x + 1;$$

 $\varphi(x, y + 1) = \varphi(\varphi(x, y), y).$

Согласно ему, экспонента вычисляется в линейное число шагов.

Если теперь задать определение функции второго порядка

$$\Phi_{1}(\varphi, x, 0) = \varphi(0, x);$$

$$\Phi_{1}(\varphi, x, y + 1) = \Phi_{1}(\varphi, \Phi_{1}(\varphi, x, y), y),$$

то оно вычисляет уже экспоненту экспонент, третий уровень башни дает

$$\Phi_2(\varphi, x, 0) = \Phi_1(0, x);$$

$$\Phi_2(\Phi_1, x, y + 1) = \Phi_2(\Phi_1, \Phi_2(\Phi_1, x, y), y)$$

и т. д.

Пример 2. В λ -исчислении, которое сейчас стало главным теоретическим аппаратом информатики, есть две базовых конструкции: применение функции f к аргументу (xf) и образование функции, вычисляющей при каждом заданном x_0 значение выражения t, содержащего переменную x: $\lambda x.t[x]$. Функции и объекты не различаются: функция может быть аргументом и результатом другой функции (почти как в теории множеств, где «всё есть множество»; только здесь «всё есть функция»). Тогда можно построить последовательность определений, уже не рекурсивных:

$$\Phi_1 = \lambda f. \lambda x. ((x f) f);$$

$$\Phi_{n+1} = \lambda \Psi_n. \lambda \Psi_{n-1}. ((\Psi_{n-1} \Psi_n) \Psi_n);$$

$$(x(f(\Phi_1...(\Phi_{n-2}(\Phi_{n-1}\Phi_n^k))...))) = \left(xf^{(2^{(2^{(...2)})})}\right)(k \text{ pas}). \bullet$$

В дальнейшем для определения функций мы свободно пользуемся λ -обозначениями.

Заметим, что в обоих определениях использовались функции над функциями. Это неизбежно, поскольку теоремы Хао Вана, Оревкова и их многочисленные применения и усиления, развившиеся сейчас в целое направление «обратной математики», используют иерархию сложности и уровней объектов. Уровень функции или множества выше уровней аргументов (элементов). Уровень логической формулы определяется в первую очередь уровнями представленных в ней объектов, по которым навешиваются кванторы, а во вторую очередь количеством перемен вложенных друг в друга кванторов.

Выразимость существенно повышается при подъеме на новый уровень. Но одновременно понятия становятся более абстрактными.

Таким образом, сколь угодно сложные рекурсивные процедуры стандартного языка программирования типа С — объекты первого уровня. Классы С или Java — где-то между первым и вторым. А для существенного сокращения текстов и борьбы с объемом программ нужны понятия высших уровней.

Явное введение сложных понятий необходимо, но нужно искать обходные пути, поскольку выразительная сила растет, как минимум, суперэкспоненциально, однако одновременно экспоненциально убывает число людей, способных воспринять понятия соответствующего уровня².

Здесь вступает в силу фундаментальное свойство алгебраических систем. Рассмотрим произ-

² Автор может пересчитать по пальцам тех, кто способен понять конструкции седьмого уровня, порою встречающиеся в информатике и в качественной теории систем (работы школы В.М. Матросова и С.Н. Васильева).



Таким образом, алгебраические модели дают возможность атаковать и второй барьер.

Пример 3. В частном случае возможность представления функций высших уровней при помощи элементарной алгебраической системы была открыта М. Шейнфинкелем [30] в 1920 г. Но общность этого метода не была осознана.

Комбинаторная алгебра Шейнфинкеля — группоид с операцией * и двумя константами: \mathbf{S} и \mathbf{K} . Постулируются следующие свойства этих констант:

Комбинаторные алгебры обладают следующими свойствами. Как известно, любая алгебраическая система, определяемая тождествами, имеет тривиальную модель из одного элемента. Все нетривиальные модели комбинаторных алгебр [31]:

- бесконечны;
- неассоциативны;
- некоммутативны;
- нерекурсивны;
- позволяют выразить внутри себя натуральные числа и любую вычислимую функцию как действие некоторого элемента на натуральные числа;
- позволяют выразить вычислимые функции над функциями и сколь угодно высоко вверх по иерархии понятий.

Таким образом, алгебраические модели в принципе позволяют атаковать и барьер сложности программ. Это косвенным образом подтверждено в работах по категорному программированию [32].

3. ПРОГРАММНЫЕ АЛГЕБРЫ

В работах по алгебрам программ (в качестве представительных примеров см. работы [1—6]) немедленно предполагается, что программа — это функция, и, соответственно, базовая алгебраическая структура этих моделей — полугруппа. Основанием такого выбора служит классическая теорема о представлении любого пространства функций, замкнутого относительно композиции, как полугруппы и любой полугруппы как полугруппы функций. Даже в книге Митчелла [7], как только

переходят к алгебраическим моделям, они рассматриваются на базе полугрупп. Есть еще один принципиальный недостаток всех рассматриваемых алгебраических конструкций. Они ставят телегу впереди лошади: сразу моделируются конкретные конструкции языков программирования и включаются как операторы в алгебраические модели. Таким образом, теоретическое исследование оказывается в плену существующей практики, которая уже не полностью адекватна запросам развития информатики, как показано ранее. Не может быть описано как функция, в частности, действие обращения операции, естественно возникающее в обратимом программировании. В самом деле, пусть действием элемента М является обращение операции. Тогда, поскольку $(a \circ b)^{-1} = b^{-1} \circ a^{-1}$, имеем $a \circ (b \circ M) \neq (a \circ b) \circ M$.

На самом деле аналогичная конструкция возникает на каждом шагу в практической информатике. Рассмотрим операцию отмены последнего действия Undo. Тогда

S;
$$T$$
; Undo \neq {S; T }; Undo.

Таким образом, действия операторов не обязательно являются функциями.

Но синтаксическая структура программ, которая важна при их преобразованиях, полугруппой является. А именно, полугруппа дает возможность соединить действия нескольких элементов в действие одного элемента, тем самым позволяя агрегировать блоки модели. Эта операция необходима для декомпозиции и синтеза систем моделирования.

Самый нейтральный способ ввести операцию, свертывающую систему элементов в элемент, аналогичен используемому в комбинаторных алгебрах. Добавим к группоиду константу ${\bf B}$ (композиция) и аксиому

$$((x * f) * g) = (x * (f * (g * B))).$$
(3)

Такие группоиды назовем композиционными алгебрами.

Предложение 1. Для каждой последовательности f_1 , ..., f_n элементов группоида ее действие $\lambda x.((...(x*f_1)...)*f_n)$ равно действию элемента $(f_1*(...(f_n*B)...))$.

Определение 1. Обогащение алгебраической системы — добавление новых операций и констант без изменения носителя и старых операций³. Расширение алгебраической системы — вложение ее в другую алгебраическую систему. Факт — равен-

³ Тем самым значения новых констант задаются в старом носителе



ство или неравенство постоянных выражений над алгебраической системой: t = u либо $t \neq u$.

Предложение 2. *Каждый группоид* **G** может быть расширен до композиционной алгебры.

Доказательство. Рассмотрим свободный группоид, образующийся добавлением к элементам \mathbf{G} константы \mathbf{B} , и хорнову теорию, образуемую всеми позитивными фактами о \mathbf{G} и тождеством (3). Фактор-алгебра по отношению выводимого равенства будет искомой композиционной алгеброй. \bullet

Однако полугруппы действий не получается: композиция в композиционных алгебрах не обязательно ассоциативна. Сделать ее ассоциативной можно, задав аксиому

$$(f * ((g * (h * B)) * B)) = = ((f * (g * B)) * (h * B)).$$
(4)

Итак, приходим к алгебре, являющейся бигруппоидом с двумя операциями. Одна из них — полугрупповая операция композиции °, вторая неассоциативная операция применения программы *. Получающуюся полугруппу по композиции будем обозначать прописной жирной буквой, ее носитель — прописной буквой; 0 и е — нулевой и единичный элемент соответствующей полугруппы, если они имеются. Операции связаны тождествами:

$$((x * f) * g) = (x * (f * g));$$

 $(0 * x) = (x * 0) = 0;$
 $(x * e) = x.$

Как бы мы ни определили понятия композиции и применения, они должны быть связаны приведенными тождествами. В остальном мы, как и всегда при алгебраическом подходе, свободны в интерпретации. Во всех известных автору программных, человеко-машинных и аппаратных системах эти свойства выполнены.

Такие алгебры назовем GAPS (generic algebraic program structure).

Первое тождество выполняется всегда, оно является определением последовательного исполнения либо агрегации элементов. Если хоть одна из констант 0 или e присутствует, она сохраняет свой смысл и для операции *.

Пусть имеется некоторый гомоморфизм α : $G \to (G \to G)$ нашей полугруппы G в полугруппу отображений ее носителя, такой, что

$$(e \ \alpha) = \lambda x.x; \quad (0 \ a) = \lambda x.0.$$

С точки зрения программиста, этот гомоморфизм задает интерпретатор или транслятор наших программ, перерабатывающий запись программы в исполняемый модуль. Синтаксически различные

программы могут порождать один и тот же исполняемый модуль, так что считать отображение α изоморфизмом нельзя.

Предложение 3. Каждая полугруппа **G** может быть обогащена до *GAPS* таким образом, что $(x * f) = (f \alpha)$.

Доказательство. Второе и третье тождества выполнены по построению.

$$((x * f) * g) = ((x(f \alpha))(g \alpha)) = (x((f \alpha) \circ (g \alpha))) = (x((f \circ g)\alpha)) = (x * (f \circ g)).$$

Такой гомоморфизм всегда существует. Достаточно определить $(x \alpha) = \lambda y.0$ для всех делителей нуля, и $\lambda x.x$ для всех остальных. Второй тривиальный вариант — просто определить $(x * y) = (x \circ y)$. \blacklozenge

Пример 4. Рассмотрим известную игру «колодец, ножницы, бумага». Бинарная операция определения выигравшего в этой игре задает коммутативный неассоциативный группоид КНБ с тремя элементами:

Для того чтобы задать GAPS, содержащую этот группоид, имеется несколько способов. Первый из них — чисто синтаксический. Отождествим элементы группоида с их одноэлементными кортежами, а композиции действий зададим как кортежи из κ , κ , κ , κ , κ — соединение кортежей. Но даже в таком пополнении имеется много способов задания операции *. Проиллюстрируем их на примере кортежей [κ , κ , κ] и [κ , κ , κ]:

$$\begin{split} [\kappa, \, H, \, \delta] * [H, \, \kappa, \, \delta] &= [\kappa, \, H, \, (((\delta * H) * \kappa) * \delta)]; \\ [\kappa, \, H, \, \delta] * [H, \, \kappa, \, \delta] &= [(((\kappa * H) * \delta) * H), \, \kappa, \, \delta]; \\ [\kappa, \, H, \, \delta] * [H, \, \kappa, \, \delta] &= \\ &= [(((\kappa * H) * \kappa) * \delta), \, (((H * H) * \kappa) * \delta), \, (((\delta * H) * \kappa) * \delta)]. \end{split}$$

Теоретически простейший способ получения GAPS — построить свободную алгебру над константами $\{\kappa, \kappa, \delta\}$ с бинарными операциями $\{*, \circ\}$ и факторизовать ее по отношению равенства *t = u выводимо в теории со следующими аксиомами»:

$$(K * K) = K;$$
 $(K * H) = K;$ $(K * G) = G;$
 $(H * K) = K;$ $(H * H) = H;$ $(H * G) = H;$
 $(G * K) = G;$ $(G * G) = G;$ $(G * H) = H;$
 $\forall x, f, g((x * f) * g) = (x * (f \circ g));$
 $\forall x, f, g((x \circ f) \circ g) = (x \circ (f \circ g)).$

Есть способ получить и конечную GAPS. Для этого сначала пополним КНБ до полугруппы действий последовательностей его элементов. Эта полугруппа оказывается некоммутативной. Приведем все различные действия последовательностей элементов игры. Перед двоеточием стоит последовательность ходов, после него — ее действие:

к: ккб, н: кнн, б: бнб, кн: ккн, нк: ккк, бн: ннн, нб: бнн, кб: ббб, бк: бкб,кнб: ббн, нбк: бкк, бкн: нкн.



Полугруппа действий состоит из двенадцати элементов. Но даже здесь операция * может быть продолжена, по крайней мере, двумя способами:

$$(KH * HK) = KHHK = KHK = KH;$$

$$(KH * HK) = (KH)*(H*K) = (KH) K = (K*K)(H*K) = KK = K.$$

Пример 5. Некоторые «естественные» предположения о GAPS разрушают систему. Например, из тождества e * f = f следует, что обе операции совпадают:

$$(x * y) = ((e * x) * y) = (e * (x \circ y)) = x \circ y.$$

Рассмотрим возникшую ситуацию содержательно: почему 0 может оставаться нулем, а единица при действиях оставаться сама собой не всегда может? Нуль в смысле интерпретации — ошибка. Дальше с нею внутри алгебры делать уже нечего: система вылетела. Константа e может быть интерпретирована как программа, ничего не делающая и не тратящая ресурсов (например, пустая). Но преобразователь программ вполне может выдать какой-то сложный код, исходя из пустых данных. См. далее пример $8. \$

Центральная проблема: когда можно некоторую систему преобразований программ соединить с данной синтаксической полугруппой программ, тем самым образовав единый язык метапрограммирования? Эта задача естественно разбивается на три подпроблемы: соединение языка с системой преобразований без изменения языка (т. е. полугруппы); сохранение некоторого подъязыка (подполугруппы) и превращение ранее не интерпретированных или не нужных нам более синтаксических конструкций в действия преобразователей; пополнение существующего языка до метаязыка путем расширения полугруппы без изменения исходной.

Определение 2. Полугруппа **F**, порожденная совокупностью функций F — минимальный тип функций, включающий F и замкнутый относительно композиций:

$$\forall f, g(f \in F \land g \in F \supset f \circ g \in F). \blacklozenge$$

Содержательно это означает, что **F** состоит из всех функций, представимых как

$$f_1 \circ \dots \circ f_n, f_1, \dots, f_n \in F$$
.

Определение 3. Пусть задана некоторая совокупность действий **A** как функции на носителе полугруппы **G**. Совокупность **A** допустима, если имеется GAPS **AG**, которая является обогащением **G**, такая, что каждое действие $g \in A$ представляется как $\lambda x.(x*\alpha)$ для некоторого α . Она консервативна над подполугруппой G_0 , если она допустима и $a*b=a\circ b$ для всех элементов G_0 . Она слабо допустима, если имеется полугруппа G_1 , подполугруппой которой является ${\bf G}$, и функции из ${\bf A}$ могут быть продолжены на G_1 таким образом, что полученная A_1 консервативна над ${\bf G}$. ullet

Очевидно, что самое сильное из этих понятий — консервативность над подполугруппой. Далее идут допустимость и слабая допустимость. Установим критерии возможности этих обогащений и расширений.

Предложение 4. Совокупность действий **A** допустима тогда и только тогда, когда имеется гомоморфизм ее замыкания в полугруппу **G**.

Доказательство. По предложению 2, если обогащение удалось осуществить, то любой элемент полугруппы, порожденной \mathbf{A} , представляется как действие некоторого элемента G. Значит, мы гомоморфно вложили замыкание \mathbf{A} в \mathbf{G} . \blacklozenge

Обратно, если замыкание **A** гомоморфно отображается в **G**, то его образ изоморфен некоторой фактор-полугруппе \hat{G} полугруппы **G** и имеется отображение $\Phi: G \to (G \to G)$, сопоставляющее каждому элементу полугруппы функцию, соответствующую его классу эквивалентности в \hat{G} .

Заметим, что важно вложение структуры полугруппы действий, а не самих действий.

Предложение 5. Совокупность действий **A** консервативна над G_0 тогда и только тогда, когда имеется такой гомоморфизм ψ ее замыкания в **G**, что для каждого действия f, такого, что $(f \psi) \in G_0$, выполнено $\forall a \in G(a f) = (a \circ (f \psi))$.

Доказательство. Если такое обогащение удалось осуществить, то действие, попавшее в G_0 , обязано по условию сохранения подполугруппы удовлетворять

$$\forall a \in G(a f) = (a \circ (f \psi)). \blacklozenge$$

Обратно, если такой гомоморфизм имеется, то применим построение из предыдущего доказательства.

Пример 5. Например, пользуясь этим критерием, можно проверить возможность обогащения языка до языка метавычислений, рассматривая строки как программы, а подъязык, работающий с числами, оставляя неизменным.

Теорема 1. Любая совокупность действий слабо допустима над любой **G**. ◆

Эта теорема доказана в статье [33]. Несмотря на простоту формулировки, доказательство требует методов теории моделей и теории алгебраических систем. Получающаяся алгебра не всегда вычислима, даже если исходная полугруппа и совокупность действий конечны.

Пример 6. Рассмотрим группу вычетов по модулю 3 Z_3 и интерпретируем ее объекты $\{0, 1, 2\}$ как коды колодца, ножниц и бумаги в игре из примера 4. Кодируя



функции тройками чисел, задаем действия игры как 002, 011, 212. Пополним группоид до двенадцатиэлементной полугруппы. Для согласованности с Z_3 необходимо добавить единичный элемент, который не может быть получен композициями остальных. Возьмем прямое произведение получившегося тринадцатиэлементного моноида C с Z_3 и зададим действия следующим образом:

$$(\langle c, x \rangle * \langle d, y \rangle) = \langle c \circ d, (x + y d) \rangle.$$

Построенная алгебра содержит Z_3 . Видно, что элементы C можно рассматривать как команды, а Z_3 — как данные; (xd) — применение последовательности действий к элементу, закодированному x,и взятие кода получившегося элемента. Команды преобразуются как полугруппа, но их эффект на Z_3 — группоид.

Имеется и другой способ определения GAPS над тем же носителем:

$$(\langle c, x \rangle * \langle d, y \rangle) = \langle (c \circ d), (x + y(c * d)) \rangle.$$

Если же в качестве пополнения КНБ возьмем бесконечную GAPS, то и получившаяся система окажется бесконечной. ◆

Дадим полный теоретический критерий пополнимости некоторой системы операций над программами с заданными свойствами операций до GAPS над уже имеющейся у нас программной системой с сохранением свойств операций и программ. Он обобщает теорему из статьи [33], где сохранялись только свойства операций и факты о программах.

Определение 4 [8]. $\Phi a \kappa m$ — выражение вида $P(t_1, ..., t_n)$ либо $\neg P(t_1, ..., t_n)$, не содержащее переменных, в частности, t = u или $t \neq u$. Факт *позитивный*, если он не начинается с отрицания. Диаграмма системы — совокупность всех истинных в ней фактов, при условии, что сигнатура пополнена константами для всех элементов носителя системы. Тождество — формула вида $\forall x_1, ..., x_k P(t_1, ..., t_n)$, где x_i — все свободные переменные элементарной формулы. Ограничение формулы A на одноместный предикат P — формула A_p , образующаяся заменами всех кванторов $\forall x B[x]$ на $\forall x (P(x) \supset B[x])$, а всех $\exists x B[x]$ на $\exists x (P(x) \land B[x])$. Если Σ — алгебраическая система, то Th^{Σ} — теория Th, пополненная всеми фактами о Σ .

Теорема 2. Пусть система действий **A** описывается теорией Th_1 , и Th — теория, моделью которой является полугруппа **G**, в которой выводится аксиома ассоциативности, а также свойства единицы и нуля, если они присутствуют в **G**, и одноместный предикат P не принадлежит объединению сигнатур. Th_{P} означает теорию Th, к сигнатуре которой добавлен предикат P без новых аксиом. Тогда G можно пополнить действиями A с сохранением их свойств тогда и только тогда, когда теория

$$Th_1^A \cup Th_P^G \cup \{ \forall x, f, g(((x * f) * g) = (x * (f \circ g))), \forall x(0 * x) = (x * 0) = 0, x(x * e) = x \}$$

непротиворечива.

Доказательство. По теореме полноты классической логики, теория имеет модель тогда и только тогда, когда она непротиворечива. Если теория содержит матрицу алгебраической системы Σ , эта система изоморфно вложима в любую ее модель [8, $c.\ 172$]. Тем самым существующая модель является искомой GAPS. Обратно, если есть такая GAPS, то она является моделью нашей теории. •

Теоретически данное решение полное. Но на практике оно редко применимо, поскольку чаще всего расширения оказываются невычислимы. Однако есть один практический критерий, который следует из данной теоремы.

Практическое следствие. Если система преобразований программ разрушает свойства каких-то уже имеющихся действий программ или заставляет отождествить разные действия, она принципиально некорректна.

Пример 7. Поскольку \(\lambda\)-исчисление лежит в основе современной математический теории полных по Тьюрингу программных систем, построим модель стратифицированного λ-исчисления как GAPS. Напомним некоторые понятия: λ-выражение называется стратифицированным, если в нем можно корректно расставить типы всех переменных и констант; λ-выражение имеет нормальную форму, если оно не может быть преобразовано далее. Выражения в нормальной форме считаются окончательными результатами вычислений. Обе возможности: рассматривать полугруппу как множество λ-термов в нормальной форме или как множество отображений нормальных форм t в нормальные формы (ft), происходящих при конверсии выражений, не дают решений. При первом подходе нет композиции, поскольку нет операции соединения λ-выражений. При втором композиция естественно определяется, но теряется суть программной системы, поскольку разные программы могут производить одну и ту же функцию преобразования.

Чтобы получить GAPS, перейдем к эквивалентному представлению λ -термов: комбинаторной логике. Комбинаторная логика естественно и взаимнооднозначно формулируется как группоид действий: применение терма к терму $(t\ r)$ записывается в наших обозначениях $(r\ *\ t)$.

Определяющие равенства для констант \mathbf{K} и \mathbf{S} переписываются в форме (2). В комбинаторной алгебре определяется комбинатор композиции

$$\mathbf{B} = (\mathbf{K} * ((\mathbf{S} * \mathbf{K}) * \mathbf{S})).$$



Преобразуя это выражение согласно определяющим равенствам для **К** и **S**, получаем:

$$(x * (f * (g * (\mathbf{K} * ((\mathbf{S} * \mathbf{K}) * \mathbf{S})))) =$$

$$= (x * (f * ((g * \mathbf{K}) * (f * (\mathbf{S} * \mathbf{K}))))) =$$

$$= (x * (f * ((g * \mathbf{K}) * \mathbf{S}))) = ((x * f) * (x * (g * \mathbf{K}))) =$$

$$= ((x * f) * g).$$

Таким образом, $(f*(g*\mathbf{B}))$ определяет композицию f и g.

Можно было перейти к другому известному базису комбинаторной логики, включающему в себя ${\bf B}$, но мы проиллюстрируем путь, показывающий способ перехода от теории или системы, в которой композиции как корректно определенной операции не было, к эквивалентной теории или системе, представимой как алгебра GAPS.

Прежде всего, добавим В как новую константу с определяющим равенством (3). Такое расширение теории с привычной точки зрения эквивалентно исходной теории, поскольку добавили определимое понятие, но оно избавляет от нежелательного эффекта: как бы мы ни определили **B**, $(f * (g * \mathbf{B}))$ не будет нормальной формой, поскольку последняя константа в определении заведомо применится (даже если она S, требующая три аргумента, то, поскольку определение не сводится к самой S, третий аргумент будет взят из текста определения). Добавляем равенство (4). Это тождество может быть эффективно реализовано как одностороннее и заведомо фундированное правило переписывания левой части в правую. Носитель алгебры — множество нормальных форм выражений в комбинаторной логике, пополненной комбинатором В, нормализованных применением правила, соответствующего равенству (4). Композицией двух выражений считается ($f * (g * \mathbf{B})$). Это определение корректно, поскольку стратифицированные λ-выражения всегда имеют нормальную форму, и композиция стратифицированных выражений стратифицирована.

Построение завершено. Таким образом, GAPS позволяют описать одну из самых общих теоретических концепций современной информатики. Для бестипового λ -исчисления эта конструкция также работает. Значением выражения, не имеющего нормальной формы, считается 0. В стратифицированной алгебре нуля нет. В построенных GAPS нет единицы. Она может быть добавлена как новая константа с правилами конверсии

$$(t * e) = t, (f * (e * \mathbf{B})) = f, (e * (f * \mathbf{B})) = f.$$

Итак, представление «конкретных» алгоритмических языков становится излишним: их методы сведения к λ -исчислению детально проработаны и изложены, в частности, в работе [7].

Пример 8. Рассмотрим другой пример превращения программной системы в абстрактную алгебру. Пусть у нас есть полный по Тьюрингу алгоритмический язык, в котором программы и данные являются строками символов в одном и том же алфавите. Соединение строк дает композицию функций программ. Пустая строка — ничего не делающая программа. Примером такого языка может служить Brainfuck [34].

Добавим константы 0 и e. Первая из них является значением вылетающей или зацикливающейся программы, вторая — пустая строка. Композиция программ определяется как соединение строк. Действие строки (a*f) состоит в вычислении строки, являющейся значением f на a. Если f выдала ошибку, значением является 0. В противном случае — получившаяся программа. На этом примере видно, что результатом действия над пустой программой может быть любая другая программа.

Пример 9. Еще в 1972 г. теория вплотную подошла к общему понятию алгебры программ [35]. Рассмотрим алфавит {K, S, (,)}. Определим следующую трансляцию его символов в комбинаторы. Выражение, сопоставляемое строке, определяется по рекурсии (a — символ, σ — строка):

$$(\mathbf{K}\phi) = \mathbf{K}; \quad (\mathbf{S} \ \phi) = \mathbf{S}; \quad (\text{`('} \ \phi) = \mathbf{B}; \ (\text{(')'} \ \phi) = \mathbf{I};$$

$$(a\sigma \ \phi) = ((\sigma \ \phi) * (a \ \phi)).$$

Для традиционной комбинаторной логики, где функцией считается первый элемент списка и скобки группируются влево, получим еще более простую запись: образы символов пишутся один за другим. Такая интерпретация превращает комбинаторную логику в язык типа Brainfuck. Но получившаяся алгебра не является алгеброй для комбинаторной логики. В нее входят образы синтаксически неправильных конструкций, скажем,)))) SK((.

4. АЛГЕБРЫ ОБРАТИМЫХ ВЫЧИСЛЕНИЙ

Обратимые действия составляют группу, и любая группа представляется как группа биективных отображений. Группа сохраняет основное свойство группоида: возможность рассматривать любое данное как функцию. В ней имеется дополнительная операция взятия обратного действия (которая может применяться как до, так и после исходного действия): a^{-1} . Таким образом, кажется, что в данном случае GAPS основана на группе.

Преобразования программ также должны быть обратимы, согласно принципу Беннета. Отсюда вытекает конструкция алгебр полностью обратимых программ (AFIP — algebras of fully invertible programs) как GAPS с дополнительной константой M (обращение программ) и дополнительными аксиомами:

$$(x \circ (x * M)) = e, ((x * M) \circ x) = e.$$
 (5)

Отсюда следует, что (M*M)=M; $(M\circ M)=e$. Для практических нужд обычно достаточно более слабого свойства: тождественности действий (слабые алгебры обратимых программ, weak AFIP, WAFIP):

$$((y * x) * (x * M)) = y, ((y * (x * M)) * x) = y.$$
 (6)

Выражение (6) следует из выражения (5), но не наоборот. В статье [33] изучались AFIP. Здесь ос-



новным понятием являются WAFIP, которых достаточно для представления обратимых программ и их аппаратных моделей. Из аксиом WAFIP автоматически следует биективность операции применения программ *:

Предложение 6. Отображение $\lambda x.(x * f)$ биективно для любого f.

Доказательство. Инъективность. Пусть (x*f) = (y*f). Тогда ((x*f)*(f*M)) = ((y*f)*(f*M)), и x = y по выражению (6).

Сюръективность. Найдем для каждого x и f такое y, что (y*f)=x: x=((x*(f*M))*f). \blacklozenge

В WAFIP полугруппа не обязательно становится группой, так как биективность каждого действия не означает различия всех действий как функций.

Пример 10. Рассмотрим обогащение Z_6 до AFIP:

$$\lambda x.(x * 3) = \lambda x.(x * 1) = \lambda x.(x * 5) = \lambda x.x^{-1},$$

 $\lambda x.(x * 0) = \lambda x.(x * 2) = \lambda x.(x * 4) = \lambda x.x.$

В качестве M нужно взять 3. Если мы возьмем другое нечетное число, то получится WAFIP.

Аналогичным образом можно пополнить до WAFIP группу целых чисел по сложению:

$$\lambda x.(x * (2 \cdot n + 1)) = \lambda x.x^{-1}, \ \lambda x.(x * (2 \cdot n)) = \lambda x.x.$$

Но здесь до AFIP не продолжишь, поскольку нет такого числа, не равного 0, что n+n=0.

Теперь рассмотрим циклический моноид, порождаемый a и определяемый равенством $a^4 = a^8$. На его носителе всем нечетным степеням a^{2n+1} сопоставим $\lambda a^n.a^{-n}$, где обратный элемент берется по модулю 8, четным — тождественное преобразование. Но данная алгебра не является WAFIP, потому что степени, большие 3, уже не могут быть обращены до степеней, меньших 3.

И завершим тривиальным примером. Любая полугруппа правых единиц, т. е. такая, что $\forall x, y \circ y = x$, может быть сделана WAFIP, а в качестве M можно взять любой ее элемент.

В последней WAFIP полугруппа группой не является. Но, тем не менее, получен следующий результат:

Предложение 7. Действия элементов WAFIP образуют группу. ◆

Действие (x * M) является обратным к действию x.

В статье [33] доказано, что групповая операция может образовывать AFIP в том и только в том случае, когда все элементы группы имеют порядок 2, т. е. группа является прямым произведением групп Z_2 . Следовательно, если операция в полугруппе образует WAFIP, то действия ее элементов образуют группу порядка 2.

Предложение 8. Каждая полугруппа G, разлагающаяся в прямое произведение $Z_2 \times G_0$, где действия элементов G_0 взаимно однозначные и для каждого действия есть элемент, производящий обратное действие, может быть обогащена до WAFIP.

Доказательство. Сначала установим несколько более абстрактное утверждение о полугруппах. Пусть \mathbf{G} — носитель полугруппы. Зададим для M в качестве значения какой-нибудь элемент, такой, что действие M^2 тождественное и M не является квадратом никакого элемента. Пусть $\mathbf{G}\mathbf{G}$ — группа всех взаимно однозначных отображений G. Пусть ϕ — гомоморфизм \mathbf{G} в $\mathbf{G}\mathbf{G}$, такой, что $(M\phi) = \lambda x.x^{-1}$. Определив $(x*f) = (x(f\phi))$, получаем искомую алгебру. \bullet

Теперь построим такой гомоморфизм: $\langle 1, e \rangle = M$, $(M \phi) = \lambda x.x^{-1}$, $(\langle 0, a \rangle \phi) = \lambda x.x$, $(\langle 1, a \rangle \phi) = \lambda x.x^{-1}$. Необходимое условие обогащаемости полугруп-

Необходимое условие обогащаемости полугруппы взаимно однозначных отображений до WAFIP заключается в наличии элемента, который:

- не является квадратом никакого другого элемента:
- имеет бесконечный порядок или является корнем нечетной степени из самого себя.

Установление необходимых и достаточных условий обогащаемости группы либо полугруппы до WAFIP, подобно тому, как было сделано для AFIP в статье [33], остается открытой проблемой.

Пример 11. Рассмотрим общую и важную для приложений конструкцию. Пусть есть произвольная полугруппа G, действия элементов которой обратимы, и для каждого действия есть обратное. Пополним ее до алгебры $Z_2 \times G$ с сохранением G.

Положим $\langle x, a \rangle * \langle 0, b \rangle = \langle x, a \circ b \rangle, \langle x, a \rangle * \langle 1, e \rangle = \langle x, a^{-1} \rangle,$ где e — произвольный элемент, действие которого тождественно. Продолжим на остальные элементы следующим образом:

$$(\langle x, a \rangle * \langle 1, b \rangle) = (\langle x, a \rangle * \langle 1, e \rangle \circ \langle 0, b \rangle)) =$$

$$= ((\langle x, a \rangle * \langle 1, e \rangle * \langle 0, b \rangle) = \langle x, a^{-1} \rangle * \langle 0, b \rangle = \langle x, a^{-1} \circ b \rangle.$$

Полученное решение может быть интерпретировано как простейшее управление. Элемент Z_2 является управляющим сигналом, изменяющим направление работы системы с прямого на обратное и наоборот.

Заметим, что не всегда обратимость действий означает обратимость алгоритма и самой задачи. Рассмотрим два характерных случая.

Пример 12. В задаче сборки кубика Рубика каждое действие обратимо. Но сама задача принципиально необратима, поскольку во время решения теряется информация об исходном состоянии кубика и о промежуточных состояниях.

Задача упорядочения массива принципиально необратима, так как при ее решении массив приводится к стандартному состоянию и невозможно восстановить его исходное состояние. В данном случае формально она может быть сделана обратимой, если каждый элемент x, стоящий на i-м месте, представлять как пару $\langle x, i \rangle$, но



алгоритм может быть все равно необратим, поскольку теряется информация о промежуточных состояниях массива.

5. НАБРОСКИ АЛГЕБРАИЧЕСКОГО ПОДХОДА К МОДЕЛЯМ СИСТЕМ

В предыдущих параграфах подводился итог исследованиям в данной области. Далее в основном приводятся примеры формализаций, которые обосновывают постановку задач для дальнейшей работы.

Пример 13. Рассмотрим условную простейшую модель беспилотника с зарядом, нацеленного на некоторый объект. Во время полета беспилотник можно рассматривать как недиссипативную динамическую систему, и в любой момент ему можно отдать приказ на возвращение назад, а возвращающемуся — на возобновление атаки (конечностью горючего пренебрегаем). После атаки состояние беспилотника уже не может быть изменено (он действует как камикадзе). На стартовой позиции его состояние не меняется, пока не будет отдан приказ на взлет.

Таким образом, фазовое пространство системы делится на три подпространства: подпространство старта, подпространство полета, подпространство завершенной миссии. Управляющих воздействий в реальности три: команда на взлет, команда на прекращение атаки, команда на возврат к атаке. Но формально мы можем свести их к двум: назначение база, назначение цель.

Динамическая система без диссипации порождает группу сдвигов. В данном случае группа становится полугруппой, поскольку движения системы ограничены. Пусть d — подлетное время к цели, x и y меняются в диапазоне [0, d). Определим операции:

$$x + + y = \begin{cases} x + y, x + y < d \\ 0, \quad x + y \ge d \end{cases},$$

$$x - - y = \begin{cases} x - y, x - y > 0 \\ 0, & x - y \le 0 \end{cases}.$$

Теперь определим алгебру управления беспилотником:

$$(\langle 0, 0 \rangle * \langle a, x \rangle) = \langle a, 0 \rangle,$$

$$(\langle 1, x \rangle * \langle a, y \rangle) = \langle a, x + y \rangle,$$

$$(\langle 0, x \rangle * \langle a, y \rangle) = \langle a, x - y \rangle (x > 0, x < d),$$

$$(0 * z) = (z * 0) = \langle a, 0 \rangle = 0.$$

Получившаяся алгебра может быть дискретизирована. Если подлетное время разбить на целое число равных интервалов, значения x и y могут быть сделаны натуральными числами, и получается конечное приближение к непрерывной алгебре управления.

Этот пример показывает принципиальные возможности формализации в алгебрах логико-динамического управления. Для того, чтобы подойти к этой проблеме с

несколько более общих позиций, рассмотрим еще пару примеров.

Пример 14. Пусть у нас есть недиссипативная динамическая система, фазовые траектории которой не пересекаются. Тогда, дискретизировав пространство состояний в момент времени 0, получаем дискретно-непрерывную группу, качественно характеризующую область, в которой находится система, и количественно время ее движения от начальной точки. Дискретизировав еще и время, получаем двумерную дискретную группу приближенных характеристик системы. Ограничив время, получаем полугруппу. Ограничивая пространство начальных состояний, получаем подгруппу (соответственно, подполугруппу) ранее построенной группы (полугруппы). Ограничивая область в фазовом пространстве, в которой может находиться система, сохраняя при этом связность оставшихся частей фазовых траекторий, получаем полугруппу, в которой, начиная с некоторого момента, $x \circ y = x$. Любая из этих (полу)групп может быть пополнена до алгебры управлений добавлением бинарного управляющего элемента, переключающего движение с прямого на обратное; тернарного элемента, включающего приостановку системы; элемента с четырьмя состояниями, где четвертое состояние означает ликвидацию системы (0 в примере 13); и т. д.

Подмечено, что идемпотенты (элементы, для которых $x \circ x = x$) играют в алгебрах роль аналогов условий завершения или переключения процессов в традиционном описании [36].

Пример 15. Пусть имеется диссипативная система. Тогда исходная алгебра — полугруппа, и операции обращения нет. Другие управляющие операции по-прежнему законны.

Пример 16. Пусть имеется логико-динамическая система (диссипативная или нет), поведение которой описывается набором операторов на фазовом пространстве и условий переключения между различными операторами. Для иллюстрации метода превращения такой системы в алгебру пусть у нас система описывается двумя операторами без диссипации и одним условием переключения. Тогда каждый из операторов задает группу сдвигов. Обозначим эти группы G_1 и G_2 . Определим полугруппу, которая может рассматриваться как частичная прямая сумма двух групп. Их элементы в большинстве случаев совпадают, потому что они действуют над одним и тем же фазовым пространством, но операции разные.

$$G = (\{1\} \times (G_1 \setminus \{e\})) \cup (\{2\} \times (G_2 \setminus \{e\})) \cup \{0, e\}.$$

Таким образом, задаем размеченное объединение двух групп, склеивая единицы и добавляя ноль. Будем считать, что $\langle 1, e \rangle = \langle 2, e \rangle = e$,

$$\langle a, x \rangle \circ \langle a, y \rangle = \langle a, x \circ y \rangle,$$

 $\langle 1, x \rangle \circ \langle 0, y \rangle = \langle 0, y \rangle \circ \langle 1, X \rangle = 0,$
 $0 \circ z = z \circ 0 = 0.$

Произведение элементов из разных групп считается ошибкой.

Действия можно определить, по крайней мере, двумя способами. При первом мы должны переключиться в



точности на поверхности переключения, иначе система ломается. При втором — на поверхности переключения система задерживается до следующей команды. Получаем две различных GAPS над одной и той же полугруппой. В $G_{\it err}$ действие задается следующим образом:

 $(\langle a, x \rangle * \langle a, y \rangle) = \langle a, x \circ y \rangle$ — траектория не проходит через точку переключения;

 $(\langle 0, x \rangle * \langle 0, y \rangle) = \langle 1, x \circ y \rangle$ — траектория приходит в точку переключения;

 $(\langle 1, x \rangle * \langle 1, y \rangle) = \langle 0, x \circ y \rangle$ — траектория приходит в точку переключения;

0 — в остальных случаях.

В G_{stp} определение следующее:

 $(\langle a, x \rangle * \langle a, y \rangle) = \langle a, x \circ y \rangle$ — траектория не проходит через точку переключения

$$(\langle 0, x \rangle * \langle 0, y \rangle) = \langle 1, x \circ z \rangle;$$

 $x \circ z$, z — первая точка переключения на траектории ($\langle 1, x \rangle * \langle 1, y \rangle$) = $\langle 0, x \circ z \rangle$;

 $x \circ z$, z — первая точка переключения на траектории;

0 — в остальных случаях.

Из этих алгебр корректно дискретизируется только вторая, поскольку в первой требуется переключение точно в заданный момент. Это соответствует тому, что в конструктивном анализе переключение точно в заданный момент нереализуемо.

Таким образом, мы приближаемся к другому вычислительному подходу к логико-динамическим системам, проанализированным, в частности, в работах [37, 38]. ◆

Качественные состояния даже необратимых систем с бесконечным числом особых точек в нынешней теории систем сплошь и рядом описываются комплексами групп. В качестве примера весьма изощренных конструкций такого рода, в совокупности с развиваемой здесь техникой дающих возможность описать системы как алгебры, можно привести как описание метода [39] и как его приложение [40].

ЗАКЛЮЧЕНИЕ: ВЫВОДЫ И НАПРАВЛЕНИЯ ДАЛЬНЕЙШИХ ИССЛЕДОВАНИЙ

Отметим, что при общем высоком уровне понятий информатика принципиально отставала от теории динамических систем в уровне абстрактности описания систем. Представляете себе, если бы в теории систем рассматривались не общие уравнения и операторы над фазовыми пространствами, а системы из рычагов, шестеренок и т. п. Переход к GAPS позволил избавиться в теории программирования от «шестеренок»: конкретных операторов алгоритмических языков.

Далее, GAPS показали свою способность описывать классы программных систем, различающиеся по свойствам их функционалов. В статье [33] охарактеризованы классы обратимых программ, полуобратимых программ (возможно взятие назад действия, но не его предотвращение), завершающихся за конечное время автоматных программ. Тем самым алгебраический аппарат еще раз продемонстрировал свою общность и гибкость.

Развитие алгебр программ потребовало исследования ранее не изучавшихся алгебраических структур, прежде всего, неассоциативных алгебр с композицией. Их применение к конкретным классам программ и первые опыты по описанию динамических систем поставили несколько проблем.

Прежде всего, в чистом виде алгебры обратимых вычислений описывают ограниченный класс систем. В реальных программах обязательно присутствуют необратимые действия, например, задание исходных данных (инициализация системы) и чтение получившихся результатов. Далее, класс описываемых систем резко расширяется и становится перспективным для описания многих реальных задач, если рассматривать почти всюду обратимые вычисления, которые обратимы на больших участках, а на стыках между ними могут делаться необратимые действия и приниматься необратимые решения. В этом случае тепловой предел все равно обходится, а выразительные возможности принципиально усиливаются. Но почти всюду обратимые алгебры — еще один неисследованный класс алгебраических систем. Их исследование актуальная и практически важная задача.

Далее, выразительные средства алгебр мощны для решения сложнейших качественных задач (как показано, например, в работах [39, 40]). Но строящиеся в теории комплексы групп непрерывны. Нет (не требуется для чистой теории) никаких методов их дискретизации и сведения к конечным структурам. Построение конечных приближений непрерывных алгебр — столь же принципиальная задача, какой явилось в свое время осознание топологического характера приближений непрерывных пространств дискретными сетками. Поэтому теория приближений групп, полугрупп и программных алгебр — новое направление, открываемое алгебраическим подходом.

И, наконец, операции конструирования алгебр из простых компонентов не сводятся к широко применяемым в теоретической алгебре прямым, подпрямым и полупрямым произведениям. А переход к фактор-структурам слишком часто разрушает вычислимость. Развитие новых способов пос-



троения алгебр (прежде всего, конечных, но и бесконечные тоже важны) — еще одно направление, исследования в котором необходимы.

Предварительные варианты данной работы докладывались на конференции PACO'2012 [41], школе молодых ученых ИПУ РАН и на семинаре ИПМ РАН.

