

ISSN 1819-3161

ПРОБЛЕМЫ УПРАВЛЕНИЯ

1/2014

CONTROL  SCIENCES

РЕДКОЛЛЕГИЯ

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

РУКОВОДИТЕЛИ РЕГИОНАЛЬНЫХ РЕДСОВЕТОВ (РРС)

**ВЛАДИВОСТОК — О.В. Абрамов (4232) 31-02-02
ВОЛГОГРАД — А.А. Воронин (8442) 40-55-06
ВОРОНЕЖ — С.А. Баркалов (4732) 76-40-07
КРАСНОЯРСК — В.В. Огурцов (3912) 27-63-89
КУРСК — С.Г. Емельянов (4712) 50-48-00
ЛИПЕЦК — А.К. Погодаев (4742) 32-80-44
ПЕРМЬ — В.Ю. Столбов (3422) 39-12-97
МИНСК — А.В. Тузиков (10-37517) 284-21-40
САМАРА — В.Г. Засканов (8463) 35-64-07
САРАТОВ — В.А. Твердохлебов (8452) 22-10-42
ТАГАНРОГ — В.М. Курейчик (8634) 39-32-60
ТИРАСПОЛЬ — С.И. Берилл (10-373553) 9-44-87
УФА — Б.Г. Ильясов (3472) 73-78-35**



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. Пашенко

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. Бабилова

Отпечатано в ИПУ РАН

Заказ № 9

Подписано в печать
28.01.2014 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

1.2014

СОДЕРЖАНИЕ

Обзоры

Бреер В.В. Модели конформного поведения. Ч. 1. От философии
к математическим моделям 2

Математические проблемы управления

Жуковский В.И., Солдатова Н.Г. Гарантированные риски и исходы
в «игре с природой» 14

Управление в социально-экономических системах

Бычков И.В., Казаков А.Л., Лемперт А.А. и др. Интеллектуальная система
управления развитием транспортно-логистической инфраструктурой
региона 27

Сидельников Ю.В. Четырехэтапная мозговая атака. 36

Беляева А.В., Гребенюк Е.А. Построение моделей массовой оценки
объектов недвижимости с учетом пространственной зависимости 45

Угольников Г.А., Усов А.Б. Моделирование коррупции в трехуровневых
системах управления 53

Информационные технологии в управлении

Степаненко С.А., Южаков В.В. Эксафлопные супер-ЭВМ: контуры
архитектуры 63

Управление подвижными объектами

Потехин А.И., Браништов С.А., Кузнецов С.К. Дискретно-событийные
модели железнодорожной сети 73

Блюмин С.Л., Сысоев А.С. Применение теоремы Лагранжа о конечных
приращениях для решения проблем управления транспортными
системами 82

* * *

Contents & abstracts 88

МОДЕЛИ КОНФОРМНОГО ПОВЕДЕНИЯ.

Ч. 1. От философии к математическим моделям

В.В. Бреев

Дан краткий обзор изучения конформного поведения с точки зрения различных наук — философии, культурологии, теоретической, практической и социальной психологии. Предпринята попытка классифицировать математические модели конформного поведения как частный случай моделей социального взаимодействия. Подробно рассмотрены модели Шеллинга и Грановеттера, давшие начало другим математическим моделям.

Ключевые слова: конформное поведение, социальное взаимодействие, модели критической массы, пороговые модели социального взаимодействия, социофизические модели, управление толпой.

ВВЕДЕНИЕ

Конформное поведение проявляется как взгляды, убеждения и, в конечном итоге, действия, подчиняющиеся правилам поведения в группе. Одним из проявлений конформного поведения являются социальные нормы. Как правило, эти нормы носят неявный характер и регулируют взаимодействия внутри социальной группы или в обществе в целом. Эта тенденция может проявляться в виде бессознательных влияний или в виде прямого и открытого социального давления. Конформность может проявляться как в присутствии других людей, так и когда человек находится в одиночестве и не испытывает непосредственного влияния извне. Например, люди имеют тенденцию следовать социальным нормам даже в одиночку, когда едят или смотрят телевизор.

Люди зачастую ведут себя конформно из желания безопасности в группе. Обычно это происходит в группах сверстников либо группах, представляющих единую культуру, религию или образовательный статус. Это часто называют групповым мышлением, подразумевая под этим такой образ мышления, который характеризуется самообманом, принудительным согласием и соответствием групповым ценностям и групповой этике поведения, которая игнорирует реалистичную оценку других вариантов действий. Нежелание соответствовать несет в себе риск социального отторжения.