ЛИТЕРАТУРА

- 1. *Глушков В. М.* Синтез цифровых автоматов. М.: Физматгиз, 1962. 476 с.
- Глушков В. М., Цейтлин Г. Е., Ющенко Е. Л. Алгебра, языки, программирование. Киев: Наукова думка, 1978. 320 с.
- 3. *Иванов П. М.* Алгебраическое моделирование сложных систем. М.: Наука, 1996. 274 с.
- 4. *Maurer W. D.* A theory of computer instructions // Journal of the ACM. 1966. Vol. 13, N 2. P. 226—235.
- Alban Ponse and Mark B. van der Zwaag. An Introduction to Program and Thread Algebra. — LNCS 3988. — 2006. — P. 445—488.
- Bergstra J. A., Bethke I., Ponse A. Program Algebra and Thread Algebra. — Amsterdam, 2006. — 114 p.
- Митиелл Дж. Основания языков программирования. М. — Ижевск: РХД, 2009. (Mitchell J.C. Foundation for programming languages, MIT, 1996.)
- Мальцев А.И. Алгебраические системы. М.: Наука, 1970. 392 с.
- 9. *Von Neumann J.* The Role of High and of Extremely High Complication, Fourth University of Illinois lecture, in Theory of Self-Reproducing Automata / A.W. Burks (ed.). Champaign, Illinois: University of Illinois Press, 1966. P. 64—73.
- Landauer R. Irreversibility and Heat Generation in the Computing Process // IBM J. Res. Develop. 1961. Vol. 5, N 3. P. 183—191.
- Halliday D. and Resnick R. The Fundamentals of Physics / Extended Third Edition. N.-Y.: John Wiley and Sons, 1988. 947 p.
- 12. *De Benedictis E.* Will Moore's Law be Sufficient // Proc. of the 2004 ACM/IEEE Conference on Supercomputing, 2004.
- 13. *De Benedictis Erik P.* How to Reach Zettaflops, SAND2008—2492C, Sanda National laboratories, 2008.
- 14. Izydorczyk J. Three steps to the thermal noise death of Moore's law // IEEE Trans. on Very Large Scale Integration (VLSI) Systems. 2010. Vol. 18, N 1. P. 161—165.
- Frank M.P. Introduction to Reversible Computing: Motivation, Progress, and Challenges // Proc. of the 2nd Conference on Computing Frontiers. — P. 385—390.
- 16. Frank, M.P. Towards a More General Model of Reversible Logic Hardware. Superconducting Electronics Approaching the Landauer Limit and Reversibility (SEALeR) Workshop, Annapolis, MD, March 15—16, 2012. Invited talk.
- 17. Bennett C.H. Logical reversibility of computation // IBM J. Res. Develop. 1973. Vol. 17, N 6. P. 525—532.
- Шубников А.В., Копцик В.А. Симметрия в науке и искусстве / Изд. 2-е, перераб. и доп. М., 1972.
- Егоров-Тисменко Ю.К., Литвинская Г.П., Загальская Ю.Г. Кристаллография. — М.: Изд-во МГУ, 1992.
- 20. *Егоров-Тисменко Ю.К., Литвинская Г П.* Теория симметрии кристаллов. М.: ГЕОС, 2000.

- Копцик В.А. Шубниковские группы. М.: Изд-во МГУ, 1966
- 22. Якушин А.А. Кризис среднего возраста? Свободное программное обеспечение в высшей школе. М.: Альт-Линукс, 2013. С. 17—19.
- Chaitin G.J. Information-theoretic limitations of formal systems // Journal of the ACM. 1974. Vol. 21, N 3. P. 403—424.
- 24. *Колмогоров А.Н.* Три подхода к определению понятия «количество информации» // Проблемы передачи информации. 1965. Т. 1, вып. 1. С. 3—11.
- Nepejvoda N.N. Abstract Incompleteness Theorems and Their Influence in Methodology // Studia Humana. — 2012. — Vol. 1: 3/4. — P. 43—58.
- Гильберт Д. Избранные труды: в 2 т. / Под ред. А. Н. Паршина. Т. 1: Теория инвариантов. Теория чисел. Алгебра. Геометрия. Основания математики. М.: Факториал, 1998. 575 с. ISBN 5—88688—029—1.
- 27. *Пойа Д.* Как решать задачу. Пособие для учителей. М.: Учпедгиз, 1959. 208 с.
- Wang Hao. Undecidable sentences generated by semantic paradoxes // Journal of Symbolic Logic. 1955. Vol. 20. P. 31—43.
- Оревков В.П. Нижние оценки увеличения сложности выводов после устранения сечений. Зап. научн. сем. ЛОМИ. 1979. Т. 88. С. 137—162.
- 30. Schoenfinkel M. Uber die Bausteine der mathematischen Logik // Mathematische Annalen 1924. B. 92. S. 305—316.
- 31. *Барендреет X.* Ламбда-исчисление. Его синтаксис и семантика. М.: Мир, 1985. 606 с.
- 32. *Косиков С.В.* Информационные системы: категорный подход. М.: АО «Центр Юр ИнфоР», 2005. 92 с.
- Непейвода Н.Н. Абстрактные алгебры различных классов программ // Аппликативные вычислительные системы. — М.: ЮрИнфоР-МГУ, 2012. — С. 103—128.
- URL: http://www.muppetlabs.com/~breadbox/bf/ (дата обрашения 23.10.2013).
- Buhm C., Dezani-Ciancaglini M. Can syntax be ignored during translation? In: Nivat M. (ed.) Automata, languages and programming. — North-Holland, Amsterdam, 1972. — P. 197—207.
- Атаманов В.В. Алгебраические модели программ, не содержащие условий и операторов. — URL: https://edu.botik. ru/mod—erl/upload/a2eb05b1b73c797396e9262f09e9d40f.pdf (дата обращения 23.10.2013).
- 37. Васильев С.Н., Жерлов А.К., Федосеев Е.А., Федунов Б.Е. Интеллектное управление динамическими системами. М.: Физматлит, 2000. 352 с.
- 38. Васильев С.Н., Косов А.А. Анализ динамики гибридных систем с помощью общих функций Ляпунова и множественных гомоморфизмов // Автоматика и телемеханика. 2011. № 6. С. 27—47.
- Krasil'shchik I.S., Lychagin V.V., Vinogradov A.M. Geometry of Jet Spaces and Nonlinear Partial Differential Equations. — N.-Y.: Gordon and Breach, 1986.
- Lychagin V., Yumaguzhin V. Cohomological uniqueness of the Cauchy problem solutions for the Einstein equation // Journal of Geometry and Physics. — 2012. — Vol. 62. — P. 2099—2120.
- Непейвода Н.Н. От численного моделирования к алгебраическому. Тр. конф. PACO'2012 / ИПУ РАН. М., 2012. Т. 1. С. 93—103.

Статья представлена к публикации членом редколлегии академиком РАН С.Н. Васильевым.

Николай Николаевич Непейвода — д-р физ.-мат. наук, гл. науч. сотрудник Института программных систем им. А.К. Айламазяна РАН, г. Переславль-Залесский,

(48535) 6-80-68,
перејуоdann@gmail.com.

14



П.А. Новиков

БОЛЬШИЕ ДАННЫЕ: ОТ БРАГЕ — К НЬЮТОНУ

Проанализированы и структурированы вызовы, которые технологии оперирования большими данными ставят перед разработчиками соответствующих технических средств, специалистами в области прикладной математики и искусственного интеллекта и учеными-предметниками из различных прикладных областей.

Ключевые слова: большие данные, анализ данных, большое управление.

Природа устроена просто. Надо лишь уметь находить надежные средства раскрытия этой осложненной подробностями простоты.

Э. Резерфорд

ВВЕДЕНИЕ

Термин «большие данные», обозначающий неструктурированные данные, объем которых превосходит существующие возможности оперирования ими в требуемые сроки, появился совсем недавно [1], тем не менее, он уже стал сверхпопулярным (запрос в Google возвращает десятки миллионов ссылок) как среди специалистов в ІТ-сфере, так и среди ученых, бизнес-аналитиков и многих других. Какие возможности и опасности несут большие данные? Какие вызовы они формулируют и какие проблемы ставят перед учеными, специалистами в различных предметных областях, системой образования?

1. БОЛЬШИЕ ДАННЫЕ. БОЛЬШАЯ АНАЛИТИКА. БОЛЬШАЯ ВИЗУАЛИЗАЦИЯ

Большие данные (Big Data, первое упоминание, — по-видимому, в специальном выпуске журнала «Nature» [1]) в информационных технологиях — направление в науке и практике, связанное с разработкой и применением методов и средств оперирования большими объемами неструктурированных данных.

Oперирование Big Data включает их¹:

- сбор (получение);
- передачу;
- ¹ Иногда говорят о конструкции «4D» выявление (Discovery), отбор (Discrimination), переработка (Distillation), доведение в нужном представлении (Delivery/Dissemination).

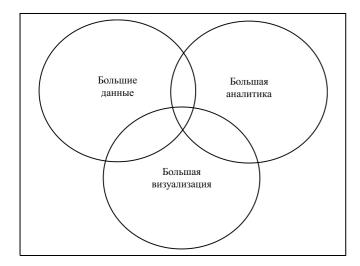
- хранение (включая запись и извлечение);
- обработку (преобразование, моделирование, вычисления и анализ данных);
- использование (включая визуализацию) в практической, научной, образовательной и других видах человеческой деятельности.

Иногда «большими данными» в узком смысле называют только технологии сбора, передачи и хранения Big Data. Тогда обработку больших данных, включая построение и анализ моделей на их основе, называют большой аналитикой (в том числе и большие вычисления), а визуализацию (учитывающую когнитивные возможности пользователя) соответствующих результатов — большой визуализацией (рис. 1).

Универсальный цикл оперирования большими (да и, вообще, любыми) данными приведен на рис. 2. Ключевые элементы в данном цикле — объект и субъект («потребитель»), которому требуются знания о состоянии (и закономерностях его изменения) первого. Но между данными, собираемыми об объекте, и знаниями, необходимыми субъекту, иногда лежит целая «пропасть». Первичные данные должны быть предобработаны — превращены в более или менее структурированную информацию, из которой в зависимости от задачи, стоящей перед субъектом, должны быть извлечены требуемые знания.

Эти знания, в частности, могут быть использованы субъектом для управления объектом — осуществления целенаправленных воздействий, обеспечивающих требуемое его поведение. В частном случае (при неодушевленном субъекте) управление может быть автоматическим. Наверное, скоро





#Рис. 1. «Большая триада»²

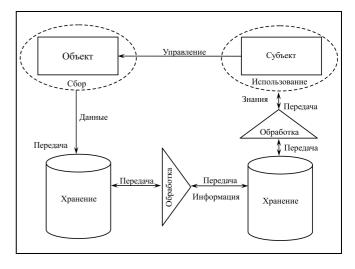
в обиход войдет термин «большое управление³» (Big Control) как управление на основе больших данных, большой аналитики и, быть может, большой визуализации⁴.

Качественный анализ огромного потока текущих публикаций по Big Data позволяет сделать субъективную (авторскую) экспертную оценку текущего распределения внимания исследователей и разработчиков (но не пользователей!) к проблемам оперирования Big Data, представленную на рис. 3.

Другими словами, подавляющее большинство усилий в области Big Data направлено на разработку технологий сбора, передачи, хранения и предобработки больших данных, в то время как большой аналитике и визуализации уделяется гораздоменьшее внимание.

2. ЦИВИЛИЗАЦИОННЫЕ ПРОБЛЕМЫ

Можно ли считать сложившееся состояние дел (см. рис. 3) нормальным? С одной стороны — да. Ведь эволюционное развитие технологий шло



Puc. 2. Универсальный цикл оперирования Big Data

именно этим путем, да и для того, чтобы анализировать и визуализировать данные, их надо сначала собрать и сохранить (естественно, таким образом, чтобы иметь возможность быстрого доступа и обработки). С другой стороны, существующий «перекос» вызван отчасти тем, что сегодня человечество, хотя и сознает, что, наверное, любые данные небесполезны, но не до конца понимает, что делать и как использовать их нарастающую лавину.

Проблема эта не нова, так как за последнее время возник целый класс подобных проблем, и носят они цивилизационный характер. Условно их можно назвать *проблемами опережающего развития технологий*: если рассмотреть соотношение между наукой, технологиями и практикой (рис. 4), то в

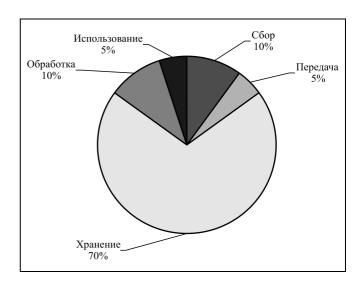


Рис. 3. Текущее распределение внимания исследователей и разработчиков к проблемам оперирования Big Data

16

² Еще одну сверхмодную триаду — большие данные, высокопроизводительные вычисления и облачные технологии — обсуждать мы не будем.

³ Справедливости ради отметим, что специалисты по теории управления в последние полтора десятилетия все чаще говорят о совместном решении задач управления, вычислений и связи — так называемая проблема С³ (Control, Computation, Communication) — решения задач синтеза управляющих воздействий в реальном времени с учетом задержек в каналах связи и временных затрат на обработку информации (включая вычисления). Кроме того, существует устойчивое словосочетание «управление большими системами» (Large-scale Systems Control), однако большие данные могут порождаться и «маленькими» в этом смысле системами.

⁴ Возможна и другая трактовка термина «большое управление» — как управление процессами оперирования Від Data, что представляет собой самостоятельную и нетривиальную проблему.



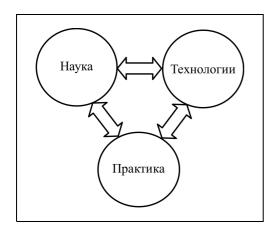


Рис. 4. Наука, технологии, практика

различные периоды развития человечества иногда наука инициировала развитие и последующее внедрение тех или иных технологий, а иногда последовательность была (и является в наше время!) «обратной».

Действительно, обратимся к истории. Сто лет, начиная примерно с середины XIX в., наблюдались триумфальное развитие науки и основывающийся на нем взрывной рост технологий (наука опережала и формировала технологии, которые массово внедрялись в практику; отметим, что до этого, как правило, опережающий запрос к науке формировала практика) — электричество, связь, атомная энергетика, электроника и т. д., причем технологий, понятных и доступных обывателю⁵. Затем, где-то с последней трети XX в. ситуация стала изменяться — уже накопленных фундаментальной наукой результатов хватало, чтобы обеспечить рост технологий (технологии начали в чем-то «опережать» науку), что привело к снижению «спроса на науку» со стороны последних (исключение представляют собой, наверное, лишь живые системы).

Но технологии продолжают развиваться, причем все более быстрыми темпами. Более того, например, в последнее десятилетие скорость развития информационно-коммуникационных технологий (ИКТ) стала опережать и практику⁶, в том

числе способности человечества по осознанию новых технологических возможностей, перспектив развития и соответствующих угроз⁷. Другими словами, на пороге третьего тысячелетия произошел определенный перелом — до сих пор человечество целенаправленно развивало науку и технологии под свои потребности, сейчас же технологии все больше опережающе навязывают направления, ограничения и условия развития, причем как на уровне отдельной личности, так и на уровне государств и человечества в целом⁸. Именно этот эффект мы и называем «опережающим развитием технологий». Осознать соответствующие цивилизационные проблемы и научиться реагировать на них еще предстоит.

То же самое происходит с большими данными — человечество обрело технологические возможности накопления гигантских объемов данных, но не готово их обрабатывать и использовать, причем основная проблема заключается не в непонимании того, как их обрабатывать, а в том, зачем это делать. Для того чтобы понять, какие вызовы стоят перед учеными и инженерами, обсудим кратко, какие данные являются «большими», где они возникают, как сегодня применяются, как повысить эффективность их использования в будущем и чем в этом может помочь наука.

3. ИСТОЧНИКИ И «ПОТРЕБИТЕЛИ» ВІ**G** DATA

Среди них такие крупные группы, как:

— наука — астрономия и астрофизика, метеорология, ядерная физика, физика высоких энергий, геоинформационные и навигационные системы, дистанционное зондирование Земли, геология и геофизика, аэродинамика и гидродинамика, генетика, биохимия и биология и др.

⁵ Забавно, но героями фантастической литературы о технологических рывках в развитии, будь то благодаря изоляции (например, Сайрес Смит в «Таинственном острове» Ж. Верна) или переносу в прошлое (например, герой М. Твена из «Янки из Коннектикута при дворе короля Артура»), выступают, как правило, инженеры или просто обыватели, но не ученые.

⁶ В качестве позитива, необходимо признать, что развитие ИКТ существенно стимулировало развитие прикладной математики, в том числе таких разделов «сетевой математики», как случайные графы, графы большой размерности, сетевые игры и многие другие — см. обзоры в работах [2—8].

⁷ Отдельные проявления этого явления встречались и раньше — например, об этической ответственности ученых задумывались и А. Нобель, и участники Манхэттенского проекта и др.

⁸ Проблемы информационной безопасности уже стали всем привычны. Но сейчас пора задуматься не только о технологической (кибербезопасности), но в более широком смысле о социальной, экономической и другой безопасности информационных технологий — защищенности пользователей ИКТ, их групп и общества в целом от информационных воздействий (ярким примером являются социальные медиа, в том числе онлайновые социальные сети [9]). А поскольку все существенные решения (начиная с малого предприятия и заканчивая страной) принимаются на основе информации, которая откуда-то поступает, как-то (и не всегда известным лицу, принимающему решение, образом) передается и обрабатывается, то придется признать и важность сопиально-экономической безопасности ИКТ — защищенности личности, экономики, общества и государства от последствий решений, принимаемых с использованием современных ИКТ (в том числе, систем поддержки принятия экономических, военных, политических и других ре-



- Интернет (в широком смысле) и другие телекоммуникационные системы;
- бизнес, торговля и финансы, а также маркетинг и реклама (включая трейдинг, таргетирование и рекомендательные системы, CRM-системы, RFID радиочастотные идентификаторы, все чаще используемые в торговле, транспорте и логистике и др.);
- мониторинг (гео-, био-, эко-; космический, авиа- и др.);
- безопасность (системы военного назначения, антитеррористическая деятельность и др.);
- электроэнергетика (включая атомную), Smart Grid:
 - медицина;
 - госуслуги и госуправление;
- производство и транспорт (объекты, узлы и агрегаты, системы управления и др.).

Многочисленные примеры приложений Від Data в этих областях можно найти в научно-популярной (иногда даже в «глянцевой») литературе, свободно доступной в Интернете (повторять эти типовые примеры и «пугать» читателя «зетта-» и «йоттабайтами» мы не будем).

Современный уровень автоматизации практически всех перечисленных отраслей таков, что в них большие данные, де факто, автоматически генерируются. Поэтому все чаще задаются вопросом — сколько потоковых данных мы «теряем» (из-за того, что не можем или не успеваем их сохранить или обработать)? Вопрос этот корректен для инженера по ИКТ, но не для ученого и, тем более, не для пользователя результатов обработки Big Data — они бы спросили соответственно: «что мы потеряли существенного» и «что изменилось бы для нас, если бы мы успели все собрать и обработать».

4. КАКИЕ ДАННЫЕ ЯВЛЯЮТСЯ «БОЛЬШИМИ». ВЫЗОВЫ НАУКЕ

Традиционно, большими считаются, как минимум, данные, объем которых превосходит существующие возможности оперирования ими в требуемые сроки. Такое определение несколько «лукаво» — данные, считающиеся «большими» сегодня, перестанут быть таковыми завтра с развитием ме-

тодов и средств работы с ними. Данные, казавшиеся «большими» несколько столетий или даже десятилетий назад (в отсутствии возможности их автоматической обработки), сегодня легко обрабатываются на бытовом компьютере. Соревнование между вычислительными потребностями (гипотетическими) человечества и соответствующими технологическими возможностями существует давно, и, естественно, потребности всегда опережали и будут опережать возможности. И несоответствие между ними служит колоссальным стимулом развития науки — приходится искать более простые (но адекватные) модели, придумывать более эффективные алгоритмы и т. д.

Иногда в определение Big Data добавляют такие их свойства, как 5V — объем (Volume), скорость (Velocity), разнообразие (Variety), достоверность (Veracity) и обоснованность (Validity); или говорят, что от большого объема обычных данных большие данные отличаются наличием большого потока (здесь учитывается и объем, и скорость — объем в единицу времени) неструктурированных 10 данных.

Неструктурированность (в широком понимании) Від Data (текст, видео, аудио, структура коммуникаций и т. п.), действительно, является их характерной чертой и вызовом для прикладной математики, лингвистики, когнитивных наук и искусственного интеллекта — разработка технологий обработки в реальном времени, в том числе с возможностью выявления «скрытой» информации, больших потоков текстовой, аудио-, видео- и другой информации составляется мэйнстрим приложений перечисленных научных областей 12 к ИКТ.

Тем самым, мы наблюдаем прямой (и явный) запрос от технологий к науке. Второй (и столь же явный) «запрос» заключается в адаптации к анализу больших данных традиционных методов статистического анализа, оптимизации и т. п. Более того, помимо адаптации, необходима разработка новых методов, учитывающих специфику Big Data—сейчас модно рекламировать средства аналитики (как правило, бизнес-аналитики) для больших данных—см., например, критический анализ в рабо-

18

⁹ Основная идея использования Big Data заключается в попытке выявить «скрытые закономерности» — найти ответы на нетривиальные вопросы, например: прогноз эпидемий по информации из соцсетей или о продажах в аптеках; задачи диагностики (медицинской и технологической); удержание клиентов благодаря анализу поведения покупателей в магазине по перемещению в пространстве RFID-меток товаров и пр.

¹⁰ Неструктурированность данных может порождаться также их пропусками и/или разномасштабностью (в пространстве и времени — так называемых multi-scale systems) анализируемых явлений и процессов.

¹¹ Эти технологии, в первую очередь, должны предусматривать агрегирование данных (например, фиксация изменений в технологических данных или хранение агрегированных показателей), ведь не всегда нужно использовать все данные (особенно, если они «однородны»).

¹² Математика может хорошо работать со структурированными данными, поэтому преобразование неструктурированных данных в структурированные — отдельная важная задача.



те [10], но список этих средств почти совпадает с хрестоматийным набором статистических инструментов (и даже уже этого набора, так как не все методы применимы в условиях большой размерности). То же самое относится и к:

- методам машинного обучения (нейронные сети, байесовы сети, нечеткий логический вывод и т. п.);
- оптимизационным задачам большой размерности (как «альтернатива», помимо ставших привычными технологий параллельных вычислений, активно развивается распределенная оптимизация, l_1 -оптимизация см., например, обзоры и результаты в работах [11—15]);
- методам дискретной оптимизации (здесь «альтернативой» служит применение мультиагентных программных систем см., например, [16]) и др.

Общее для перечисленных «запросов» технологий к науке заключается в том, что речь идет об адаптации или небольшой модификации известных, уже хорошо зарекомендовавших себя методов. Нужно понимать, что автоматическое построение (с помощью традиционного аппарата 13) модели по сырым данным — в общем случае представляет собой не более чем модную иллюзию 14 — мы придумаем алгоритмы, «напустим» их на большие массивы неструктурированной (и в большинстве случаев нерелевантной) информации и благодаря этому будем принимать более эффективные решения. Подобные заблуждения уже встречались в истории науки — на ранних стадиях развития кибернетики и искусственного интеллекта¹⁵ — и, приведя к множеству разочарований, очень сильно затормозили развитие этих научных направлений. Чудес на свете не бывает: как правило, для получения качественно новых выводов нужна новая модель, новая парадигма (см. работы по методологии науки [17, 18]).

5. ОБЩИЙ ВЫЗОВ

Сложность окружающего нас мира растет не так быстро, как возможности фиксации («измерения») и хранения данных, которые, похоже, опередили возможности человечества по осознанию возможности и целесообразности их использования, т. е., мы «захлебываемся» в данных и судорожно пытаемся придумать, что с ними можно делать.

Но на эту ситуацию можно посмотреть и с другой стороны: основной тезис заключается в следующем — получить большие (сколь угодно большого мыслимого объема) данные можно и достаточно просто (лежащие на поверхности примеры нам дают комбинаторная оптимизация, нелинейная динамика или термодинамика — см. далее), нужно понимать, что с ними делать (Природе нужно задавать правильные вопросы). Более того, можно придумать сколь угодно сложную модель, использующую Big Data, а затем пытаться достичь в ее рамках все более высокой точности. Проблема в том, получим ли мы при этом, кроме массы новых проблем¹⁶, качественно новые результаты. Математикам и физикам давно известно, что увеличение размерности модели и ее «усложнение» (стремление учесть все больше факторов и связей между ними) далеко не всегда ведет к адекватном росту «качества» результатов моделирования, а иногда и вовсе приводит к абсурду 17 .

Рассмотрим ряд примеров.

Пример 1. В книге нобелевского лауреата Г. Саймона [19] рассматривается следующий пример. Предположим, что мы наблюдаем за тем, как муравей движется по песку. Целью муравья может быть стремление минимизировать затраты своей энергии по перемещению из одной точки в другую, поэтому он огибает горки песка, иногда поворачивает назад и т. д. Если мы наблюдаем только проекцию на горизонтальную плоскость траектории муравья, а рельеф, по которому двигался муравей, неизвестен, то объяснить поведение муравья (сложную, петляющую траекторию) довольно непросто. Г. Саймон делает вывод, что наблюдаемое разнообразие и сложность поведения людей объясняются не сложностью принципов принятия ими решений, которые сами по себе просты, а разнообразием ситуаций, в которых принимаются решения. С этим мнением вполне можно согласиться. Действительно, нетривиальные результаты может давать как сложная модель на простых входных данных, так и простая модель на сложных данных. Желательно (в иде-

¹³ Обычно ситуация дополнительно обременена предшествующим опытом исследователя/разработчика и традициями его научной школы — успешное решение некоторой одной задачи приводит к формированию вполне естественного убеждения, что этими же методами (ими и только ими!) можно решить все остальные нерешенные задачи.

¹⁴ Хотя в некоторых случаях увеличение объема данных может (при правильной обработке) дать дополнительную информацию.

¹⁵ Кибернетическая система не может продемонстрировать поведения, отличного от являющегося результатом заложенных в нее алгоритмов (которые могут быть «стохастическими», «недетерминированными» и т. д.), несмотря на кажущуюся генерацию новых знаний или проявления качественно нового («неожиданного») поведения, особенно при взаимодействии нескольких и, тем более, значительного числа элементов.

 $^{^{16}}$ О проблемах адекватности моделей и устойчивости результатов моделирования мы, осознавая их важность, пока забудем.

¹⁷ Не говоря уже о ситуациях, когда в рамках существующих научных парадигм принципиально невозможно моделирование поведения системы на достаточно большом горизонте времени (примером может служить «точное» прогнозирование погоды).



але) уметь получать нетривиальные результаты в рамках простой модели, правильно выбрав для нее релевантные простые входные данные (недаром математики говорят, что простота — критерий истины).

Пример 2. Предположим, что в руки ученых, например, XVIII века волшебным образом попали современный ноутбук и компьютерный томограф (с инструкциями по эксплуатации). Сделав томограмму ноутбука и сохранив ее в последнем, они, вряд ли, анализируя полученные данные (очень подробные и многочисленные) о внутреннем «физическом» устройстве ноутбука, поняли бы хоть что-то о том, как он работает: т. е., правильная парадигма, правильная концептуальная модель являются необходимым (но, к сожалению, не достаточным) условием успеха.

Пример 3. Если взять любую NP-трудную задачу комбинаторной оптимизации [20], например — задачу коммивояжера, то в ней существует порядка n! подлежащих анализу (в общем случае для поиска точного решения) вариантов, что уже для $n \sim 100$ превышает вычислительные возможности человечества. Это свойство NP-трудных задач известно давно - уже не одно десятилетие оно стимулирует специалистов на разработку методов поиска приближенных решений (с оценкой гарантированной «точности») за разумное время. Аналогичным примером могут служить модели нелинейной динамики — результатами «наблюдений» за динамическим хаосом, демонстрируемым даже достаточно простой (небольшой размерности) нелинейной динамической системой, можно занять память всех компьютеров Земли, но новых знаний эти данные содержать не будут.

Пример 4. Хрестоматийным примером из истории физики служит открытие закона всемирного тяготения. Тихо Браге (1546—1601) в течение двух десятилетий регулярно наблюдал за движением планет Солнечной системы. Его записи представляют собой Большие (по тем временам) данные. Иоганн Кеплер (1571—1630), на основе данных Браге, сформулировал свои эмпирические (!) законы движения планет.

Три закона Кеплера агрегировали информацию Браге, и движение каждой конкретной планеты могло быть рассчитано по ним (а не по многотомным записям Браге) с высокой точностью. Другими словами, Браге научился описывать¹⁸ движение планет; Кеплер — описывать и предсказывать это движение. Но законы Кеплера ничего не говорят о том, почему планеты движутся в соответствии с этими законами. Ответ на этот вопрос (т. е. объяснение) дал закон всемирного тяготения И. Ньютона (1643—1727). Вывести законы Кеплера, наверное, мог бы любой современный компьютер (если в него заложить адекватные задаче алгоритмы и исходные данные), а сформулировать закон обратных квадратов — ни один (если в него не заложить соответствующую модель взаимодействия масс). Но, законы Кеплера являются «следствиями» закона тяготения (и могут быть выведены из

него), как, в свою очередь, результаты Браге могут быть получены из законов Кеплера. Таким образом, закон всемирного тяготения «сделал ненужными» (точнее — гносеологически избыточными) и законы Кеплера, и большие данные Браге.

Сегодняшний опыт оперирования большими данными пока свидетельствует, что мы в большинстве случаев находимся на уровне Браге, предпринимая титанические попытки достичь уровня Кеплера. Но качественный скачок возможен только тогда, когда появляются обобщения (уровня Ньютона), радикально упрощающие ситуацию. Повторим: искусство заключается в том, чтобы «задавать Природе правильные вопросы».

Пример 5. Вторым хрестоматийным примером из истории физики служит создание молекулярно-кинетической теории, который свидетельствует, что породить большой поток данных — не проблема, вопрос в том, что мы хотим с этими данными делать и на какие вопросы отвечать.

Рассмотрим следующий мысленный эксперимент — задачу детального описания поведения идеального газа. Предположим, что при нормальных условиях находится один кубический метр воздуха. В нем содержится примерно 10^{25} молекул, движение каждой из которых и их соударения исчерпывающе (в рамках модели идеального газа) описываются кинематикой и динамикой, т. е. с точки зрения физики никаких принципиальных проблем описания их движения и взаимодействия не существует. Каждая молекула за одну секунду испытывает порядка 10^9 столкновений с другими молекулами. Описание поведения такой системы (координаты и скорости всех молекул) в реальном времени породит поток данных порядка 10^{35} байт/с.

Такой поток данных превосходит технологические возможности человечества даже сегодня! Наверное поэтому, еще в середине XIX в. при создании молекулярно-кинетической теории газов физики поняли бессмысленность детального анализа (при четком осознании его принципиальной возможности) и перешли к макроописанию в терминах агрегированных характеристик — температуры, объема, давления, а в дальнейшем в рамках статистической физики — к описанию в терминах вероятностных распределений. А ведь, если бы тогда смогли все «посчитать», остались бы мы без статистической физики! ◆

Завершая рассмотрение примеров, отметим, что большие данные по своему источнику можно условно разделять на естественные и искусственные. В первом случае данные порождает некоторый существующий независимо от нас объект, а мы (как «исследователи») решаем, что и сколько «измерять» и т. д. (см. примеры 1, 2 и 4). Во втором случае источником данных служит модель, которая может порождать большие данные (см. примеры 3 и 5), при этом сложность (поток данных) отчасти управляема и определяется в процессе моделирования.

¹⁸ Напомним основные функции научного познания (в том числе и моделирования) [18]: описание (феноменологическая функция) — объяснение — прогноз (прогностическая функция) — управление (нормативная функция).



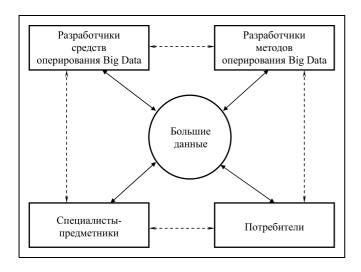


Рис. 5. Субъекты, оперирующие большими данными

«Рецепты». Можно выделить четыре большие группы субъектов (рис. 5), оперирующих (явно или косвенно) большими данными в своей профессиональной (научной и/или практической) деятельности:

- разработчики средств оперирования Big Data (производители соответствующего программного и аппаратного обеспечения, а также его продавцы, консультанты, интеграторы и др.);
- разработчики методов оперирования Big Data (специалисты по прикладной математике и компьютерным наукам);
- специалисты-предметники ученые, исследующие реальные объекты (или их модели), служащие источниками больших данных;
- потребители, использующие или собирающиеся использовать результаты анализа больших данных в своей практической деятельности.

Каждый из представителей перечисленных групп взаимодействует с другими (см. штриховой контур на рис. 5) — нормативное (к которому надо стремиться) разделение «зон ответственности» приведено на рис. 6, где толщина стрелок условно отражает степень вовлеченности).

Не претендуя на конструктивность, даже исходя из здравого смысла, можно сформулировать следующие общие «рецепты» для перечисленных групп субъектов.

Для разработчиков средств оперирования Big Data: продавать решения в области Big Data (в том числе и аналитические) станет все труднее, если их не пополнять новыми адекватными математическими методами и не предусматривать возможность работы потребителя в тесном взаимодействии с разработчиками методов и специалистами-предметниками.

Для собратьев-математиков: актуален запрос на адаптацию известных и развитие новых (в первую очередь — обладающих линейной сложностью!) методов обработки больших потоков неструктурированных данных, которые представляют собой хороший полигон тестирования новых моделей, методов и алгоритмов (желательно за счет разработчиков и/или потребителей).

Для специалистов-предметников: технологии Big Data дают новые возможности получения и хранения огромных массивов «экспериментальной» информации, постановки так называемых вычислительных экспериментов, а развиваемые методы прикладной математики дают возможность системной генерации и быстрой верификации гипотез (выявления скрытых закономерностей).

Для потребителей: дорогие технологии сбора и хранения Big Data вряд ли дадут эффект без привлечения специалистов по методам и по предмету

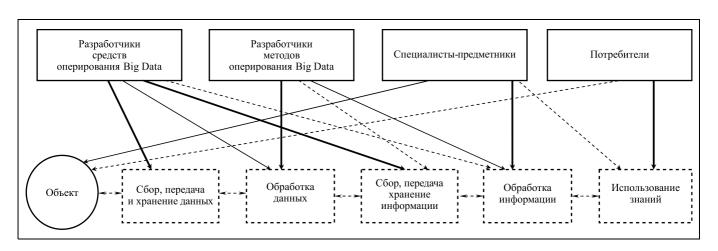


Рис. 6. Разделение «зон ответственности»



(при обязательном четком понимании, на какие вопросы потребитель хочет получить ответы от Big $Data^{19}$).

Некоторые опасности. Помимо упомянутых необходимости поиска адекватных простых моделей и настораживающего тренда опережающего развития технологий, можно предположить будущую актуальность следующих проблем (их список неструктурирован и открыт).

- *Информационная безопасность Big Data*. Здесь потребуется и адаптация известных, и разработка принципиально новых методов и средств.
- Энергетическая эффективность Big Data. Уже сейчас центры обработки данных представляют собой существенный класс потребителей электроэнергии. Чем больше данных мы хотим обрабатывать, тем больше потребуется энергии.
- Принцип дополнительности давно известен в физике измерение изменяет состояние системы. А как обстоит дело в социальных системах, элементы которых (люди) активны обладают своими интересами и предпочтениями, способны самостоятельно выбирать свои действия и пр. [21]?

Одно из проявлений заключается в так называемом манипулировании информацией. В теории коллективного выбора давно известно, что активный субъект сообщает информацию, прогнозируя результаты ее использования, и в общем случае не будет сообщать достоверную информацию [3, 22].

Другой пример — так называемый активный прогноз, когда система меняет свое поведение на основании новых знаний, полученных о себе [23].

Снимаются или усугубляются эти и подобные (примеры — краудсорсинг [24], конформное поведение [25] и др. — см. работу [3]) проблемы в области Big Data?

• Если уж был упомянут принцип дополнительности, то необходимо вспомнить и о «принципе неопределенности» в следующем (гносеологическом) варианте [18]: текущий уровень развития науки характеризуется определенными совместными ограничениями на «обоснованность» результатов и их области применимости. Применительно к Від Data принцип неопределенности означает, что существует рациональный баланс между степенью детальности описания состояния интересующей нас системы и обоснованностью тех результатов и выводов, которые мы хотим сделать на основании этого описания.

Традиционно при построении и эксплуатации информационных систем (будь то корпоративные системы или системы поддержки госуслуг, межведомственного документооборота и т. п.) считается, что содержащаяся в них информация должна быть максимально полной, унифицированной и общедоступной (с учетом разделения прав доступа). Но ведь возможно показывать каждому пользователю реальность, искаженную в своем «кривом зеркале» — создавать для каждого свою индивидуальную информационную картину²⁰, осуществляя тем самым информационное управление [21, 23]. Стремиться к этому или бороться с этим в области Big Data?

ВМЕСТО ЗАКЛЮЧЕНИЯ

Итак, данные всегда были «большими». Интенсивно появляются новые, все более совершенные инструменты их сбора, хранения и обработки. Хотелось бы уметь делать это эффективно и в реальном времени — для этого необходимо развитие соответствующих отраслей прикладной математики и компьютерных наук (что является актуальным запросом от технологий и практики к современной науке). Хочется надеяться, что российские ученые (специалисты по ИКТ, прикладной математике, управлению) не окажутся в стороне от этой мировой тенденции — сейчас интенсивно появляются новые журналы [26—28 и др.] и конференции [29—32 и др.], организуемые ведущими мировыми ассоциациями (ІЕЕЕ и др.), посвященные проблематике Big Data; все больше государственных и коммерческих грантов выделяется по этой тематике.

Также необходима массовая подготовка специалистов по большим данным, большой аналитике и большой визуализации (со специализацией в конкретных предметных областях).

Но этого мало — необходимо получение знаний (в рамках соответствующих отраслей науки) и развитие моделей, позволяющих компактно и адекватно (с учетом решаемой задачи) описывать интересующие нас явления и процессы. Другими словами, в каждой из областей возможных приложений Big Data желательно стремиться сделать шаг от «Браге» до соответствующего «Ньютона», иначе мы обречены оперировать частностями, за деревьями не видя леса (см. также цитату, вынесенную в эпиграф).

22

 $^{^{19}}$ Правда, можно «складировать» данные на всякий случай на будущее — вдруг когда-то возникнет мысль, что с ними делать — захочется, например, проверить ту или иную гипотезу, а данные уже под рукой.

 $^{^{20}}$ Как минимум — часть «объективной» картины (правду, только правду, но не всю правду), как максимум — произвольную непротиворечивую систему представлений о реальности.



Кроме того, опережающее развитие технологий стало общецивилизационной проблемой, которая должна учитываться как учеными и инженерами, в том числе, в области больших данных, так и потребителями разрабатываемых ими методов и средств.

ЛИТЕРАТУРА

- 1. *Nature.* 2008. September 3 (Special Issue).
- Dorogovtsev S. Lectures on Complex Networks. Oxford: Oxford University Press, 2010.
- 3. *Губанов Д.А., Коргин Н.А., Новиков Д.А., Райков А.Н.* Сетевая экспертиза. М.: Эгвес, 2010.
- 4. *Jackson M.* Social and Economic Networks. Princeton: Princeton Univ. Press, 2010.
- Polyak B.T., Tremba A.A. Regularization-based Solution of the PageRank Problem for Large Matrices // Automation and Remote Control. — 2012. — Vol. 73, N 11. — P. 1877—1894.
- Райгородский А.М. Модели случайных графов и их применения // Тр. МФТИ. 2010. Т. 2, № 4. С. 130—140.
- 7. Сетевые модели в управлении // Управление большими системами (спец. вып.). 2010. Вып. 30.1.
- Словохотов Ю.Л. Физика и социофизика. Ч. 1 3 // Проблемы управления. 2012. № 1. С. 2—20; № 2. С. 2—31; № 3. С. 2—34.
- 9. *Губанов Д.А., Новиков Д.А., Чхартишвили А.Г.* Социальные сети: модели информационного влияния, управления и противоборства. М.: Физматлит, 2010.
- 10. Викторов Д. Большая проблема Big Data в России // Компьютерра. 7 февраля 2013. URL: http://www.computerra.ru/52840/bolshaya-problema-big-data-v-rossii/ (дата обращения: 8.09.2013).
- Algorithmic Game Theory / Ed. N. Nisan, T. Roughgarden, E. Tardos, and V. Vazirani). — N.-Y.: Cambridge University Press, 2009.
- 12. Boyd S., Parikh N., Chu E., et al. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers // Foundations and Trends in Machine Learning. — 2011, — N 3 (1). — P. 1—122.
- Granichin O.N., Pavlenko D.V. Randomization of Data Acquisition and I1-Optimization (Recognition with Compression) // Automation and Remote Control. 2010. Vol. 71, N 11. P. 2259—2282.

- Nesterov Y. Efficiency of Coordinate Descent Methods on Huge-scale Optimization Problems // SIAM Journal on Optimization. — 2012. — Vol. 22, N 2. — P. 341—362.
- Shoham Y., Leyton-Brown K. Multiagent systems: Algorithmic, Game-Theoretical and Logical Foundations. — Cambridge: Cambridge University Press, 2009.
- Wooldridge M. An Introduction to MultiAgent Systems. N.-Y.: John Wiley & Sons, 2002.
- 17. *Kuhn T.* The Structure of Scientific Revolutions. Chicago: University of Chicago Press, 1962.
- Novikov A., Novikov D. Research Methodology: From Philosophy of Science to Research Design. Leiden: CRC Press, 2013.
- 19. Simon H. The Sciences of the Artificial / 3rd Edition. The MIT Press, 1996.
- Garey M., Johnson D. Computers and Intractability: A Guide to the Theory of NP-Completeness. — San Fancisco. W. H. Freeman and company, 1979. — 14 p.
- 21. *Novikov D.* Theory of Control in Organizations. N.-Y.: Nova Science Publishers, 2013.
- 22. Aizerman M., Aleskerov F. Theory of Choice. Amsterdam: Elsevier, 1995.
- 23. Novikov D., Chkhartishvili A. Reflexion and Control: Mathematical Models. Leiden: CRC Press, 2014.
- 24. Surowiecki J. The Wisdom of Crowds. N.-Y.: Anchor, 2005.
- Breer V., Novikov D. Models of Mob Control // Automation and Remote Control. — 2013. — Vol. 74 (in press).
- 26. *URL*: http://www.journalofbigdata.com (дата обращения: 8.09.2012).
- 27. URL: http://www.hipore.com/ijbd (дата обращения: 8.09.2012).
- 28. URL: http://cci.drexel.edu/bigdata/bigdata2013 (дата обращения: 8.09.2012).
- 29. *URL*: http://www.ieeebigdata.org/2014 (дата обращения: 8.09.2012).
- 30. *URL*: http://www.liebertpub.com/overview/big-data/611 (дата обращения: 8.09.2012).
- 31. URL: http://www.swinflow.org/confs/bdds2013 (дата обращения: 8.09.2012).
- 32. *URL*: https://theinnovationenterprise.com/summits/big-data-innovation-boston (дата обращения: 8.09.2012).

Статья представлена к публикации членом редколлегии Ф.Т. Алескеровым.

Дмитрий Александрович Новиков — чл.-корр. РАН, зам. директора, Институт проблем управления им. В.А. Трапезникова РАН, г. Москва,

☎ (495) 334-75-69, ⊠ novikov@ipu.ru.



Если Вы не успели подписаться

Подписку на журнал «Проблемы управления» можно оформить через редакцию с любого месяца, при этом почтовые расходы редакция берет на себя. Позвоните по телефону (495) 330-42-66 или обратитесь по e-mail: datchik@ipu.ru, и подписка будет оформлена за один день. Подписаться на журнал можно и в любом отделении связи (подписной индекс 81708 в каталоге Роспечати и 38006 в объединенном каталоге «Пресса России»). Отдельные номера редакция высылает по первому требованию.



УДК 37.018.43:004.75:519.872

УПРАВЛЕНИЕ ПОТОКАМИ ЗАПРОСОВ ПРИ ДОСТУПЕ К ШИРОКОПОЛОСНЫМ МУЛЬТИМЕДИЙНЫМ ОБРАЗОВАТЕЛЬНЫМ РЕСУРСАМ СИСТЕМЫ ДИСТАНЦИОННОГО ОБУЧЕНИЯ¹

И.П. Болодурина, Д.И. Парфенов

Представлены модель обслуживания заявок пользователей широкополосных мультимедийных образовательных ресурсов и модель доступа к данным хранилища гибридной облачной системы. Описано их применение для повышения эффективности использования вычислительных ресурсов в системе дистанционного обучения путем распределения потоков запросов и балансировки нагрузки между узлами облака с помощью разработанного дополнительного управляющего модуля для контроллера системы «OpenStack».

Ключевые слова: облачные вычисления, мультимедийные образовательные ресурсы, распределение нагрузки, гибридная облачная система «OpenStack».

ВВЕДЕНИЕ

Развитие информационных технологий и возрастающие потоки передаваемой информации требуют построения масштабируемых решений, способных обслуживать большое количество одновременных запросов пользователей. Одна из возникающих при этом задач состоит в обеспечении гибкого управления и эффективного использования выделенных для этих целей вычислительных ресурсов. Наиболее перспективное направление ее решения заключается в применении технологий облачных вычислений. Они позволяют унифицировать доступ не только к конечным данным, но и к ресурсу в целом, что очень важно для построения приложений, требующих поддержания высокого качества обслуживания и круглосуточной доступности сервиса. На рынке облачных вычислений присутствуют не только проприетарные решения, такие как «VMware ESX», «Xen» и др., но и хорошо документированные комплексы с открытым исходным кодом, такие как «OpenStack» [1]. Одним из направлений развития таких сервисов — создание мультимедийных ресурсов, осуществляющих трансляцию видео как в режиме реального времени, так и по запросу пользователя. Наиболее широкое распространение эти ресурсы получили при организации обучения с помощью дистанционных образовательных технологий. Как правило, образовательные ресурсы представляют собой многокомпонентную систему, решающую одновременно множество вычислительных задач для обеспечения работы различных информационных сервисов. Поэтому возникает необходимость разработки особых подходов как в организации управления инфраструктурой в целом, так и каждым компонентом в отдельности.

В результате исследования нами установлен ряд особенностей потребления программно-аппаратных ресурсов, используемых для обеспечения работы системы дистанционного обучения (СДО) Оренбургского государственного университета:

- нагрузка на ключевые ресурсы носит периодический и неравномерный характер;
- пропускная способность внешних каналов связи ограничена и не позволяет предоставлять доступ к мультимедийному контенту с должным качеством обслуживания;
- одновременно происходят обращения к нескольким типам ресурсов;
- интенсивность обращения к каждому ресурсу может изменяться в зависимости от внешних условий:
- ввиду отсутствия распределения нагрузки между ресурсами при пиковой нагрузке оборудование не всегда позволяет обслужить все запросы;

¹ Работа выполнена при финансовой поддержке Федеральной целевой программы «Научные и научно-педагогические кадры инновационной России» (гранты № 14.В37.21.1881 и 14.132.21.1801), а также РФФИ (гранты № 13-07-00198 № 13-07-00198 и 13-01-97050).



— до 90 % нагрузки предопределены, поскольку для доступа к ресурсам используется предварительная регистрация.

Отметим, что 80 % ресурсов востребованы лишь 20 % времени работы сервисов.

В существующих решениях на базе облачных сервисов применяется универсальный подход к организации доступа к размещаемым в них ресурсам. Особенности каждого сервиса не учитываются, что приводит к увеличению потребляемых ресурсов и неэффективному их использованию. Цель нашего исследования заключается в определении ключевых параметров, влияющих на работу каждого из ресурсов, задействованных при построении системы дистанционного обучения, и оптимизации потребления ресурсов с учетом решаемой ими вычислительной задачи.

1. СХЕМА ПРЕДОСТАВЛЕНИЯ ДОСТУПА К МУЛЬТИМЕДИЙНЫМ ОБРАЗОВАТЕЛЬНЫМ РЕСУРСАМ ПОЛЬЗОВАТЕЛЯМ СИСТЕМЫ ДИСТАНЦИОННОГО ОБУЧЕНИЯ

Ограничение пропускной способности выходного канала связи создает трудности применения широкополосных мультимедийных образовательных услуг в дистанционном обучении, особенно для пользователей, обращающихся к веб-приложениям из сети Интернет. Для анализа эффективности имеющейся архитектуры нами разработана трехуровневая модель подсистем СДО: уровень — подсистема контроля знаний, второй — подсистема предоставления учебно-методических комплексов (электронная библиотека) и третий — подсистема трансляции и публикации видео- и аудиоматериалов (видеопортал дистанционного обучения).

Комплекс подсистем, обеспечивающий работу мультисервисного набора услуг для физически *распределенных пользователей*, предъявляет различные требования к прикладному программному обеспечению оборудования и качеству обслуживания на каждом из уровней модели.

Для повышения надежности и улучшения качества предоставляемых сетевых мультимедийных услуг требуется внедрение эффективных методов обеспечения распределения нагрузки аппаратнопрограммных ресурсов. Проанализировав интенсивность использования каждого из компонентов в СДО, нами получен рейтинг востребованности ключевых ресурсов:

- 1) канал связи;
- 2) система хранения данных;
- 3) система управления базами данных.

Для представленных в рейтинге ресурсов могут быть применены методы, позволяющие оптимизировать и повысить эффективность обслуживания запросов, поступающих от пользователей. При этом следует учитывать индивидуальные характе-

ристики выбранного ресурса и алгоритмы его работы для обеспечения необходимого качества предоставляемого сервиса.

Эффективным считается прогнозирование поведения клиентов. Используя механизмы предварительной регистрации (подписки на сервисы), а также статистику потребления ресурсов по каждой из подсистем, можно предопределить объем необходимых вычислительных мощностей, требуемых для обслуживания поступающего потока заявок [2].

Поскольку требуется одновременное обслуживание нескольких типов заявок, поступающих в разные каналы обслуживания, необходимо эффективное управление потоками запросов, поступающих на широкополосные мультимедийные ресурсы системы дистанционного обучения. Так как представленные подсистемы являются веб-сервисами, их можно описать как систему массового обслуживания (СМО) с ограниченным временем пребывания в очереди и пуассоновским потоком заявок λ, причем длительность процедуры обслуживания каждым из каналов является случайной величиной, подчиненной экспоненциальному закону распределения [3, 4]. Отличительная особенность имитационной модели, построенной для исследования процесса обслуживания заявок в данной предметной области, состоит в неоднородном потоке событий, поступающем на вход системы. Это связано с рядом признаков, характеризующих каждую из поступающих заявок:

- ресурсоемкость оценивается по рейтингу востребованности основных ресурсов системы;
- предполагаемое время выполнения оценивается с помощью статистики обслуживания однотипных заявок в зависимости;
- рейтинг конечного исполнителя заявки учитывается в качестве весового коэффициента для рационального распределения ресурсов в соответствии с приведенной ранее уровневой модели полсистем.

Каждая из заявок во входном потоке данных получает динамический приоритет, в зависимости от представленных признаков и текущего состояния всей СМО в целом. Все каналы K обслуживания, в рамках выбранного класса решаемой задачи, идентичны и любая заявка может быть обслужена любым свободным каналом. В каждом из каналов для эффективного обслуживания заявок применяются относительные приоритеты.

Несмотря на то, что в модели можно четко классифицировать поступающие заявки, групповой режим обработки в качестве дисциплины обслуживания не эффективен, так как он не позволяет обеспечить равноценное качество для всех представленных мультимедийных сервисов.

Учитывая особенности каждого сервиса, формализуем характеристики построенной модели. Число



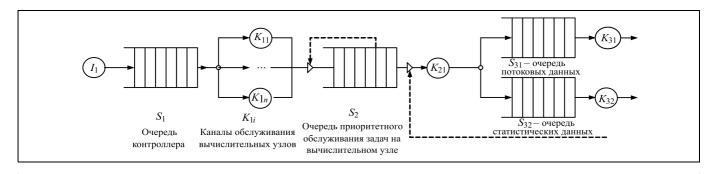


Рис. 1. Схема управления потоками запросов в системе дистанционного обучения

источников I и их интенсивности μ_n , n = 1, ..., I, напрямую зависит от числа пользователей, обращающихся в данный момент к облаку СДО. В случае одновременного обращения одного клиента к разным уровням подсистем будем считать заявки как поступившие от двух независимых друг от друга источников. Учитывая это, интенсивность и поступления заявок в облачную систему в целом будет неравномерной независимо от выбранного интервала времени моделирования; поэтому моделирование будем проводить в переходном режиме функционирования СМО. Кроме того, в СМО облака можно выделить несколько фаз Гобслуживания заявок. Это обусловлено архитектурой технического решения, позволяющего масштабировать вычислительные мощности в зависимости от поставленных задач. Облачный контроллер, управляющий размещением вычислительных задач на запущенных экземплярах приложений, а также запуском/остановкой вычислительных узлов, способен определять классы задач, что дает возможность гибкого управления потоками запросов.

Определим схему управления потоками запросов (рис. 1) и выделим три фазы обслуживания заявок: накопление заявок в контроллере облачной системы первая фаза, приоритетное обслуживание заявок на выбранном вычислительном узле вторая фаза, генерация пакетов данных, запрашиваемых пользователями третья фаза.

Рассмотрим каждую из фаз обслуживания потока заявок более подробно. При поступлении заявки в систему дистанционного обучения она попадает в очередь S_1 облачного контроллера для последующего распределения по каналам K_1 , i=1,...,n, обслуживания вычислительных узлов (n — число узлов). Длина очереди S_1 в данном случае не ограничена, так как время пребывания заявки в очереди на обслуживание фиксировано, что обусловлено принципом работы веб-приложений. Число узлов, выступающих каналами обслуживания K_1 , напрямую зависит от текущей загруженности облака, объема решаемых задач, а также числа экземпляров каждой из подсистем, запущенных в данный момент.

После того, как контролер выбрал доступный вычислительный узел, используя алгоритм минимизации потребляемых ресурсов и максимизации обслуживания клиентов, заявка поступает в очередь приоритетного обслуживания S_2 . В соответствии с алгоритмом приоритетного обслуживания заявки поступают в канал K_{21} для выполнения запрошенных вычислительных операций. Для генерации и передачи обработанного пакета данных запросы пользователей направляются в одну из двух очередей: для потоковых данных или для статических данных. Откуда они в соответствии алгоритмом предоставления доступа к данным в системе хранения облака предаются в каналы обслуживания K_{31} или K_{32} соответственно.

Выбор и расстановка приоритетов для каждой заявки, поступающей в облако, базируется на модели обслуживания запросов пользователей мультимедийных образовательных ресурсов, построенной в рамках настоящего исследования.

2. МОДЕЛЬ ОБСЛУЖИВАНИЯ ЗАЯВОК ПОЛЬЗОВАТЕЛЕЙ ШИРОКОПОЛОСНЫХ МУЛЬТИМЕДИЙНЫХ ОБРАЗОВАТЕЛЬНЫХ РЕСУРСОВ

Отличительная особенность облачных вычислений заключается в возможности управлять порядком выполнения вычислительных задач, используя различные алгоритмы обработки очередей. Это позволяет эффективно распределять запросы в высоконагруженных приложениях с критическим временем отклика.

Нами предложен алгоритм выбора и расстановки приоритетов для обработки поступающих заявок на второй фазе обслуживания запроса. В основу алгоритма положен расчет востребованности базовых ресурсов каждого из выделенных ранее уровней подсистем, входящих в СДО.

Численные показатели использования базовых ресурсов можно определить по формуле $R_{i\text{исп}} = R_i 100/(R_1 + ... + R_n)$, где R_1 , ..., R_n — численные показатели использования ресурса по каж-

26



дому из классификационных признаков, полученные в результате измерений на интервале времени ΔT [5].

Индикаторы приоритета обслуживания уровней модели определим на основе рейтинга востребованности ресурсов системы в целом. Общую ресурсоемкость системы дистанционного обучения определим как суммарную площадь $U_{\rm CДO}$, занимаемую всеми уровнями модели (U_i). Максимально возможные ресурсы сервера обозначим как площадь, полученную при использовании 100~% всех ключевых сервисов (рис. 2).

Так как система работает непрерывно, поступление заявок к ее ресурсам можно описать в дискретном времени: $I_j(T_j) = \{j:t \in (0, T_i)\}$ — множество номеров заявок, пришедших в интервал времени $(0, T_i)$ на подсистему i (i — уровень подсистемы, i = 1, ..., M).

Статус обработки j-й заявки поступившей на i-й уровень обозначим x_{ij} . Будем считать, что отказу соответствует $x_{ij}=0$, а успеху $x_{ij}=1$.

Интенсивность поступления и обработки заявок на каждый из уровней модели обозначим λ_i , она напрямую зависит от ресурсоемкости подсистемы. Введем показатель приоритета P_i для каждого из уровней, распределение которого зависит от числа одновременно используемых ресурсов. Тогда на нагрузку, создаваемую каждым из уровней, можно наложить ограничение

$$\sum_{I_{i}(T_{i})}U_{i}x_{ij}\leq H_{i},\quad I=1,\;...,\;M.$$

При задании целевой функции введены ограничения, связанные с предметной областью исследования:

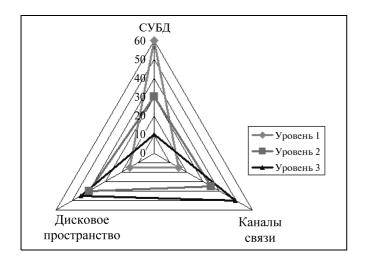


Рис. 2. Диаграмма приоритетов востребованности ресурсов системы дистанционного обучения

- время обработки T любого запроса ограничено;
 - мощность сервера H фиксирована.

Ввиду неравномерности использования основных ресурсов каждым из уровней системы дистанционного обучения следует определить условия максимальной загрузки каждого из вычислительных узлов облака, при которой возможна безотказная работа всех запущенных экземпляров приложений:

$$\sum_{i=1}^{M} \sum_{j \in I_{f}(T_{i})} U_{i} x_{ij} \leq H, \quad x_{ij} = \{0, 1\}.$$

Таким образом, для обработки максимального числа запросов пользователей в единицу времени получим целевую функцию вида

$$\sum_{i=1}^{M} \sum_{I_{i}(T_{i})} \lambda_{i} x_{ij} P_{i} \to \max.$$

При выборе приоритетов оцениваются характеристики заявки:

- время нахождения заявки в очереди;
- текущая длина очереди заявок;
- интенсивность обращения к каждому из компонентов ресурса, необходимых для выполнения заявки.

Выбор приоритетов и оценка текущей ресурсоемкости задачи производятся на основе компонентов ресурса, имеющих индивидуальные пороговые значения, связанные с физическими ограничениями оборудования. В ходе исследования выполнено моделирование обслуживания заявок в облачной системе в рамках описанной модели. Была принята приоритетная обработка запросов на основе оценки востребованности ресурсов системы. Это позволило повысить эффективность работы компонентов системы благодаря сокращению времени пребывания заявки в очереди, что привело к сокращению ее длины. Эффективность предложенного решения оценивалась по значению отношения числа обслуженных заявок на выходе третий фазы к общему числу поступивших заявок на вход первой фазы, на интервале времени моделирования $\Delta T = 60$ с. В результате получен прирост на 12-15 % по сравнению со стандартными средствами обработки очередей (рис. 3).

Установлено, что единой точкой агрегации трафика выступает система хранения данных (СХД), обеспечивающая обработку потока запросов, поступивших от потребителей мультимедийных образовательных услуг. Следовательно, эффективность работы всей системы дистанционного обучения, а также качество предоставляемых услуг напрямую зависит от производительности хранилища данных. Поэтому для эффективного управления потоком запросов необходимо разработать модель доступа



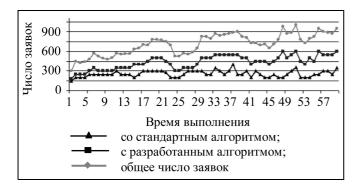


Рис. 3. Диаграмма обслуживания заявок в облачной системе

к мультимедийным данным хранилища гибридной облачной системы.

3. МОДЕЛЬ ДОСТУПА К ДАННЫМ ХРАНИЛИЩА ГИБРИДНОЙ ОБЛАЧНОЙ СИСТЕМЫ

Ключевое отличие хранилищ мультимедийных данных состоит в неоднородности размещаемой информации (текстовые, аудио или видеоданные), и, как следствие, существуют разные подходы к организации доступа к ней. Помимо методов доступа к данным, существенна интенсивность обращения к тем или иным элементам, значение которой может быть получено с помощью внутрисистемных алгоритмов идентификации пользователей, что, в свою очередь, позволяет оценить востребованность и спрогнозировать нагрузку на устройства системы хранения. В связи с этим важный аспект управления ресурсами системы, при значительном увеличении числа одновременных запросов, состоит в грамотной организации процесса размещения и распределения элементов данных по устройствам [6, 7].

Отличительной характеристикой облачных хранилищ является реконфигурируемость их структуры в зависимости от потребляемых ресурсов. Это позволяет внедрять алгоритмы оптимизации размещения данных внутри дискового пространства, а также управлять изменением числа используемых системой устройств. Процесс оптимизации размещения не должен приводить к снижению качества обслуживания клиентов СХД, для чего в алгоритмах необходимо учитывать пропускную способность сети и максимальный объем данных, который можно передавать в один момент времени [3]. Кроме того, необходимо учитывать текущую загрузку самих устройств, а также их расположение относительно друг друга и клиентов, подключаемых к ним.

Для оптимизации механизмов доступа к данным необходимо построить общую модель доступа к данным системы хранения. Пусть R = (U, M, Q),

где $U = \{u_1, u_2, ...\}$ — множество пользователей; $M = \{m_1, m_2, ...\}$ — множество уникальных элементов данных, размещаемых на устройствах хранения. Минимальной единицей данных m_i будем считать файл, имеющий обязательное свойство h — размер.

Для обеспечения безопасного хранения данных и балансировки нагрузки между устройствами хранения определим функцию распределения элементов данных, для этого введем множество $M_c = \{m_1^{j_1}, m_1^{j_2}, m_1^{j_3}, ..., m_2^{j_1}, m_2^{j_2}, m_2^{j_3}\}$, где $m_i^{j_k} - k$ -я копия элемента размещаемых данных (m_i) на j_k -м устройстве хранения, при условии $k \geq 3$ (не менее трех копий минимальной единицы хранения на различных устройствах). Тогда функция распределения элементов данных по устройствам хранения принимает вид $P: M_c \to D$.

Исходя из изложенного, запишем требование пользователя к элементам данных. $Q: U \to X \subseteq M_c$, где X — множество данных запрошенных множеством пользователей U. Тогда хранилище данных можно описать кортежем $S = (M_c, D, P, L, C, R, G)$, где $D = \{d_1, d_2, ...\}$ — множество устройств хранения; $L = \{l_1, l_2, ...\}$ — множество значений, характеризующее загрузку каждого устройства хранения (число одновременных обращений пользователей к конкретному устройству); $C = \{c_1, c_2, ...\}$ — множество значений, характеризующее объем каждого из устройств в хранилище; $G \in N$ — натуральный коэффициент, характеризующий географический (топологический) приоритет использования хранилища.

Как правило, для крупных облачных структур используются консолидированные хранилища, состоящие из ферм, объединяющих несколько хранилищ в единый массив. Представим его как $S_{farm} = \{S_1, S_2, ...\}.$

Так как характеристики требований пользователей меняются во времени, преобразуем кортеж требований $R(t)=(U,\,M_c,\,Q(t))$. Тогда $Q(t)\colon U\to X\subseteq M_c$ — требования пользователя к элементам данных, меняющиеся во времени. Так как кроме активности пользователя изменяются свойства хранилища, запишем кортеж хранилища в зависимости от времени $S(t)=(M_c(t),\,D(t),\,P(t),\,L(t),\,C,\,R(t),\,G)$, где $D(t)=\{d_1,\,d_2,\,...\}$ — множество устройств хранения, меняющихся во времени, таких что $\forall t\ D(t)>0;\,P(t)\colon M_c\to D$ — функция распределения элементов данных по устройствам хранения, меняющаяся во времени.

Для оптимизации затрат на аппаратные ресурсы и сокращения одновременно используемых устройств введем кортеж отношений $S_{cloud}(t) = \{S(t), D(t), D_{use}(t)\}$, где $\forall t \ D_{use}(t) \subseteq D(t)$ — множество уст-

28



ройств хранения, используемых в масштабируемом хранилище S в момент времени t. При масштабировании хранилища и миграции данных должно выполняться условие $\forall t,\ i,\ j,\ i\neq j \Rightarrow D_i(t) \cap D_j(t) = 0$, т. е. при миграции данных хранилища не должны использоваться одни и те же устройства. Это позволит как гарантировать скорость обработки информации, так и обеспечить приемлемое время реконфигурации.

Таким образом, для минимизации числа одновременно используемых устройств хранения в рамках одного масштабируемого хранилища и максимизации числа обработанных запросов пользователей в единицу времени введем целевую функцию вида

$$\sum_{i=1}^{N} P_i(t) \to \min; \quad \sum_{i=1}^{N} L_i P_i(t) R_i(t) \to \max.$$

На основе модели доступа к данным хранилища нами разработан алгоритм балансировки нагрузки между устройствами, реализованный в виде программного модуля для компонента Swift облачной системы «OpenStack». Выбор данной облачной системы обусловлен открытостью ее архитектуры и возможностью ее модификации под поставленные задачи. Основной недостаток системы «OpenStack» заключается в неэффективном алгоритме распределения вычислительных задач между узлами хранения данных. Стандартный алгоритм, предложенный в системе, не учитывает маршрутизацию виртуальной и топологию локальной сети, а также удаленность виртуальных машин, выполняющих обработку запросов пользователей, и хранилищ данных, обеспечивающих передачу данных. Все это негативно влияет на время отклика как самой облачной системы, так и запущенных в ней экземпляров приложений. Кроме того, сами алгоритмы распределения данных, применяемые в хранилище облачной системы, не позволяют эффективно размещать информацию и предоставлять доступ к востребованным данным по сети [5].

При исследовании алгоритма, применяемого в хранилище данных, нами получен ряд закономерностей, оказывающих существенное влияние на производительность СХД.

- При увеличении числа копий данных значительно снижается нагрузка на основных устройствах хранения. Однако при этом возрастает число задействованных устройств, что не соответствует поставленной задаче.
- При одновременном доступе к нескольким устройствам, содержащим разный объем данных, возникает дисбаланс производительности хранилища, что приводит к отказам в обслуживании запросов пользователя. Основная причина неравномерное размещение больших и малых

- по объему данных, что увеличивает время занятости устройств.
- При многократном обращении к одним и тем же данным устройства, содержащие востребованные элементы, не в состоянии обслужить запросы, так как отсутствует распределение нагрузки между узлами. Применяемые в СХД алгоритмы кеширования не могут эффективно предоставить доступ к таким данным.

Как правило, для распределения нагрузки и повышения эффективности работы масштабируемых хранилищ, помимо дублирования и перемещения данных между устройствами, также применяют систему кеш-областей (массивы устройств, обеспечивающих возможность быстрой обработки операций чтения/записи), построенных с использованием твердотельных SSD-накопителей или больших объемов оперативной памяти [5]. Однако, алгоритмы и методы использования таких ресурсов недостаточно эффективны. Чаще всего устройства кеш-области заполняются наиболее востребованными данными, при этом не учитывается модель поведения пользователя. Как правило, при обращении к мультимедийному сервису клиент отправляет последовательно несколько запросов для получения данных. В рамках мультимедийного образовательного сервиса можно предсказать набор запрашиваемых данных и порядок их получения, что позволяет построить прогноз и зарезервировать вычислительные ресурсы для решения поставленной задачи.

Учитывая изложенное, представим алгоритм, позволяющий оптимизировать доступ пользователя к мультимедийным данным.

Шаг 1. Получение входных параметров. При регистрации нового запроса, выделяются узлы (устройства хранения D), содержащие необходимые данные и анализируется их загрузка (L) и географический приоритет относительно клиента (G). Определяется тип (статические, динамические) и рейтинг востребованности запрошенных данных, составленный на основе статистики обращений.

Шаг 2. Обработка запроса. Для статических данных, используя полученные на шаге 1 показатели (G, L), определяется оптимальный узел. Для динамических данных осуществляется поиск необходимого элемента данных в кеш-области. Если он не найден, то производится процедура кеширования данных с оптимального узла, полученного с использованием показателей (G, L).

Далее, применяя алгоритм поиска связей, учитывающий рейтинг востребованности ресурсов, формируется перечень элементов, которые могут быть запрошены клиентом в ближайшее время. Для эффективной работы алгоритма осуществляется поиск наименее нагруженных узлов системы, содержащих необходимые данные, что в свою оче-



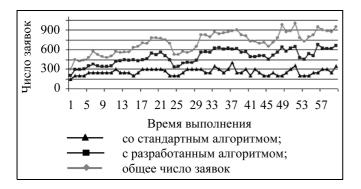


Рис. 4. Диаграмма обслуживания заявок в облачной системе при использовании алгоритма интеллектуального размещения данных

редь позволяет частично изолировать процесс кеширования от основных операций, производимых системой. Используя полученные данные, производится процедура кеширования. Число элементов зависит от востребованности начальных данных и общей нагрузки на систему.

Шаг 3. Передача данных. Запрошенные в текущий момент времени данные направляются пользователю из выбранного источника.

Шаг 4. Постобработка результатов. По окончанию работы алгоритма в базе данных хранимых ресурсов обновляется рейтинг востребованности использованных в обработке элементов.

Разработанный нами алгоритм позволяет снизить время отклика, используя информацию о топологии и маршрутизации основных потоков данных, а гибкое управление их размещением и кэшированием данных позволяет сократить накладные расходы вычислительных мощностей при миграции данных и виртуальных машин [8, 9].

Предложенный алгоритм учитывает перечисленные ранее недостатки работы стандартных средств системы управления хранением данных, что с учетом динамически формируемых приоритетов в каналах обслуживания дает прирост производительности облака и решаемых в нем задач. Относительно средств, используемых по умолчанию в «OpenStack», на 5—9 % увеличено число обслуженных заявок (рис. 4).

ЗАКЛЮЧЕНИЕ

Для оценки эффективности результатов paspaботанного модуля облачной системы «OpenStack» в качестве балансировщика нагрузки, проведено комплексное моделирование работы гибридной облачной системы с учетом особенностей работы компонентов мультимедийных ресурсов системы дистанционного обучения. Прогнозирование нагрузки и приоритезация очереди запросов позволяет управлять масштабированием облака, снижая при этом объемы задействованных в работе ресурсов, а применение алгоритмов оптимизации в системе хранения позволяет предоставлять эффективный доступ к пользователям независимо от запрошенного объема и типа данных. В результате комплексного моделирования процесса обслуживания потока заявок получен прирост производительности от 14 до 21 % по сравнению со стандартными средствами, что весьма эффективно при большой интенсивности запросов.

Разработанные оптимизационные модели управления могут применяться для дальнейшего исследования эффективности использования аппаратных и программных ресурсов в целях повышения качества предоставления услуг не только в распределенных информационных системах дистанционного обучения, но и для разработки мультимедийных ресурсов в целом.

ЛИТЕРАТУРА

- OpenStack Open Source Cloud Computing Software. [Электронный ресурс]. — Режим доступа: http://www.openstack. org/ (дата обращения 29.09.2013).
- Болодурина И.П., Решетников В.Н., Парфенов Д.И. Распределение ресурсов в информационной системе дистанционной поддержки образовательного процесса // Программные продукты и системы. 2012. № 3. С. 151—155.
- 3. *Гусев О.В., Жуков А.В., Поляков В.В., Поляков С.В.* Проблема адекватной оценки производительности веб-серверов в корпоративных сетях на предприятиях ЦБП // Материалы 6-й науч.-техн. конф. «Новые информационной технологии в ЦБП и энергетике». Петрозаводск, 2004. С. 84—87.
- 4. *Жуков А.В.* Некоторые модели оптимального управления входным потоком заявок в интранет-системах // Там же. С. 87—90.
- Математические модели облачного вычислительного центра обработки данных с использованием OpenFlow / В.Н. Тарасов, П.Н. Полежаев, А.Е. Шухман и др. // Вестник Оренбургского гос. ун-та. 2012. № 9. С. 150—155.
- 6. *Петров Д.Л.* Оптимальный алгоритм миграции данных в масштабируемых облачных хранилищах // Управление большими системами. 2010. № 30. С. 180—197.
- Петров Д.Л. Динамическая модель масштабируемого облачного хранилища данных // Известия ЛЭТИ. 2010. № 4. С. 17—21.
- 8. *Парфёнов Д.И.* Сравнение эффективности алгоритмов динамического распределения данных в облачных хранилищах системы дистанционного обучения // Системы управления и информационные технологии. 2012. № 4.1 (50). С. 163—168.
- 9. Парфёнов Д.И. Сравнение эффективности алгоритмов динамического распределения данных в гибридных облачных системах дистанционного обучения // Информационные технологии моделирования и управления. 2012. № 6 (78). С. 491—498.

Статья представлена к публикации членом редколлегии A.C. Манделем.

Ирина Павловна Болодурина — д-р техн. наук, зав. кафедрой,

☎ (3532) 37-25-36, ⊠ prmat@mail.osu.ru,

Денис Игоревич Парфенов — аспирант,

☎ (3532) 37-59-32, ⊠ parfenovdi@mail.ru,

Оренбургский государственный университет.

30

УДК 332.14:004.94

КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ ВЗАИМОДЕЙСТВИЯ МЕЖДУ МУНИЦИПАЛИТЕТАМИ, РЕГИОНАМИ, ОРГАНАМИ ГОСУДАРСТВЕННОГО УПРАВЛЕНИЯ¹

В.Л. Макаров, А.Р. Бахтизин, Е.Д. Сушко

Дан обзор современных мультиагентных систем, разработанных для анализа типичных проблем территориальных образований. Представлена мультиагентная модель муниципальных образований в составе региона, в которой агентами являются люди и организации, в которых они работают. Основное внимание уделено обеспечению реалистичности имитации поведения таких агентов в качестве участников производства, а также взаимозависимости поведения и состояний агентов разных типов и изменений внешней среды. Приведен пример использования модели для апробации различных вариантов вертикальных межбюджетных отношений.

Ключевые слова: агент-ориентированное моделирование, апробация региональной политики, поведение человека в социально-экономической среде, трудовой потенциал.

ВВЕДЕНИЕ

Одной из важнейших управленческих задач, стоящих перед властями разного уровня — как перед местными властями на уровне муниципальных образований, так и перед региональными на уровне субъектов Российской Федерации, а также перед федеральными органами государственного управления, — является задача разработки планов территориального развития. Для того чтобы подобные планы были реализованы и принесли ожидаемые экономические и социальные эффекты, каждый такой план должен быть не только тщательно подготовлен, выверен по срокам и ресурсам, но и согласован с планами соседей на одном уровне иерархии и планами субъектов других уровней. А кроме того, при планировании реализации тех или иных проектов, особенно крупных, необходимо проанализировать, как поведет себя объект управления при возможных изменениях внешней среды, особенно неблагоприятных, и какие из этих изменений являются для реализации проекта критичными. Именно с целью получения качественных прогнозов параметров управляемой социально-экономической системы в зависимости от принимаемых мер, а также оценок динамики этих параметров при возможных изменениях внешней среды, разработчики различных территориальных планов все чаще обращаются к такому инструменту, как имитационное моделирование. Очевидно, что для правомерности применения компьютерной модели в процессе территориального управления необходимо, чтобы она, как минимум:

- обеспечивала максимально возможное приближение к реальности, с тем, чтобы позволять распространять выводы, полученные в результате компьютерных экспериментов, на существующий в действительности объект моделирования;
- обладала развитым и удобным интерфейсом, обеспечивающим пользователя наглядным представлением процессов, происходящих в модели (с обязательной привязкой их к карте территории моделируемого объекта), а также возможностью управлять этими процессами.

¹ Работа выполнена в рамках Программы фундаментальных исследований Президиума РАН «Роль пространства в модернизации России: природный и социально-экономический потенциал».



Главная отличительная особенность таких крупных территориальных объектов, как муниципальные образования, а тем более субъекты РФ, в состав которых они входят, состоит в том, что они представляют собой большие системы со сложной иерархией разного рода включенных в них самостоятельных акторов, таких, например, как жители региона или же предприятия на его территории, т. е. являются активными системами [1]. Самостоятельность акторов означает, что они обладают некоторыми ресурсами (возможностями), миссией и соответствующими критериями ее выполнения, а также способностью принимать какие-то решения, реализация которых отражается на результатах их деятельности. Причем, даже акторы одного уровня (например, предприятия) могут существенно различаться между собой характеристиками, значимыми в смысле их участия в экономической жизни территорий.

Для имитации поведения сложных социально-экономических систем наиболее адекватным инструментом представляются активно развивающиеся в последнее время агент-ориентированные (мультиагентные) модели (АОМ) [2, 3]. Появление АОМ можно рассматривать как результат эволюции методологии моделирования: переход от моно-моделей (одна модель — один алгоритм) к мульти-моделям (одна модель — множество независимых алгоритмов). Мультиагентные модели способны имитировать поведение больших систем на основе реконструкции их внутренней структуры, а также структуры и поведения включенных в них более мелких экономических акторов, демонстрируя динамику социально-экономических характеристик этих систем как результат действия данных самостоятельных акторов. От объектноориентированных моделей AOM отличаются «активностью» своих элементов, каждый из которых обладает не только заданным набором личностных характеристик («ресурсов»), но и целевой функцией («интересами»), на основе чего имитируется его реакция на изменения внешней среды, затрагивающие сферу его интересов («поведение»). Таким образом, АОМ представляет собой искусственное общество, состоящее из взаимодействующих между собой самостоятельных агентов, причем важная особенность мультиагентных моделей заключается в том, что конкретные значения характеристик агентов различаются, поэтому полученное сообщество состоит из похожих, но не одинаковых членов. А возможности современных вычислительных средств позволяют создавать искусственные сообщества, количество агентов которого сопоставимо с численностью членов реального моделируемого общества.

Современные вычислительные средства предоставляют также и другие возможности, которые могут быть востребованы при создании инструментов имитационного моделирования территориальных образований, — в первую очередь это относится к геоинформационными системам (ГИС). По своей сути ГИС — это системы, позволяющие создавать базы данных, которые сочетают в себе графическое и атрибутивное представления разнородной информации. Кроме того, эти системы обеспечивают возможность пространственного анализа данных и представление его результатов в привычной для пользователей форме (в виде графиков, диаграмм, таблиц, карт и т. п.). Сочетание мультиагентного подхода с возможностями ГИС позволяет обеспечить не только правдоподобную имитацию процессов, происходящих в реальном моделируемом объекте, но и наглядное отображение этих процессов на карте его территории.

1. ОБЗОР ТЕРРИТОРИАЛЬНЫХ МУЛЬТИАГЕНТНЫХ МОДЕЛЕЙ

Мультиагентный подход на базе ГИС применялся многими исследователями для разработки инструментов территориального прогнозирования и планирования. Причем спектр проблем, для анализа и поиска решения которых создавались АОМ, очень широк. Остановимся на некоторых из них.

1.1. Управление в условиях стихийных бедствий

В работе японских ученых [4] представлена интегрированная система моделирования землетрясений и вызванных ими последствий, предназначенная для проектирования и апробации систем ликвидации этих последствий. Речь идет о разработке систем оповещения жителей, о выборе схем размещения пунктов оказания первой помощи и приютов для пострадавших, об определении первоочередности аварийно-восстановительных работ, об организации вывоза мусора и т. д. Одним из ключевых компонентов данной интегрированной системы является мультиагентная модель, позволяющая имитировать развитие последствий катастроф во взаимодействии с деятельностью большого множества людей. Агенты-люди в модели — это не только пострадавшие, но и специалисты, участвующие в спасательных работах, такие как пожарные, спасатели, медицинские работники, возможности которых по оказанию помощи ограничены, что также в модели учитывается.

1.2. Управление природными ресурсами

В коллективной монографии [5] отдельная глава посвящена мультиагентным системам (MAC), разработанным для имитации и анализа различ-



ных процессов, связанных с природопользованием. Представлены следующие МАС:

- модель для изучения взаимосвязи между диверсификацией сельскохозяйственных культур и риском деградации земель в северном Таиланде [5, p. 167—190];
- модель для анализа зависимости биофизических условий выращивания риса от различных стратегий водопользования фермеров [5, p. 191—210];
- модель для анализа экономических эффектов и рисков при различных пропорциях двух видов сельскохозяйственной деятельности в дельте реки Меконг — выращивания риса и разведения креветок [5, p. 211—236];
- модель для анализа влияния изменений в методах землепользования на ландшафтную эволюцию пригородных сельскохозяйственных систем [5, p. 237—254];
- модель лесопользования, в которой заинтересованными сторонами являются те, кто выращивает лес на плантации в штате Сабах, и те, кто занимается лесозаготовками и обработкой древесины [5, p. 275—294].

Все эти модели разработаны на примере конкретных территорий, для экономики которых именно моделируемые процессы являются ключевыми, и во всех этих моделях агентами являются люди-жители данных территорий. Агенты-люди в моделях в стремлении максимизировать свои доходы выбирают тот или иной вид (или способ) деятельности, каждый из которых характеризуется определенным уровнем использования основных природных ресурсов. Указанные модели позволяют анализировать взаимное влияние социальных процессов и динамики экологических параметров территорий и предназначены для поиска рациональных компромиссов между интересами различных видов деятельности (и занятых ею людей) для предотвращения деградации земель, справедливого использования возобновляемых ресурсов и смягчения конфликтов.

1.3. Управление расширением городской территории

Разрастание городов и сопровождающий его кризис развития муниципалитетов в сфере предоставления услуг — таких, как чистая вода, утилизация отходов, аварийные службы и охрана правопорядка, — является одной из основных проблем, стоящих перед регионами по всему миру, не случайно среди описанных в литературе мультиагентных моделей довольно много посвящено этому вопросу. Отмеченные далее наиболее известные из подобных моделей создавались как инструменты, позволяющие в комплексе изучать процессы разрастания городов с учетом поведения отдельных

экономических акторов, действующих на территории этих городов, а также как инструменты апробации различных стратегий местных органов власти и оценки их влияния на эволюцию своих городов.

Интерактивная имитационная модель развития города CityDev [6] включает в себя агентов, товары и рынки и работает на 3D пространственной структуре, которая организована в кубические ячейки. Типы агентов, присутствующих в модели:

- семьи (группа жителей из ста человек, проживающих в одной 3D ячейке);
 - производственные и коммерческие фирмы;
 - частные фирмы, производящие услуги;
 - государственные учреждения,
 - застройщики.

Каждый агент производит товары, используя другие товары, и торгует произведенными товарами на рынках. Агенты-семьи могут выбирать место жительства и место работы, а также товары для потребления, минимизируя свои расходы. А агенты-организации — место расположения производства и объем этого производства, причем коммерческие организации максимизируют свою прибыль, а государственные учреждения ориентированы на максимизацию общественной полезности. Агенты-девелоперы также максимизируют свою прибыль, выбирая, какой земельный участок купить, где и какого типа здания строить — жилые, промышленные или коммерческие — в зависимости от спроса и цен.

Кроме симуляции результатов активности агентов, модель CityDev позволяет также и пользователям активно принимать участие в своей работе, управляя развитием города в соответствии с планом строительства новых дорог и размещения объектов общественного назначения. Администрирование осуществляется с помощью управления государственными услугами и, возможно, изменения градостроительного кодекса.

Модель разрастания пригородов [7] включает в себя агентов четырех классов, представляющих различные типы стейкхолдеров, заинтересованных в развитии земельных участков:

- агенты-землевладельцы соответствуют владельцам крупных свободных земельных участков, пригодных для жилищного, коммерческого или промышленного развития;
- агенты-девелоперы играют роль строителей жилых, коммерческих или промышленных сооружений;
- агенты-домовладельцы соответствуют множеству жителей отдельных территорий;
- агенты-правительственные учреждения имитируют деятельность муниципальных органов власти.



Процедура принятия решений относительно характера развития земельных участков, выполняющаяся на каждом шаге работы модели (одному году в реальной действительности), включает в себя действия агентов каждого типа, производимые в определенной последовательности. Вначале агенты-землевладельцы решают, выгодно ли им продать свои участки. Затем для поступивших в продажу участков выбирается наиболее перспективное направление развития и соответствующая категория агентов-девелоперов, которые представляют свои предложения правительственному агенту и информируют о нем агентов-домовладельцев, которых данный проект затрагивает. Агенты-домовладельцы решают, будут ли они протестовать против данного проекта, что влияет в модели на решение правительственных агентов о том, что с проектом делать: утверждать, утверждать с изменениями или отклонять. Причем влияние агентов-домовладельцев на решение властей зависит не только от количества недовольных, но и от значимости недовольных домовладельцев (их «весов»), возрастающей с увеличением их дохода.

После принятия решений по всем участкам земли имитируется воздействие выполнения этих проектов на природную среду. Агенты-домовладельцы снова определяются с лояльностью действующей власти и как бы «голосуют» за тот или иной тип правительственного агента, который будет действовать на следующем временном интервале (каждому типу соответствует своя политика в развитии территорий).

Модель для прогнозирования расширения Большого Торонто [8] работает на 2D пространственной структуре, в которой каждая клетка соответствует участку земли на местности и имеет несколько экзогенных характеристик, таких как тип землепользования, качество природной среды, рельеф местности. Кроме того, каждая клетка характеризуется доступностью общественных благ для жителей, проживающих на этом участке. Под доступностью понимается близость от транспортных сетей — шоссе, железных дорог, а также от таких объектов, как парки, больницы, школы и торговые центры.

В модель входят:

- агенты-жители;
- агенты-девелоперы;
- агенты-местные власти.

Агенты-жители подразделяются на две категории: коренные жители и вновь прибывшие, так как их требования к жилой застройке резко различаются. Так, для всех агентов-жителей привлекательность земельных участков для проживания пропорциональна качеству природной среды и доступности общественных благ. Но для агентов-новичков, выбирающих место для застройки, важно

минимизировать стоимость строительства (которая зависит от рельефа местности), а агенты-старожилы предпочитают жить в малонаселенной местности и имеют возможность препятствовать появлению новых жителей (протестовать). Агенты-девелоперы разрабатывают проекты застройки, максимизируя свою ожидаемую прибыль, а правительственные агенты выдают или не выдают разрешения на строительство, учитывая и уровень протеста местных жителей.

1.4. Управление транспортной инфраструктурой города

В монографии [2, гл. 4] представлена модель работы транспортной системы города Москвы, в которой присутствуют агенты трех типов:

- агенты-люди;
- агенты-легковые автомобили;
- агенты-единицы общественного транспорта.

Агенты-транспортные средства движутся по дорогам города, причем агент-автомобиль способен передвигаться по любым дорогам, но перевозит в среднем двух человек, а агенты, относящиеся к наземному городскому транспорту, передвигаются по заданным маршрутам, перевозя до 150-ти пассажиров. В модели агенты-люди, которым необходимо добраться из одного пункта на территории Москвы в другой, выбирают маршрут, используя или автомобиль, или общественный транспорт, возможно, с пересадками. Модель наглядно демонстрирует образование автомобильных пробок и предназначена для апробации различных мер для улучшения транспортной ситуации в городе.

1.5. Управление сегрегацией населения города

Городская сегрегация — разделение жилых районов города по локализации в них различных групп населения — широко распространенное явление, вызывающее напряжение в обществе и представляющее собой значительное препятствие для достижения социальной интеграции в городах. Чтобы смягчить остроту этой проблемы, необходимо проводить политику, основанную на лучшем понимании процессов сегрегации и прогнозировании ее динамики. Моделирование механизмов, формирующих городскую сегрегацию в зависимости от изменения структуры населения и с учетом поведения отдельных экономических акторов, действующих на территории городов, позволяет создавать инструменты для апробации различных стратегий местных органов власти. Рассмотрим примеры применения агент-ориентированного подхода для создания подобных инструментов.

Модель расселения жителей города по этническому признаку [9]. Агенты в модели — жители го-



рода Яффа (Израиль), которые делятся на три группы:

- агенты-иудеи (евреи);
- агенты-мусульмане (арабы);
- агенты-христиане (арабы).

Агенты, недовольные своим жильем, подыскивают себе другое, стремясь снизить дискомфорт, который может быть вызван как непривычным архитектурным стилем здания (возможны традиционный восточный дом или современная постройка), так и нежелательными соседями. При этом все агенты предпочитают проживать в окружении себе подобных, и проявляют разной степени нетерпимость к соседям из других групп, к тому же вкусы агентов из разных групп относительно архитектурного стиля жилья также различаются.

Модель работает на двумерной пространственной структуре, в которой территория города разбивается на земельные участки по правилу многоугольников Вороного с заданным пороговым значением расстояния между центрами соседних многоугольников. Затем для каждого агента-жителя в качестве окрестности, занимаемой значимыми для него соседями, рассматривается многоугольник Вороного с центром в доме проживания этого агента. Поскольку границами многоугольников в модели могут служить и дороги, то полученная таким образом окрестность, в основном, состоит из домов по одной стороне улицы, которые видны от данного дома.