Конформность часто ассоциируется с культурой поведения в период отрочества и юности, но также может сильно проявляться у людей всех возрастов.

Несмотря на то, что давление со стороны сверстников может проявляться негативно, конформность может иметь хорошие или плохие эффекты в зависимости от ситуации. Вождение по правильной стороне дороги, и вообще по правилам, можно рассматривать как полезный эффект конформности. При правильном влиянии родителей и окружения в раннем детстве ребенок впоследствии намного легче адаптируется и разделяет поведение, принятое в обществе. Конформность влияет на формирование и поддержание социальных норм, а также помогает обществу функционировать нормально и предсказуемо через самоустранение таких видов поведения, которые находятся в противоречии с неписаными правилами. В этом смысле конформность может быть позитивной силой, которая предотвращает разрушительные и опасные акты поведения.

Так как конформность — групповой феномен, такие факторы, как размер группы, единодушие, сплоченность, социальный статус и общественное мнение, оказываются определяющими для степени конформности отдельных индивидуумов.

Явление конформности изучается разными науками. Цель настоящего обзора состоит в рассмотрении математических моделей, описывающие данное явление, и в попытке их классификации.



Первая часть обзора имеет следующую структуру. В § 1 конформное поведение кратко описывается с точки зрения гуманитарных наук — *философии* (И. Кант [1] — о природе моды, Ф. Ницше [2] — о стадном инстинкте), *классической психологии* (К. Юнг [3] — об идентификации и имитации) и *мифологии* (М. Элиаде [4] — исследование мифологии архаических народов).

В § 2 явление конформности рассматривается с точки зрения *прикладной психологии*. Отмечаются положительные (единство в кризисных ситуациях и др.) и отрицательные (снижение способности к самостоятельности и др.) факторы конформности, вводятся типы социального влияния при конформности: информационное и нормативное влияние. Здесь же приводятся результаты хрестоматийных психологических экспериментов (М. Шерифа [5], Р. Барона [6], С. Эша [7, 8]), подтверждающих наличие явления конформности при разных типах влияния на испытуемых.

В § 3 сделана попытка объяснить, какое место занимают *модели социального взаимодействия* в теориях, описывающих массовое поведение больших социальных групп. Для этого приводится классификация математических моделей социального взаимодействия по составу, основным параметрам и принципу функционирования моделей, а также по применяемому математическому аппарату (теория игр, теория динамических систем, теория вероятностей и др.). При рассмотрении математических моделей в настоящей работе по возможности принято единообразное описание по таким аспектам (пунктам): состав участников, принцип функционирования, основные параметры и основные результаты. Если требуются специфические аспекты, то они добавляются в каждом конкретном случае.

Основным эффектом, который изучается в математических моделях социального взаимодействия, является *эффект критической массы*, который заключается в том, что при непрерывном изменении определенного параметра («массы») поведение системы меняется кардинальным образом. Например, при превышении плотностью населения данного типа определенной величины число его представителей будет расти, тогда как, если плотность ниже этой величины, то число представителей этого типа уменьшается. В § 3 делается обзор первых моделей критической массы, основоположниками которых в социологии являются Т. Шеллинг [9–13] и М. Грановеттер [14]. Две базовые модели Шеллинга [9] — *модель пространственного соседства* (Spatial Proximity Model) и *модель ограниченного окружения* (Bounded-Neigh-

borhood Model), в основном, интерпретируются в терминах сегрегации и других проявлений расового разделения.

Ставшая хрестоматийной модель М. Грановеттера [14] основывается на социологических исследованиях генезиса погромов и революций. В отличие от Шеллинга, Грановеттер исследовал устойчивость равновесия в поведении одного типа агентов. Работа [14] стала отправной точкой для дальнейшего развития моделей *порогового социального поведения* (Threshold Models) [15], а также для приложений этой модели к внедрению инноваций (новых технологий и т. п.) [16], потребительскому спросу [17], сегрегации [18] и задачам распределения общественных благ (public goods) [19]. Более подробный список приложений моделей порогового поведения приведен далее в табл. 3.

Развитие классических моделей Шеллинга и Грановеттера будут рассмотрены в следующей части обзора.