В модели также учитывается миграция, причем миграционный приток задается экзогенно как параметр модели, а миграционный отток формируется из агентов-жителей, не нашедших подходящего свободного жилья.

Модель расселения жителей города по уровню доходов [10]. Агенты в модели — жители города Сан-Жозе-дус-Кампус (Бразилия), которые делятся на группы по уровню дохода. Доступные для агента действия (варианты выбора):

- остаться в текущем местоположении;
- сменить место жительства в том же районе;
- переехать в район того же типа (например, из одного неоднородного в социальном плане района в другой);
- переехать в район другого типа (например, из неоднородного в социальном плане района в однородный).

Каждый агент обладает личностными характеристиками, связанными с его предпочтениями при выборе места проживания, а также характеристиками, присущими той группе, к которой он принадлежит. Так же, как и в описанной ранее модели, агенты предпочитают проживать в окружении себе подобных и подыскивают себе другое жилье, если недовольны своими соседями, в данном случае — их социально-экономическим статусом.

Кроме того, в модели присутствует блок, отвечающий за симуляцию расширения городской территории, причем выбор направления разрастания города, включения в его территорию тех или иных земельных участков происходят случайным образом и описываются с помощью цепей Маркова.

1.6. Управление деловой активностью населения

В работе [11] представлена межрегиональная модель, которая позволяет в комплексе имитировать эволюцию сложных взаимодействий домохозяйств и фирм с учетом их расположения на данной территории и возможных транспортных решений. В модели присутствуют несколько базовых производственных секторов, города (городские районы), включенные в более широкие регионы (или регионы внутри страны), а также агенты нескольких типов:

- агенты-домохозяйства (люди);
- агенты-фирмы;
- агенты-муниципальные (или региональные)
 органы власти;
 - агенты-центральные правительства.

Агенты-люди могут выбирать регион-место жительства, регион-место работы и конкретный вид деятельности для максимизации своего дохода. Агенты-фирмы для максимизации прибыли выбирают локализацию производства (как распределение инвестиций между регионами) и способ распространения своей продукции (рынки сбыта).

Регионы в модели различаются по таким экономическим характеристикам, как стоимость земли, рабочей силы и производственных затрат, а также размеры рынка и спроса. Кроме того, существует явное пространственное представление транспортных связей между регионами, чьи характеристики зависят от конкретной топологии и конфигурации транспортной сети. Отрасли производства отличаются друг от друга значимостью фактора труда и капиталоемкостью, требованиями к уровню трудовых навыков работников и отдачей от масштаба производства. Все это влияет на решения агентов-людей, касающихся поездок на работу, выбора места работы и миграции, а также на решения агентов-фирм.

Агенты-власти разного уровня в пределах своих полномочий могут управлять такими переменными, как ставки различных налогов, а также осуществлять государственные инвестиции, включая инвестиции в развитие транспортной сети системы городов. Меняя значения этих параметров, можно апробировать различные варианты налоговой и инвестиционной политики.

Все рассмотренные мультиагентные модели были реализованы в программных системах, откалиброваны на реальных данных и использованы при проведении вычислительных экспериментов



по апробации возможных управляющих воздействий. Как видно даже из приведенного обзора, этот подход оказывается продуктивным при моделировании самых разнообразных процессов, участниками которых служат самостоятельные экономические акторы, способные принимать сигналы из внешней среды и действовать в соответствии со своими интересами.

Отдельно отметим комплексную имитационную модель муниципального образования (МО) Н.А. Бегунова [12], назначение и концепция которой очень близки к тем задачам и подходам к их решению, которые реализуем мы при создании своей модели МО. Речь идет об имитации в рамках одной модели экономической жизни территориального объекта с учетом сложной иерархии действующих на данной территории самостоятельных экономических акторов, а также сложной конструкции самих агентов и их взаимозависимости. Такая модель призвана позволить одновременно рассматривать динамику нескольких процессов, поскольку каждый экономический актор играет в реальной жизни не одну роль, и его действия, соответственно, влияют на множество характеристик территории — и на экономические, и на экологические, и на социальные.

В модели муниципального образования Н.А. Бегунова рассматриваются типы агентов:

- агенты-люди (домашние хозяйства), максимизирующие полезность от приобретаемых ими товаров и услуг;
- агенты-фирмы, максимизирующие свою прибыль;
- агенты-государственные органы управления, с помощью налоговых механизмов пытающиеся создать благоприятную социально-экономическую среду для развития территории.

Моделирование производственной деятельности агентов-фирм состоит из двух этапов: первый — получение потенциального выпуска на основе использования производственных функций, включающих в себя фактор труда (в качестве которого выступают агенты-люди, работающие в фирме), второй — корректировка полученного значения на основе межотраслевого баланса для учета реального спроса на продукцию фирм. Отдельно выделяются предприятия отраслей, ориентированных на экспорт продукции, и отраслей, спрос (и цены) на продукцию которых формируется внутри МО.

Поведение агентов в модели определяется стадиями их жизненного цикла, но механизмы выбора у агентов разных типов различаются. Так, действия агентов-людей основаны на матрице решений, задающей выбор агента на множестве ситуаций, а действия агентов-фирм — на целевых функциях, определяющих логику их поведения, когда конкретные ситуации могут быть заранее не определены. Взаимодействуя между собой, агенты образуют рыночную инфраструктуру: рынок товаров и услуг, рынок труда, рынок недвижимости, финансовый рынок и др., что влияет на доходы бюджета МО и бюджетов других уровней. Интерфейс модели позволяет пользователю экспериментировать, апробируя реализацию различных проектов путем ввода новых промышленных агентов и/или изменения характеристик уже существующих.

2. ОПИСАНИЕ МОДЕЛИ МУНИЦИПАЛЬНОГО ОБРАЗОВАНИЯ

Цель настоящей работы состоит в создании компьютерной АОМ, настолько реалистично имитирующей внутреннее устройство экономической жизни муниципального образования, чтобы позволять органам государственного управления и/или органам местной власти апробировать различные варианты управляющих воздействий, оценивать возможные социально-экономические последствия этих воздействий, а также позволять сравнивать варианты стратегий для выбора наиболее приемлемой. В основу мультиагентной модели положено детализированное описание ее агентов как участников процессов формирования, накопления и реализации трудового потенциала жителей муниципальных образований, что и обусловливает благополучие экономики региона в целом. К муниципальным образованиям относятся и сельское поселение, и городской округ, и муниципальный район, поэтому очевидно, что существенно разнятся не только масштабы таких объектов, но и отраслевая структура их экономик, и наиболее злободневные проблемы, и средства, доступные администрации для решения этих проблем. Кроме того, состояние любого МО в значительной мере зависит от состояния и действий более крупных административных единиц, на территории которых оно располагается, а также от состояния и действий (в том числе, совместных) соседних муниципальных образований, поэтому компьютерная модель рассматривает МО в составе региона (субъекта Российской Федерации).

В создаваемой АОМ присутствуют агенты двух типов, соответствующих основным видам экономических акторов, действующих на территории МО и образующих иерархию агенты-физические лица (люди — жители региона) \rightarrow агенты-юридические лица (предприятия, организации). Агентами третьего уровня иерархии служат МО на территории региона-субъекта РФ, а на последнем уровне находится сам регион. Автоматами в модели представлены только агенты первого типа — люди, а изменение характеристик агентов остальных типов доступно пользователю-эксперту для



имитации различных вариантов управляющих возлействий.

Агенты-люди и агенты-организации обладают определенными наборами характеристик, обусловливающих их ценность в смысле предназначенной им в модели роли в общественном производстве, причем часть из этих характеристик сохраняют постоянное значение, а часть — изменяются во времени и/или вследствие каких-либо процессов, происходящих в модели. К таким характеристикам агентов-людей относятся уровень образования, уровень здоровья и трудовой стаж, а агентов-организаций — трудовой потенциал работников, основные фонды и располагаемые финансовые средства. Агрегирование этих частных характеристик позволяет оценить общую «полезность» каждого отдельного агента, которая в модели для агентов-людей соответствует их трудовому потенциалу [13], а для агентов следующих уровней представляет собой функцию от суммарного трудового потенциала включенных в них агентов-людей.

Роль внешней для агентов среды играют агенты вышестоящих уровней иерархии (агенты одного уровня в модели напрямую не контактируют друг с другом). Так, для агентов-людей ими выступают агенты-организации, агенты-муниципальные районы и регион в целом, т. е., в соответствии с общепринятой терминологией [14], описываемая модель относится к классу SNP^n -моделей (Social Norm Promotion with n influential groups) — arehtориентированных моделей поощрения социальных норм с n группами влияния, где n = 3. Внешняя среда выполняет в модели и ограничительную функцию, задавая, с одной стороны, условия, необходимые для реализации возможностей агентов, а с другой — имитируя спрос на производимую ими «продукцию» (содержание понятия «продукция» определяется ролью, которая предназначена в модели тому или иному агенту).

Перечислим условия внешней среды, влияющие на состояние агентов-людей.

Условия, зависящие от места жительства — рабочее место; налоговая нагрузка; инфраструктура; системы здравоохранения и образования. Работа системы здравоохранения в модели влияет на состояние здоровья населения, которое служит одним из элементов трудового потенциала, а система образования обеспечивает необходимый уровень квалификации работников, т. е. также влияет на величину их трудового потенциала. Спрос определяет наличие рабочих мест на рынке труда.

Условия, зависящие от места работы, основаны на понятии организационной культуры как специфического для каждой организации набора целей и соответствующих критериев эффективности деятельности, а также ключевых факторов, определяющих эту эффективность [15]. Организационная

культура определяет стабильность организации, а также спрос на такие свойства агентов-работников, как активность и квалификация — если агент соответствует требованиям организации, в которой он работает, то возрастают результаты его труда и заработная плата.

Агенты-организации разделяются на две категории.

Агенты-предприятия реального сектора, т. е. предприятия, производящие продукцию — промышленность и сельское хозяйство, а также предприятия сферы услуг, отнесенные к инфраструктуре. Продукция, производимая этими предприятиями, может поступать как на муниципальный и региональный рынки, так и экспортироваться за пределы региона, соответственно подразделяется и спрос на продукцию.

Агенты-организации социальной сферы — образование и здравоохранение, а также наука и государственное управление. Качество выполнения этими системами своих функций оценивается по степени охвата их услугами целевых групп, а также обеспечения соответствующих нормативов финансирования по сравнению с базовым годом.

Условия внешней среды, влияющие на состояние агентов-организации:

региональные — естественное движение населения (смертность, рождаемость); налоговая нагрузка (для предприятий реального сектора); денежные средства, выделяемые из бюджета; спрос на продукцию;

зависящие от муниципального района, в котором находится организация — трудовые ресурсы; инфраструктура; спрос на продукцию.

Поведение агентов-людей и агентов-предприятий имитируется на основе целевых функций, для чего агенты наделяются «системой ценностей» (предпочтений). Так, агенты-люди трудоспособного возраста стремятся к стабильной занятости и высокому заработку, а агенты-организации — к увеличению «продукции» и поддержанию требуемого уровня квалификации работников. Агентылюди могут менять место работы и место жительства, а агенты-организации — нанимать или увольнять работников и управлять оплатой их труда.

На каждом шаге работы модели (шаг соответствует одному году в реальной действительности) агенты, занятые в производстве, анализируют доступную им информацию о характеристиках среды и собственном положении в ней и в соответствии со своей системой ценностей переходят в то или иное состояние: «нормальное» (стартовое), «недовольство» и «критическое» (необходимость срочных действий). Если состояние какого-либо агента отлично от нормального, то это побуждает его к реакции, которая будет зависеть не только от самого состояния и от значений остальных характе-



ристик агента, но и от возможностей, которые предоставляет ему внешняя среда. Реализованный в модели алгоритм поведения агентов основан на структуре поведенческого акта из теории функциональных систем физиолога П.К. Анохина [16], включающей в себя стадии идентификации ситуации взаимодействия с внешней средой как требующей принятия каких-либо мер; постановки целей действия и формирования его программы; реализации действия и оценки достигнутого результата. Таким образом, поведение агентов является функцией структуры их личности и характеристик внешней среды, что для агентов-людей согласуется с теорией личности Р. Кеттела [17].

Самостоятельную задачу представляет собой создание механизмов, позволяющих связать трудовой потенциал агентов-людей и реальную отдачу от их трудовой деятельности, для чего в модели используются не производственные функции, а некие аналоги их производных; т. е. задается первоначальный объем производства базового года (для предприятий реального сектора — в денежном выражении, а для организаций социальной сферы как уровень выполнения соответствующей функции) и экспертно задаются коэффициенты, указывающие процент, на который увеличивается/уменьшается объем производства агента-организации при росте/снижении каждого факторного показателя и каждого влияющего показателя внешней среды на 1 %. Далее на следующих шагах работы модели результат деятельности рассчитывается как изменение объема базового года, вызванное изменением влияющих внутренних факторов и внешних условий с учетом их значимости.

В модели формируется консолидированный бюджет региона с выделением бюджетов входящих в регион муниципальных образований. Доходы бюджетов складываются из налоговых отчислений агентов-предприятий реального сектора и налога на доходы физических лиц — как доли от заработной платы агентов-людей, занятых во всех видах деятельности. Ставки налоговых отчислений как доли от объема произведенной продукции, а также пропорции распределения налоговых поступлений между бюджетами двух уровней задаются отдельно для каждого вида деятельности. Кроме того, за счет части налоговых поступлений образуются два региональных фонда, используемые для выравнивания бюджетной обеспеченности различных муниципальных образований: «Региональный фонд финансовой поддержки муниципальных районов (городских округов)» и «Региональный фонд компенсаций». Процедуры формирования и распределения средств этих фондов между муниципальными образованиями региона в модели соответствуют существующему законодательству в сфере межбюджетных отношений [18].

Далее поступившие в бюджет средства распределяются между такими видами деятельности, как государственное управление, образование, здравоохранение, наука и культура. В модели эти сферы полностью финансируются из бюджета в соответствии с нормативами бюджетных расходов. Оставшиеся бюджетные средства могут вкладываться в реализацию различных проектов для развития инфраструктуры и других отраслей реального сектора.

3. АПРОБАЦИЯ МОДЕЛИ

Прототип модели реализован в виде пользовательского программного продукта, разработанного в среде AnyLogic [19] на примере муниципальных районов Вологодской области с привязкой к ее карте [20]. Симулятор настроен на апробацию различных стратегий управления доходами и расходами бюджетов двух уровней: бюджета области и бюджетов муниципальных районов при некоторых предположениях о развитии общей экономической ситуации в стране, что и обусловило требования к его интерфейсу, такие как набор управляемых параметров, а также характер и форма отображения результатов работы модели. Основными параметрами модели, с помощью которых экспериментатор может управлять протекающими в симуляторе процессами (так же, как и реальными рычагами воздействия администрации МО на его социально-экономическое состояние), служат показатели налогового бремени и нормативы бюджетных расходов на отрасли социальной сферы, а также средства, вкладываемые в развитие отраслей реального сектора. Выбор отображаемых в интерфейсе результатов симуляций связан с критериями оценки благополучия социально-экономического состояния МО и осуществлялся с учетом требований, предъявляемых к органам исполнительной власти МО со стороны государства [21]. Таким образом, модель можно использовать в качестве инструмента планирования при бюджетировании, ориентированном на результат, т. е. при распределении бюджетных ресурсов в согласовании с целями, задачами и функциями органов управления.

Для калибровки модели использовались данные о состоянии Вологодской области в базовом (2004) и последующих годах, представленные в сборниках Росстата. Кроме того, были использованы результаты обследований, проведенных Институтом социально-экономического развития территорий РАН (г. Вологда) [22]. Модель откалибрована таким образом, чтобы статистические данные, рассчитанные на множестве агентов, совпадали с реальными данными официальной статистики о численности населения каждого муниципального района и его половозрастном составе,



а также о структуре занятости по видам деятельности, полу, возрасту и уровню образования работников с соответствующей дифференциацией заработной платы. Адекватность модели подтвердили проведенные с ее использованием численные эксперименты по имитации динамики состояния населения Вологодской области, ее муниципальных районов и предприятий на их территории при разных значениях управляемых параметров.

Рассмотрим, например, эксперимент по управлению межбюджетными отношениями. Вопрос о распределении полномочий между властями разного уровня и соответствующих финансовых средств между бюджетами этих уровней весьма актуален, и вокруг него не прекращается дискуссия. С одной стороны, государство стремится к выравниванию уровней социального обеспечения граждан независимо от места их проживания, а для этого требуется концентрация средств в федеральном и региональных бюджетах, так же, как и для реализации крупных инфраструктурных проектов. С другой стороны, уверенность местных властей в том, что проблемы их муниципалитетов будут решены благодаря дотациям из различных фондов, не стимулирует эти власти к эффективной работе. Эксперимент, проведенный на нашей модели, заключался в перераспределении налоговых поступлений между областным бюджетом и бюджетами муниципальных районов при неизменной общей налоговой нагрузке на предприятия различных отраслей, а также на работающих жителей области. Целью эксперимента было проследить влияние перераспределения налоговых поступлений на такие характеристики муниципальных районов, как самообеспеченность их бюджетов и бюджетная обеспеченность на одного жителя.

Эксперимент показал, что хотя средний объем бюджета МО, приходящийся на одного жителя, возрастает при увеличении доли налоговых поступлений, остающихся в местном бюджете, но одновременно возрастает и дифференциация районов по этому показателю. На рисунке представлено сопоставление самообеспеченности различных вариантов бюджетов МО в предположении, что расходы остаются на уровне варианта 1 бюджета. (Вариант 1 соответствовал фактическим пропорциям, а варианты 2 и 3 отличались от него все возрастающей долей оставляемых в муниципальном бюджете собираемых средств).

На рисунке видно, что при третьем варианте доходы семи муниципальных образований из 28 превышают расходы базового варианта (самообеспеченность бюджетов свыше 100 %), а доходы бюджетов еще пяти районов почти покрывают расходы (самообеспеченность больше 90 %). Однако даже и такое перераспределение налоговых поступлений большинству муниципальных райо-



Сопоставление самообеспеченности расходов бюджетов МО варианта 1 при разных вариантах доходов, %

нов не позволяет обойтись без дотаций. Иными словами, эксперимент показал, что с помощью одного только перераспределения налоговых поступлений не удалось добиться самообеспеченности муниципальных бюджетов.

ЗАКЛЮЧЕНИЕ

Разработанная мультиагентная модель представляет собой попытку создания инструмента для апробации механизмов регионального управления с учетом человеческого фактора. Вот те основные особенности модели, которые делают ее таким инструментом.

Прежде всего, это моделирование структуры региона как сложной иерархической системы, включающей в себя самостоятельных экономических агентов разного уровня, где основой служат агенты — жители региона.

Далее, это моделирование структуры личности и поведения агентов с учетом их дифференциации, что позволяет отследить различное влияние управленческих мер на отдельных агентов и группы таких агентов. Общество неоднородно, и это может оказаться существенным для прогнозирования его реакции на изменение экономической ситуации и/или правил игры, задаваемых сверху. Это перекликается с подходом поведенческой экономики [23], а, по мнению С. Делавиньи, «нужно шире применять идеи поведенческой экономики в государственной политике» [23, с. 106].

Наконец, это построение цепочек зависимостей, позволяющих смоделировать также и ответное влияние состояния и трудовой деятельности



агентов-людей на экономические и социальные показатели региона, такие, например, как удельный трудовой потенциал работающих и доля удовлетворенных жизнью жителей. В модели такое влияние учитывается с помощью различных процедур агрегирования характеристик агентов нижестоящих уровней для получения характеристик агентов следующих уровней.

Таким образом, с помощью процедур расчета доходных и расходных статей бюджетов двух уровней в соответствии с заданными уровнем налоговой нагрузки и параметрами межбюджетных отношений, процедур расчета состояний агентов-жителей региона и моделирования их поведения в качестве работников, процедур агрегирования трудовых потенциалов агентов и оценки результатов их труда выстраиваются цепочки влияния действий пользователя на выходные параметры модели. В частности, конструкция модели призвана наглядно демонстрировать зависимость интегральных показателей региона от бюджетных расходов на социальную сферу как инвестиций в человеческий капитал.

ЛИТЕРАТУРА

- Бурков В.Н., Новиков Д.А. Теория активных систем: состояние и перспективы. М.: СИНТЕГ, 1999. 128 с.
- 2. *Макаров В.Л., Бахтизин А.Р.* Социальное моделирование новый компьютерный прорыв (агент-ориентированные модели). М.: Экономика, 2013. 295 с.
- 3. *Тарасов В.Б.* От многоагентных систем к интеллектуальным организациям: философия, психология, информатика. М.: Эдиториал УРСС, 2002. 352 с.
- 4. Goto Y., Takeuchi I. and Kakumoto S. Integrated earthquake disaster simulation systems for the highly-networked information society // Proc. of the 13th World Conference on Earthquake Engineering, Vancouver (Canada), 2004. Paper No. 2793. URL: http://www.iitk.ac.in/nicee/wcee/article/13_2793.pdf (дата обращения: 09.08.13).
- Companion modeling and multi-agent systems for integrated natural resource management in Asia / Ed. by F. Bousquet, G. Trébuil, and B. Hardy. — Los Baños (Philippines): International Rice Research Institute, 2005. — 360 p.
- Semboloni F., Assfalg J., Armeni S., Gianassi R. and Marsoni F.
 CityDev, an interactive multi-agents urban model on the web //
 Computers, Environment and Urban Systems. 2004. —
 Vol. 28, N 1. P. 45—64.
- Monticino M.G., Brooks E., Cogdill T., Acevedo M. and Callicott B.
 Applying a Multi-Agent Model to Evaluate Effects of Development Proposals and Growth Management Policies on Suburban Sprawl // Proc. of the International Environmental Modelling and Software Society, Summit on Environmental Modelling and Software. Burlington (USA), 2006. URL: http://www.math.unt.edu/~monticino/papers/mult-agent_development.pdf (дата обращения: 09.08.13).
- 8. Rui Y., Ban Y. Multi-agent Simulation for Modeling Urban Sprawl In the Greater Toronto Area // Proc. of the 13th AGILE International Conference on Geographic Information Science. Guimarães (Portugal), 2010. URL: http://plone.itc.nl/agile_

- old/Conference/2010-guimaraes/ShortPapers_PDF%5C124_DOC.pdf (дата обращения: 09.08.13).
- 9. Benenson I., Omer I., Hatna E. Entity-based modeling of urban residential dynamics: the case of Yaffo, Tel Aviv // Environment and Planning B: Planning and Design. 2002. Vol. 29. P. 491—512.
- Feitosa F.F., Le Q.B., Vlek P.L.G. Multi-agent simulator for urban segregation (MASUS): A tool to explore alternatives for promoting inclusive cities // Computers, Environment and Urban Systems. 2011. Vol. 35, N 2. P. 104—115.
- 11. *Tsekeris T., Vogiatzoglou K.* Multi-Regional Agent-Based Economic Model of Household and Firm Location and Transport Decisions // Proc. of the 10th STRC Swiss Transport Research Conference. Monte Verità, 2010. URL: http://www.strc.ch/conferences/2010/Tsekeris.pdf (дата обращения: 09.08.13).
- 12. *Бегунов Н.А*. Моделирование развития муниципальных образований на основе агентного подхода // Современные исследования социальных проблем. 2011. № 4. С. 1—18.
- 13. *Гулин К.А., Шабунова А.А., Чекмарева Е.А.* Трудовой потенциал региона / Под рук. В.А. Ильина. Вологда: ИСЭРТ РАН, 2009. 84 с.
- Bloomquist K.M. A comparison of agent-based models of income tax evasion // Social Science Computer Review. — 2006. — Vol. 24, N 4. — P. 411—425.
- Камерон К., Куинн Р. Диагностика и измерение организационной культуры. — СПб.: Питер, 2001. — 320 с.
- Судаков К.В. Общие принципы построения поведенческих актов на основе теории функциональных систем // Системные механизмы поведения / Под ред. К.В. Судакова, М. Баича. — М.: Медицина, 1990. — 939 с.
- 17. *Холл К., Линдсей Г.* Теории личности / Пер. с англ. М.: Изд-во Ин-та психотерапии, 2008. 672 с.
- 18. Закон Вологодской области «О межбюджетных трансфертах в Вологодской области» от 04.07.2005 г. № 1308-ОЗ (с последующими изменениями). URL: http://www.df35.ru/index.php?option=com_content&view=article&id=241:-q-q-4-2005-1308-&catid=86:2012-04-16-07-46-48&Itemid=116 (дата обращения: 09.08.13).
- 19. XJ Technologies. Simulation Software and Services. URL: http://www.xjtek.com/ (дата обращения: 09.08.13).
- Сушко Е.Д. Мультиагентная модель региона: концепция, конструкция и реализация / Препринт # WP/2012/292. — М.: ЦЭМИ РАН, 2012. — 54 с.
- 21. Указ Президента РФ от 28.04.2008 г. № 607 «Об оценке эффективности деятельности органов местного самоуправления городских округов и муниципальных районов» (в ред. Указа Президента РФ от 13.05.2010 № 579).
- 22. *Амелин Д.Е.* Местное самоуправление в региональном развитии. Вологда: ВНКЦ ЦЭМИ РАН, 2006. 182 с.
- 23. Делавинья С. Психология и экономика: результаты эмпирических исследований. Часть III. Нестандартное принятие решений и реакция рынка // Вопросы экономики. 2011. № 4. С. 82—106.

Статья представлена к публикации членом редколлегии чл.-корр. РАН Д.А. Новиковым.

Валерий Леонидович Макаров — директор, академик РАН, **☎** (499) 129-10-11, ⊠ makarov@cemi.rssi.ru,

Альберт Рауфович Бахтизин — д-р экон. наук, вед. науч. сотрудник, **☎** (499) 129-07-44, \bowtie cgemodel@yandex.ru,

Елена Давидовна Сушко — канд. экон. наук, ст. науч. сотрудник, **☎** (499) 129-38-22, ⊠ sushko_e@mail.ru.

Центральный экономико-математический институт РАН, г. Москва.



УДК 35.073.5

О НЕКОТОРЫХ ОЦЕНКАХ УСТОЙЧИВОСТИ ФОНДОВОГО РЫНКА И ВЛИЯНИИ НА НИХ ИНФОРМИРОВАННОСТИ ИНВЕСТОРОВ

В.А. Горелик, Т.В. Золотова

Предложены показатели устойчивости и риска для отдельного инвестора и фондового структурированного рынка. Проведено исследование рыночной модели на устойчивость с помощью предложенных показателей. Показано, что инвесторы, придерживаясь различных прогнозов развития фондового рынка, как следствие различной их информированности, и выбирая соответственно разные стратегии (портфели), способствуют тем самым повышению устойчивости данной системы.

Ключевые слова: коллективный риск, коэффициент риска, средняя ковариация, устойчивость.

ВВЕДЕНИЕ

В работе [1] предложен подход к оценке устойчивости сложных систем с использованием понятия коллективного риска. Под коллективным риском понимается непредсказуемость состояния некоторой системы как результат индивидуального поведения ее подсистем, обладающих различной информированностью и различными интересами. В данной статье эти общесистемные аспекты рассматриваются на примере фондового рынка, а именно, исследуется влияние на устойчивость этой системы различной информированности инвесторов и различного отношения их к риску.

Вопросам поведения инвесторов на фондовом рынке и принятия ими решения о составе своих портфелей ценных бумаг посвящена обширная литература (см., например, библиографические списки в работах [2, 3]). Современные исследования посвящены в основном динамическим моделям управления портфелем и использованию производных финансовых инструментов (фьючерсов, опционов и т. д.). Статическая постановка задачи формирования портфеля, впервые сформулированная Г. Марковицем, в принципе, исследована достаточно полно. Однако это относится в первую очередь к индивидуальному поведению инвесторов. Насколько известно авторам, модели их взаимодействия и влияния на устойчивость рынка отсутствуют. В связи с этим возникает идея воспользоваться понятием коллективного риска для оценки устойчивости фондового рынка.

В работах [1, 4, 5] получены оценки риска на фондовом рынке с использованием корреляцион-

ных моментов случайных величин доходностей, проведено исследование коллективного риска на фондовом рынке с разными структурными характеристиками и ограничениями инвестиционных портфелей, рассмотрен вопрос оценки устойчивости фондового рынка с использованием понятия энтропии как меры разнообразия поведения инвесторов. В настоящей статье рассмотрен ряд новых вопросов, относящихся к этой области. Предлагаются показатели (меры) риска и устойчивости стратегии отдельного инвестора и фондового рынка в целом, проведено исследование рыночной модели на устойчивость с помощью предложенных показателей.

1. ОЦЕНКА КОРРЕЛЯЦИОННОЙ ЗАВИСИМОСТИ ДОХОДНОСТЕЙ ОПТИМАЛЬНЫХ ПОРТФЕЛЕЙ РАЗНЫХ ИНВЕСТОРОВ

1.1. Случай одинаковой информированности инвесторов

В основе рассматриваемой нами модели фондового рынка лежит предположение, что теоретически существует вероятностное распределение n-мерного вектора случайных величин доходностей r_i финансовых инструментов на фондовом рынке. При этом известно, что доходности представляют собой взаимосвязанные случайные величины и мерой, определяющей эту взаимосвязь, служит ковариация (или корреляционный момент) доходностей. Будем считать, что фондовый рынок характеризуется вектором математических ожиданий доходностей финансовых инструментов



 $\bar{r} = (\bar{r}_1, ..., \bar{r}_i, ..., \bar{r}_n)$ и ковариационной матрицей $\sigma = (\sigma_{ii})_{n \times n}$. Это — объективная информация, доступная исследователю (биржевому аналитику). Под объективной информацией о векторе \bar{r} и матрице о здесь понимаются значения этих характеристик, полученные исследователем и не зависящие от влияния участников фондового рынка. Инвесторы могут при принятии решений использовать эту объективную информацию или руководствоваться своей собственной, т. е. субъективной информацией. Предположим сначала, что инвесторы основывают свое поведение на единой объективной информации. Различие между ними заключается в отношении к риску, выражающееся в значении коэффициента в целевой функции, представляющей собой линейную свертку двух критериев: математического ожидания и дисперсии случайных доходностей портфелей.

Рассмотрим индивидуальное поведение инвестора, управление которого есть вектор x (портфель инвестиций), компоненты которого x_i — доли средств, вкладываемых в финансовые инструменты из конечного списка (i=1,...,n). Определим оптимальный портфель как решение задачи на экстремум линейной свертки критериев математического ожидания доходности портфеля и дисперсии доходности портфеля:

$$\max_{x \in X} \left[\sum_{i=1}^{n} \overline{r}_{i} x_{i} - \alpha \sum_{i,j=1}^{n} \sigma_{ij} x_{i} x_{j} \right], \tag{1}$$

где
$$X = \{x | x_i \ge 0, \ i = 1, ..., \ n, \ \sum_{i=1}^n x_i = 1\}, \ \alpha \ge 0$$
 — ве-

совой коэффициент, определяющий отношение инвестора к риску (коэффициент риска). Заметим, что впервые рассмотревший эту проблему Г. Марковиц [6] не формулировал ее в виде (1), а говорил об эффективных портфелях, а в книге [7] задача поиска оптимального портфеля была поставлена им как задача минимизации разности дисперсии и математического ожидания доходности портфеля (коэффициент риска при дисперсии равен 1). Кроме того, в той же книге рассмотрена задача на максимум доходности при ограничении на дисперсию, а наиболее распространена сейчас задача минимизации дисперсии при ограничении по доходности. Однако любая задача, решением которой является эффективный портфель, эквивалентна задаче (1) (в силу свойств выпуклости она представляет собой необходимые и достаточные условия Парето-оптимальности). Поэтому выбор линейной свертки критериев «математическое ожидание дисперсия» для нахождения оптимального портфеля не принципиален; любые другие принципы оптимального выбора приводят к одному из эффективных портфелей, соответствующему определенному $\alpha > 0$ в задаче (1) (подробнее об этом см. в работе [8]). Задача выбора оптимального портфеля (1) предполагает отсутствие коротких продаж, которые совершаются путем займа ценных бумаг и затем погашения займа такими же ценными бумагами (в задаче (1) это отражается в наличии условия неотрицательности на компоненты вектора x), безрискового заимствования и кредитования (эти случаи рассмотрены в работе [5]).

Будем называть портфель полноразмерным, если у составляющего его вектора x все компоненты больше нуля. Решение задачи (1) приведено в работе [9], а именно, состав оптимального полноразмерного портфеля имеет вид

$$x^0(\gamma) = C_0 + C_1 \gamma, \tag{2}$$

где $\gamma=1/(2\alpha),\ \gamma\in(0;\infty),\ e=(1,...,1),\ a\ C_0=(C_{01},...,C_{0j},...,C_{0n}),\ C_1=(C_{11},...,C_{1j},...,C_{1n})$ определяются по формулам

$$C_0 = \frac{\sigma^{-1}e}{e\sigma^{-1}e}, \quad C_1 = \sigma^{-1}\bar{r} - \frac{e\sigma^{-1}\bar{r}}{e\sigma^{-1}e}\sigma^{-1}e.$$
 (3)

Здесь и далее мы не делаем различия в обозначении вектора-строки и вектора-столбца, считая их соответствующими требованиям операций умножения матриц и векторов, а в качестве коэффициента риска используется γ . Согласно формуле (2), составы оптимальных портфелей x^{01} и x^{02} инвесторов имеют вид $x^{01} = C_0 + C_1\gamma_1$ и $x^{02} = C_0 + C_1\gamma_2$ соответственно. В работе [5] приведены необходимые и достаточные условия полноразмерности портфеля, а также показано, что ковариация случайных величин доходностей r_{x^1} и r_{x^2} двух произвольных портфелей, имеющих составы x^1 и x^2 , вычисляется через составы этих портфелей по формуле

$$cov(r_{1}, r_{2}) = x^{1} \sigma x^{2}.$$
 (4)

Теорема 1. Если определитель ковариационной матрицы $\det \sigma \neq 0$, то ковариация $\cot (r_{x^{01}}, r_{x^{02}})$ доходностей двух полноразмерных оптимальных портфелей положительна. Если дополнительно ковариационная матрица σ строго положительно определена, то ковариация доходностей любых двух оптимальных портфелей положительна. \bullet

Доказательство теоремы 1 приведено в работе [5]. Как известно, ковариационная матрица σ всегда неотрицательно определена. Далее предполагается, что ковариационная матрица σ строго положительно определена, т. е. $x\sigma x > 0$ для любых действительных значений $x \neq 0$, тогда существует и строго положительно определена обратная матрица σ^{-1} .



Из выражений (2)—(4) следует, что ковариация двух полноразмерных портфелей есть

$$cov(r_{x^{01}}, r_{x^{02}}) = x^{01}\sigma x^{02} = (C_0 + C_1\gamma_1)\sigma(C_0 + C_1\gamma_2) =$$

$$= C_0\sigma C_0 + (C_1\sigma C_1)\gamma_1\gamma_2 + (C_0\sigma C_1)\gamma_2 + (C_1\sigma C_0)\gamma_1.$$

Учитывая свойства скалярного произведения, имеем

$$C_{0}\sigma C_{1} = \frac{\sigma^{-1}e}{e\sigma^{-1}e}\sigma\left(\sigma^{-1}\bar{r} - \frac{e\sigma^{-1}\bar{r}}{e\sigma^{-1}e}\sigma^{-1}e\right) =$$

$$= \frac{\sigma^{-1}e}{e\sigma^{-1}e}\left(\bar{r} - \frac{e\sigma^{-1}\bar{r}}{e\sigma^{-1}e}e\right) =$$

$$= \frac{1}{e\sigma^{-1}e}\left(e\sigma^{-1}\bar{r} - \frac{(e\sigma^{-1}\bar{r})(e\sigma^{-1}e)}{e\sigma^{-1}e}\right) = 0. \quad (5)$$

Так как матрица σ — симметрическая, то $C_0 \sigma C_1 = C_1 \sigma C_0$ и

$$cov(r_{v_01}, r_{v_02}) = C_0 \sigma C_0 + (C_1 \sigma C_1) \gamma_1 \gamma_2.$$
 (6)

Отметим, что ковариация доходностей двух произвольных портфелей вычисляется с помощью объективной ковариационной матрицы σ , характеризующей рынок, т. е. по формуле (4), независимо от того, какой субъективной информацией пользуются инвесторы при формировании своих портфелей.

1.2. Случай разной информированности инвесторов

Традиционно предполагается, что инвесторы одинаково информированы (например, используют объективную информацию) о ситуации на финансовом рынке и их различное поведение связано с различным отношением к риску (выбор параметра у). Предположим, что инвесторы обладают различной информированностью (субъективной информацией) о ситуации, складывающейся на финансовом рынке, которая выражается в том, что они по-разному оценивают математические ожидания доходностей финансовых инструментов (пока мы не конкретизируем способ коррекции ими объективной информации). Ковариационная матрица здесь для простоты считается объективной и единой для всех (в дальнейших исследованиях мы планируем отказаться от этого предположения).

Рассмотрим двух инвесторов, оптимальные портфели x^{01} и x^{02} которых определены из решения задачи (1) при различных значениях параметра γ и различных значениях математических ожиданий доходностей. Пусть первый инвестор имеет вектор математических ожиданий доходностей \bar{r}^1 , а второй — \bar{r}^2 . Составы оптимальных портфелей x^{01} и x^{02} инвесторов имеют вид $x^{01} = C_0 + C_1^1 \gamma_1$ и

 $x^{02} = C_0 + C_1^2 \gamma_2$ соответственно, где C_1^1 и C_1^2 определяются по формулам

$$C_0 = \frac{\sigma^{-1}e}{e\sigma^{-1}e}, \quad C_1^1 = \sigma^{-1}\bar{r}^1 - \frac{e\sigma^{-1}\bar{r}^1}{e\sigma^{-1}e}\sigma^{-1}e,$$

$$C_1^2 = \sigma^{-1} \bar{r}^2 - \frac{e \sigma^{-1} \bar{r}^2}{e \sigma^{-1} e} \sigma^{-1} e.$$

Теорема 2. Если определитель ковариационной матрицы $\det \sigma \neq 0$, то ковариация $\cot (r_{x^{01}}^1, r_{x^{02}}^2)$ двух полноразмерных оптимальных портфелей отрицательна для \overline{r}^1 , \overline{r}^2 и σ , удовлетворяющих условию

$$(e\sigma^{-1}\bar{r}^1)(e\sigma^{-1}\bar{r}^2) - (\bar{r}^1\sigma^{-1}\bar{r}^2)(e\sigma^{-1}e) > 1/\gamma_1\gamma_2. \bullet (7)$$

Доказательство теоремы 2 см. в работе [9]. Условие (7) характеризует степень различия оценок, которая приводит к отрицательной ковариации.

2. ИНФОРМАЦИОННЫЕ АСПЕКТЫ И УСЛОВИЕ УСТОЙЧИВОСТИ СТРАТЕГИИ ИНВЕСТОРА

Рассмотрим одного инвестора, который на основе своей собственной субъективной информации может давать оценки математических ожиданий доходностей финансовых инструментов при различных сценариях развития экономики.

Пусть $\overline{r}(y)$ — вектор математических ожиданий доходностей финансовых инструментов, зависящий от значения внешних (неконтролируемых) факторов y, описание которых включает в себя указание вида неконтролируемых факторов и информированности о них инвестора (например, законы распределения случайных параметров, область значений неопределенных факторов, схемы передачи информации в системе, процедуры обработки информации). Тогда задача (1) примет вид

$$\max_{x \in X} \left[\sum_{i=1}^{n} \overline{r}_i(y) x_i - \alpha \sum_{i,j=1}^{n} \sigma_{ij} x_i x_j \right], \tag{8}$$

а состав оптимального полноразмерного портфеля вид

$$x^{0}(\gamma, y) = C_{0} + C_{1}(y)\gamma,$$
 (9)

где
$$C_0 = \frac{\sigma^{-1}e}{e\sigma^{-1}e}$$
, $C_1(y) = \sigma^{-1}\bar{r}(y) - \frac{e\sigma^{-1}\bar{r}(y)}{e\sigma^{-1}e}\sigma^{-1}e$.

Согласно выражениям (8) и (9) оптимальное управление инвестора (состав портфеля) зависит теперь не только от отношения инвестора к риску (выбора параметра γ), но и от предполагаемого инвестором сценария развития экономической ситуации, характеризуемой информацией о значениях внешних факторов y (например, цены на нефть,



валютного курса, роста ВВП и др.). Для двух сценариев имеем два вектора значений внешних факторов y^1 и y^2 и соответственно две оценки вектора математических ожиданий доходностей $\bar{r}(y^1)$ и $\bar{r}(y^2)$ (примером может служить многофакторная модель [2, $c.\ 295$]). Тогда составы двух оптимальных полноразмерных портфелей одного инвестора согласно выражению (9) есть $x^0(\gamma, y^1) = C_0 + C_1(y^1)\gamma$ и $x^0(\gamma, y^2) = C_0 + C_1(y^2)\gamma$.

Оценим ковариацию случайных величин доходностей двух разных портфелей одного инвестора, имеющих составы $x(\gamma, y^1)$ и $x(\gamma, y^2)$. По формуле (4)

$$cov(r_{x(\gamma, y^1)}(y^1), r_{x(\gamma, y^2)}(y^2)) = x(\gamma, y^1)\sigma x(\gamma, y^2). \quad (10)$$

Можно считать оптимальное управление $x^0(y)$ (состав портфеля) инвестора устойчивым, если при рассматриваемых сценариях развития ковариация случайных значений доходностей портфелей, определяемая формулой (10), положительна. Отметим, что в случае оценки риска фондового рынка в целом положительная ковариация доходностей портфелей разных инвесторов является фактором неустойчивости рынка [8]. Положительная же ковариация доходностей разных портфелей одного инвестора говорит об устойчивости управления конкретного инвестора, т. е. при рассматриваемых сценариях развития экономической ситуации случайные значения доходностей его портфелей имеют тенденцию меняться в одну и ту же сторону.

Следствие (из теоремы 2). Если определитель ковариационной матрицы $\det \sigma \neq 0$, то ковариация $\operatorname{cov}(r_{x^0(\gamma,y^1)}(y^1), r_{x^0(\gamma,y^2)}(y^2))$ двух полноразмерных оптимальных портфелей $x^0(\gamma,y^1)$ и $x^0(\gamma,y^2)$ инвестора положительна для $\bar{r}(y^1)$, $\bar{r}(y^2)$ и σ , удовлетворяющих условию

$$(e\sigma^{-1}\bar{r}(y^{1}))(e\sigma^{-1}\bar{r}(y^{2})) - (\bar{r}(y^{1})\sigma^{-1}\bar{r}(y^{2}))(e\sigma^{-1}e) < (1/\gamma)^{2}. \bullet$$

Доказательство следствия непосредственно вытекает из формулы (7).

3. ОЦЕНКА УСТОЙЧИВОСТИ РЫНКА

3.1. Влияние отношения инвестора к риску на устойчивость рынка

Как вытекает из результатов анализа, доходности оптимальных портфелей при одинаковой информированности имеют положительную корреляцию, причем, чем ближе значения коэффициентов у инвесторов, т. е. сходно их отношение к

риску, тем ближе к единице коэффициенты корреляции доходностей их портфелей. Так как инвесторы не обязаны вести себя на фондовом рынке оптимально (согласно задаче (1)) и использовать одинаковую информацию, то доходности их портфелей могут быть коррелированны как положительно, так и отрицательно. Таким образом, однотипное поведение инвесторов может вызывать большие колебания рынка, а разнотипное гасит колебания рынка, т. е. снижает уровень коллективного риска.

В модели САРМ исследуется коллективное поведение инвесторов в ситуации равновесия, когда спрос равен предложению и цены стабилизируются. К числу основных предположений этой модели относится наличие безрискового актива и, главное, однородных ожиданий инвесторов, т. е. одинаковых оценок ими ожидаемых доходностей, среднеквадратических отклонений и ковариаций доходностей ценных бумаг [2, с. 259]. Поэтому в нашем исследовании коллективного поведения инвесторов модель САРМ не применима, а рынок может находиться вне ситуации равновесия. В зависимости от характера переходных процессов (скорости, амплитуды) могут возникнуть значительные колебания рынка, а в поведении инвесторов — известный в теории игр «эффект толпы» (например, массовый переход в кеш), что может привести (как и было недавно) к обвалу. Поэтому положительная корреляция случайных величин доходностей служит одним из существенных факторов неустойчивости. Наличие безрискового актива принципиально картину не меняет [5].

Как известно, систематический (или рыночный) риск фондового рынка определяется предельным значением средней ковариации доходностей финансовых инструментов (см., например, работу [3, с. 61, 62]). Эта величина не зависит от распределения финансовых инструментов (ценных бумаг) по инвестиционным портфелям. На реальном фондовом рынке все финансовые инструменты принадлежат инвесторам, т. е. распределены по портфелям. Такой рынок будем называть структурированным. Поэтому предельное значение средней ковариации ценных бумаг можно считать оценкой риска неструктурированного рынка, т. е. первичного рынка, на котором эмитенты размещают новые выпуски финансовых инструментов.

Мерой (показателем) риска структурированного фондового рынка может служить средняя ковариация доходностей портфелей. Пусть N — общее число портфелей (или инвесторов), присутствующих на рынке. Вычислим дисперсию средней



доходности портфелей, т. е. дисперсию величины

$$r_{\rm cp}^p = \frac{1}{N} \sum_{m=1}^{N} r^m$$
. Имеем

$$Dr_{cp}^{p} = \sum_{m=1}^{N} \frac{1}{N^{2}} Dr_{x^{m}} + \sum_{s, m=1, s \neq m}^{N} \frac{1}{N^{2}} cov(r_{x^{s}}, r_{x^{m}}) =$$

$$= \frac{1}{N} \left(\frac{1}{N} \sum_{m=1}^{N} Dr_{x^{m}} \right) +$$

$$+ \frac{1}{N^{2}} (N^{2} - N) \left(\sum_{s, m=1, s \neq m}^{N} \frac{1}{N^{2} - N} cov(r_{x^{s}}, r_{x^{m}}) \right) =$$

$$= \frac{1}{N} \overline{\sigma}_{N}^{2} + \frac{1}{N^{2}} (N^{2} - N) \overline{cov}_{N},$$

где $\bar{\sigma}_N^2 = \frac{1}{N} \sum_{m=1}^N Dr_{x^m}$ — средняя дисперсия доходностей портфелей,

$$\overline{\text{cov}}_N = \frac{1}{N^2 - N} \sum_{s, m=1}^{N} \text{cov}(r_{x^s}, r_{x^m}) - (11)$$

средняя ковариация доходностей портфелей.

Если $\overline{\sigma}_N^2 \le A$ и $\overline{\text{cov}}_N \to B$ при $N \to \infty$, то имеем $\overline{\sigma}_N^2/N \to 0$ и $Dr_{\rm cp}^p \to B$ при $N \to \infty$. Значит, если B = 0, то диверсификацией с достаточно большим N значение Dr_{cp}^p можно сделать сколь угодно малым. Однако при однотипном (оптимальном по Марковицу) поведении инвесторов доходности их портфелей положительно коррелированы, поэтому B > 0, и риск структурированного рынка неустраним. При разнотипном поведении доходности портфелей инвесторов могут быть коррелированны отрицательно, что приводит к хеджированию коллективного риска: колебания рынка гасятся, и он возвращается в прежнее положение равновесия. Отрицательное значение средней ковариации доходностей портфелей можно считать условием устойчивости рынка, а положительное значение условием неустойчивости, причем, чем больше значение средней ковариации, тем менее устойчив фондовый рынок.

Согласно формуле (4) выражение (11) можно записать в виде

$$\overline{\operatorname{cov}}_{N} = \frac{1}{N^{2} - N} \sum_{s = 1}^{N} \sum_{s \neq m}^{N} x^{s} \sigma x^{m}.$$
 (12)

Рассмотрим на примере полноразмерных оптимальных портфелей вопрос, как изменится значение средней ковариации, если число портфелей на рынке увеличилось на единицу, т. е. исследуем знак разности $\Delta \overline{\text{cov}} = \overline{\text{cov}}_{N+1} - \overline{\text{cov}}_N$.

Теорема 3. Если σ строго положительно определена, то при добавлении еще одного (N+1)-го полноразмерного оптимального портфеля имеет

место равенство $\operatorname{sign}(\Delta \overline{\operatorname{cov}}) = \operatorname{sign}(\gamma_{N+1} \frac{1}{N} \sum_{m=1}^{N} \gamma_m -$

$$-\frac{1}{(N-1)N}\sum_{s=m-1}^{N}\gamma_{s}\gamma_{m}). \blacklozenge$$

Доказательство. Согласно выражению (6) ${\rm cov}(r_{v^{0s}},\ r_{v^{0m}})=C_0\sigma C_0+C_1\sigma C_1\gamma_s\gamma_m$. Так как

$$C_0 \sigma C_0 = \frac{\sigma^{-1} e}{e \sigma^{-1} e} \sigma \frac{\sigma^{-1} e}{e \sigma^{-1} e} = \frac{\langle \sigma^{-1} e, e \rangle}{(e \sigma^{-1} e)^2} = \frac{1}{e \sigma^{-1} e} > 0,$$

то по формуле (12)
$$\overline{\text{cov}}_N = \frac{1}{e\sigma^{-1}e} + \frac{C_1\sigma C_1}{N^2 - N} \sum_{s, m = 1, s \neq m}^N \gamma_s \gamma_m.$$

При N+1 портфелей их средняя ковариация $\overline{\operatorname{cov}}_{N+1} =$

$$= \frac{1}{e\sigma^{-1}e} + \frac{C_1\sigma C_1}{(N+1)^2 - (N+1)} \sum_{s, m=1, s \neq m}^{N+1} \gamma_s \gamma_m.$$
 Преобразуем последнее выражение к виду

$$\overline{\operatorname{cov}}_{N+1} = \frac{1}{e\sigma^{-1}e} + \frac{C_{1}\sigma C_{1}}{(N+1)N} \left(\sum_{s, m=1, s \neq m}^{N} \gamma_{s} \gamma_{m} + \frac{1}{2 \sum_{m=1}^{N} \gamma_{N+1} \gamma_{m}} \right) = \frac{1}{e\sigma^{-1}e} + \frac{C_{1}\sigma C_{1}}{(N+1)N} \frac{N-1}{N+1} \sum_{s, m=1, s \neq m}^{N} \gamma_{s} \gamma_{m} + \frac{2 C_{1}\sigma C_{1}}{(N+1)N} \sum_{m=1}^{N} \gamma_{N+1} \gamma_{m} = \frac{1}{e\sigma^{-1}e} + \frac{C_{1}\sigma C_{1}}{(N-1)N} \sum_{s, m=1, s \neq m}^{N} \gamma_{s} \gamma_{m} + \frac{2 C_{1}\sigma C_{1}}{(N-1)N(N+1)} \left((N-1) \sum_{m=1}^{N} \gamma_{N+1} \gamma_{m} - \sum_{s, m=1, s \neq m}^{N} \gamma_{s} \gamma_{m} \right).$$

Значит,
$$\overline{\text{cov}}_{N+1} = \overline{\text{cov}}_N + \frac{2C_1\sigma C_1}{(N+1)} \left(\gamma_{N+1} \frac{1}{N} \sum_{m=1}^N \gamma_m - \frac{1}{N} \gamma_m \right)$$

$$-\frac{1}{(N-1)N}\sum_{s,\ m=1,\ s\neq m}^{N}\gamma_{s}\gamma_{m}$$
). Учитывая, что $C_{1}\sigma C_{1}>0$, по-

лучаем утверждение теоремы. •

Таким образом, для того чтобы при добавлении (N+1)-го полноразмерного оптимального портфеля средняя ковариация совокупности портфелей уменьшалась, необходимо и достаточно, чтобы коэффициент риска (N+1)-го портфеля был меньше отношения среднего попарных произведений коэффициентов риска первых N портфелей к среднему их коэффициентов риска.

Следствие 1. Для того чтобы при добавлении (N+1)-го полноразмерного оптимального портфеля средняя ковариация совокупности портфелей уменьшалась, достаточно выполнения неравенства

$$\gamma_{N+1} < \frac{1}{N-1} \min_{m=1, ..., N} \sum_{s=1, s \neq m}^{N} \gamma_{s}.$$



Доказательство. Из доказательства теоремы 3 имеем $\overline{\text{cov}}_{N+1} < \overline{\text{cov}}_N$, если

$$(N-1)\sum_{m=1}^{N}\gamma_{N+1}\gamma_{m} - \sum_{s, m=1, s \neq m}^{N}\gamma_{s}\gamma_{m} < 0$$

или
$$\sum_{m=1}^{N} \gamma_m \left((N-1) \gamma_{N+1} - \sum_{s=1, s \neq m}^{N} \gamma_s \right) < 0.$$

Для того чтобы последнее соотношение имело место достаточно, чтобы $\forall m=1,...,N$ $(N-1)\gamma_{N+1}<\sum_{s=1,\,s\neq\,m}^{N}\gamma_s$. Это эквивалентно $\gamma_{N+1}<\frac{1}{N-1}\min_{m=1,\,...,\,N}\sum_{s=1,\,s\neq\,m}^{N}\gamma_s$, что и требовалось доказать. lacktriangle

Таким образом, появление оптимального полноразмерного портфеля, коэффициент риска ко-

торого
$$\gamma_{N+1} < \frac{1}{N-1} \min_{m=1,...,N} \sum_{s=1,s\neq m}^{N} \gamma_s$$
, уменьшает

среднюю ковариацию портфелей и рынок становится более устойчивым.

Следствие 2 из теоремы 3. Для того чтобы при добавлении (N+1)-го полноразмерного оптимального портфеля средняя ковариация совокупности портфелей не уменьшалась, достаточно выполнения не-

равенства
$$\gamma_{N+1} \geqslant \frac{1}{N-1} \sum_{m=1}^{N} \gamma_m$$
.

Доказательство. Из доказательства теоремы 3 имеем $\overline{\text{cov}}_{N+1} \geqslant \overline{\text{cov}}_N$, если

$$(N-1)\sum_{m=1}^{N}\gamma_{N+1}\gamma_m - \sum_{s, m=1, s \neq m}^{N}\gamma_s\gamma_m \ge 0,$$

т. е.
$$(N-1)\gamma_{N+1} \geqslant \sum_{s, m=1, s \neq m}^{N} \gamma_s \gamma_m / \sum_{m=1}^{N} \gamma_m$$
 или, используя

тождество
$$\sum_{m=1}^{N} \sum_{s=1}^{N} \gamma_s \gamma_m = \sum_{m=1}^{N} (\gamma_m)^2 - \sum_{m=1}^{N} (\gamma_m)^2$$
, имеем

$$(N-1)\gamma_{N+1} \geqslant \sum_{m=1}^{N} \gamma_m - \sum_{m=1}^{N} (\gamma_m)^2 / \sum_{m=1}^{N} \gamma_m$$
. Для выполнения

последнего неравенства достаточно $\gamma_{N+1} \geqslant \frac{1}{N-1} \sum_{m=1}^{N} \gamma_m$

что и требовалось доказать.

Таким образом, если на фондовом рынке появляется оптимальный полноразмерный портфель, коэффициент риска которого γ_{N+1} не меньше смещенной средней величины коэффициентов риска уже присутствующих на рынке портфелей (т. е. риск портфеля относительно высок), то средняя ковариация полноразмерных оптимальных портфелей по крайней мере не уменьшается (т. е. фондовый рынок не становится более устойчивым).

3.2. Влияние оценок доходности рыночного индекса на устойчивость рынка

Рассмотрим в качестве примера широко используемую рыночную модель [2, с. 293], которая предполагает, что случайные величины доходностей финансовых инструментов за данный период времени связаны со случайной доходностью рыночного индекса за данный период соотношением $r_i = a_{iI} + \beta_{iI}r_I + \varepsilon_i, i = 1, ..., n$, где α_{iI} — коэффициент смещения, β_{iI} — коэффициент наклона (бета-коэффициент), r_I — случайное значение доходности рыночного индекса I, ε_i — случайная погрешность (белый шум). Тогда математические ожидания доходностей $\bar{r}_i = a_{iI} + \beta_{iI}\bar{r}_I, i = 1, ..., n$, где \bar{r}_I — математическое ожидание доходности рыночного индекса I. Оптимальный полноразмерный портфель, определяемый из решения задачи (1), имеет вид

$$\begin{split} x^0(\gamma,\ \overline{r}_I\) &=\ C_0 +\ C_1\gamma = \\ &=\frac{\sigma^{-1}e}{e\sigma^{-1}e}\ + \left(\sigma^{-1}(a_I+\beta_I\overline{r}_I\) - \frac{e\sigma^{-1}(a_I+\beta_I\overline{r}_I)}{e\sigma^{-1}e}\sigma^{-1}e\right)\!\gamma, \end{split}$$
 где $a_I = (a_{II},\ ...,\ a_{nI}),\ \beta_I = (\beta_{II},\ ...,\ \beta_{nI}).$

Пусть инвесторы руководствуются субъективной информацией относительно величины \bar{r}_I . Согласно формуле (6) получаем выражение для ковариации случайных величин доходностей любых двух портфелей $x^0(\gamma_s, \bar{r}_{I,s}) = C_0 + C_{1, s}\gamma_s$ и $x^0(\gamma_m, \bar{r}_{I,m}) = C_0 + C_{1, m}\gamma_m$, $s, m = 1, ..., N, s \neq m$ (из выражения (5) нетрудно видеть, что $\forall m \ C_0 \sigma C_{1, m} = 0$):

$$\begin{aligned} \operatorname{cov}(r_{x^{0s}}, \, r_{x^{0m}}) &= x^{0}(\gamma_{s}, \, \bar{r}_{I, \, s}) \sigma x^{0}(\gamma_{m}, \, \bar{r}_{I, \, m}) = \\ &= C_{0} \sigma C_{0} + C_{1, \, s} \sigma C_{1, \, m} \gamma_{s} \gamma_{m} = \\ &= \frac{1}{e \sigma^{-1} e} + \left((a_{I} + \beta_{I} \bar{r}_{I, \, s}) - \frac{e \sigma^{-1} (a_{I} + \beta_{I} \bar{r}_{I, \, s})}{e \sigma^{-1} e} e \right) \times \\ &\times \sigma^{-1} \left((a_{I} + \beta_{I} \bar{r}_{I, \, m}) - \frac{e \sigma^{-1} (a_{I} + \beta_{I} \bar{r}_{I, \, m})}{e \sigma^{-1} e} e \right) \gamma_{s} \gamma_{m}. \end{aligned}$$

Тогда согласно формуле (12) средняя ковариация портфелей

$$\overline{\text{cov}}_{N} = \frac{1}{N^{2} - N} \sum_{s, m = 1, s \neq m}^{N} x^{0}(\gamma_{s}, \overline{r}_{I, s}) \sigma x^{0}(\gamma_{m}, \overline{r}_{I, m}) =
= \frac{1}{e\sigma^{-1}e} + \frac{1}{N^{2} - N} \sum_{s, m = 1, s \neq m}^{N} \left((a_{I} + \beta_{I}\overline{r}_{I, m}) - \frac{e\sigma^{-1}(a_{I} + \beta_{I}\overline{r}_{I, s})}{e\sigma^{-1}e} e \right) \sigma^{-1} \left((a_{I} + \beta_{I}\overline{r}_{I, m}) - \frac{e\sigma^{-1}(a_{I} + \beta_{I}\overline{r}_{I, m})}{e\sigma^{-1}e} e \right) \gamma_{s} \gamma_{m}. \tag{13}$$



Преобразуем разность
$$(a_I + \beta_I \overline{r}_{I,s}) - \frac{e\sigma^{-1}(a_I + \beta_I \overline{r}_{I,s})}{e\sigma^{-1}e}e$$
 к виду $\left(\beta_I - \frac{e\sigma^{-1}\beta_I}{e\sigma^{-1}e}e\right)\overline{r}_{I,s} + \left(a_I - \frac{e\sigma^{-1}a_I}{e\sigma^{-1}e}e\right) = \xi \overline{r}_{I,s} + \eta$. Аналогично для раз-

ности
$$(a_I + \beta_I \bar{r}_{I,m}) - \frac{e\sigma^{-1}(a_I + \beta_I \bar{r}_{I,m})}{e\sigma^{-1}e}e$$
 имеем

$$\left(\beta_{I} - \frac{e\sigma^{-1}\beta_{I}}{e\sigma^{-1}e}e\right)\bar{r}_{I,\,m} + \left(a_{I} - \frac{e\sigma^{-1}a_{I}}{e\sigma^{-1}e}e\right) = \xi\bar{r}_{I,\,m} + \eta.$$
 Tor-

стве (13), примет вид $(\xi \bar{r}_{I,s} + \eta)\sigma^{-1}(\xi \bar{r}_{I,m} + \eta)\gamma_s \gamma_m$. После преобразования получим выражение $(\xi \bar{r}_{I,s} + \eta)\sigma^{-1}(\xi \bar{r}_{I,m} + \eta)\gamma_s \gamma_m = [(\xi \sigma^{-1} \xi) \bar{r}_{I,s} \bar{r}_{I,m} + \eta)\gamma_s \gamma_m = [(\xi \sigma^{-1} \xi) \bar{r}_{I,s} \bar{r}_{I,m} + \eta)\gamma_s \gamma_m = [(\xi \sigma^{-1} \xi) \bar{r}_{I,s} \bar{r}_{I,m} + \eta)\gamma_s \gamma_m = [(\xi \sigma^{-1} \xi) \bar{r}_{I,s} \bar{r}_{I,m} + \eta)\gamma_s \gamma_m = [(\xi \sigma^{-1} \xi) \bar{r}_{I,s} \bar{r}_{I,m} + \eta)\gamma_s \gamma_m = [(\xi \sigma^{-1} \xi) \bar{r}_{I,s} \bar{r}_{I,m} + \eta)\gamma_s \gamma_m = [(\xi \sigma^{-1} \xi) \bar{r}_{I,s} \bar{r}_{I,m} + \eta)\gamma_s \gamma_m = [(\xi \sigma^{-1} \xi) \bar{r}_{I,s} \bar{r}_{I,m} + \eta)\gamma_s \gamma_m = [(\xi \sigma^{-1} \xi) \bar{r}_{I,s} \bar{r}_{I,m} + \eta)\gamma_s \gamma_m = [(\xi \sigma^{-1} \xi) \bar{r}_{I,s} \bar{r}_{I,m} + \eta)\gamma_s \gamma_m = [(\xi \sigma^{-1} \xi) \bar{r}_{I,s} \bar{r}_{I,m} + \eta)\gamma_s \gamma_m = [(\xi \sigma^{-1} \xi) \bar{r}_{I,s} \bar{r}_{I,m} + \eta)\gamma_s \gamma_m = [(\xi \sigma^{-1} \xi) \bar{r}_{I,s} \bar{r}_{I,m} + \eta)\gamma_s \gamma_m = [(\xi \sigma^{-1} \xi) \bar{r}_{I,s} \bar{r}_{I,m} + \eta)\gamma_s \gamma_m = [(\xi \sigma^{-1} \xi) \bar{r}_{I,s} \bar{r}_{I,m} + \eta)\gamma_s \gamma_m = [(\xi \sigma^{-1} \xi) \bar{r}_{I,s} \bar{r}_{I,m} + \eta)\gamma_s \gamma_m = [(\xi \sigma^{-1} \xi) \bar{r}_{I,s} \bar{r}_{I,m} + \eta)\gamma_s \gamma_m = [(\xi \sigma^{-1} \xi) \bar{r}_{I,s} \bar{r}_{I,m} + \eta)\gamma_s \gamma_m = [(\xi \sigma^{-1} \xi) \bar{r}_{I,s} \bar{r}_{I,m} + \eta)\gamma_s \gamma_m = [(\xi \sigma^{-1} \xi) \bar{r}_{I,s} \bar{r}_{I,s} \bar{r}_{I,m} + \eta)\gamma_s \gamma_m = [(\xi \sigma^{-1} \xi) \bar{r}_{I,s} \bar{r}_{I,s} \bar{r}_{I,s} + \eta)\gamma_s \gamma_m = [(\xi \sigma^{-1} \xi) \bar{r}_{I,s} \bar{r}_{I,s} + \eta)\gamma_s \gamma_m + \eta \gamma_s \gamma_m +$ $+ ησ^{-1}η + (ξσ^{-1}η)(\bar{r}_{I,s} + \bar{r}_{I,m})]γ_sγ_m],$ с учетом которого равенство (13) примет вид

$$\overline{\text{cov}}_{N} = \frac{1}{e\sigma^{-1}e} + \frac{1}{N^{2} - N} \sum_{s, m = 1, s \neq m}^{N} [(\xi\sigma^{-1}\xi)\bar{r}_{I,s}\bar{r}_{I,m} + \eta\sigma^{-1}\eta + (\xi\sigma^{-1}\eta)(\bar{r}_{I,s} + \bar{r}_{I,m})]\gamma_{s}\gamma_{m} = \frac{1}{e\sigma^{-1}e} + \frac{1}{N^{2} - N} \left[(\xi\sigma^{-1}\xi) \sum_{s, m = 1, s \neq m}^{N} \bar{r}_{I,s}\bar{r}_{I,m}\gamma_{s}\gamma_{m} + (\eta\sigma^{-1}\eta) \times \sum_{s, m = 1, s \neq m}^{N} \gamma_{s}\gamma_{m} + (\xi\sigma^{-1}\eta) \sum_{s, m = 1, s \neq m}^{N} (\bar{r}_{I,s} + \bar{r}_{I,m})\gamma_{s}\gamma_{m} \right].$$

Так как здесь $e\sigma^{-1}e > 0$, $\gamma_s \gamma_m > 0 \ \forall s, m, \ (\xi \sigma^{-1} \xi) \ge 0$ и $(\eta \sigma^{-1} \eta) \leq 0$, то средняя ковариация $\overline{\text{cov}}_N$ может быть отрицательной только тогда, когда $(\xi \sigma^{-1} \eta) < 0$. Пусть ξ и η не равны нулю. При $\overline{r}_{I,s}$ и $\overline{r}_{I,m} \to \infty$ произведение $\bar{r}_{I,\,s}\,\bar{r}_{I,\,m} \to \infty$ быстрее, чем $\bar{r}_{I,\,s}\,+\,\bar{r}_{I,\,m}$, и при больших значениях $\bar{r}_{I,s}$ и $\bar{r}_{I,m}$ имеем $\overline{\text{cov}}_N \gg 0$; а при $\bar{r}_{I,\,s}$ и $\bar{r}_{I,\,m} \to 0$ имеем $\overline{\mathrm{cov}}_N \geq 0$. Поэтому при прогнозе быстро растущего рынка поведение инвесторов приводит к высокой вероятности кризисных явлений. При прогнозе падения рынка вероятность кризисных явлений также значительна. При умеренных значениях \bar{r}_{Ls} , \bar{r}_{Lm} и условии $(\xi \sigma^{-1} \eta) < 0$ состояние рынка относительно устойчивое.

Нетрудно видеть, что при одинаковых оценках всеми инвесторами значений доходности индекса $\bar{r}_{I,s} = \bar{r}_{I,m} = \bar{r}_{I,s}, m = 1, ...,N, s \neq m$ имеем $(\xi \bar{r}_{I,s} + \eta)\sigma^{-1}(\xi \bar{r}_{I,m} + \eta)\gamma_s \gamma_m = (\xi \bar{r}_I + \eta)\sigma^{-1}(\xi \bar{r}_I + \eta)\gamma_s \gamma_m \ge 0$. Значит, если различное поведение инвесторов связано только лишь с различным отношением к риску (выбор параметра у), то средняя ковариация случайных значений доходностей портфелей, определяемая соотношением (13), всегда положительная, что согласуется с утверждением теоремы 1.

ЗАКЛЮЧЕНИЕ

Проведенное исследование показывает, что однотипное поведение инвесторов (оптимизация портфеля при одинаковой информированности) даже при разном отношении к риску служит фактором неустойчивости фондового рынка. С другой стороны, инвесторы, придерживаясь различных прогнозов развития фондового рынка, как следствие различной их информированности, будут выбирать разные стратегии (портфели) и, тем самым, повышать устойчивость всей системы (фондового рынка). При этом средняя ковариация доходностей портфелей инвесторов является приближенным значением дисперсии рыночной доходности и может быть принята в качестве оценки риска структурированного рынка (состоящего из большого числа портфелей). Дальнейшие исследования мы связываем с рассмотрением вопросов коллективного риска и устойчивости фондового рынка при использовании других моделей оценки математических ожиданий и ковариаций доходностей финансовых инструментов, а также иных моделей формирования инвестиционных портфелей.

ЛИТЕРАТУРА

- 1. Горелик В.А., Золотова Т.В. Модели оценки коллективного и системного риска. Научное издание. - М.: ВЦ РАН,
- Шарп У., Александер Г., Бейли Дж. Инвестиции: Пер. с англ. М.: ИНФРА-М, 2004. Т. XII. 1028 с.
- Ширяев А.Н. Основы стохастической финансовой математики. Т. 1. Факты. Модели. — М.: ФАЗИС, 1998. — 512 с.
- *Горелик В.А., Золотова Т.В.* Оценка коллективного риска на фондовом рынке // Моделирование, декомпозиция и оптимизация сложных динамических процессов / ВЦ PAH. — M., 2010. — C. 55—65.
- Горелик В.А., Золотова Т.В. О некоторых обобщениях результатов исследования корреляции доходностей инвестиционных портфелей // Моделирование, декомпозиция и оптимизация сложных динамических процессов / ВЦ РАН. — М., 2012. — С. 40—52.
- 6. Markowitz H.M. Portfolio selection // Journal of Finance. 1952. — N 7. — P. 77—91.

 Markowitz H. M. Portfolio Selection: Efficient Diversification
- of Investment. N.-Y.: Wiley, 1959. 344 c.
- Горелик В.А., Золотова Т.В. Оценка корреляции доходности инвестиционных портфелей и устойчивость фондового рынка // Государственный университет Минфина России.
- Финансовый журнал. 2012. № 3. С. 43—52. *Горелик В.А., Золотова Т.В.* Некоторые вопросы оценки корреляции доходностей инвестиционных портфелей // Проблемы управления. — 2011. — № 3. — С. 36—42.

Статья представлена к публикации членом редколлегии Р.М. Нижегородцевым.

Виктор Александрович Горелик — д-р физ.-мат. наук, вед. науч. сотрудник, Вычислительный центр им. А.А. Дородницына РАН, **☎** (499) 135-62-04, ⊠ vgor16@mail.ru,

Татьяна Валерьяновна Золотова — д-р физ.-мат. наук, проф., Финансовый университет при Правительстве РФ,

☎ (499) 277-21-02, ⊠ tgold11@mail.ru.



УДК 004.8

ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ ВАЛЮТНОГО ХЕДЖИРОВАНИЯ НА ОСНОВЕ РЕЗУЛЬТАТОВ НЕЙРОСТРУКТУРНОГО ПРОГНОЗИРОВАНИЯ¹

П.В. Сараев, Ю.Е. Сяглова

Дан анализ эффективности применения результатов нейроструктурного прогнозирования временных рядов (курсов валют) в процессе хеджирования валютных рисков с применением производных финансовых инструментов. Описан разработанный программный продукт. Рассмотрена методика расчетов по исследованию эффективности применения нейроструктурных прогнозов. Приведены результаты вычислительных экспериментов.

Ключевые слова: нейроструктурное моделирование, прогнозирование, временные ряды, хеджирование, валютные риски.

ВВЕДЕНИЕ

Компании, существенная часть выручки которых номинирована в иностранной валюте, регулярно сталкиваются с необходимостью продажи валюты для исполнения рублевых обязательств. Изменения обменных курсов служат источником валютных рисков. Часто, когда такой риск существует, его следует хеджировать, в частности, с помощью производных финансовых инструментов. Поэтому компании часто пользуются прогнозами обменных курсов, чтобы решить, следует ли страховать риски [1]. Цель данной работы состоит в анализе применения методов нейроструктурного моделирования для повышения эффективности валютного хеджирования.

1. ПОСТАНОВКА ЗАДАЧИ

Для управления валютным риском рассматривается применение стратегий:

- отказ от хеджирования (принятие валютного риска в полном объеме без мероприятий по хеджированию);
- заключение трехмесячных форвардов (стратегия позволяет сократить потери на укреплении рубля).

Так как прогнозы банков могут в значительной степени отличаться друг от друга, то у ряда компаний возникает потребность в собственном инструменте прогнозирования. Его применение не исключает прогнозы банков, он служит дополнительным источником информации о возможном движении валютных курсов. Заметим, что нет оснований считать прогноз, полученный с помощью такой системы, более точным, чем прогноз банка. Поэтому вырабатываемые на основе анализа собственного прогноза инструкции руководству компании носят рекомендательный характер.

Для оценки реального эффекта от прогнозирования курсов валют по каждой из перечисленных стратегий в качестве расчетного периода рассматривается 2012 г. Исходные данные для исследования представляют собой временные ряды спотового и форвардного курсов валют, полученные с помощью информационной системы «Bloomberg», а также потребность в рублях (объем продажи валюты). В качестве форвардного курса берется цена трехмесячного форвардного контракта, рассчитанная системой «Bloomberg», со страйком, равным спотовой цене валюты на каждую среду расчетного периода.

В случае отказа от хеджирования весь объем валюты, предназначенный для продажи, реализуется по спот-курсу. В случае заключения трехмесячных форвардов весь объем валюты, предназначенный для продажи, реализуется по соответствующему форвардному курсу. Эффект от стратегии рассчитывается как разница между прибылью или убыт-

¹ Работа выполнена при финансовой поддержке РФФИ (грант № 11-07-97504-р_центр_а).



ком, полученными от реализации форвардов, и прибылью или убытком, полученными от продажи валюты по спот-курсу (без хеджирования).

Предполагается, что компания заключает форвардные контракты еженедельно. Таким образом, построение прогнозов курсов валют осуществляется каждую неделю с горизонтом прогнозирования в 13 значений, соответствующих еженедельным значениям валютных курсов. Для получения прогноза на каждое следующее измерение исходный временной ряд дополняется уже известным фактическим значением. В том случае, если полученное прогнозное значение оказывается больше форвардного курса, рекомендуется отказаться от заключения форвардного контракта и реализовать валюту по спотовому курсу. В противном случае принимается решение о заключении форвардного контракта. Эффект от применения прогнозов рассчитывается по фактическим значениям курсов валют.

Задача исследования состоит не в учете и оценке возникающих валютных рисков, а в иллюстрации возможности и эффекта от применения нейроструктурного прогнозирования в процессе управления валютными рисками. В этой связи в работе используется подход, учитывающий лишь сравнение стратегий хеджирования без учета и оценки возможных рисков.

2. НЕЙРОСТРУКТУРНОЕ МОДЕЛИРОВАНИЕ ДИНАМИЧЕСКИХ СИСТЕМ

Пусть значения временного ряда доступны в дискретные моменты времени $t=1,\,2,\,...,\,T$. Обозначим временной ряд $Y(t)=Y(1),\,Y(2),\,...,\,Y(T)$. В момент времени T необходимо определить значения процесса Y(t) в будущие моменты времени $T+1,\,...,\,T+P$. Для прогнозирования значений временного ряда требуется определить функциональную зависимость, отражающую связь между прошлыми и будущими значениями этого ряда — модель прогнозирования:

$$\hat{Y}(t) = F(Y(t-1), Y(t-2), Y(t-3), ...),$$

где $\hat{Y}(t)$ — прогнозные (расчетные) значения временного ряда Y(t).

Для прогнозирования элементов временных рядов широко применяют метод скользящих окон. Его основная идея состоит в использовании двух окон, входного и выходного, фиксированных размеров. Входное окно формирует данные для входов модели, а выходное, соответственно, для выходов (оценок). Обучающий пример формируется

путем наложения на массив исходных данных временного окна.

При моделировании динамических зависимостей сложных систем хорошо зарекомендовали себя методы нейросетевого моделирования и прогнозирования [2, 3]. Более мощный математический аппарат по сравнению с применяемыми моделями нейронных сетей прямого распространения (НСПР) представляют собой нейроструктурные модели (НСМ), которые обобщают и развивают нейросетевое моделирование. Они позволяют описывать более широкий класс моделей, чем НСПР, благодаря применению различных функций активации (Φ A) в нейроноподобных элементах (НПЭ) и возможности передачи выходных сигналов НПЭ на любой из последующих слоев аналогично сетям каскадной корреляции Фальмана [2]. Это позволяет описывать различные архитектуры нейросетевого типа — нейронные сети с радиальными базисными функциями, вероятностные нейронные сети, нелинейные нейронные сети Вольтерра, а также нечеткие модели Такаги-Суджено, гибридные нейро-нечеткие модели структуры ANFIS и др. [4, 5]. При идентификации таких моделей необходимо дополнительно решить проблемы выбора оптимальных ФА в НПЭ, которая для каждого НПЭ может быть своей. В этом случае целесообразно задавать некоторое множество допустимых ФА для применения в НСМ. Применение различных ФА позволяет повысить вычислительные способности НСМ. Применение периодических тригонометрических функций, например $\sigma(net) = \sin(net)$, позволяет при моделировании динамических процессов выявлять не только трендовую, но и сезонную составляющую. Исследование целесообразности и эффективности применения различных ФА для моделирования динамических процессов представлено в работе [6].

Функционирование НПЭ — базового элемента — подобно искусственному нейрону. Нейроподобный элемент преобразует векторный вход x в скалярный выход y с возможным использованием векторов весов w и априори задаваемых параметров α по формуле $y = \sigma(net(x, w); \alpha)$, где веса w или параметры α могут отсутствовать, net — уровень активности НПЭ — дифференцируемая функция по весам w; σ — дифференцируемая ФА по аргументу net(x, w). В частном случае

$$net(x, w) = \sum_{i=1}^{n} x_i w_i,$$

где n — число входов. Если параметры α присутствуют, то Φ A — параметризованная.



Нейроструктурная модель — совокупность связанных послойно организованных НПЭ. В отличие от НСПР на входы НПЭ слоя (i+1) могут поступать выходы НПЭ от любого из предыдущих слоев. Функция активации выбирается из множества Ω допустимых Φ A, которое может ограничиваться конкретным классом применяемых НСМ.

Задача обучения НСМ, как и НСПР, заключается в выборе размера входного окна и настройке весов НПЭ. Перед применением НСМ необходимо обучить на основе данных обучающего множества. Интегральная степень соответствия НСМ данным из обучающего множества, как правило, задается квадратичным функционалом качества:

$$J(w) = \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{r} (y_{ij}(w) - \tilde{y}_{ij})^{2},$$

где k — число примеров обучающего множества, r — число выходов НСМ, $y_{ii}(w)$ — j-й выход модели на

i-м примере, \tilde{y}_{ij} — известный выход (указание учителя) j-го выхода для i-го примера. Для решения задачи обучения могут быть применены различные методы оптимизации, учитывающие специфику задачи обучения нейроструктурных моделей [2, 7].

Для конструктивного построения НСМ предлагается алгоритм, приведенный на рис. 1. На начальном этапе выбирается тип НСМ, что определяет специфику связей в НСМ и ограничения на выбор Φ А. Для сохранения гарантий монотонности конструирования моделей множество Ω должно включать в себя единичную Φ А σ (net) = net.

При наращивании структуры для сохранения монотонности убывания ошибки обучения необходимо применять специальный способ задания некоторых вновь добавляемых весов. При добавлении нового НПЭ в последний скрытый слой его выход должен идти на выходные НПЭ с нулевыми весами. При формировании нового скрытого слоя перед выходным его выход должен поступать на выходные НПЭ с единичными весами, а при наличии фиктивных входов для выходных НПЭ соответствующие им веса должны быть нулевыми.

Для прогнозирования динамических систем на основе нейроструктурных моделей разработано специальное программное обеспечение. Разработанный программный продукт решает задачи:

- загрузки исходных данных из текстового файла;
- разбиения исходного массива данных на примеры обучающего множества;
- генерации начальных весов для обучения модели;

- определения необходимого числа НПЭ в скрытом слое;
- подбора ФА для каждого НПЭ скрытого слоя из списка заданных пользователем в соответствии с минимумом ошибки обучения;
- визуализации полученного прогноза и сохранения прогнозных значений в текстовом файле;
 - просмотра структуры обученной сети.

Программа реализована на языке C++ в среде программирования Borland C++ Builder 6 [8].

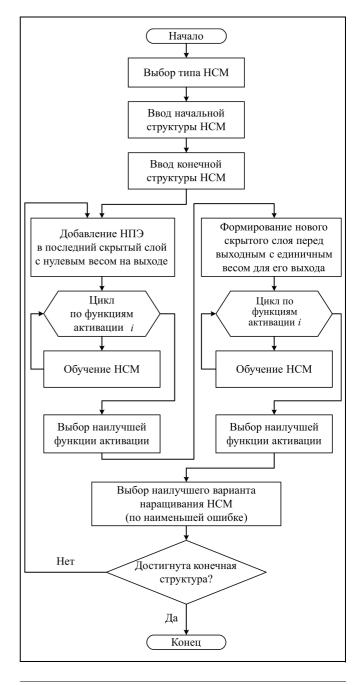


Рис. 1. Алгоритм конструктивного наращивания структуры модели



3. РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

В данной работе рассматривается двухслойная НСМ по типу НСПР. В работе применяется алгоритм, позволяющий снизить ошибку обучения НСМ благодаря использованию различных ФА и подбору оптимального числа нейронов скрытого слоя.

При проведении исследования применялся метод ретроспективного анализа для построения прогноза курсов доллара США и евро по еженедельным данным за период с января 2000 г. по декабрь 2011 г. По результатам экспериментов наилучший прогноз курса доллара США получен при использовании десяти входов и четырех НПЭ скрытого слоя со следующими Φ A: $\sigma_1(net) = 1/(1 + e^{-net})$,

$$\sigma_2(net) = 1/(1 + e^{-net}), \ \sigma_3(net) = e^{-net^2} \ \text{if } \sigma_4(net) =$$

 $=e^{-net^2}$. Среднее относительное отклонение прогнозных значений от фактических составило 3,40 %, среднее квадратическое отклонение (СКО) — 1,2972 руб., ошибка обучения составила 0,0360. График полученного прогноза приведен на рис. 2.

Наилучший прогноз курса евро получен при использовании десяти входов и двух НПЭ скрытого слоя со следующими ФА: $\sigma_1(net) = 1/(1 + e^{-net})$ и $\sigma_2(net) = 1/(1 + e^{-net})$. Среднее относительное

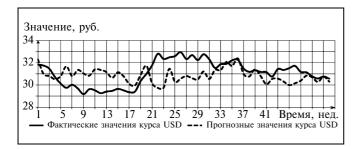


Рис. 2. Прогноз курса доллара США

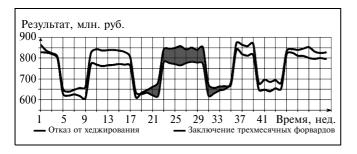


Рис. 4. Сравнение результатов стратегий (доллар США)

отклонение прогнозных значений от фактических составило 3,20 %, СКО — 1,6540 руб., ошибка обучения составила 0,0349. График полученного прогноза приведен на рис. 3.

Сравнение результатов первоначально рассматриваемых стратегий приведено на рис. 4 и 5. Предполагается, что компания заключает форвардные контракты постоянно (еженедельно) без возможности остановки, поскольку нет достоверной информации об изменении конъюнктуры валютного рынка. Закрашенные области на графиках иллюстрируют отрицательный эффект (убыток) от реализации форвардов по сравнению с отказом от хеджирования.

На рис. 6 и 7 отражен результат от некоторой возможной идеальной стратегии, позволяющей не только достигнуть безубыточности посредством заключения форвардных контрактов, но и получить прибыль.

На рис. 8 и 9 отражен результат от применения стратегии с использованием полученных прогнозов курсов валют. При этом форвардные контракты заключаются только в том случае, если, согласно прогнозу, сделка выгодна. Как видно из приведенных рисунков, предлагаемая стратегия существенно снижает отрицательный эффект от реализации форвардов.

Суммарный эффект (для доллара США и евро) от применения стратегии с учетом прогноза курсов валют составил 1440 млн руб., а без использования

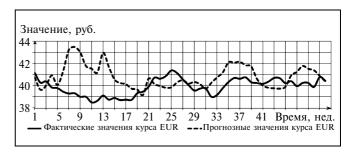


Рис. 3. Прогноз курса евро

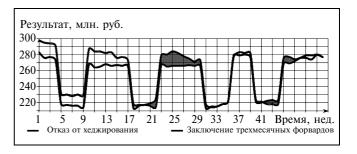


Рис. 5. Сравнение результатов стратегий (евро)



Рис. 6. Сравнение результатов стратегий (доллар США)

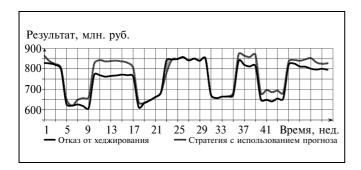


Рис. 8. Сравнение результатов стратегий (доллар США)

прогноза — 706 млн руб. Таким образом, можно заключить, что учет прогнозирования на основе НСМ курсов валют значительно повышает эффективность применения форвардных контрактов в процессе управления валютными рисками посредством хеджирования.

ЗАКЛЮЧЕНИЕ

Выполнен анализ эффективности применения результатов нейроструктурного прогнозирования курсов валют в процессе хеджирования валютных рисков с применением производных финансовых инструментов. Результаты вычислений позволили сравнить эффективность применения форвардных контрактов с использованием результатов прогнозирования и без них. Анализ показал, что использование результатов нейроструктурного прогнозирования курсов валют повышает эффективность применения производных финансовых инструментов в процессе управления валютными рисками компании.

ЛИТЕРАТУРА

1. *Рэдхэд К., Хьюс С.* Управление финансовыми рисками / Пер. с англ. — М.: ИНФРА-М, 1996. — 288 с.

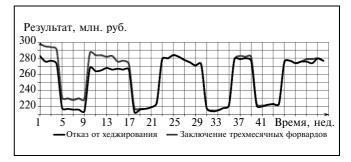


Рис. 7. Сравнение результатов стратегий (евро)

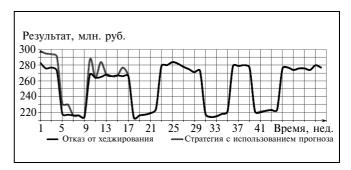


Рис. 9. Сравнение результатов стратегий (евро)

- Осовский С. Нейронные сети для обработки информации. М.: Финансы и статистика, 2002. — 344 с.
- 3. *Хайкин С.* Нейронные сети: полный курс / Пер. с англ., 2-е изд. М.: Изд. дом «Вильямс», 2006. 1104 с.
- 4. *Сараев П.В.* Развитие нейросетевого моделирования сложных систем на основе нейроструктурного подхода // Вести вузов Черноземья. 2012. № 2 (28). С. 30—35.
- 5. *Погодаев А.К.*, *Блюмин С.Л.*, *Сараев П.В.* Нейроструктурное моделирование: некоторые результаты и направления развития // Вести вузов Черноземья. 2012. № 4 (30). С. 30—37
- 6. Сараев П.В., Сяглова Ю.Е. Анализ эффективности выбора функций активации в нейросетевом прогнозировании // Системы управления и информационные технологии. 2012. № 3.1 (49). С. 165—169.
- 7. *Сараев П.В.* Идентификация нейросетевых моделей. Липецк: ЛГТУ, 2011. 94 с.
- 8. *Сяглова Ю.Е.*, *Сараев П.В.* Нейросетевое моделирование и прогнозирование динамики временных рядов. М.: Фонд алгоритмов и программ ВНТИЦ, 2013. Зарег. № 50201350432 от 14.05.2013.

Статья представлена к публикации членом редколлегии P.M. Нижегородцевым.

Павел Викторович Сараев — канд. техн. наук, декан, Липецкий государственный технический университет,

(4742) 32-80-02, ⋈ psaraev@yandex.ru,

Юлия Евгеньевна Сяглова — инженер-программист, ООО «НЛМК — информационные технологии», г. Липецк, **☎** (4742) 44-21-54, ⊠ yuliyasyaglova@yandex.ru.