1. ОБЩЕНАУЧНЫЕ ИССЛЕДОВАНИЯ КОНФОРМНОСТИ

Философия. Конформное поведение проявляется у человека в большей или меньшей степени всегда. Одним из множества примеров служит следование *моде*. Некоторые исследователи (см., например, работы [20–22]) считают, что следование моде в большой степени зависит от компромисса между индивидуальным желанием выделиться и социальным желанием быть похожим на других. Высказывания некоторых философов также иллюстрируют эту гипотезу. Так, И. Кант говорит, что следование моде не соответствует разумному понятию пользы [1, с. 277]: «Человеку присуща естественная склонность сравнивать себя в своем поведении с тем, кто более значителен (ребенку — со взрослым, простому человеку — с более знатными людьми), и подражать его образу действий. Закон такого подражания, цель которого состоит в том, чтобы не казаться менее значительным, чем другие, причем в той области, где совершенно не принимаются во внимание соображение пользы, называется модой».

Другим примером конформного поведения служит *мораль*. Одним из самых известных исследователей и критиков морали является Ф. Ницше [2, 23, 24]. В своем афоризме «Страх перед одиночеством» он приводит один из компромиссных внутренних мотивов конформного поведения человека в обществе [2, с. 84]: «Упреки совести и у самого совестливого человека слабы по сравнению с чувством: «вот это и вон то противно хорошему

тону твоего общества». Даже сильнейший все еще боится холодного взгляда, искривленного гримасой рта, со стороны тех, среди которых и для которых он воспитан. Чего же тут, собственно, бояться? Одиночества! — этого аргумента, перед которым отступают даже наилучшие аргументы в пользу какой-нибудь личности или дела! — Так вещает в нас стадный инстинкт.»

Для краткого определения моделей конформного поведения термин Ницше «*стадный инстинкт*» подходит наилучшим образом (существуют и современные термины — *herd behavior* [25], *nerd behavior* [26]). В работе [27] предпринята попытка объяснить, почему социальные традиции, а именно стадный инстинкт, так долго сохраняются, передаваясь от поколения к поколению.

Психология. Еще к большему обобщению явлений конформного поведения приходит К.Г. Юнг. Для описания он определяет это явление через термин *диссимилиация*, «как уподобляющее приспособление субъекта к объекту и отчуждение субъекта от самого себя в пользу объекта, будь то внешний объект или же объект «психологический», как, например, какая-либо идея» [3, с. 569]. Соответствующий психологический процесс Юнг называет *идентификацией* или *имитацией*. Здесь более понятен термин имитация как сознательное подражание. Идентификация есть бессознательное подражание («стадный инстинкт» у Ницше). Важно то, что «имитация есть необходимое вспомогательное средство для развивающейся, еще юной личности. Она способствует развитию до тех пор, пока не служит для простого удобства и не задерживает развития подходящего индивидуального метода. Подобно этому и идентификация может содействовать развитию, пока индивидуальный путь еще не проложен. Но как только открывается лучшая индивидуальная возможность, так идентификация обнаруживает свой патологический характер тем, что оказывается в дальнейшем настолько же задерживающей развитие, насколько до этого она бессознательно содействовала подъему и росту ... субъект под ее влиянием расщепляется на две частичные личности, чуждые одна другой» [3, с. 586].

Приведенное определение показывает, что субъект развивается с помощью идентификации и имитации (в нашем определении — конформности), т. е. подражания другим, и, достигнув определенного *уровня*, развивается далее благодаря индивидуальным возможностям. Этот уровень зависит от многих факторов — возраста (молодые люди подвержены конформности в большей степени, чем зрелые), пола (женщины, как правило, более

конформны, чем мужчины), воспитания и образования (менее образованные люди охотнее повторяют поведение друг друга), среды обитания (монокультурное окружение приводит к большей конформности поведения). Зачастую при определенных обстоятельствах субъекту необходим *выбор* поведения, которое характерно для каждой из его двух частичных личностей, описанных Юнгом. Этот выбор зависит от уровня индивидуальности субъекта. Все психологические типы Юнг строит на основании степеней идентификации и имитации.

Мифология. Историко-мифологические корни конформного поведения обнаружил известный французский культуролог Мирче Элиаде. В результате всестороннего исследования *мифологии* архайческих народов, населяющих различные уголки нашей планеты, он делает следующий вывод: «...человек традиционных культур представлял себя реальным лишь в той мере, в какой он переставал быть самим собой (с точки зрения современного наблюдателя), довольствуясь *имитацией* и *повторением* кого-то другого. Иными словами, он признавал себя реальным, «действительно самим собой» именно в той мере, в какой переставал им быть» [4, с. 46].