УДК 519.862.7

АКТИВНЫЙ ПОТРЕБИТЕЛЬ: ЗАДАЧА ОПТИМИЗАЦИИ ПОТРЕБЛЕНИЯ ЭЛЕКТРОЭНЕРГИИ И СОБСТВЕННОЙ ГЕНЕРАЦИИ¹

И.О. Волкова, М.В. Губко, Е.А. Сальникова

Отмечено, что для решения задач управления спросом в условиях интеллектуальной энергетической системы (SmartGrid) необходима математическая модель принятия решений активным потребителем, однако существующие модели либо не учитывают важных аспектов поведения потребителя, либо слишком сложны для их учета в мультиагентном моделировании. Предложена математическая модель активного потребителя, на основе которой сформулирована и решена задача оптимизации режимов работы электроприборов и оборудования потребителя, а также определения условий загрузки собственной генерации. Приведены условия, при которых задача потребителя допускает простое и эффективное решение. Предложенный подход проиллюстрирован на примере оптимизации режима работы оборудования отдельно взятого домохозяйства.

Ключевые слова: активный потребитель в энергетике, интеллектуальная энергосистема, управление спросом.

ВВЕДЕНИЕ

В последние годы в России начался процесс инновационных преобразований электроэнергетики в рамках концепции развития отрасли, получившей название «Интеллектуальной электроэнергетической системы России» и основывающейся на принципиальных положениях концепции SmartGrid. Часть этой концепции состоит в «активизации» потребителей, т. е. предоставлении потребителям возможности самостоятельного управления объемом получаемой электроэнергии и ее функциональными свойствами (уровнем надежности, качеством и др.) на основании баланса своих потребностей (минимизации затрат на электроэнергию и получении дохода от продажи электроэнергии и мощности) и возможностей энергосистемы с использованием информации о характеристиках цен, объемах поставок электроэнергии, надежности, качестве и др. [1].

Основные функции активного потребителя в электроэнергетической системе заключаются в:

— управлении собственным энергопотреблением в соответствии с необходимостью выполнения

¹ Работа выполнена при частичной финансовой поддержке РФФИ (проект № 13-07-00491).

своих производственных планов по выпуску продукции или обеспечению энергией домохозяйства путем оптимизации своих затрат на покупку электроэнергии с внешних рынков;

- определении условий загрузки собственной мощности (при ее наличии), для формирования заявки на участие в покупке/продаже электроэнергии на оптовом и розничном рынках;
- определении степени своего участия в предоставлении дополнительных услуг, заключающихся в предоставлении управляемых активных и реактивных нагрузок (мощностей) для управления со стороны системного оператора.

Реализация перечисленных функций требует создания системы внешних условий для функционирования активного потребителя и инструментов, которые позволят потребителям реализовывать свои возможности. Такие инструменты должны отвечать следующим требованиям:

- отражать экономические интересы активного потребителя;
- обеспечивать формирование оптимальной загрузки электроприборов и оборудования потребителя исходя из прогнозируемых ценовых сигналов, а также с учетом распределения нагрузки между различными типами оборудования потребителя;



— обеспечивать работу собственной генерации потребителя: электроснабжения внутри домохозяйства/предприятия или выдачи электроэнергии в сеть.

В настоящей работе представлена модель активного потребителя, на основе которой формулируется и решается задача оптимизации режимов работы электроприборов и оборудования потребителя, а также определения условий загрузки собственной генерации.

Разработанная авторами модель может использоваться как самим активным потребителем для формирования стратегии своего энергопотребления, так и энергосбытовой компанией и регулятором для выработки механизмов влияния на активных потребителей. Во втором случае предлагаемая модель представляет собой модель поведения объекта управления, позволяющую оценить отклик потребителя на ценовые сигналы, которые являются основным механизмом воздействия на управляемую систему (на совокупность активных потребителей), с целью обеспечения требуемого ее поведения. Требуемое поведение активных потребителей заключается в привлечении их к участию в оптимизации режимов функционирования энергосистемы, в том числе в сглаживании пиков потребления в рамках программ управления спросом (demand response) [2].

1. ОБЗОР ЛИТЕРАТУРЫ

Задачи управления поведением потребителя в энергосистеме рассматривались в различные периоды времени как отечественными, так и зарубежными учеными. В 1980-е гг. основная задача управления поведением потребителей электроэнергии в СССР состояла в снижении неравномерности графика нагрузки энергосистемы путем привлечения потребителей к регулированию потребления электроэнергии в целях снижения суммарных народно-хозяйственных затрат. Возможность участия в регулировании нагрузки рассматривалась только для крупных промышленных потребителей (потребителей-регуляторов), технологический процесс которых позволял гибко регулировать свою нагрузку в оперативном режиме. Задачи, рассматриваемые авторами, можно отнести к трем основным направлениям:

— оптимизация режимов электропотребления для действующих и проектируемых предприятий в целях снижения суммарных затрат предприятия при сохранении плана выпуска продукции при ограничении по технологически допустимому уровню снижения потребляемой мощности при тарифе на электроэнергию, дифференцированном по зонам суток, дням недели и сезонам года;

- оптимизация режимов электропотребления по критерию минимального суммарного расхода энергетических ресурсов (электроэнергия и прочие энергоносители) при заданных ограничениях допустимого колебания суточного плана выпуска продукции;
- распределение ограниченной мощности (объема электроэнергии) в условиях возникновения дефицита в энергосистеме (в том числе при ограничениях по пропускной способности электрических сетей) между различными предприятиями или технологическими установками одного предприятия, обеспечивающее минимальный народно-хозяйственный ущерб от снижения потребляемой мощности.

В работе [3] предложено решать эти задачи методами нелинейного программирования. Возможность применения методов непрерывной оптимизации обусловлена тем, что автор не рассматривает возможность отключения электроустановок, разрешая только непрерывное изменение режима. Предложенный им подход не полностью отражает интересы потребителя при моделировании его поведения (поскольку не все графики потребления, позволяющие выполнять план, равнозначны) и не учитывает распределенную генерацию. Кроме того, требуется идентификация сложных нелинейных зависимостей.

Позднее в статье [4] рассматривалась задача управления поведением потребителя в целях сглаживания графика нагрузки энергосистемы, способствующего достижению максимальной экономической эффективности энергоблоков, снижению вредных выбросов в атмосферу и пр. Если в перечисленных выше задачах предполагалось изменение режимов работы самого предприятия для достижения эффекта снижения затрат или снижения расхода энергоресурсов, то здесь предлагается методика расчета тарифа на электроэнергию, учитывающая поведение потребителей и стимулирующая различные группы потребителей распределять собственные нагрузки в течение суток таким образом, чтобы в энергосистеме фактическая нагрузка стремилась к величине, соответствующей максимальной экономической эффективности — точке излома энергетической характеристики, а среднее значение тарифа для каждой группы и объем производимой энергии оставались неизменными. Разработанная методика учитывает, прежде всего, энергетические характеристики генерирующего оборудования и, затем, вид суточного графика нагрузки потребителей. Предлагаемый метод расчета тарифов с учетом потребительского поведения решает задачу стимулирования потребителей к выравниванию суточного графика нагрузок энергосистемы при сохранении среднего тарифа неизменным. Данный подход в большей степени учитывает



интересы генерации и регулятора, поскольку отражает технологические особенности различных режимов работы генерации, и учитывает только обобщенный вид графика нагрузки потребителя без проведения оценки приемлемости изменения его конфигурации.

Новый интерес к управлению энергопотреблением возник с развитием концепции SmartGrid, в частности предусматривающей, как уже было сказано, «активизацию» потребителей. Изучение проблем управления спросом в настоящее время актуально во всем мире, причем в основном рассматриваются математические модели бытовых потребителей (домохозяйств) [5—9].

В работе [5] предложена математическая модель, учитывающая такие основные параметры, как денежный эквивалент выгодности графика потребления, стоимость покупки электроэнергии из сети, а также выгоду, получаемую от продажи электроэнергии собственной генерации. По итогам оптимизации по критерию минимума затрат на энергопотребление формируется график нагрузки каждого электроприбора потребителя (с учетом заданных характеристик работы, например, продолжительности работы). Авторы предложили генетический алгоритм приближенного решения задачи планирования потребления и собственной генерации на сутки вперед с возможностью оперативного перепланирования. Алгоритм апробирован на примере домохозяйства, расположенного в г. Сарагоса (Испания).

Аналогичная задача по формированию оптимального профиля энергопотребления рассмотрена в работе [9], но в отличие работы [5] в модели учитывается не денежный эквивалент выгодности, а время ожидания до начала работы электроприбора. Авторы предложили решать задачу с помощью алгоритма выпуклого программирования. Результаты решения апробированы на условных данных для отдельного домохозяйства, но взяты реальные цены с учетом индексов инфляции.

К задаче формирования оптимального графика нагрузки оборудования в работе [7] добавлена задача эффективного распределения электроэнергии, вырабатываемой возобновляемыми источниками энергии (ВИЭ), расположенными у потребителя, между зарядкой накопителя электроэнергии и потреблением электроприборов. В отличие от работ [5] и [9], авторы работы [7] не рассматривают приемлемость графика для потребителя в смысле уровня удовлетворенности потребителя и предлагают решать оптимизационную задачу с помощью метаэвристического алгоритма, апробированного на условных данных.

В работе [6] рассмотрено объединение нескольких потребителей, имеющих общий генерирующий источник. Предложен подход, ключевая осо-

бенность которого состоит в предположении, что потребителям выгодно взаимодействовать между собой в целях максимизации своего индивидуального результата, поскольку размер платы за электроэнергию для каждого потребителя зависит от графика энергопотребления других потребителей. Это приводит к игровому взаимодействию участников. Рассмотрены две задачи в целом для всего объединения потребителей: минимизация коэффициента неравномерности графика нагрузки и минимизация затрат на электроэнергию. Поскольку обе задачи взаимосвязаны, авторы рассмотрели задачу минимизации затрат на электроэнергию и предложили алгоритм решения сна основе аппарата теории игр. По итогам решения определяется профиль энергопотребления каждого участника, при котором достигается равновесие Нэша. Результаты апробированы на условных данных.

В работе [8] рассмотрена задача минимизации затрат на энергопотребление для всего MicroGrid объединения нескольких домохозяйств. В отличие от работы [7], не предполагается взаимодействие между потребителями в формировании графика энергопотребления. Авторами предложен последовательный алгоритм оптимизации энергопотребления. На первом этапе решается задача, аналогичная рассмотренным в работах [5, 7, 9]: определяется график энергопотребления для электроприборов каждого отдельного домохозяйства независимо от режима работы других домохозяйств. Минимизируются затраты на энергопотребление с учетом коэффициентов приемлемости, отражающих уровень удовлетворения потребителя заданным графиком энергопотребления, а также с учетом штрафов за прерывание работы электроприборов. На втором этапе осуществляется управление распределенной генерацией (которая, в отличие от рассмотренной в работах [5, 7, 9], является общей для всей сети), решается задача максимизация использования распределенной генерации и минимизации затрат энергопотребления для всей сети. На третьем этапе осуществляется управление работой накопителя электроэнергии, исходя из технических характеристик самого устройства, а также внешних параметров, определенных на первых двух этапах: энергопотребление и объема выработки ВИЭ. Перечисленные задачи авторы предлагают решать с помощью метаэвристического алгоритма. Результаты апробированы на условных данных. Все рассмотренные подходы обеспечивают снижение затрат на энергопотребление от 8 до 25 %.

Проведенный анализ позволяет сделать вывод об отсутствии в настоящее время модели управления потреблением и генерацией активного потребителя, которая бы подходила для описания домохозяйств, позволяла учесть потери потребителя в случае различных графиков нагрузки, включала



	Время																								
Возможные графики нагрузки $\alpha_2 = 1,, R_2$		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3 4	$\begin{vmatrix} 0 \\ 0 \end{vmatrix}$	0	0	1	1 1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	α_2^i	0	0	1	0	0	1	0	0	1	0	0	1	. 0	0	0	0	0	0	0	0	0	0	0	0
	$\alpha_2^{R_2}$	0	0	0	1	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Пример матрицы графиков нагрузки кондиционера

в себя управление собственной генерацией и, в то же время, была вычислительно достаточно простой, чтобы лечь в основу поведения агентов-потребителей в рамках мультиагентной системы моделирования энергосистемы при решении задачи управления спросом путем тарифного регулирования. Такая модель предлагается далее, приводятся условия, при которых задача потребителя допускает простое и эффективное решение, показывается, что эти условия позволяют описать существенную долю случаев, встречающихся при моделировании поведения отдельных домохозяйств, по крайней мере, в рамках решения задачи определения рациональных тарифов на электроэнергию.

1. МОДЕЛЬ УПРАВЛЕНИЯ НАГРУЗКОЙ И СОБСТВЕННОЙ ГЕНЕРАЦИЕЙ

1.1. Общая модель активного потребителя

В работе [10] предложена общая экономико-математическая модель активного потребителя. Считается, что активный потребитель минимизирует функцию операционных затрат, учитывающих:

- цену на покупаемую с рынка электроэнергию;
- профиль энергопотребления;
- потери из-за отклонения текущего профиля энергопотребления (потребности в электроэнергии) от желаемого;
- затраты на перенастройку генерирующих мощностей потребителя;
- цену на передаваемую в сеть электроэнергию собственной генерации;
- объем собственной генерации, передаваемой в сеть;
- себестоимость производства электроэнергии собственной генерации.

Модель позволяет выделить основные экономические факторы, влияющие на поведение активного потребителя, помогает оценить порядки их значений в финансовом выражении и может быть использована для описания его поведения при разработке как механизмов мотивации потребителя к участию в регулировании, так и механиз-

мов управления спросом. Настоящая статья фактически детализирует и адаптирует эту модель к описанию поведения активного потребителя в интересах решения задачи управления спросом.

1.2. Модель поведения активного потребителя

Модель активного потребителя, отражающая его экономические интересы на временном промежутке, соответствующем оперативной деятельности [10], детализируется следующим образом.

Разобьем промежуток времени, для которого осуществляется планирование, на T периодов (для определенности на 24 периода по одному часу). Пусть потребитель обладает N единицами энергопотребляющего оборудования. Работа типичной единицы оборудования моделируется графиком нагрузки, определяющим мощность, потребляемую оборудованием в каждый из 24-х плановых периодов.

Матрица A_n всех возможных графиков нагрузки для оборудования $n \in \{1, ..., N\}$ формируется на каждый час последующих суток. Матрица содержит T столбцов и R_n строк (число возможных графиков нагрузки оборудования n). Появление нулей в строке означает, что оборудование должно быть выключено в соответствующие часы.

Пример матрицы A_n для кондиционера (n=2) приведен в таблице:

- возможна как непрерывная работа в течение заданного времени, так и прерывистая работа с заданным интервалом;
- параметры² работы: порядковый номер оборудования n=2, электропотребление оборудования (номинальная мощность) $P_2=1$ кВт · ч.

Обозначим через $a_n^{\alpha}(t)$ элемент матрицы A_n , соответствующий строке (режиму работы $\alpha \in \{1, ..., R_n\}$) и столбцу $t \in \{1, ..., T\}$. Тогда если для обору-

² Здесь и далее в качестве примера используются данные о параметрах работы оборудования из работы [7].



дования n потребитель выбрал график загрузки $\alpha_n \in \{1, ..., R_n\}$, то суммарное потребление в момент

времени
$$t$$
 запишется как $a(t) = \sum_{n=1}^{N} a_n^{\alpha_n}(t)$.

Не все графики нагрузки оборудования одинаково предпочтительны для пользователя. Обозначим через d_n^{α} денежный эквивалент выгодности для пользователя графика потребления $\alpha \in \{1, ..., R_n\}$ объекта оборудования $n \in \{1, ..., N\}$. Если среди режимов работы оборудования есть режим, соответствующий его полному отключению, то логично считать для него этот параметр равным нулю, и тогда для других режимов d_n^{α} будет означать денежную сумму, которую потребитель готов заплатить за возможность эксплуатировать электрооборудование n в режиме α по сравнению с ситуацией неиспользования объекта оборудования³.

На практике оценка денежного эквивалента выгодности представляет собой сложную социально-экономическую задачу по выявлению предпочтений потребителей. В экономической теории ее эквивалентом служит количественная оценка полезности. Денежные оценки выгодности того или иного графика потребления носят исключительно индивидуальный, субъективный характер: для одного человека, который не работает и сидит дома, безразлично, в какое время смотреть телевизор (денежный эквивалент выгодности будет одинаков в течение дня), для другого, который работает ценность имеет просмотр телевизора за пару часов до сна (денежный эквивалент выгодности в течение дня будет нулевым, и только за два часа до сна будет положительным). Предполагается, что только конкретный потребитель может дать оценку денежного эквивалента выгодности электропотребления для конкретного типа электрооборудования, и для определения выгодности необходимо провести довольно трудоемкий опрос⁴. В рамках более грубого, но более практичного подхода все технически реализуемые графики нагрузки делятся на допустимые для потребителя и недопустимые. Недопустимые графики исключаются из мат-

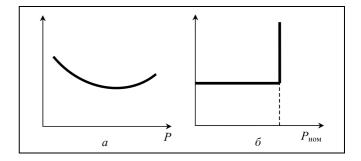


Рис. 1. Примеры графиков зависимости себестоимости от объема выработки P электроэнергии для дизель-генератора (a) и солнечной панели (δ)

рицы A, а всем допустимым графикам назначается одинаковая полезность.

Предположим далее, что потребитель обладает M источниками собственной генерации — солнечной и/или ветряной установкой, дизельным или газовым электрическим генератором. С учетом возможности различных режимов работы генераторов для каждого генератора $n \in \{1, ..., M\}$ можно построить зависимость себестоимости генерации единицы электроэнергии $c_n(g_n)$ от выдаваемой в каждый момент времени мощности g_n . Точка на графике соответствует минимальной себестоимости, с которой возможна выдача требуемой мощности на данной энергетической установке.

Примеры соответствующих графиков для солнечной установки и дизель-генератора представлены на рис. 1.

Если обозначить через $g_n(t)$ мощность генера-

ции установки
$$n \in \{1, ..., M\}$$
, то $g(t) \coloneqq \sum_{n=1}^{M} g_n(t)$ —

это полный объем генерации в период времени, который делится между внутренним потреблением $g^I(t)$ и объемом электроэнергии $g^E(t)$, выдаваемым

в сеть, а
$$C(t) := \sum_{n=1}^{M} c_n(g_n(t))$$
 — затраты на произ-

водство активным потребителем объема электроэнергии g в период времени t. Для упрощения расчетов в рамках настоящей статьи затраты на пуск/остановку генерирующих мощностей считаются нулевыми.

Обозначим через ξ_g параметры тарификации для передаваемой в сеть электроэнергии, например, накопленное потребление при кумулятивной зависимости цены от переданного с начала отчетного периода объема электроэнергии, включая оговоренные в контракте с энергетической компанией ограничения на генерацию. Ниже состав и функции этих параметров будут уточняться. Аналогично обозначим через ξ_a аналогичные парамет-

³ В исходной модели [10] аналогичные функции выполняли так называемые «потери из-за отклонения потребления от потребности». Здесь терминология несколько изменена для удобства сравнения результатов с другими моделями.

⁴ Интересной представляется автоматизация этого опроса путем изучения поведения потребителя в условиях информационных сигналов от «умного счетчика электроэнергии» о денежном эквиваленте принимаемых решений (если потребитель включает посудомоечную машину, получив информацию о стоимости электроэнергии в данный момент времени, логично считать, что для него денежный эквивалент полезности превышает эти затраты). После периода «настройки на потребителя» необходимость в сигналах пропадает.



ры тарификации для потребляемой электроэнергии, включая оговоренные в контракте ограничения на потребление. Через η обозначим внешние условия на горизонте планирования, такие как среднесуточная температура или длительность светового дня.

Обозначим через $p_a(t, a(t), \xi_a, \eta)$ цену на потребляемую электроэнергию в зависимости от временного промежутка, объема потребления и прочих параметров, через $p_g(t, g^E, \xi_g, \eta)$ — цену на передаваемую в сеть электроэнергию в зависимости от временного промежутка, объема внешней генерации и прочих параметров.

С учетом введенных обозначений запишем целевую функцию активного потребителя

$$f = \sum_{n=1}^{N} d_n^{\alpha_n} - \sum_{t=1}^{T} p_a(\cdot) \left[\sum_{n=1}^{N} a_n^{\alpha_n}(t) - g^I(t) \right] + \sum_{t=1}^{T} \left[p_g(\cdot) g^E(t) - \sum_{n=1}^{M} c_n(g_n(t)) \right]$$

как сумму «прибыли» от потребления (разницы выгоды от потребления и стоимости отобранной из сети электроэнергии) и прибыли от собственной генерации (разницы дохода от продажи электроэнергии в сеть и себестоимости генерации).

Формально задача активного потребителя (задача оптимизации потребления и собственной генерации) состоит в том, чтобы максимизировать свою целевую функцию f выбором для каждого объекта электрооборудования n=1,...,N графика потребления $\alpha_n \in \{1,...,R_n\}$, для каждого из имеющихся генераторов n=1,...,M его графика генерации (т. е. для каждого из периодов t=1,...,T выбора неотрицательного числа — мощности генерации $g_n(t)$), а также объема передаваемой в сеть

электроэнергии
$$g^{E}(t) \leq g(t) = \sum_{n=1}^{M} g_{n}(t)$$
.

1.3. Классификация задач

В общем случае исследование сформулированной модели сводится к решению сложной задачи смешанной оптимизации. Основная цель настоящей статьи состоит в выявлении практически важных случаев, когда эта задача, тем не менее, имеет простое, а зачастую даже аналитическое, решение. Для этого введем следующую классификацию частных задач, основаниями которой служит сложность различных компонентов общей задачи.

Основание классификации 1. Формула расчета гибкого тарифа $p_a(\cdot)$.

• Зонный тариф (Time-Of-Use Pricing), $p_a(t)$, при котором цена на потребляемую электроэнергию зависит только от времени суток, включая

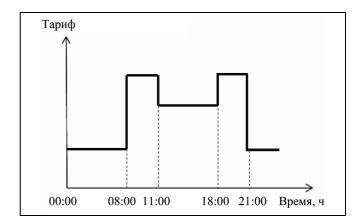


Рис. 2. Пример зонных тарифов (time-of-use tariff)

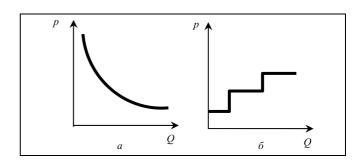


Рис. 3. Модификации графика цены p в зависимости от условий ценообразования в рамках программ управления спросом: a — наличие скидок при увеличении объема потребления Q; δ — установление цены для каждого диапазона потребления

обычно от двух до четырех ценовых периодов (рис. 2).

• Зонный тариф с учетом ограничения на максимальную потребляемую мощность, $p_a(t, a(t), \xi_a)$, предполагает зависимость цены не только от времени, но и от потребляемой мощности a(t), а именно,

$$p_a(t, a(t), \xi_a) = \begin{cases} p_a(t), a(t) \le \xi_a, \\ +\infty, a(t) > \xi_a. \end{cases}$$

Такая зависимость цены от мощности фактически запрещает потребителю превышать норму ξ_a объема потребляемой одномоментно электроэнергии. На практике такой тариф реализуется через комбинацию зонного тарифа и автоматического выключателя по току.

• Тариф с нормами потребления (социальная норма потребления) предполагает зависимость цены от суммарного потребления за период.

Так, в документе [11] утвержден комплекс мер, направленных на переход к установлению социальной нормы потребления коммунальных услуг в РФ, т. е. в рамках которой потребляемая электро-



энергия будет оплачиваться по регулируемым тарифам (сейчас рассматриваются различные варианты норм потребления от 50 до 100 кВт·ч), а сверх нее — по рыночной цене.

• Более сложные формулы расчета цены. В рамках программ управления спросом могут использоваться более сложные формулы расчета цены. Наиболее распространенные примеры (рис. 3) — установление скидок при увеличении объема потребления электроэнергии и установление цены для каждого диапазона потребления, когда задаются цены для определенных диапазонов электропотребления, при этом при переходе на новый диапазон потребления цена увеличивается.

Основание классификации 2. Собственная генерация.

- Отсутствуют возможности использования распределенной генерации.
- Возможна продажа в сеть электроэнергии, вырабатываемой генерацией потребителя, по рыночной цене $(p_g(\cdot)) \equiv p_a(\cdot)$.
- Существует возможность использования распределенной генерации только для собственного потребления (без выдачи в сеть).
- Возможна продажа электроэнергии, вырабатываемой распределенной генерацией потребителя, в сеть по цене, отличной от рыночной⁵.

Далее будет изучена модель, когда тариф зависит от времени суток (Time-Of-Use Pricing) без ограничения на потребляемую мощность, а распределенная генерация допускает продажу произведенной электроэнергии по рыночной цене. Оказывается, что в этом случае задачи оптимизации графиков потребления и собственной генерации разделяются, и их можно решать независимо друг от друга. Также понятно, что ситуация, когда отсутствует распределенная генерация — частный случай решаемой задачи.

3. РЕШЕНИЕ ЗАДАЧИ

3.1. Оптимизация собственной генерации

Предположим, что тариф на электроэнергию зависит только от времени суток и энергетическая компания приобретает электроэнергию у пот-

ребителя по той же цене, что и продает, т. е. $p_a(t) = p_g(t) = p(t)$.

В этом случае целевую функцию потребителя можно записать как

$$f = \sum_{n=1}^{N} d_n^{\alpha_n} + \sum_{t=1}^{T} [p(t)g^{E}(t) - p(t)(a(t) - g^{I}(t)) - C(t)] = \sum_{n=1}^{N} d_n^{\alpha_n} + \sum_{t=1}^{T} [p(t)g(t) - a(t)) - C(t)].$$

Из формулы видно, что с экономической точки зрения потребителю неважно, пускать ли произведенную электроэнергию в первую очередь на собственные нужды или поставлять ее целиком в сеть. Таким образом, можно считать, что $g^E(t) = g(t)$, $g^I(t) = 0$.

Тогда

$$f = \sum_{n=1}^{N} \sum_{t=1}^{T} \left[d_n^{\alpha_n} / T - p(t) a_n^{\alpha_n}(t) \right] +$$

$$+ \sum_{n=1}^{M} \sum_{t=1}^{T} \left[p_g(t) g_n(t) - c_n(g_n(t)) \right].$$

В задаче отсутствуют ограничения, связывающие выбор переменных α_n и $g_n(t)$, поэтому отдельно можно проводить максимизацию и части целевой функции, связанной с потреблением,

$$f_a = \sum_{n=1}^{N} \sum_{t=1}^{T} [d_n^{\alpha_n} / T - p(t) a_n^{\alpha_n}(t)], \tag{1}$$

и части, связанной с собственной генерацией

$$f_g = \sum_{n=1}^{M} \sum_{t=1}^{T} [p_g(t)g_n(t) - c_n(g_n(t))].$$

Это, между прочим, означает, что наличие собственной генерации при возможности поставки ее в сеть по цене закупки не влияет на стремление потребителя максимально снижать собственное энергопотребление в периоды высокой цены (что обычно предполагается программами тарифного управления спросом) — расходы на электроэнергию замещаются упущенной прибылью от недопоставки электроэнергии в сеть.

Если, как предполагалось ранее, потребитель может независимо управлять мощностью источников собственной генерации, то задача максимизации функции f_g также распадается на M независимых задач выбора мощности генерации для каждой из M имеющихся энергетических установок.

Мощность $g_n(t)$ собственной генерации установки, $n \in \{1, ..., M\}$, в момент времени t выбирается, исходя из цены на электроэнергию в этот период.

⁵ За рубежом важная тенденция состоит в увеличении доли производства электроэнергии посредством ВИЭ, в связи с чем разработана система поддержки такого производства с помощью специальных закупочных тарифов (feed-in tariffs): установление специальных закупочных тарифов на энергию, вырабатываемую ВИЭ, или надбавок к рыночной цене на электроэнергию в зависимости от объемов проданной электроэнергии и типа ВИЭ. В России законодательством предусмотрены обе формы поддержки ВИЭ [12—14].

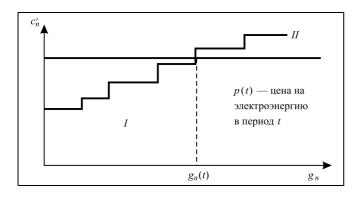


Рис. 4. Выбор режима работы генератора

Так, если маржинальная себестоимость генерации $c'_n(g_n)$ возрастает (что соответствует возрастанию расходов на масштаб), то оптимальная мощность определяется из уравнения $c'_n(g_n(t)) = p(t)$, т. е. $g_n(t) = [c'_n]^{-1}(p(t))$. На рис. 4 представлен типичный график маржинальной себестоимости электроэнергии собственной генерации, который имеет ступенчатый вид, при этом каждая новая ступень означает включение менее экономного режима генерации. На участке І, где маржинальная себестоимость производства электроэнергии собственной генерации потребителя ниже рыночной цены p(t), потребителю выгодно загружать собственную генерацию (в объеме, g_n^I), как для покрытия собственной нагрузки, так и для продажи электроэнергии в сеть. Маржинальная себестоимость генерации на участке II выше рыночной цены p(t), и, следовательно, потребителю экономически нецелесообразно загружать соответствующие генерирующие мощности при заданной рыночной цене.

Если же себестоимость с ростом мощности изменяется более сложным образом, имеется эффект экономии на масштаб, скачкообразное изменение себестоимости и т.п., задача сводится к задаче одномерной нелинейной оптимизации, оставаясь при этом задачей, типичной для экономики производства [15].

3.2. Выбор оптимального профиля нагрузки потребителя

Если, как предполагается, пользователь независимо может управлять графиками потребления различных приборов, то, как видно из формулы (1), задача максимизации целевой функции

$$f_a = \sum_{n=1}^N \left[d_n^{\alpha_n} - \sum_{t=1}^T p(t) a_n^{\alpha_n}(t) \right]$$
 распадается на N задач

выбора оптимального варианта загрузки оборудования (строки матрицы A_n) для каждого прибора $n=1,\,...,\,N$.

Для этого:

- 1) каждая строка матрицы A_n умножается на вектор [p(1), ..., p(T)] график изменения цены в течение всего планируемого промежутка времени;
- 2) все элементы каждой строки по-отдельности суммируются, давая столбец $[c_n^1,...,c_n^{R_n}]$ расходов на электроэнергию для каждого варианта загрузки оборудования;
- 3) из столбца $[d_n^1, ..., d_n^{R_n}]$ выгодности вариантов поэлементно вычитается столбец $[c_n^1, ..., c_n^{R_n}]$ расходов на электроэнергию, получается столбец $[\phi_n^1, ..., \phi_n^{R_n}]$ оценок вариантов;
- 4) выбирается вариант, имеющий максимальную оценку $\phi_n^1, ..., \phi_n^{R_n}$.

По итогам расчетов формируется почасовой график работы оборудования потребителя и загрузки собственной генерации.

Таким образом, задача решается в случае, когда цена зависит только от временного периода (от времени суток), отсутствуют ограничения по мощности, продажа электроэнергии собственной генерации возможна по рыночной цене. Понятно, что описанный алгоритм решает задачу и в случае отсутствия у потребителя возможностей распределенной генерации.

4. ЧИСЛЕННЫЙ ПРИМЕР РЕШЕНИЯ ЗАДАЧИ

Для апробации предложенной модели без учета оптимизации работы собственной генерации были использованы условные данные и частично данные работы [5].

Рассматривалась работа следующих устройств:

- кондиционер: порядковый номер оборудования n=1, электропотребление (номинальная мощность) $P_1=1$ кВт;
- электромобиль: порядковый номер оборудования n=2, электропотребление (номинальная мощность при зарядке) неравномерно (на базе усредненных данных [5]) $P_2=0.79...3,56$ кВт, продолжительность зарядки 19 ч;
- стиральная машина: порядковый номер оборудования n=3, электропотребление (номинальная мощность) $P_3=0.95~\mathrm{kBT};$
- телевизор: порядковый номер оборудования n=4, электропотребление (номинальная мощность) $P_4=0.3~\mathrm{kBt}.$

Принималась цена на электроэнергию для розничных потребителей ОАО «Мосэнергосбыт», утвержденная на I полугодие 2013 г. (многотарифный учет с применением тарифа, дифференциро-



ванного по зонам суток)⁶. Денежный эквивалент выгодности графика потребления d_n принят на основе условных данных.

Исходные данные для расчетов и итоговый расчет оптимальных графиков нагрузки доступны в Интернете (http://www.mtas.ru/upload/library/VGS2013.xls). В частности, оптимальный вариант для потребителя графика работы выглядит так:

- кондиционер: вариант 3 (работа с 11.00 по 12.00, с 15.00 по 16.00 и с 19.00 по 20.00);
- электромобиль: вариант 4 (зарядка с 16.00 по 10.00);
- стиральная машина: вариант 3 (работа с 3.00 до 5.00);
 - телевизор: вариант 5 (работа с 11.00 до 15.00).

ЗАКЛЮЧЕНИЕ

Разработанная модель активного потребителя может использоваться для автоматизации управления нагрузкой потребителя (для категории «активно-пассивных» потребителей, которые управляют своей нагрузкой с помощью автоматически настроенных программ). Кроме того, разработанная модель позволяет оценить экономический эффект для потребителя от участия в управлении спросом. Показано, что, по крайней мере, для домохозяйств может быть предложена реалистичная модель оптимального поведения активного потребителя, адекватно описывающая основные мотивы принятия решений по энергопотреблению и собственной генерации. Также показано, что во многих важных с практической точки зрения ситуациях (зонные тарифы в условиях некритичности ограничений по единовременно потребляемой мощности, равенство цены потребляемой и передаваемой в сеть энергии) оптимизационная задача потребителя допускает чрезвычайно простое решение, существенно более эффективное в смысле вычислений, чем общие методы, предлагаемые в литературе.

Этот алгоритм поведения отдельного потребителя, как раз в силу его простоты, может браться за основу при мультиагентном моделировании отклика потребителей на тарифные механизмы управления спросом. В процессе моделирования популяция агентов, соответствующих отдельным потребителям, с описанным в статье алгоритмом принятия решений, но с различными параметрами (профилями предпочтений, набором электроприборов и режимов их работы, мощностей локальной генерации) погружается в условия общей тарифной политики. В результате моделирования выявляется изменение графика совокупного потребления при фиксированном меню тарифов. Этот «отклик спроса» (demandresponse) используется за-

тем для поиска оптимальной тарифной политики с помощью, например, алгоритмов локального поиска. В целом такое моделирование позволяет сделать надежные предсказания об эффективности планируемых мер тарифного регулирования.

ЛИТЕРАТУРА

- 1. Кобец Б.Б., Волкова И.О. Инновационное развитие электроэнергетики на базе концепции SMART GRID. М.: ИАЦ Энергия, 2010.
- 2. Основные положения концепции интеллектуальной энергосистемы с активно-адаптивной сетью. 2012. URL: http://www.fsk-ees.ru/upload/docs/ies_aas.pdf (дата обращения: 08.04.2013).
- Михайлов В.В. Тарифы и режимы электропотребления. М.: Энергоатомиздат, 1986.
- Гребенюк Г.Г., Соловьев М.М. Непрерывное тарифное регулирование для формирования желаемого графика нагрузки энергосистемы // Автоматика и телемеханика. 2004. № 5. С. 166—173.
- Juan M. Lujano-Rojas, Claudio Monteiro, Rodolfo Dufo-Lopez, Jose L. Bernal-Agustin. Optimum residential load management strategy for real time pricing demand response programs // Energy Policy. — 2012. — Vol. 45. — P. 671—679.
 Amir-Hamed Mohsenian-Rad, Vincent W.S. Wong, Juri Jatskev-
- Amir-Hamed Mohsenian-Rad, Vincent W.S. Wong, Juri Jatskevich, Robert Schober, Alberto Leon-Garcia. Autonomous Demand Side Management Based on Game-Theoretic Energy Consumption Scheduling for the Future Smart Grid // IEEE Trans.onSmart Grid. 2010. Vol. 1, N 3. P. 320—331.
- Nikhil Gudi, Lingfeng Wang, Vijay Devabhaktuni. A demand side management based simulation platform incorporating heuristic optimization for management of household appliances // Electrical Power and Energy Systems. — 2012. — Vol. 43. — P. 185—193.
- Bingnan Jiang, Yunsi Fei. Dynamic Residential Demand Response and Distributed Generation Management in Smart Microgrid with Hierarchical Agents // Energy Procedia. 2011. Vol. 12. P. 76—90.
- Amir-Hamed Mohsenian-Rad, Alberto Leon-Garcia. Optimal Residential Load Control With Price Prediction in Real-Time Electricity Pricing Environments // IEEE Trans. on Smart Grid. – 2010. – Vol. 1, N 1. – P. 120–133.
- Бурков В.Н., Губко М.В., Новиков Д.А. Организационные механизмы управления в электроэнергетике // Управление развитием крупномасштабных систем / Под ред. А.Д. Цвиркуна. М., 2012. С. 261—278.
- 11. Распоряжение Правительства РФ от 10 сентября 2012 г., № 1650-р «Об утверждении комплекса мер, направленных на переход к установлению социальной нормы потребления коммунальных услуг в Российской Федерации».
- 12. *Федеральный* закон от 26.03.2003, № 35-ФЗ «Об электроэнергетике».
- 13. *Основы* ценообразования в области регулируемых цен (тарифов) в электроэнергетике (утверждены постановлением Правительства РФ от 29.12.2011, № 1178).
- 14. *Основные* положения функционирования розничных рынков электрической энергии (утверждены постановлением Правительства РФ от 4 мая 2012 г., № 442).
- Hackman S.T. Production Economics. Heidelberg: Springer-Verlag, 2008.

Статья представлена к публикации членом редколлегии чл.-корр. РАН Д.А. Новиковым.

Ирина Олеговна Волкова — д-р экон. наук, профессор, Научно-исследовательский университет — Высшая школа экономики, **≅** (495) 725-30-73, ⊠ iovolkova@hse.ru,

Михаил Владимирович Губко — канд. техн. наук, ст. науч. сотрудник, Институт проблем управления им. В.А. Трапезникова РАН, **≅** (495) 334-90-51, ⊠ mgoubko@mail.ru,

Евгения Александровна Сальникова — аспирант, Научно-исследовательский университет — Высшая школа экономики, **☎** (495) 725-30-76, ⋈ esalnikova@hse.ru.

⁶ Данные сайта ОАО «Мосэнергосбыт».



УДК 519.874

ПОСТРОЕНИЕ РАСПИСАНИЙ ВЫПОЛНЕНИЯ ЗАДАНИЙ ДЛЯ ПАРАЛЛЕЛЬНО РАБОТАЮЩИХ СИСТЕМ МАШИН И ПОТОЧНЫХ ЛИНИЙ

Ю.А. Зак

В развитие Flow-Shop-Problem рассмотрены постановки и математическая модель задачи построения расписаний для параллельно работающих систем машин в условиях заданной системы ограничений на сроки выполнения заданий. Исследованы свойства допустимых и оптимальных расписаний и предложены методы решения задачи с помощью последовательных алгоритмов оптимизации. Приведены числовые примеры.

Ключевые слова: система машин, параллельная работа, Flow-Shop-Problem, ограничения, сроки выполнения, допустимые и оптимальные расписания, оптимизация, последовательные алгоритмы.

ВВЕДЕНИЕ

Рассматриваемая задача построения расписаний для параллельно работающих систем машин и автоматических (поточных) линий в некоторой степени обобщает классические задачи теории расписаний выполнения работ на параллельных машинах и Flow-Shop-Problem (см., например, работы [1-4]). В отличие от рассматриваемых в литературе обобщений в виде Hybrid Flow-Shop Scheduling Problem (см., например, работы [5—7]), где каждая ј-я операция любого задания может выполняться на любой из параллельно стоящих на j-й стадии обработки M машинах (т. е. все последовательные стадии обработки всех операций одного задания не объединены в одну технологическую линию), в рассматриваемой постановке каждое из N заданий должно выполняться только на одной технологической линии, состоящей из Kпоследовательно работающих машин. В публикациях [5—7]) были предложены только приближенные и эвристические методы решения этих задач без учета всего комплекса ограничений на сроки выполнения заданий и времена работы машин. В настоящей работе предложены точные методы решения сформулированной ниже важной для практических приложений проблемы.

На M параллельно работающих системах машин (участках, автоматических линиях), каждая из которых включает в себя K последовательно работающих машин, необходимо выполнить N заданий. Каждое задание состоит из одного и того же числа работ (операций), равном K, и может выполняться только на одной технологической линии.

Операции всех заданий выполняются в одной и той же и одинаковой для всех заданий последовательности. Выполнение каждой операции может быть назначено только на соответствующей машине; т. е. первая операция выполняется на машине 1, вторая — на машине 2 и т. д., и К-я операция — на машине K. Заданы t_{ik}^l — времена выполнения k-й операции i-го задания на l-й технологической линии. Если в i-м задании k-я операция отсутствует, то полагаем значение $t_{ik}^l = 0$. Следовательно, все перемещения, связанные с окончанием выполнения задания на одной машине и началом выполнения его на другой, должны следовать в одном направлении. Рассматриваемая совокупность последовательно работающих машин каждой технологической линии представляет собой систему конвейерного типа.

Необходимо определить непересекающиеся подмножества заданий, выполняемых на каждой технологической линии, а также оптимальные последовательности выполнения заданий каждого из сформированных подмножеств, удовлетворяющие заданной системе ограничений на сроки выполнения заданий и времена работы машин и доставляющие экстремальное значение некоторому критерию оптимальности.

Рассматриваемая далее обобщенная постановка задачи Flow-Shop-Problem для параллельно работающих систем машин (технологических линий) может быть сформулирована в условиях следующих предпосылок.

• Система состоит из l = 1, ..., M параллельно функционирующих технологических линий,



каждая из которых включает в себя k=1,...,K последовательно работающих машин (рабочих станций). Каждая машина во всех технологических линиях в любой момент времени может выполнять не более одной операции, номер которой равен номеру этой машины, и может начать выполнение операции с тем же номером другого задания только после завершения выполнения операции задания, стоящего в последовательности выполнения на данной технологической линии непосредственно перед ним. Временами переналадок, необходимыми при переходе машин от выполнения операции предыдущего задания к последующему, можно пренебречь.

- Необходимо выполнить *N* заданий. Выполнение каждого из заданий может быть назначено только на одну из технологических линий.
- Каждое задание состоит из некоторой последовательности работ (операций) k = 1, ..., K, которые должны выполняться в строго установленной и одинаковой для всех заданий последовательности друг за другом. При этом операция с индексом k может выполняться только на определенной технологическим регламентом машине с соответствующим индексом k.
- Длительности выполнения каждой из этих операций $t_{ij}^l(t_{ik}^l)$ известны и могут быть различны для соответствующих машин различных технологических линий, они зависят от последовательности выполнения остальных заданий на этой или других машинах. Никакая из этих операций не допускает прерываний в процессе ее выполнения.
- Для каждого задания и для каждой машины всех рабочих станций могут быть заданы ограничения на начальные и конечные сроки его выполнения и времени работы, которые не могут быть нарушены. Если такие ограничения отсутствуют, то для начальных сроков эти граничные значения устанавливаются равными нулю, а для конечных — ∞.
- Все граничные значения и времена выполнения операций предполагаются целочисленными величинами.

Необходимо найти подмножества заданий, выполняемых на каждой технологической линии, а также последовательность выполнения заданий каждого из этих подмножеств, в условиях ограничений, что сроки начала и окончания работы машин, начала и завершения выполнения отдельных заданий удовлетворяют всей сформулированной системе ограничений. В качестве критериев оптимальности могут быть выбраны различные показатели эффективности построенного расписания. Как частный случай сформулированной задачи

могут рассматриваться задачи, в которых ограничения на допустимые сроки выполнения заданий и работы машин отсутствуют.

В то время как для Flow-Shop-Problem и задачи построения расписаний выполнения работ на параллельных машинах разработаны алгоритмы получения точных и приближенных решений (см., например, работы [2-4, 8, 9]), исследованию свойств этой важной для практических приложений задачи не уделялось достаточного внимания в литературе. Автору не известны публикации, в которых были бы предложены эффективные методы решения сформулированной задачи для различного вида критериев оптимальности и в условиях заданной системы ограничений. В данной работе рассматривается математическая модель задачи построения расписаний на параллельно работающих системах машин, исследуются свойства допустимых и оптимальных расписаний и предлагаются точные и приближенные методы решения задачи последовательными алгоритмами оптимизации и эвристическими алгоритмами. Алгоритмы решения иллюстрируются числовым примером.

1. МАТЕМАТИЧЕСКАЯ ФОРМУЛИРОВКА ЗАДАЧИ

Так как номер операции однозначно определяет и номер машины, на которой она выполняется, то множество индексов (i, k, l) однозначно идентифицируют каждую операцию любого задания.

Введем обозначения:

 $\Pi = \{1, ..., k, ..., K\}$ — последовательность выполнения операций, одинаковая для каждого задания;

t(i, k, l) — время выполнения k-й операции i-го задания на k-й машине l-й технологической линии:

 b_i , B_i — соответственно граничные сроки на начало и завершение выполнения i-го задания;

 h^{lk} , H^{lk} — соответственно граничные начальные и конечные времена возможности использования k-й рабочей станции l-й технологической линии.

Если величины b_i или h^{lk} не заданы, то они полагаются равными нулю. В случае отсутствия граничных значений B_i или H^{lk} они полагаются равными ∞ .

Обозначим:

 $\tilde{I}^l = \{i_1^l, ..., i_p^l, ..., i_N^l\}$ и N^l — соответственно подмножество и число заданий, выполняемых на l-й технологической линии, $\tilde{I}^l \cap \tilde{I}^\lambda = \emptyset$, l, $\lambda = 1, ..., M$, $l \neq \lambda$; $\bigcup_{l=1}^M \tilde{I}^l = \tilde{I} = \{1, ..., N\}$; $\sum_{l=1}^M N^l = N$;



x(i, k, l) — время начала выполнения операции (i, k, l);

 $\theta(i, k, l)$ — допустимый наиболее ранний срок начала выполнения операции (i, k, l);

 $\sigma(i, k, l) = x(i, k, l) + t(i, k, l) - 1$ — время завершения выполнения операции (i, k, l);

 $T_i = \sigma(i, K, l)$ — время завершения выполнения всех операций *i*-го задания на *l*-й поточной линии;

$$T = \max_{1 \le l \le M} \max_{i \in \tilde{I}_l} T_i = \sigma(i, K, l)$$
 — время окончания

выполнения всего множества заданий.

Ограничения на начало выполнения каждой операции первого из выполняемых заданий на каждой из технологических линий имеют вид:

$$x(i_1^l, 1, l) \ge \theta(i_1^l, 1, l) = \max(b_i, h^{l}), \quad l = 1, ..., M;$$

$$x(i_1^l, k, l) \ge \theta(i_1^l, 1, l) = \max(\sigma(i_1^l, k-1, l) + 1, h^{lk}),$$

 $k = k + 1, ..., K, l = 1, ..., M.$ (1)

Допустимый наиболее ранний срок начала выполнения операций (i, k, l) определяется выражением

$$x(i, 1, l) \ge \theta(i, 1, l) = \sigma(i - 1, 1, l) + 1,$$

$$x(i, k, l) \ge \theta(i, k, l) =$$

$$= \max\{\sigma(i - 1, k, l), \sigma(i, k - 1, l)\} + 1,$$

$$(i - 1) = i_1^l, \quad i = i_2^l, ..., i_N^l, \quad k = k + 1, ..., K,$$

$$l = 1, ..., M,$$

$$(2)$$

где

$$\sigma(i_1^l - 1, k, l) = 0, \quad i = i_1^l, i_2^l, ..., i_N^l.$$
 (3)

Ограничения на времена завершения выполнения заданий могут быть представлены в виде

$$T_i = \sigma(i, K, l) \le \min(B_i; H^{lK}),$$

 $i = 1, ..., N, l = 1, ..., M.$ (4)

В качестве критериев оптимальности могут рассматриваться следующие показатели:

— выполнение расписания в кратчайшие сроки

$$F_1 = \min \max_{1 \le l \le M} \max_{i \in \tilde{I}_l} \sigma(i, K, l), \tag{5}$$

 минимальное средневзвешенное время, затраченное на выполнение всех заданий

$$F_2 = \sum_{i=1}^{N} w_i \sigma(i, K, I);$$
 (6)

 минимальное суммарное или средневзвешенное время технологических линий

$$F_{3} = \sum_{l=1}^{M} \beta_{l} \sigma(i^{N_{l}}, K, l);$$
 (7)

 минимальное суммарное или средневзвешенное время работы машин всех технологических линий

$$F_4 = \sum_{l=1}^{M} \sum_{k=1}^{K} \eta_{lk} \sigma(i^{N^l}, k, l).$$

В выражениях (6) и (7) $w_i \ge 0$, i=1,...,N, и $\beta_l \ge 0$, l=1,...,M, $\eta_{lk} \ge 0$ — некоторые весовые коэффициенты, удовлетворяющие условиям нормировки

$$\sum_{i=1}^{N} w_i = 1, \quad \sum_{l=1}^{M} \beta_l = 1, \quad \sum_{l=1}^{M} \sum_{k=1}^{K} \eta_{lk} = 1.$$

2. СВОЙСТВА ДОПУСТИМЫХ И ОПТИМАЛЬНЫХ РАСПИСАНИЙ

Обозначим: $\bar{t}(i, k) = \bar{t}(i, k, l) = \min_{1 \le l \le M} t(i, k, l)$;

$$\bar{h}^k = \min_{1 \le l \le M} h^{lk}; \ \hat{H}^k = \max_{1 \le l \le M} H^{lk}; \ k = 1, ..., K.$$

Пусть известны подмножества выполняемых заданий на каждой поточной линии \tilde{I}^l , l=1,...,M. Вычислим значения ранних сроков завершения

Вычислим значения ранних сроков завершения выполнения работ на k-й машине (\overline{g}^{lk}), на последних (K-k) машинах ($k=1,2,...,\gamma,...,K-1$) каждой технологической линии \overline{G}^{lk} , а также оценку нижней границы времени выполнения всех заданий на k-й машине на всех линиях — \overline{R}^k

$$\bar{g}^{lk} = \min_{i \in \tilde{I}_l} \left[\theta(i, 1, l) + \sum_{\gamma = 1}^{k} t(i, \gamma, l) \right],$$

$$l = 1, ..., M, \quad k = 1, ..., K;$$
(8)

$$\overline{G}^{lk} = \min_{i \in I_l} \sum_{j=1}^k t(i, \gamma, l), \quad l = 1, ..., M, \quad k = 1, ..., K;$$

$$\overline{R}^k = E \left[\frac{1}{N} \left(\overline{h}^k + \sum_{i=1}^N \overline{t}(i, k) \right) \right], \quad k = 1, ..., K. \quad (9)$$

Здесь и в дальнейшем E[a] — наибольшее целое значение величины в квадратных скобках, т. е.

$$E[a] = \begin{cases} |a|, \text{ если } |a| = a, \\ |a| + 1, \text{ если } |a| < a \le |a| + 1, \end{cases}$$

где |a| — целая часть числа a.

Утверждение 1. Время завершения выполнения всех заданий на параллельно работающих системах машин (значение критерия оптимальности F_1) не может быть меньше величины

$$F_1 \ge \xi_0(F_1) = \max_{1 \le l \le M} \max_{1 \le k \le K} (\overline{g}^{lk} + \overline{G}^{lk} - 2t(i, k, l)). \bullet$$



Определим времена начала \hat{L}^{lk} и завершения \overline{Q}^{lk} выполнения всех работ на каждой из машин всех поточных линий k=1,...,K,l=1,...,L:

$$\hat{L}^{I1} = \max[h^{I1}, \min_{1 \le i \le N} b_i],$$

$$\overline{Q}^{I1} = \max[h^{I1}, \min_{1 \le i \le N} b_i] + \sum_{i \in \tilde{I}^I} t(i, 1, I),$$

$$I = 1, ..., L; \tag{10}$$

$$\hat{L}^{lk} = \max\{h^{lk}; \min_{i \in \tilde{I}^{l}} [\Theta(i, (k-1), l) + t(i, (k-1), l)]\},$$

$$k = 2, ..., K, \quad l = 1, ..., L;$$

$$\overline{Q}^{lk} = \hat{L}^{lk} + \sum_{i \in \tilde{I}^{l}} t(i, k, l),$$

$$k = 2, ..., K, \quad l = 1, ..., L,$$
(12)

где значения $\theta(i, (k-1), l)$ вычисляются по формулам (1) и (2).

Утверждение 2. Время завершения выполнения всех заданий не может быть меньше величины

$$F_{1} \ge \xi_{1}(F_{1}) =$$

$$= \max_{1 \le l \le M} \max_{1 \le k \le K} \left(\overline{Q}^{lk} + \overline{G}^{lk} - 2 \sum_{i \in I} t(i, l, k) \right). \bullet$$

В случае отсутствия ограничений на времена работы машин и сроки начала и завершения выполнения заданий ($h^{kl}=b_i=0,\ H^{kl}=B_i=\infty;$ $k=1,\ ...,\ K,\ l=1,\ ...,\ L,\ i=1,\ ...,\ N$) выражения (8)—(10) упрощаются:

$$\bar{g}^{lk} = \min_{i \in \tilde{I}^l} \sum_{\gamma=1}^k t(i, \gamma, l), \ \bar{R}^k = E \left[\frac{1}{N} \sum_{i=1}^N \bar{t}(i, k) \right],
\hat{L}^{l1} = \sum_{i \in \tilde{I}^l} t(i, 1, l), \quad k = 1, ..., K, \quad l = 1, ..., L,$$

и оценки значения критерия оптимальности $\xi_0(F_1)$ и $\xi_1(F_1)$ вычисляются по более простым формулам.

Обозначим: $\tilde{J}_1^{l\omega}$ и $\tilde{V}_1^{l\omega}$, $\tilde{J}_2^{l\omega}$ и $\tilde{V}_2^{l\omega}$ — соответственно подмножества и подпоследовательности выполнения заданий на l-й поточной линии в ω -м подмножестве анализируемых вариантов, развитие которых в дальнейшем соответственно целесообразно, и после того, как построены все допустимые последовательности для данной линии; $\sigma_k(\tilde{V}_1^{l\omega})$, k=1,...,K, — времена завершения выполнения всех заданий на k-й машине для последовательности $\tilde{V}_1^{l\omega}$; $f_p(\tilde{V}_1^{l\omega})$ — значение соответствующего критерия оптимальности для последовательности $\tilde{V}_1^{l\omega}$. В частном случае $f_1(\tilde{V}_1^{l\omega}) = \sigma_k(\tilde{V}_1^{l\omega})$.

Утверждение 3. Если для двух вариантов ω и π в процессе построения последовательностей выполнения заданий на некоторой l-й поточной линии справедливы соотношения

$$\tilde{J}_{1}^{l\pi} \subseteq \tilde{J}_{1}^{l\omega}, \quad \sigma_{k}(\tilde{V}_{1}^{l\omega}) \leq \sigma_{k}(\tilde{V}_{1}^{l\pi}), \quad k = 1, ..., K;
f_{p}(\tilde{V}_{1}^{l\omega}) \leq f_{p}(\tilde{V}_{1}^{l\pi}), \quad p = 1, ..., P,$$
(13)

и хотя бы одно из неравенств выполняется как строгое неравенство, то последовательность $\tilde{V}_1^{l\omega}$ предпочтительнее последовательности $\tilde{V}_1^{l\omega}$ $\tilde{V}_1^{l\pi}$, и последняя может быть исключена из рассмотрения как неперспективная. lack

Следствие. Если справедливы соотношения

$$\tilde{J}_{2}^{l\pi} \subseteq \tilde{J}_{2}^{l\omega}, \quad \sigma_{K}(\tilde{V}_{2}^{l\omega}) \leq \sigma_{K}(\tilde{V}_{2}^{l\pi}), \quad f_{p}(\tilde{V}_{2}^{l\omega}) \leq f_{p}(\tilde{V}_{2}^{l\pi}),
p = 1, \dots, P,$$
(14)

то последовательность $\tilde{V}_2^{l\omega}$ предпочтительнее последовательности $\tilde{V}_2^{l\pi}$, и последняя может быть исключена из рассмотрения как неперспективная. lacktriangle

Правило отсева (14) может применяться, когда построены все допустимые последовательности \tilde{V}_2^{lo} , а правило (13) — в процессе их построения.

Обозначим $\Omega^{w}(\lambda) - w$ -й рассматриваемый вариант объединения λ -последовательностей $\tilde{V}_{2}^{l\omega}$ подмножества заданий $\tilde{J}_{2}^{l\pi}$, $l=1,\,2,\,...,\,\lambda$, выполняемых на первых λ -поточных линиях, $\lambda=2,\,...,\,L$; $\tilde{J}[\Omega^{w}(\lambda)]$ — подмножество заданий, входящих в $\Omega^{w}(\lambda)$;

$$\tilde{J}[\Omega^{w}(\lambda)] = \bigcup_{l=1}^{\lambda} \tilde{J}_{2}^{l\omega}, \text{ где } \tilde{J}_{2}^{l_{1}^{\omega}} = \varnothing,$$

$$l_{1}, l_{2} = 1, \dots, \lambda, \quad l_{1} \neq l_{2}. \tag{15}$$

 $F_p[\Omega^w(\lambda)]$ — значение соответствующего критерия оптимальности для $\Omega^w(\lambda)$. Так, например,

$$F_{1}[\Omega^{w}(\lambda)] = \max_{1 \le l \le \lambda} \sigma_{K}(\tilde{J}_{1}^{l\omega}),$$

$$F_{2}[\Omega^{w}(\lambda)] = \sum_{l=1}^{\lambda} \sigma_{K}(\tilde{J}_{1}^{l\omega}).$$
(16)

Утверждение 4. Если для двух вариантов w_1 и w_2 объединения последовательностей $\Omega^{w_1}(\lambda)$ и $\Omega^{w_2}(\lambda)$ справедливы соотношения

$$\tilde{J}[\Omega^{w_2}(\lambda)] \subseteq \tilde{J}[\Omega^{w_1}(\lambda)];$$

$$F_p[\Omega^{w_1}(\lambda)] \leqslant F_p[\Omega^{w_2}(\lambda)], \tag{17}$$

и хотя бы одно из неравенств выполняется как строгое неравенство, то объединение $\Omega^{w_1}(\lambda)$ предпочти-



тельнее $\Omega^{w_2}(\lambda)$, и объединение $\Omega^{w_2}(\lambda)$ может быть исключено из рассмотрения как неперспективное.

Оценки возможности выполнения заданной системы ограничений. Положив $\sigma(i, 0, l) = 0$, вычислим времена начала и завершения выполнения работ на каждой машине

$$x(i, k, l) = \max(h^k, \sigma(i, k - 1, l) + 1),$$

 $\sigma(i, k, l) = x(i, 1, l) + t(i, k, l) - 1,$
 $i = 1, ..., N, k = 1, ..., K, l = 1, ..., L.$

Утверждение 5. Если для некоторой пары индексов (\hat{i}, \hat{l}) справедливо неравенство

$$\sigma(\hat{i}, K, \hat{l}) > B_{\hat{i}}, \tag{18}$$

то выполнение \hat{i} -го задания на \hat{l} -й поточной линии недопустимо. Если для некоторого \hat{i} -го задания неравенство (18) справедливо для всех значений индексов $\hat{l}=1,...,L$, то система неравенств задачи несовместимая. \blacklozenge

Пусть на некотором s-м шаге решения задачи определены \tilde{I}^{sl} — подмножества заданий, выполняемых на l-й поточной линии.

Упорядочим все задания $i \in \tilde{I}^{sl}$ в последовательность

$$\begin{split} \tilde{U}^{sl} &= \{u_1^{ls}, \ u_2^{ls}, ..., \ u_p^{ls}, ..., \ u_P^{ls}|u_p^{ls} \in \tilde{I}^{ls}, \\ p &= 1, ..., \ P; \ B(u_p^{ls}) \leq B(u_{p+1}^{ls}), \ p = 1, ..., \ P-1\}, \end{split}$$

где P — число элементов подмножества \tilde{I}^{sl} . Вычислим значения

$$x(u_1^{ls}, 1, l) \geq \theta(u_1^{ls}, 1, l) = \max(b_i, h^{ll}),$$

$$x(i, k, l) \geq \theta(i, k, l) = \max\{\sigma(i - 1, k, l),$$

$$\sigma(i, k - 1, l)\} + 1, \quad i = u_2^{ls}, ..., u_p^{ls}, ..., u_p^{ls}, ..., u_p^{ls},$$

$$\text{и, если } i = u_p^{ls}, \text{ то } (i - 1) = u_{p-1}^{ls};$$

$$\sigma(i, 1, l) = x(i, 1, l) + t(i, 1, l) - 1,$$

$$i = u_1^{ls}, ..., u_2^{ls}, ..., u_p^{ls}, ..., u_p^{ls}.$$

$$Y^s(k, l) = \sum_{i \in \tilde{I}^{ls}} t(i, k, l), \ \overline{g}^s(k, l) =$$

$$= \min_{i \in \tilde{I}^{ls}} \left[\theta(i, 1, l) + \sum_{\gamma = 1}^{k-1} t(i, \gamma, l) \right],$$

$$\overline{G}^s(k, l) = \min_{i \in \tilde{I}^{ls}} \sum_{\gamma = k+1}^{k} t(i, \gamma, l),$$

$$\tilde{I}^{l,s} = \{i_1^{ls}, i_2^{ls}, ..., i_p^{ls}, ..., i_p^{ls}\}.$$

Утверждение 6. Если выполняется хотя бы одно неравенство из системы неравенств

$$\sigma(i,\,L,\,l) \geq B_i, \quad i \in \, \tilde{I}^{l,\,s},$$

то подмножество заданий \tilde{I}^{sl} не может быть выполнено на l-й поточной линии в установленные ограничениями сроки, и строящееся расписание не содержит допустимых решений. \blacklozenge

Утверждение 7. Если выполняется хотя бы одно из неравенств системы неравенств

$$Y^{s}(k, l) + \overline{g}^{s}(k, l) + \overline{G}^{s}(k, l) > B_{i},$$

 $k = 1, 2, ..., K, i \in \tilde{I}^{l, s}.$

то подмножество заданий \tilde{I}^{sl} не может быть выполнено на l-й поточной линии в установленные ограничениями сроки, и строящееся расписание не содержит допустимых решений. \blacklozenge

Доказательства утверждений 6 и 7 приведены в работе автора [4].

3. АЛГОРИТМЫ РЕШЕНИЯ ЗАДАЧ

Рассматриваемые задачи относятся к классу NP-сложных задач. Рассмотрим методы решения задачи с помощью алгоритмов динамического программирования [10] путем последовательного анализа с отсевом неперспективных вариантов [11]. Алгоритмы решения задачи состоят из двух этапов. На первом из них строятся различные допустимые и перспективные последовательности выполнения заданий на каждой из поточных линий

$$\tilde{V}_{1}^{l\omega} = \{ v_{1}^{l\omega}, v_{2}^{l\omega}, ..., v_{P(l\omega)}^{l\omega} \},$$

$$l = 1, ..., L, \quad \omega = 1, ..., W,$$

где $v_p^{l\omega}$ — номера заданий, входящих в последовательность $\tilde{V}_1^{l\omega}$, а $P(l,\omega)$ — число заданий в этой последовательности

В процессе построения этих последовательностей вычисляются времена начала и завершения выполнения всех работ каждого из заданий на каждой машине по формулам и отсеиваются варианты, не удовлетворяющие системе ограничений задачи. В процессе построения на основе правил отсева (13) и (14) (см. утверждение 3) производится отсев бесперспективных вариантов. В результате выполнения первого этапа будут построены последовательности $\tilde{V}_1^{l\omega}$ и $\tilde{V}_2^{l\omega}$, объединение которых в множество последовательностей $\Omega^r(\lambda)$, $r=1,\dots,R$, обеспечит получение допустимых и оптимальных по различным критериям оптимальности решений.



На втором этапе решения последовательно $\lambda=2,...,L$ рассматриваем различные r-е варианты $\Omega'(\lambda)$ объединения последовательностей $\tilde{V}_2^{l\omega}$, $l=1,2,...,\lambda$, удовлетворяющие соотношениям (15) с отсевом неперспективных продолжений на основе правила отсева (17), сформулированного утверждением 4.

Пусть $P\{\Omega^r(\lambda)\}$ — число заданий во множестве $\Omega^r(\lambda)$. При построении $\Omega^r(\lambda+1)$ рассматриваются только такие последовательности выполнения заданий на линиях $I=(\lambda+1), ..., L$, для которых справедливо соотношение

$$P\{\Omega^{r}(\lambda)\} + \max_{\substack{(\lambda+2) \le \mu \le L \ 1 \le \omega \le W}} P(\mu, \omega) + P(\lambda+1, w) \ge N, \quad w = 2, ..., W.$$
 (19)

Далее приводится формальная схема алгоритма. Ha nepsom smane решения для каждой поточной линии l=1,...,L строятся последовательно все допустимые последовательности $\tilde{V}_1^{l\omega}$ выполнения n различных заданий $p=2,3,...,P(l,\omega)$.

Времена начала и завершения выполнения всех работ каждого задания вычисляются по формулам (1)—(4), (10)—(12). Если время завершения выполнения p-го задания $T_p^{l\omega} = \sigma^{l\omega}(p,\,K,\,l) > \max(B_p,\,H^{Kl})$ или $\sigma^{l\omega}(p,\,K,\,l) > H^{kl}$, то строящаяся последовательность отбрасывается как недопустимая.

Кроме того, в процессе построения на основе соотношений (13) и (14) исключаются последовательности, не содержащие оптимальных решений. Для дальнейшего рассмотрения оставляются в каждой поточной линии только те последовательности выполнения заданий $\tilde{V}_2^{l\omega}$, число заданий в которых удовлетворяют соотношению (19).

На втором этапе решения последовательно строятся различные варианты объединений последовательностей:

$$\Omega^{2, r} = \{ \tilde{V}_{2}^{1, \omega_{1}}, \tilde{V}_{2}^{2, \omega_{2}} \},$$

$$\Omega^{3, s} = \{ \Omega^{2, r}, \tilde{V}_{2}^{3, \omega} \}, ..., \Omega^{l, s} =$$

$$= \{ \Omega^{l-1, r}, \tilde{V}_{2}^{l, \omega} \}, ..., \Omega^{L, s} = \{ \Omega^{L-1, r}, \tilde{V}_{2}^{L, \omega} \}.$$

Допустимые множества объединений последовательностей $\Omega^{l,\,s}=\{\Omega^{l-1,\,r},\,\,\tilde{V}_2^{l,\,\omega}\},\,\,l=2,\,...,\,\,L,\,\,w=1,\,...,\,W,\,s,\,r=1,\,...,\,R,\,$ должны удовлетворять соотношениям (15). Для построенных множеств $\Omega^{l,\,s}$ вычисляется значение соответствующего критерия оптимальности (например, по формулам (16)). Среди построенных подмножеств объединений последовательностей $\Omega^{l,\,s}$ на основе правила отсева (17), сформулированного утверждением 4, осу-

ществляется отсев объединений, не содержащих оптимальных решений.

После выполнения l-го шага построения допустимых и перспективных объединений последовательностей выполнения заданий на первых l линиях $\Omega^{L,\ r},\ r=1,\ ...,\ R$, среди допустимых последовательностей выполнения заданий на оставшихся линиях $\lambda=(l+1),\ ...,\ L$ оставляем только последовательности $\tilde{V}_2^{\lambda,\,\omega}$, число выполняемых заданий в которых

$$P(\lambda, w) \geqslant N - \left[\max_{1 \le r \le R} P\{\Omega^{r}(l)\} + \sum_{\substack{\mu = (l+1), \\ \mu \ne \lambda}}^{L} \max_{1 \le \omega \le W} P(\mu, \omega) \right],$$

$$\lambda = (l+1), \dots, L.$$

Если существуют допустимые объединения последовательностей $\Omega^{L,\,s}=\{\Omega^{L-1,\,r},\,\tilde{V}_2^{L,\,\omega}\},\,s=1,\,...,\,S,$ то задача имеет S различных допустимых решений, значения критериев оптимальности которых соответственно равны $F_p(\Omega^{L,\,s})$. Оптимальным решением задачи является расписание, удовлетворяющее соотношению

$$F_{p}(\overline{\Omega}^{L}) = \min_{1 \leq s \leq S} F_{p}(\Omega^{L, s}),$$

где $F_p(\Omega^{L, s})$ — соответствующий вид критерия оптимальности.

Так, например,

$$F_{1}(\overline{\Omega}^{L}) = \min_{1 \leq s \leq S} F_{1}(\Omega^{L, s}) =$$

$$= \min_{1 \leq s \leq S} \max_{1 \leq l \leq L} \sigma[v_{P(l, s)}^{l, s}, K, l],$$

$$F_{3}(\overline{\Omega}^{L}) = \min_{1 \leq s \leq S} F_{3}(\Omega^{L, s}) =$$

$$= \min_{1 \leq s \leq S} \sum_{l=1}^{L} \beta_{l} \sigma[v_{P(l, s)}^{l, s}, K, l].$$

4. ИЛЛЮСТРАТИВНЫЙ ПРИМЕР

На трех поточных линиях, каждая из которых состоит из трех машин, необходимо выполнить восемь заданий. Каждое из заданий может выполняться только на одной из поточных линий и состоит из трех операций, которые должны выполнятся в строго заданной последовательности на последовательно расположенных машинах. Времена выполнения операций каждого из заданий на соответствующих машинах этих линий, допустимые времена начала и директивные сроки завершения выполнения каждого из заданий, а также



допустимые времена работы всех машин каждой поточной линии приведены в табл. 1 и 2.

В результате выполнения первого этапа алгоритма получены следующие допустимые и перспективные расписания выполнения заданий на технологических линиях.

Первая технологическая линия:

1)
$$\tilde{V}^{1,1} = \{7\}, \ \sigma(\tilde{V}^{1,1}) = 28;$$

2)
$$\tilde{V}^{1,2} = \{7, 1\}, \sigma(\tilde{V}^{1,2}) = 33;$$

3)
$$\tilde{V}^{1,3} = \{7, 2\}, \, \sigma(\tilde{V}^{1,3}) = 37;$$

4)
$$\tilde{V}^{1,4} = \{7, 5\}; \ \sigma(\tilde{V}^{1,4}) = 33;$$

5)
$$\tilde{V}^{1,5} = \{7, 1, 3\}, \sigma(\tilde{V}^{1,5}) = 36;$$

6)
$$\tilde{V}^{1,6} = \{7, 1, 5\}, \, \sigma(\tilde{V}^{1,6}) = 39;$$

7)
$$\tilde{V}^{1,7} = \{7, 3, 5\}, \, \sigma(\tilde{V}^{1,7}) = 38;$$

8)
$$\tilde{V}^{1,8} = \{7, 5, 3\}, \, \sigma(\tilde{V}^{1,8}) = 36.$$

Так как $\sigma(\tilde{V}^{1,8}) \leq \sigma(\tilde{V}^{1,7})$, а $\tilde{I}^{1,8} = \tilde{I}^{1,7}$, то на основе правил предпочтения последовательность $\tilde{I}^{1,7}$ исключается из дальнейшего рассмотрения как неперспективная.

 $Ta6 лица \ 1$ Времена выполнения операций на машинах поточных линий

первой линии второй линии третьей линии чало шени чало 1 4 8 5 6 7 6 3 9 7 5 38 2 5 7 10 6 6 12 7 8 9 10 40 3 10 2 3 8 1 4 9 5 4 3 50 4 6 7 8 7 5 8 6 7 9 1 25 5 3 5 6 2 7 4 4 4 4 60 6 6 10 3 6 9 7 5 8 6 5 27 7 7 8 8 6 9 10 10 8 8 2 28	Nō			ена і аши	Допустимые сроки выполнения заданий							
2 5 7 10 6 6 12 7 8 9 10 40 3 10 2 3 8 1 4 9 5 4 3 50 4 6 7 8 7 5 8 6 7 9 1 25 5 3 5 6 2 7 4 4 4 4 4 60 6 6 10 3 6 9 7 5 8 6 5 27 7 7 8 8 6 9 10 10 8 8 2 28		П	машинах машинах первой второй третьей				ей		Завер- шение			
3 10 2 3 8 1 4 9 5 4 3 50 4 6 7 8 7 5 8 6 7 9 1 25 5 3 5 6 2 7 4 4 4 4 4 60 6 6 10 3 6 9 7 5 8 6 5 27 7 7 8 8 6 9 10 10 8 8 2 28	1	4	8	5	6	7	6	3	9	7	5	38
4 6 7 8 7 5 8 6 7 9 1 25 5 3 5 6 2 7 4 4 4 4 4 60 6 6 10 3 6 9 7 5 8 6 5 27 7 7 8 8 6 9 10 10 8 8 2 28	2	5	7	10	6	6	12	7	8	9	10	40
5 3 5 6 2 7 4 4 4 4 4 4 60 6 6 10 3 6 9 7 5 8 6 5 27 7 7 8 8 6 9 10 10 8 8 2 28	3	10	2	3	8	1	4	9	5	4	3	50
6 6 10 3 6 9 7 5 8 6 5 27 7 7 8 8 6 9 10 10 8 8 2 28	4	6	7	8	7	5	8	6	7	9	1	25
7 7 8 8 6 9 10 10 8 8 2 28	5	3	5	6	2	7	4	4	4	4	4	60
	6	6	10	3	6	9	7	5	8	6	5	27
	7	7	8	8	6	9	10	10	8	8	2	28
8 3 6 7 5 5 9 4 5 7 3 20	8	3	6	7	5	5	9	4	5	7	3	20

Таблица 2

Допустимые времена работы машин

Вре- мена работы	(вая ли номер иашин	oa	(рая ли номер иашин	a	Третья линия (номера машин)			
машин	№ 1	№ 2	№ 3	№ 1	№ 2	№ 3	№ 1	№ 2	№ 3	
Начало Завер- шение	6 40	6 40	6 40	1 40	5 40	5 40	5 40	5 40	5 40	

Вторая технологическая линия:

1)
$$\tilde{V}^{2,1} = \{4\}, \ \sigma(\tilde{V}^{2,1}) = 20;$$

2)
$$\tilde{V}^{2,2} = \{4, 1\}, \sigma(\tilde{V}^{2,2}) = 26;$$

3)
$$\tilde{V}^{2,3} = \{4, 2\}, \, \sigma(\tilde{V}^{2,3}) = 33;$$

4)
$$\tilde{V}^{2,4} = \{4, 3\}, \ \sigma(\tilde{V}^{2,4}) = 24;$$

5)
$$\tilde{V}^{2,5} = \{4, 5\}, \, \sigma(\tilde{V}^{2,5}) = 24;$$

6)
$$\tilde{V}^{2,6} = \{4, 1, 3\}, \, \sigma(\tilde{V}^{2,6}) = 32;$$

7)
$$\tilde{V}^{2,7} = \{4, 1, 5\}, \, \sigma(\tilde{V}^{2,7}) = 32;$$

8)
$$\tilde{V}^{2,8} = \{4, 2, 3\}, \, \sigma(\tilde{V}^{2,8}) = 37$$
:

9)
$$\tilde{V}^{2,9} = \{4, 2, 5\}, \ \sigma(\tilde{V}^{2,9}) = 37;$$

10)
$$\tilde{V}^{2,10} = \{4, 3, 1\}, \, \sigma(\tilde{V}^{2,10}) = 34;$$

11)
$$\tilde{V}^{2,11} = \{4, 3, 2\}, \sigma(\tilde{V}^{2,11}) = 39$$
:

12)
$$\tilde{V}^{2,12} = \{4, 3, 5\}; \sigma(\tilde{V}^{2,12}) = 28;$$

13)
$$\tilde{V}^{2,14} = \{4, 5, 1\}, \sigma(\tilde{V}^{2,12}) = 32$$
:

14)
$$\tilde{V}^{2,14} = \{4, 5, 3\}, \sigma(\tilde{V}^{2,14}) = 28.$$

15)
$$\tilde{V}^{2,15} = \{4, 3, 5, 1\}, \sigma(\tilde{V}^{2,15}) = 37$$
:

16)
$$\tilde{V}^{2,16} = \{4, 5, 1, 3\}, \, \sigma(\tilde{V}^{2,16}) = 36$$
:

17)
$$\tilde{V}^{2,17} = \{4, 5, 3, 1\}, \sigma(\tilde{V}^{2,17}) = 36.$$

Так как $\sigma(\tilde{V}^{2,6}) < \sigma(\tilde{V}^{2,10}), \quad \tilde{I}^{2,6} = \tilde{I}^{2,10},$ $\sigma(\tilde{V}^{2,8}) < \sigma(\tilde{V}^{2,11})$ и $\tilde{I}^{2,8} = \tilde{I}^{2,11},$ а также $\sigma(\tilde{V}^{2,13}) = \sigma(\tilde{V}^{2,7})$ и $\tilde{I}^{2,13} = \tilde{I}^{2,7},$ то последовательности $\tilde{V}^{1,10}, \quad \tilde{V}^{2,11}$ и $\tilde{V}^{2,13}$ исключаются из дальнейшего рассмотрения как неперспективные. Так как $\sigma(\tilde{V}^{2,16}) = \sigma(\tilde{V}^{2,17}) < \sigma(\tilde{V}^{2,15}), \quad \tilde{I}^{2,15} = \tilde{I}^{2,16} = \tilde{I}^{2,17},$ то на втором этапе алгоритма достаточно рассмотреть только последовательность $\tilde{V}^{2,16}$.

Третья технологическая линия:

1)
$$\tilde{V}^{3,1} = \{8\}, \ \sigma(\tilde{V}^{3,1}) = 20;$$

2)
$$\tilde{V}^{3,2} = \{8, 6\}, \, \sigma(\tilde{V}^{3,2}) = 27;$$

3)
$$\tilde{V}^{3,3} = \{8, 6, 1\}, \, \sigma(\tilde{V}^{3,3}) = 37;$$

4)
$$\tilde{V}^{3,4} = \{8, 6, 2\}, \, \sigma(\tilde{V}^{3,4}) = 38;$$

5)
$$\tilde{V}^{3,5} = \{8, 6, 3\}, \, \sigma(\tilde{V}^{6,5}) = 31;$$

6)
$$\tilde{V}^{3,6} = \{8, 6, 5\}, \, \sigma(\tilde{V}^{3,6}) = 31.$$

7)
$$\tilde{V}^{3,7} = \{8, 6, 3, 5\}, \sigma(\tilde{V}^{3,7}) = 35;$$

8)
$$\tilde{V}^{3,8} = \{8, 6, 5, 3\}, \sigma(\tilde{V}^{3,8}) = 35.$$

Так как $\sigma(\tilde{V}^{3,7}) = \sigma(\tilde{V}^{3,8}), \ \tilde{I}^{3,7} = \tilde{I}^{3,8},$ то на втором этапе алгоритма достаточно рассмотреть только последовательность $\tilde{V}^{3,7}$.



На втором этапе решения задачи получены следующие допустимые и перспективные объединения последовательностей заданий, выполняемых на различных поточных линиях:

1)
$$\overline{\Omega}_1 = {\tilde{V}^{1,2}, \ \tilde{V}^{2,8}, \ \tilde{V}^{3,6}}, \ \overline{T}(\overline{\Omega}_1) = (33, 37, 31),$$

 $F_1(\overline{\Omega}_1) = 37;$

2)
$$\overline{\Omega}_2 = {\tilde{V}^{1,2}, \ \tilde{V}^{2,9}, \ \tilde{V}^{3,5}}, \ \overline{T}(\overline{\Omega}_2) = (33, 37, 31),$$

 $F_1(\overline{\Omega}_2) = 37;$

3)
$$\overline{\Omega}_3 = {\tilde{V}^{1,5}, \tilde{V}^{2,3}, \tilde{V}^{3,6}}, \overline{T}(\overline{\Omega}_3) = (36, 33, 31),$$

 $F_1(\overline{\Omega}_3) = 36;$

4)
$$\overline{\Omega}_4 = {\{\tilde{V}^{1,2}, \, \tilde{V}^{2,3}, \, \tilde{V}^{3,7}\}, \, \overline{T}(\overline{\Omega}_4)} = (33, 33, 35),$$

 $F_1(\overline{\Omega}_4) = 35.$

Так как все остальные допустимые объединения последовательностей выполнения заданий дают не лучшие, чем $F_1(\overline{\Omega}_4)$, результаты, то получено оптимальное решение задачи $\overline{\Omega}_4=\{\tilde{V}^{1,2}=(7,1),\ \tilde{V}^{2,3}=(4,2),\ \tilde{V}^{3,6}=(8,6,3,5)\}$. Времена начала и завершения выполнения операций на различных машинах каждой из линий для построенного расписания приведены в табл. 3.

5. ОЦЕНКА СЛОЖНОСТИ АЛГОРИТМА

Система состоит из M технологических линий; $M_1 \leq M$ — число различных по техническим характеристикам технологических линий. Каждая технологическая линия содержит K машин, выполняющих N заданий, предусматривающих выполнение K операций. В условиях отсутствия системы

ограничений требуемое число операций на первом этапе решения составляет Π_1 вычислений, рассчитываемых по формулам (1)—(3), а также Π_2 операций сравнения для отсева неперспективных продолжений. Значения Π_1 и Π_2 вычисляются по формулам:

$$\Pi_{1} = O\left\{M_{1}\left[R_{1} + \sum_{i=2}^{N} 0,5[R_{i-1}(R_{i-1}-1)]\right]\right\},$$

$$\Pi_{2} = M_{1} \sum_{i=1}^{N} R_{i}!,$$

где
$$R_1 = N$$
; $R_i = 0.5R_{i-1}(R_{i-1} - 1)$, $i = 2, ..., N$.

Число операций на втором этапе решения зада-

чи
$$\Pi_3 = O\left\{ \left[\sum_{i=1}^N R_i \right]^M \right\}.$$

В случае ограничений объем вычислений существенно сокращается. Так, например, если время работы технологических линий не может превышать минимального для времени выполнения n < N заданий, и при наличии ограничений на сроки выполнения заданий в результате отсева недопустимых вариантов, значения R_i , i=1,...,N, становятся существенно меньшими. Вследствие этого при вычислении значения Π_1 число элементов суммирования сокращается, что приведет соответственно к уменьшению значений Π_1 , Π_2 и Π_3 .

Автором был выполнен вычислительный эксперимент на персональном компьютере с процессором Intel Core I-5 3230M, 2,6 ГГц, до 3,2 ГГц, 3 Мб. Решались модельные задачи теории расписаний по критерию F_1 — минимизации длины распи-

Таблица 3

Оптимальные последовательности выполнения заданий на поточных линиях

		Времена выполнения операций заданий на различных машинах											
№ № линии задания		Маши	на № 1	Маши	на № 2	Маши	Граничное						
		Начало	Завершение	Начало	Завершение	Начало	Завершение	значение					
1	7 1 3	6 13 17	12 16 26	13 21 29	20 28 30	21 29 34	28 33 36	28 38 50					
2	4 2	1 10	7 15	8 16	12 21	13 22	20 33	25 40					
3	8 6 5	5 9 14	8 13 17	9 14 22	13 21 25	14 22 28	20 27 31	20 27 60					



 $Taблица \ 4$ Результаты вычислительного эксперимента

	П	Гарал Зад	метр ачи	Ы	Параметр	Время	
№	M	M_1	N	K	На время работы линий	На сроки выполнения заданий	реше- ния, с
1	2	2	8	4	До 4—6 заданий	Слабые на 4 задания	12—15
2	3	2	10	5	До 3—5 заданий	Слабые на 6 заданий	28—42
4	4	2	12	4	До 3—6 заданий	Средние на 4 и жесткие на 3 задания	110—130
5	4	2	25	8	До 3—5 заданий	Жесткие на 10, средние на 7—8	320—440
						и слабые на 5—7 заданий	
6	5	3	15	3	До 3—4 заданий	Слабые на 2, средние на 3 и жесткие на 3 задания	230—270

сания. Данные времен выполнения операций выбирались случайным образом из диапазона 5—20. Данные граничных значений времен работы технологических линий и завершения выполнения, длина заданий выбирались также случайным образом из соответствующего диапазона, ширина которого зависела от числа операций в задании и от значения лингвистической переменной («слабые», «средние», «жесткие»). Значения b_i , B_i и h^{lk} , H^{lk} выбирались также случайным образом из соответствующего диапазона, длина которого зависела от числа операций в задании и от значения лингвистической переменной, соответствующий терм которой определялся также стохастически. Результаты вычислительного эксперимента сведены в табл. 4.

Результаты вычислительного эксперимента подтверждают полученные оценки сложности предлагаемых алгоритмов, так как решение задач даже большей размерности, но с более жесткой системой ограничений или с большим числом идентичных по техническим характеристикам технологических линий требуют меньших объемов вычислений.

ЗАКЛЮЧЕНИЕ

Автором были проведены вычислительные эксперименты по решению модельных задач размерностью до пяти поточных линий, включающих в себя до восьми машин и выполняющих до 25 заданий. Эксперименты показали, что предложен-

ные точные методы эффективны при решении задач в условиях жесткой системы ограничений на сроки выполнения заданий и на времена работы машин либо в условиях применения идентичных поточных линий. В первом случае происходит интенсивный отсев недопустимых последовательностей выполнения работ на первом этапе алгоритма. Во втором случае — выполнение вычислений на первом этапе осуществляется для одной или небольшого числа систем машин. Для решения практических задач большой размерности могут успешно применяться эвристические и приближенные методы решения, основанные на свойствах допустимых и оптимальных расписаний, установленные в настоящей статье. Рассмотрение этих алгоритмов составляет предмет самостоятельной работы.

ЛИТЕРАТУРА

- 1. *Конвей Р.В., Максвелл В.А., Миллер Л.В.* Теория расписаний. М.: Наука, 1975. 359 с.
- Bruker P. Scheduling Algorithms. Leipzig: Springer, 2007. 371 p.
- 3. *Domschke W., Scholl A., Voβ S.* Produktionsplanung. Ablauforganisatorische Aspekte. Berlin Heidelberg: Springer Verlag, 2005. 456 s.
- Зак Ю.А. Прикладные задачи теории расписаний и маршрутизации перевозок. — М.: Кн. дом «Либроком», 2011. — 393 с.
- Salvador M.S. A solution to special case of flow shop scheduling problems // Elmaghraby S.E. (Hrsg): Symposium of the Theory of Scheduling and its Applications. Berlin: Springer, 1973. P. 83—91.
- Quadt D., Kuhn H. A taxonomy of flexible flow line scheduling procedures // European Journal of Operational Research. — 2007. — Vol. 178, iss. 3. — P. 686—698.
- Ribas I., Leisten R., Framinan J.M. Review and classification of hybrid flow shop scheduling problems from a production system and a solutions procedure perspective. — Computers & Operations Research. — 2010. — Vol. 37, N 8. — P. 1439—1454.
- 8. Зак Ю.А. Решение обобщенной задачи Джонсона с ограничениями на сроки выполнения заданий и времена работы машин. Ч. 1. Точные методы решения // Проблемы управления. 2010. № 3. С. 17—25. Ч. 2. Приближенные методы решения. Там же. № 4. С. 12—19.
- 9. Зак Ю.А. Распределение множества заданий и определение оптимальных очередностей их выполнения на параллельных машинах методами динамического программирования // Информационные технологии. 2012. № 8. С. 14—20.
- 10. *Беллман Р.* Динамическое программирование. М.: Иностранная литература, 1960. 400 с.
- 11. *Михалевич В.С., Трубин В.А., Шор Н.З.* Оптимизационные задачи производственно-транспортного планирования. М.: Наука, 1986. 260 с.

Статья представлена к публикации членом редколлегии В.Н. Бурковым.

Юрий Александрович Зак — д-р техн. наук, науч. консультант, г. Аахен, Германия,

≅ +49/(0) 241-54-32-55, ⊠ yuriy_zack@hotmail.com.



УДК 004.89

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ И УПРАВЛЕНИЕ ПРОЦЕССАМИ В СИСТЕМАХ АВТОМАТИЗАЦИИ ПРОМЫШЛЕННЫХ УСТАНОВОК С ПОМОЩЬЮ МНОГОЗНАЧНЫХ НЕЙРОННЫХ СЕТЕЙ

В.Н. Мещеряков, О.В. Мещерякова, П.В. Сараев

Рассмотрена многозначная пороговая логика, на которой базируются основные принципы многозначных нейронов, перечислены преимущества многозначных нейронных сетей. Для многозначного нейрона определена функция активации и приведены алгоритм обучения и алгоритм обучения многозначной нейронной сети методом обратного распространения ошибки. Предложен метод применения многозначных нейронных сетей для решения задачи моделирования технологических процессов.

Ключевые слова: математическое моделирование, многозначная нейронная сеть, нейросетевое управление.

ВВЕДЕНИЕ

При построении систем автоматизации технологических процессов возможно применение нейронных сетей в случаях, когда традиционные решения недостаточно эффективны. Искусственные нейронные сети могут быть использованы для решения различных задач, таких как моделирование технологических процессов, управление объектами, диагностика оборудования, прогнозирование ситуаций.

Нейронная сеть является математическим аппаратом, который имеет ряд преимуществ в задачах моделирования и управления по сравнению с традиционными системами: возможность реализации задач с существенными нелинейностями, высокая производительность вычислений благодаря параллельности нейронной сети, способность к обучению и обобщению накопленных знаний [1—3].

Нейросетевые технологии успешно применяются для решения задач прогнозирования значений выходных параметров сложных технологических процессов, характерных, например, для металлургического производства. В таких случаях перспективно применение многозначной нейронной сети, функционирующей на множестве комплексных чисел, обладающей некоторыми преимуществами по сравнению с классическими нейронными сетями на множестве действительных чисел.

Многозначные нейронные сети отличаются более быстрым алгоритмом обучения и более точным прогнозом, что позволяет судить о возможности их применения к моделированию технологических процессов.

Цель исследования состоит в разработке методики моделирования технологических процессов с помощью многозначных нейронных сетей. Актуальность данной работы определяется необходимостью применения наиболее современных математических методов, к которым относятся нейронные сети, для математического моделирования технологических процессов, прогнозирования значений технологических параметров и управления.

1. АНАЛИЗ МЕТОДОВ УПРАВЛЕНИЯ

В управлении промышленными процессами со сложными математическими моделями эффективно применяются ПИ- и ПИД-контроллеры, и знание точной модели процесса не обязательно. Эти контроллеры строятся на основе классической теории управления. Тем не менее, ПИД-контроллеры обладают рядом недостатков. При возникновении возмущений требуется ручная перенастройка параметров. При наличии в системе переменных параметров, временных задержек, существенных нелинейностей и помех, ПИД-контроллеры могут не обеспечить оптимального управления.

Идея управления с самонастройкой была предложена Калманом. В ее основе лежит концепция



машины, которая бы автоматически выполняла самонастройку. Продолжением этой идеи было создание ПИД-управления с самонастройкой, которое позволяет контроллерам приспосабливаться к изменениям параметров объекта управления в оперативном режиме. Винтгенмарком была предложена схема самонастраивающегося контроллера на основе схемы размещения полюсов.

Еще один метод управления — алгоритм обобщенного управления по минимальной дисперсии. Такое управление на основе долгосрочного прогнозирования предполагает применение ряда методов обобщенного управления (динамическое матричное прогнозирование, адаптивное управление по расширенному горизонту, обобщенное прогнозирующее управление) [1].

Один из наиболее эффективных — метод нечеткого управления. Логический контроллер может разрабатываться по лингвистическим правилами, и для него не требуется априорная математическая модель объекта. Он может применяться для многомерных нелинейных процессов, для систем с неизвестной динамикой.

Адаптивные методы управления требуют наличия математической модели, основанной на физических явлениях. При изменениях в объекте управления или во внешних условиях необходима перенастройка модели и определение нового закона управления. Для применения методов управления на практике они должны обладать способностью к обучению, гибкостью, устойчивостью, нелинейностью. Нейросетевые методы управления обладают этими свойствами.

Нейросетевое управление имеет ряд преимуществ по сравнению с перечисленными ранее методами. Нейронные сети обладают свойством нелинейности, способностью к самообучению, в отличие от других методов нейроуправление не требует сложного математического аппарата и большого объема априорной информации об объекте. Кроме того, нейронные сети обеспечивают параллельную обработку информации [1, 4].

2. МНОГОЗНАЧНЫЙ НЕЙРОН

Многозначные нейроны основаны на многозначной пороговой логике [5] и определены на множестве комплексных чисел. Их входы, выходы и весовые коэффициенты являются комплексными числами.

Главные преимущества многозначных нейронов по сравнению с другими типами нейронов заключаются в значительно более высоких функциональных возможностях быстро сходящихся алгоритмов обучения. Высокие функциональные возможности означают, что любая заданная многозначная функция может быть реализована на од-

ном нейроне [6]. Многозначные нейроны быстрее обучаются и лучше адаптируются, обучаются даже нелинейным функциям. Кроме того, в отличие от обычного нейрона, функционирующего на поле действительных чисел, многозначный нейрон способен решить проблему исключающего ИЛИ.

Так как функциональность отдельно взятого многозначного нейрона выше, чем обычного нейрона, функциональность многозначной нейронной сети также выше. Даже малая многозначная нейронная сеть обучается быстрее, чем нейронная сеть на действительных числах.

На рис. 1 представлена модель многозначного нейрона. Математическая модель нейрона характеризуется входными сигналами $x_1, ..., x_n$, весовыми коэффициентами $w_1, ..., w_n$, функцией активации $f(x_1, ..., x_n)$ и выходным сигналом y. Значения входных сигналов перемножаются на соответствующие весовые коэффициенты, после чего суммируются, и к этому значению применяется функция активации, преобразующая взвешенную сумму в выходной сигнал.

Особенность многозначного нейрона состоит в том, что входы, весовые коэффициенты и выход нейрона являются комплексными числами, а функция активации — пороговой многозначной функцией. Входы и выход лежат на единичной окружности в комплексной плоскости.

Рассмотрим многозначную пороговую функцию, которая определена на множестве комплексных чисел [5]. Она принимает значения корней степени k из единицы. Рассмотрим множество E(P), образованное корнями степени k из единицы:

$$E(p) = \left\{ \varepsilon_k^j = e^{\frac{2\pi i j}{k}} \right\}, \quad j = 0, 1, ..., k - 1.$$

Рассмотрим комплексную плоскость и покажем, где расположены эти корни. Разделим единичную окружность на k частей. На рис. 2 приведен пример для случая k=8.

Таким образом, в многозначной логике на поле комплексных чисел действует многозначная (k-значная) функция n переменных $f(x_1, ..., x_n)$:

$$E_k^n \to E_k$$
.

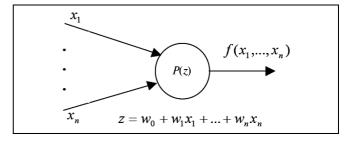


Рис. 1. Модель многозначного нейрона



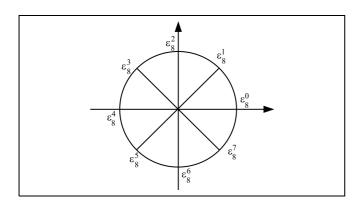


Рис. 2. Корни степени 8 из единицы

Многозначный нейрон реализует принципы многозначной пороговой логики. Он представляет переход между n входами и одним выходом. Такие нейроны функционируют на поле комплексных чисел.

Многозначная пороговая функция приведена в работах [2, 5]:

$$P(z)=\mathit{CSIGN}(z)=\,arepsilon_k^j=\,e^{rac{2\pi ij}{k}},\quad j=0,\,1,\,...,\,k-1,$$
если $rg z\in\left(rac{2\pi j}{k},rac{2\pi (j+1)}{k}
ight).$

Ее можно трактовать следующим образом: если комплексное число z расположено в j-м секторе,

так что
$$\frac{2\pi j}{k} \leqslant \arg z < \frac{2\pi (j+1)}{k}$$
, то $P(z) = e^{\frac{2\pi i j}{k}}$.

Отметим, что булева пороговая функция представляет собой частный случай многозначной пороговой функции. В самом деле, в случае k=2 функция активации может быть представлена в виде:

$$P(z) = \begin{cases} 1, 0 \le \arg z < \pi, \\ -1, \pi \le \arg z < 2\pi. \end{cases}$$

В процессе обучения многозначный нейрон корректирует свои весовые коэффициенты в соответствии с правилом корректировки весов. Другими словами, в ходе итерационного процесса обучения весовые коэффициенты изменяются каждый раз, когда для некоторой обучающей выборки фактическое значение выхода нейрона не совпадает с желаемым выходом.

Доказано, что алгоритм обучения многозначных нейронов сходится за конечное число шагов [7]. Так как обучение многозначных нейронов не сводится к решению некоторой задачи оптимизации, проблема локального минимума не возникает.

Введем обозначения: $X = (x_1, ..., x_n)$ — вектор входов; $W = (w_1, ..., w_n)$ — вектор весовых коэффициентов; $z = w_0 + x_1w_1 + ... + x_nw_n$ — взвешенная сум-

ма; $f(x_1, ..., x_n) = P(z) = P(w_0 + x_1w_1 + ... + x_nw_n)$ — многозначная пороговая функция активации; y = P(z) — выход нейрона; d — ожидаемый выход; $\delta = d - y$ — ошибка нейрона.

Алгоритм 1. Обучение многозначного нейрона

- 1. Подать входы нейрона, рассчитать выход у.
- 2. Рассчитать ошибку нейрона $\delta = d y$.
- 3. Изменить веса по правилу корректировки весов [8]:

$$W_{r+1} = W_r + \frac{C_r}{n+1}(d-y)\bar{X} = W_r + \frac{C_r}{n+1}\delta\bar{X}$$
,

где r — номер итерации; C_r — скорость обучения;

 \bar{X} — вектор, сопряженный входному.

4. Если критерий останова не выполняется, перейти на шаг 1;

$$MSE = \frac{1}{N} \sum_{s=1}^{N} \gamma_s^2 < \lambda,$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{s=1}^{N} \gamma_s^2 < \lambda},$$

где γ_s — ошибка на s-м обучающем примере, λ — точность обучения.

3. МНОГОЗНАЧНАЯ НЕЙРОННАЯ СЕТЬ

Рассмотрим многослойную нейронную сеть прямого распространения на основе многозначных нейронов (рис. 3). Входной сигнал в таких сетях распространяется в прямом направлении, от слоя к слою. Число входных и выходных элементов в многослойном нейронной сети определяется условиями задачи.

Представим алгоритм обучения многозначной нейронной сети в векторном виде, для этого введем обозначения: L — число слоев нейронной сети; X^l — вектор входов l-го слоя, l = 1, ..., L; W^l — вектор весовых коэффициентов l-го слоя; Y — век-

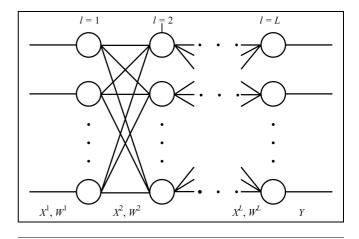


Рис. 3. Многозначная нейронная сеть



тор выходов нейронной сети; D — вектор желаемых выходов нейронной сети; δ^l — вектор ошибок l-го слоя.

Алгоритм обучения многозначной нейронной сети методом обратного распространения ошибки по сути аналогичен классическому алгоритму для нейронных сетей, действующих на множестве действительных чисел, но отличается большим быстродействием. Этот алгоритм имеет ряд преимуществ: основан на правиле корректировки ошибки и не требует вычисления производной; самоадаптация скорости обучения для всех нейронов; более высокие быстродействие и скорость распознавания/предсказания/классификации по сравнению с другими нейронными сетями, нечеткими сетями на основе метода опорных векторов.

Алгоритм 2. Обучение многозначной нейронной сети

- 1. Подать входы нейронной сети X^{l} , рассчитать вектор выходов.
 - 2. Рассчитать ошибку нейронной сети

$$\delta^l = \frac{D - Y}{N^{L-1} + 1}.$$

3. Рассчитать вектор ошибок каждого слоя

$$\delta^l = \frac{\delta^{l+1}(W^{L+1})^{-1}}{N^{l-1}+1}$$
, где $l=1, ..., L-1$.

4. Изменить веса по правилу корректировки весов:

$$W_{r+1}^l = W_r^l + \frac{C_r}{N^{l-1}+1} \delta^l \overline{X}^l$$
, где $l=1, ..., L$.

5. Если критерий останова не выполняется, перейти на шаг 1;

$$MSE = \frac{1}{N} \sum_{s=1}^{N} \gamma_s^2 < \lambda,$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{s=1}^{N} \gamma_s^2 < \lambda},$$

где γ_s — ошибка на s-м обучающем примере, λ — точность обучения.

4. МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ТЕХНОЛОГИЧЕСКИХ ПРОЦЕССОВ С ПОМОЩЬЮ МНОГОЗНАЧНЫХ НЕЙРОННЫХ СЕТЕЙ

Многозначные нейронные сети могут быть применены для прогнозирования параметров технологических процессов. В качестве входов $X = (x_1, ..., x_n)$ нейронной сети могут выступать измеряемые величины, а в качестве выхода Y — моделируемое значение. Нейронная сеть строится следующим образом: выбирается число слоев и скрытых ней-

ронов, весовые коэффициенты $W = (w_1, ..., w_n)$ — произвольные комплексные числа. Далее на обучающей выборке производится обучение нейронной сети по алгоритму, описанному в § 2. После обучения сети мы получаем математическую модель процесса в виде черного ящика: имеем входные значения и выходные значения, которые описаны некоторой зависимостью, скрытой внутри многозначной сети.

Благодаря эффективному моделированию достигается оптимальное управление различными объектами. Аппарат многозначных нейронных сетей позволяет успешно реализовывать задачу прогнозирования и моделирования технологических процессов, при этом проявляются преимущества этого типа нейронных сетей [9].

5. УПРАВЛЕНИЕ С ПОМОЩЬЮ НЕЙРОННЫХ СЕТЕЙ

Работающая и адекватная модель технологического процесса позволяет прогнозировать значения технологических параметров, что приводит к решению обратной задачи моделирования — управлению. В реальном производстве часть переменных измеряется с большой погрешностью, а некоторые возмущающие переменные и факторы вообще не поддаются измерению и контролю [10], поэтому для получения продукции требуемого качества необходим контроль технологических параметров. В данной работе для этой задачи предложен аппарат многозначных нейронных сетей.

При управлении с помощью нейронных сетей используется модель объекта управления в виде черного ящика. Состояние объекта нейроуправления считается недоступным для внешнего наблюдения, а наблюдаемыми являются текущие значения входа и выхода [4].

Основные методы нейроуправления: подражающее нейроуправление (Controller Modeling), инверсное нейроуправление (Direct Inverse Control), прогнозирующее нейроуправление (NN Predictive Control), гибридное нейроуправление, вспомогательное нейроуправление (Adaptive Inverse Control) [4] и др.

Другой подход к нейронному управлению — косвенное адаптивное управление (управление с пассивной адаптацией или косвенным измерением возмущений), в зарубежной литературе он известен под названием Internal Model Control (IMC). Этот метод позволяет компенсировать возмущения. Для оценки возмущений вычисляется разность регистрируемого выхода объекта y(t) и модели (рис. 4). Сигнал рассогласования e(t), вычитаемый из сигнала уставки r(t), используется для оценки возмущения d(t). Данная схема часто применяется в сочетании с классическими подходами к синтезу систем управления, к примеру, с ПИД-регулятором [12].



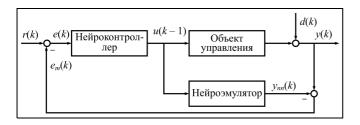


Рис. 4. Схема адаптивного косвенного управления

В этой схеме для управления объектом используются две нейронные сети — нейроконтроллер и нейроэмулятор. Нейроэмулятор служит прямой моделью объекта, а нейроконтроллер — инверсной моделью и предназначен для выработки управляющего сигнала. В такой архитектуре нейронного управления применяется алгоритм обратного распространения и обеспечивается более точное непосредственное обучение нейроконтроллера, так как ошибка может распространяться в обратном направлении через эмулятор в каждой выборке [1].

В качестве нейроконтроллера и нейроэмулятора предлагается использовать многозначные нейронные сети. Обучение прямого нейроэмулятора происходит в режиме офф-лайн. Для этого применяется алгоритм 2 — алгоритм обратного распространения для многозначной нейронной сети. Обучение происходит до тех пор, пока отличие выходных сигналов объекта и модели (нейроэмулятора) не будет незначительным. Далее, также с помощью алгоритма 2, происходит он-лайн обучение нейроконтроллера — инверсной модели. На вход нейроконтроллера поступает значение уставки r(k)для следующего такта, нейроконтроллер генерирует значение управляющего сигнала u(k-1), который идет на вход объекта управления и модели. На объект также воздействуют возмущения d(k). В результате вырабатывается выход объекта управления y(k) и выход модели — нейроэмулятора $y_{nn}(k)$. Разность этих значений является ошибкой управления $e_m(k) = y_{nn}(k) - y(k)$, которая пропускается в обратном направлении (при этом весовые коэффициенты нейроэмулятора не корректируются). После прохождения нейроэмулятора ошибка далее распространяется через нейроконтроллер, где корректируются весовые коэффициенты. Этот метод позволяет оценивать возмущения по реакции модели объекта на управляющие воздействия [11]. Он может быть применен для моделирования нагрузки [12].

ЗАКЛЮЧЕНИЕ

Интеллектуальное управление находит все более широкое применение в различных технических приложениях. Во многих случаях обучение системы

управления должно быть выполнено с учетом различных факторов, таких как динамические и статические характеристики объекта, возмущения окружающей среды. Благодаря особенностям обучения, параллельности и адаптации многозначные нейронные сети, отличающиеся более простыми и быстрыми алгоритмами обучения, более точным прогнозом выходных параметров, могут успешно применяться для решения задач моделирования и управления техническими процессами.

Предложенный подход к моделированию и управлению технологическими процессами с помощью многозначных нейронных сетей позволяет повысить точность системы управления и, в свою очередь, качество конечной продукции.

ЛИТЕРАТУРА

- 1. *Омату С., Халид М., Юсов Р.* Нейроуправление и его приложения. М: ИПРЖР, 2000. 272 с.
- 2. *Сараев П.В.* Идентификация нейросетевых моделей. Липецк: ЛГТУ, 2011. 94 с.
- 3. Леденева Т.М., Подвальный С.Л., Васильев В.И. Системы искусственного интеллекта и принятия решений. Уфа, 2005. 165 с.
- Чернодуб А.Н., Дзюба Д.А. Обзор методов нейроуправления // Проблемы программирования. 2011. № 2. С. 79—94.
- Блюмин С.Л. Пороговые многозначные функции // Известия АН СССР. Техническая кибернетика. 1972. № 1. С. 101—108.
- 6. Айзенберг Н.Н., Айзенберг И.Н., Кривошеев Г.А. Нейросети на многозначных нейроэлементах: обучение, обработка и распознавание изображений // Компьютерная оптика. 1995. № 14—15. С. 179—186.
- 7. Aizenberg I. Complex-Valued Neural Networks with Multi-Valued Neurons. Berlin: Springer, 2011. 264 p.
- Айзенберг Н.Н., Иваськив Ю.Л., Поспелов Д.А. Об одном обобщении пороговой функции // Доклады АН СССР. — 1971. — Т. 196, № 6. — С. 1287—1290.
- 9. *Мещерякова О.В.* Применение многозначных нейронных сетей // Управление большими системами: материалы IX Всерос. школы-конференции молодых ученых. Тамбов—Липецк, 2012. Т. 2. С. 58—59.
- 10. *Кудинов Ю.И., Военков А.В., Келина А.Ю.* Моделирование технологических и экологических процессов. Липецк: ЛЭГИ, 2001. 131 с.
- Змеу К.В., Марков Н.А., Ноткин Б.С. Прогнозирующее инверсное нейроуправление позиционно-следящим пневмоприводом // Информатика и системы управления. 2011. № 3. С. 104—117.
- 12. Подвальный Е.С., Тюрин С.В., Соляник А.А. Оперативное управление автоматизированными технологическими комплексами на основе графического моделирования и визуализации задач диагностики и моделирования нагрузки // Системы управления и информационные технологии. 2011. Т. 46, № 4.1. С. 171—175.

Статья представлена к публикации членом редколлегии A.C. Манделем.

Виктор Николаевич Мещеряков — д-р техн. наук, зав. кафедрой, **☎** (4742) 32-80-56, ⊠ mesherek@stu.lipetsk.ru,

Ольга Викторовна Мещерякова — мл. науч. сотрудник,

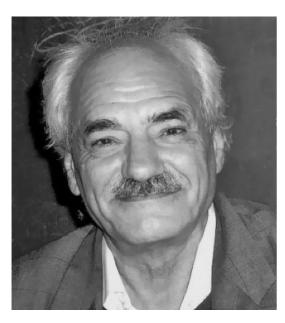
Павел Викторович Сараев — канд. техн. наук, декан,

☎ (4742) 32-80-02, ⊠ psaraev@yandex.ru,

Липецкий государственный технический университет.



СИМПОЗИУМ «МЕРЫ СЛОЖНОСТИ», ПОСВЯЩЕННЫЙ 75-ЛЕТИЮ А.Я. ЧЕРВОНЕНКИСА



В городе Пафос на Кипре 2 октября 2013 г. состоялся международный симпозиум «Меры сложности», посвященный 75-летию Алексея Яковлевича Червоненкиса. В его программе были представлены доклады, посвященные истории и современному состоянию проблемы машинного обучения и распознавания образов. Она содержала доклады Р. Дадли (США), А. Гаммермана и В. Вовка (Великобритания), Б. Шелькопфа (Германия), Л. Ботту (США) и К. Воронцова (Россия). В работе Симпозиума в режиме телеконференции принял участие многолетний коллега юбиляра В.Н. Вапник (США). С сообщением на тему «Меры сложности» выступил сам А.Я. Червоненкис. В прочитанных на Симпозиуме докладах подчеркнута важная роль, которую методы машинного обучения, разработанные А.Я. Червоненкисом вместе с В.Н. Вапником, сыграли в теории распознавания образов и машинного обучения.

В своем докладе «От классов событий к классам функций» Р. Дадли отметил, что существуют возможные расширения действия закона больших чисел на семейства функций и что исследования В. Вапника и А. Червоненкиса открыли возможности расширения действия центральной предельной теоремы на новые области. Он отметил многочис-

ленные применения такого расширения в статистике вообще, а не только в исходно поставленной задаче машинного обучения. По его словам, «после некоторых предшественников XIX в., утверждение о виде сложности класса множеств, сделанное А. Червоненкисом и В. Вапником в 1968 г., значительно расширило сферу действия законов больших чисел в теории вероятностей ».

В своем выступлении А. Гаммерман изложил основные факты научной биографии юбиляра. После окончания в 1961 г. Московского физико-технического института А.Я. Червоненкис участвовал в создании светомузыкальной установки, которая в 1961 г. демонстрировалась в Лондоне. Увлекшись задачей распознавания образов, он быстро продвинулся в ней и скоро стал авторитетом в этой области. В 1971 г. А.Я. Червоненкис защитил кандидатскую диссертацию, а с 1987 по 2005 г. по совместительству выполнял обязанности научного консультанта фирмы «Интегра», создавая компьютерные программы для управления разработкой месторождений золота и других драгоценных металлов. За эти работы А.Я. Червоненкис удостоен Государственной премии СССР. В 2000—2009 гг. он был профессором Центра компьютерного обучения Королевского университета в Лондоне и ему присвоено звание Почетного профессора Лондонского университета. С 2007 г. по настоящее время А.Я. Червоненкис, будучи старшим научным сотрудником Института проблем управления им. В.А. Трапезникова РАН, по совместительству работает консультатном фирмы «Яндекс» в России.

В своем выступлении В. Вовк рассказал об истории развития работ по распознаванию образов в Институте проблем управления. В частности, он отметил роль М.А. Айзермана и М.А. Бравермана. После опубликования монографии «Теория распознавания образов» (1974) А.Я. Червоненкис и В.Н. Вапник стали признанными авторитетами в теории распознавания образов и компьютерного обучения. Среди наиболее значимых результатов

¹ Artificial intelligence applications and innovations / H. Papadopoulos, A.S. Andreou, L. Iliadis, I. Maglogiannis (Eds.). — Heidelberg, New-York, Dordrecht, London: Springer, 2013. — P. XIX.



А.Я. Червоненкиса находится разработанный им вместе с В.Н. Вапником алгоритм распознавания образов, известный как «метод обобщенного портрета²». Развитие принципов и алгоритмов оптимального выбора сложности в задаче распознавания образов на основе доступных экспериментальных данных и сложности класса решающих правил стало основой многих алгоритмов машинного обучения. А.Я. Червоненкису принадлежит вывод необходимых и достаточных условий равномерной сходимости средних значений к математическому ожиданию, а также изучение характеристики класса множеств, вошедшей в современную математику как функция Вапника—Червоненкиса (размерность Вапника—Червоненкиса, VC-dimension).

В докладе Б. Шелькопфа «Причинные связи и статистическое обучение» поставлена общая проблема изучения причинных связей на основе их статистических «отпечатков». Автор рассмотрел задачи применения причинных структур в сценариях машинного обучения, таких как ковариационный сдвиг или обучение с частичным привлечением учителя.

Свой доклад «О возникновении леммы Вапника — Червоненкиса» Л. Ботту посвятил истории разработки теории распознавания образов. Он отметил, что переход от простого закона больших чисел к равномерному закону больших чисел основывается на важной комбинаторной лемме, доказательство которой появилось в нескольких странах почти одновременно. Он привел данные по анализу этого перехода, который он назвал потрясающим, «earth shattering result».

В докладе *К. Воронцова* «Комбинаторная теория оверфиттинга: как связность и расщепленность уменьшают локальную сложность» рассмотрен новый подход, позволяющий получить границы вероятности оверфиттинга, зависящие от исходных данных. Этот подход требует представления поискового пространства в виде направленных ациклических графов, обычно весьма большого размера. В отличие от методов бустинга, баггинга или метода случайных деревьев, изучающих большие ансамбли слабых классификаторов, в докладе рассматриваются малые ансамбли сильных классификаторов.

С докладом на тему «Меры сложности» выступил А.Я. Червоненкис. По его словам, чем сложнее восстанавливаемая зависимость, тем больше должен быть размер обучающего множества для ее

построения. Вероятно, первым теоретическим результатом в этой области была теорема Котельникова (критерий Найквиста). Теорема утверждает, что для восстановления непрерывной функции на основании некоторого числа измерений в дискретных точках необходимо количество измерений, пропорциональное ширине ее спектра. Таким образом, сама ширина спектра может служить одной из возможных мер сложности. В общем случае исследователь должен ограничить себя определенным уровнем сложности модели, который зависит от объема данных, что приводит к необходимости определения понятия сложности и к поиску способов ее количественной оценки.

Совместно с В.Н. Вапником А.Я. Червоненкис свел анализ способности системы к обучению к проблеме равномерной сходимости частот к вероятностям по классу событий (или равномерной сходимости средних к математическим ожиданиям по классу функций). Условия равномерной сходимости были сформулированы ими в терминах индекса класса событий относительно заданной выборки, функции роста и так называемой размерности Вапника—Червоненкиса. Если равномерная сходимость имеет место, то система способна обучаться. Однако противное не верно: система может сохранить способность к обучению, даже если такая сходимость отсутствует. Можно указать примеры поиска решающего правила, зависящего от очень большого числа параметров, для нахождения которого, однако, достаточно лишь небольшого числа показов. Например, в методе бустинга используются очень громоздкие формулы, но, несмотря на это, метод дает неплохие результаты даже при ограниченном количестве данных для обучения.

Таким образом, в докладе показано, что размерность Вапника — Червоненкиса является одной из возможных мер сложности. Рассмотренные в докладе примеры заставляют исследователя искать и другие меры сложности, не связанные с понятием равномерной сходимости. Вероятно, такие меры будут зависеть от распределения вероятностей в пространстве входных переменных. Но этого избежать невозможно, заключает автор.

В.Н. Новосельцев

Василий Николаевич Новосельцев — д-р техн. наук, гл. науч. сотрудник, Институт проблем управления им. В.А. Трапезникова РАН, г. Москва,

2 (495) 334-88-91, ⋈ novoselc@yandex.ru.

Редколлегия и редакция поздравляют Алексея Яковлевича с 75-летием и желают ему доброго здоровья и дальнейших творческих успехов!

² Метод обобщенного портрета послужил основой широко известного метода опорных векторов — Support Vector Machine, за разработку которого В.Н. Вапник в 2012 г. получил в США медаль им. Б. Франклина.



УКАЗАТЕЛЬ СТАТЕЙ, ОПУБЛИКОВАННЫХ В 2013 г.

Акинфиев В.К., Цвиркун А.Д. Проблемы управления инвестициями. Программный комплекс «ТЭО-ИНВЕСТ». — 2013. — № 4. — С. 32—40.

Алешин Б.С., Баженов С.Г., Лебедев В.Г., Кулида Е.Л. Оценка реализуемости и безопасности траекторий самолета с помощью бортовой математической модели. — 2013. — № 4. — С. 64—71.

Алисейчик А.П., Павловский В.Е. Модель и динамические оценки управляемости и комфортабельности движения многоколесного мобильного робота. — 2013. — N 1. — C. 70—78.

Андриенко А.Я., Тропова Е.И., Чадаев А.И. Совершенствование управления расходованием топлива ракет-носителей «Союз-2» в обеспечение повышения экологичности их эксплуатации. — 2013. — № 4. — С. 72—75.

Асратян Р.Э., Лебедев В.Н. Прокси-серверы в распределенных гетерогенных мультисетевых средах. — $2013. - N_{\odot} 2. - C. 45-50.$

Ахохов А.А., Блудян Н.О., Дорофеюк Ю.А., Чернявский А.Л. Разработка системы управления межрегиональными автобусными перевозками на базе методов коллективной многовариантной экспертизы. — 2013. — N 5. — C. 29—38.

Бабушкина Н.А. Оценка управляющих дозовых воздействий противоопухолевой вакцинотерапии с помощью математического моделирования. — 2013. — \mathbb{N}° 5. — C. 60—65.

Болодурина И.П., Парфенов Д.И. Управление потоками запросов при доступе к широкополосным мультимедийным образовательным ресурсам системы дистанционного обучения. — 2013. — 1000 6. — C. 1000 C. 10

Бурлаков Е.А. Математическое моделирование деятельности малых и средних организаций на ранней стадии кризиса. — 2013. — \cancel{N} 3. — C. 41—50.

Ведешенков В.А. Подход к диагностированию возникшего отказа в цифровых системах со структурой симметричного двудольного графа. — 2013. — № 2. — С. 36—40.

Вице-президент Академии наук СССР академик Борис Николаевич Петров. — 2013. — № 1. — С. 79—82.

Вишневский В.М., Ларионов А.А., Семенова О.В. Оценка производительности высокоскоростной беспроводной тандемной сети с использованием каналов сантиметрового и миллиметрового диапазона радиоволн в системах управления безопасностью дорожного движения. — 2013. — N 4. — C. 50—56.

Волкова И.О., Губко М.В., Сальникова Е.А. Активный потребитель: задача оптимизации потребления электроэнергии и собственной генерации. — 2013. — № 6. — С. 53—61.

Волковицкий А.К., Каршаков Е.В., Павлов Б.В. Позиционирование подвижных объектов в низкочастотном электромагнитном поле. Ч. 1. Базовый алгоритм относительного позиционирования. — 2013. — N 1. — C. 57—62.

Волковицкий А.К., Каршаков Е.В., Павлов Б.В. Позиционирование подвижных объектов в низкочас-

тотном электромагнитном поле. Ч. 2. Особенности технической реализации. — $2013. - \mathbb{N} 2. - \mathbb{C}. 58-64.$

Воробьев Г.Г., Дмитренко Л.Г. Анализ возможности построения теории интеллекта на основе аналитической психологии К.Г. Юнга. — 2013. — № 4. — С. 82-87.

Вшивкова О.А., Комаров А.С., Фролов П.В., Хлебопрос Р.Г. Роль гетерогенности среды обитания при управлении численностью популяций иксодового клеща: клеточно-автоматная модель. — 2013. — \cancel{N} 4. — С. 57—63.

Гольдовская М.Д., Дорофеюк Ю.А., Киселева Н.Е. Методы структурного анализа в прикладных задачах исследования временных рядов. — 2013. — № 3. — С. 33—40.

Горелик В.А., Золотова Т.В. О некоторых оценках устойчивости фондового рынка и влиянии на них информированности инвесторов. — 2013. — № 6. — С. 41—47.

Гребенюк Г.Г., Крыгин А.А., Никишов С.М. Подход к алгоритмизации выбора точки технологического присоединения к электрическим сетям. — 2013. — № 3. — С. 60—70.

Гусев Л.А. Интервальные оценки вероятности при наличии неразличимости. — $2013. - \mathbb{N} \cdot 4. - C. 16 - 22.$

ХХ Международная конференция «Проблемы управления безопасностью сложных систем». — 2013. — \mathbb{N}_2 3. — C. 78—83.

10 лет журналу «Проблемы управления». — 2013. — № 1. — С. 2—6.

Жарких А.А., Бычкова С.М. Два способа вычисления вероятностей распознавания направления переноса точки на плоскости на фоне случайных поворотов. — 2013. — N 2. — C. 9—15.

Жилякова Л.Ю. Управление предельными состояниями в поглощающих ресурсных сетях. — 2013. — N 2. — C. 51—59.

Зак Ю.А. Построение расписаний выполнения заданий для параллельно работающих систем машин и поточных линий. — 2013. — № 6. — C. 62—70.

Зенков С.М., Митришкин Ю.В., Фокина Е.К. Системы управления положением, формой и током плазмы в токамаке Т-15. — 2013. — N 4. — С. 2—10.

Зыкин С.В., Полуянов А.Н. Формирование многомерных данных с использованием промежуточных представлений. — $2013. - N \cdot 5. - C. 54 - 59.$

Иванов В.П. Взаимосвязь фундаментальной науки, прикладных исследований и практики проектирования систем управления. — 2013. — № 1. — С. 83—85.

Конев Д.С. Интеграция инерциально-спутниковых систем на основе инвариантной модели состояния подвижного объекта. — 2013. — \mathbb{N} 2. — \mathbb{C} . 65—72.

Корнетова А.Н., Червоненкис А.Я. Оптимизация показов рекламы в поисковых системах. — 2013. — 1. —

Корноушенко Е.К. Регрессионный подход к массовой оценке при наличии ненаблюдаемой составля-

78



ющей в зависимой переменной. — 2013. — № 4. — С. 23—31.

Косачев Ю.В. Оптимальное управление инновационным процессом в рамках интегрированной структуры. — $2013. - N_0 1. - C. 32-39.$

Косяченко С.А., Ковалевский С.С., Сомов С.К. Задача синтеза оптимальной модульной системы обработки данных реального времени с двухзвенной архитектурой «Клиент — Сервер». — 2013. — № 4. — С. 41—49.

Крутова И.Н., Суханов В.М. Сравнительный анализ особенностей динамики гиросилового управления крупногабаритным упругим спутником для трех типов базовых алгоритмов. — 2013. — N 4. — C. 76—81.

Кубышкин В.А., Постнов С.С. Исследование задачи оптимального управления для одиночного и двойного интеграторов дробного порядка с помощью метода моментов при поиске допустимых управлений в пространстве $L_p[0,\ T].-2013.-N 2.-C.9-17.$

Кузьмин О.В., Кедрин В.С. Анализ структуры гармонических рядов динамики на базе алгоритма сингулярного разложения. — 2013. — № 1. — C. 26—31.

Лакеев А.В., Русанов В.А., Козырев В.А. К реализации непрерывных квазилинейных систем с автономными операторами в гильбертовом пространстве. — 2013. — № 1. — C. 7—18.

Ларина Е.А., Сидоренко А.М., Хоботов Е.Н. Выбор порядка сборки узлов и агрегатов при планировании работ на предприятиях с дискретным характером производства. — 2013. — № 3. — С. 71—77.

Легович Ю.С., Максимов Д.Ю. Логические модели выбора решения в самоорганизующихся системах. — 2013. — N 2. — N — N 2. — N

Макаров В.Л., Бахтизин А.Р., Сушко Е.Д. Компьютерное моделирование взаимодействия между муниципалитетами, регионами, органами государственного управления. — 2013. — № 6. — С. 31—40.

Мещеряков В.Н., Мещерякова О.В., Сараев П.В. Математическое моделирование и управление процессами в системах автоматизации промышленных установок с помощью многозначных нейронных сетей. — 2013. — N 6. — C. 71—75.

Морозов М.В. Робастная устойчивость дискретных систем управления с периодическими интервальными ограничениями. — 2013. — N 4. — C. 11—15.

Непейвода Н.Н. Алгебраический подход к управлению. — 2013. — № 6. — С. 2—14.

Новиков Д.А. Большие данные: от Браге — к Ньютону. — 2013. — № 6. — C. 15—23.

Павел Павлович Пархоменко (к 90-летию со дня рождения). — 2013. — № 1. — С. 86—87.

Панкова Л.А., Пронина В.А. Модели текстового поиска на основе теории нечетких множеств. — 2013. — N 2. — C. 41—44.

Петровец Ю.О., Андрианов Д.Л. Задача оптимального планирования работ по обновлению железнодорожной линии: постановка, алгоритмы решения. — 2013. — № 1. — C. 50—56.

Петрущенко В.В. Учет неоднородности в оболочечном анализе данных. — 2013. — № 5. — С. 2—11.

Рубан А.И. Функционалы чувствительности в задаче Больца для многомерных динамических систем, описываемых интегро-дифференциальными уравнениями с запаздыванием. — 2013. — № 2. — С. 2—8.

Сараев П.В., Сяглова Ю.Е. Повышение эффективности валютного хеджирования на основе результатов нейроструктурного прогнозирования. — 2013. — $N \ge 6$. — C. 48—52.

Симпозиум «Меры сложности», посвященный 75-летию А.Я. Червоненкиса. — 2013. — № 6. — С. 76—77.

Сомов Е.И. Синтез нелинейного закона управления для перезакрутки упругого спутника слабыми внутренними моментами. — 2013. — № 1. — С. 63—69.

Сомов Е.И., Бутырин С.А. Полетная идентификация и силовая гироскопическая стабилизация слабо демпфированной конструкции крупногабаритного спутника. — 2013. — № 2. — С. 51—57.

Стецюра Г.Г. Базовые механизмы взаимодействия активных объектов цифровых систем и возможные способы их технической реализации. — 2013. — № 5. — С. 39—53.

Сулакшин С.С., Авинова А.Н., Богдан И.В. Количественная оценка индекса успешности развития страны и ее прогностический потенциал. — 2013. — N 2. — C. 16—19.

Указатель статей, опубликованных в 2013 г. — 2013. — № 6. — С. 78—79.

Фуртат И.Б. Адаптивное управление неминимально-фазовыми объектами определенного класса. — 2013. — № 1. — С. 19—25.

Хабров В.В. Многошаговая задача управления инвестиционными портфелями на основе моделей векторных авторегрессий и моделей многомерной волатильности. — 2013. - N 2. - C. 20 - 35.

Цыкунов А.М. Робастное управление объектом с распределенным запаздыванием. — 2013. — № 3. — С. 2—8.

Ченцов А.А., Ченцов А.Г., Ченцов П.А. Элементы динамического программирования в экстремальных задачах маршрутизации. — 2013. — № 5. — С. 12—21.

Шайкин М.Е. К задаче скрытности подвижного объекта на периферии зоны его контакта с обнаружителем. — 2013. — № 5. — С. 66—78.

Шестая международная конференция «Параллельные вычисления и задачи управления». — 2013. — \mathbb{N}_2 2. — C. 73—75.

Шестая международная конференция «Управление развитием крупномасштабных систем» MLSD'2012. — $2013. - N_{\odot} 2. - C. 76-79.$

Шилин А.А., Букреев В.Г. Динамическое определение траектории скольжения при релейном управлении нелинейным объектом. — 2013. — № 5. — С. 22-28.

Шомова Е.Н. Вероятностная модель влияния финансирования научного исследования на эффективность инновационного проекта. — 2013. — № 3. — С. 27—32.

Экономическая политика в период глобальных потрясений: на пути к новой парадигме. — 2013. — $N \ge 5$. — C. 79—81.



CONTENTS & ABSTRACTS

Keywords: algebraic programming, magmas, control systems, supercomputers, Landauer limit, Chaitin limit.

D.A. Novikov

Challenges of Big Data technologies to engineers, specialists in applied mathematics and artificial intelligence, as well as to scientists in various branches of applications, are analyzed and structured.

Keywords: Big Data, Data Analysis, Big Control.

I.P. Bolodurina, D.I. Parfenov

The paper presents a model of servicing requests for broadband multimedia learning resources and model of data access in hybrid cloud storage system. The models are used to improve the effectiveness of computing resources in distance learning system by distributing the flow of requests between nodes with use of developed module for the «OpenStack» system controller.

Keywords: cloud computing, multimedia learning resources, load balancing, hybrid cloud system «OpenStack».

V.L. Makarov, A.R. Bakhtizin, E.D. Sushko

The paper gives an overview of current multi-agent systems developed for analysis of typical problems of territorial entities. The multi-agent model of municipalities within a region where the agents are people and organizations in which they work is presented. The special consideration is given to providing a realistic simulation of the behavior of such agents as participants of production, as well as of the interdependence of behavior and state of agents of various types and environment changes. The example of using the model to test different options of vertical interbudgetary relations is given.

Keywords: agent-based modeling, testing of regional policy, human behavior in socio-economic environment, labor potential.

V.A. Gorelik, T.V. Zolotova

The paper proposes the measures of stability and risk for a separate investor and the structured share market. The research of the market model stability with the use of proposed measures is carried out. It is shown that investors, adhering to various forecasts of the share market development as a result of their various information awareness, and choosing different strategies (portfolios), promote the given system stability.

Keywords: collective risk, risk coefficient, average covariance, stability.

P.V. Saraev, Yu.E. Syaglova

The paper gives the analysis of effectiveness of using results of time series neurostructural prediction (exchange rates) in hedging of currency risks using derivative financial instruments. Description of the developed software is given. The technique of computation of effectiveness of use of neurostructural predictions is considered. Results of computational experiments are provided.

Keywords: neurostructural modeling, time series prediction, foreign exchange hedging.

I.O. Volkova, M.V. Goubko, E.A. Salnikova

As a part of managing behavior of an active consumer of electric power in prospective smart grids it is necessary to create a mathematical model that meets his or her economic interests. Existing models either do not take into account all relevant aspects or turn out to be too complicated for the purposes of multi-agent modeling. Mathematical model of an active consumer and its use for investigating the problem of consumption and local generation regimes optimization is proposed. The conditions when the consumer's problem has a pretty simple and efficient solution are derived. The proposed approach is illustrated by optimizing the operating modes of equipment for a single household.

Keywords: active consumer in power industry, Smart Grid, Demand-Side management, Demand Response.

THE CONSTRUCTION SCHEDULE FOR THE PERFORMANCE OF TASKS OF PARALLEL SYSTEMS OF MACHINES AND PRODUCTION LINES. . . 62

Yu.A. Zak

The development of Flow-Shop-Problem is considered. The mathematical model of scheduling for parallel working systems of machines under a given set of constraints on deadlines for separate tasks is considered. The properties of admissible and optimal schedules are examined and the methods of optimal scheduling by sequential optimization algorithms are proposed. The presented example shows the efficiency of the developed algorithms.

Keywords: parallel system of machines, Flow-Shop-Problem, limits on assignments, valid and optimal scheduling algorithms, sequential optimization.

V.N. Meshcheryakov, O.V. Meshcheryakova, P.V. Saraev

The paper considers a multi-valued threshold logic, which is the base of multi-valued neurons, discusses the advantages of complex-valued neural networks. The complex-valued activation function is defined for the multi-valued neuron and the back propagation learning algorithms for the multi-valued neuron and complex-valued neural networks are considered. Further, a method of applying complex-valued neural networks for modeling processes is proposed.

Keywords: mathematical modeling, complex-valued neural networks, neurocontrol.

80