2. КОНФОРМНОСТЬ В ПРИКЛАДНОЙ ПСИХОЛОГИИ

Конформностью в психологии называют такой процесс взаимодействия в социальной группе, когда поведение индивидуума направляется или корректируется поведением других людей. Конформность играет как положительную, так и отрицательную роли для личности и социальной группы.

Положительная роль конформности состоит в следующем:

- формируется единство группы в кризисных ситуациях, что позволяет, например, выжить в сложных условиях;

- упрощается организация совместной деятельности благодаря отсутствию размышлений по поводу поведения в стандартных обстоятельствах и получению инструкций по поведению в нестандартных обстоятельствах;

- уменьшается время адаптации человека в коллективе;

- группа приобретает единое лицо.

В отрицательном смысле (беспринципность, приспособленчество) конформность:

- снижает способность самостоятельно ориентироваться в новых и непривычных условиях;

- притупляет критичность восприятия окружающей реальности;



- способствует некритичной подмене индивидуальных морально-этических норм социальными;
- способствует развитию предрассудков и предрешений против меньшинств;
- снижает способность к оригинальным и творческим идеям.

Конформность имеет место как в малых группах, так и в обществе в целом. Она может проявляться как подсознательно, так и непосредственно с помощью социального давления или добровольно (см., например, игру [28], описанную далее). Конформность проявляется и в присутствии окружающих, и тогда, когда другие люди фактически не присутствуют.

Такая универсальность этого явления не могла пройти мимо внимания тех, кто изучает поведение человека: ранее — философов, а в настоящее время — психологов. Соответствующие процессы назывались и называются по-разному, и все названия в какой-то мере отражают его суть — конформность.

Глубинные причины и механизмы конформности пытаются исследовать психологи. Так, считается, что люди часто подвержены конформности из-за стремления почувствовать себя в безопасности внутри группы, которая обычно состоит из людей одинакового возраста, культуры, религии или уровня образования. Любое непринятие правил сопровождается реальным риском социального отвержения (см. цитату Ф. Ницше [2] ранее). В соответствии с этим конформность служит одним из способов защиты от запугивания и критики со стороны равных (peer pressure).

Так как конформность — это явление, проявляющееся в социальной группе, то индивидуальный порог конформизма зависит от многих факторов, таких как размер группы, степень единогласия в ней, индивидуальный статус субъекта в группе, априорная приверженность идеям группы, общественное мнение и т. п. [29].

В работе [30] отмечаются три главных *типа социального влияния* при конформности:

- согласованность (compliance) как общественная конформность;
- идентификация (identification) — подражание кому-то, кого субъект любит, уважает или этот кто-то пользуется всеобщим признанием (см. цитату К.Г. Юнга [3] ранее);
- интернализация (internalization) как процесс принятия индивидуумом норм, которые установлены людьми или группами, имеющими влияние на этого индивидуума.

В социальной психологии рассматриваются в основном два типа социального влияния — *норма-*

тивное и *информационное*, соответствующие первому и третьему из перечисленных типов влияния.

Информационное влияние возникает тогда, когда индивидуум обращается к членам группы, чтобы получить точную информацию. Это в основном бывает в трех случаях, когда ситуация:

- сомнительная, когда индивидуум не уверен, что нужно делать;
- критическая, когда необходимо принять быстрое решение, несмотря на панику; индивидуум повторяет действия большинства, и зачастую это не всегда правильное решение;
- «неразрешимая» без помощи экспертов; с другой стороны, мнение экспертов не всегда истина.

Информационное влияние часто приводит к *интернализации*, т. е. имеет место искренняя вера индивидуума в то, что информация верна. Информационное социальное влияние было впервые отмечено в работе М. Шерифа [5], где исследуется поведение людей, изменяющих свое мнение для того чтобы соответствовать общему мнению группы. Участники помещались в темную комнату, и им предлагали вглядываться в светлое пятно на расстоянии около пяти метров. Далее их просили оценить расстояние, на которое переместилось пятно. На самом деле пятно не двигалось, а имел место обман зрения, называемый «самодвижущимся эффектом» (auto kinetic effect). Все участники оценивали расстояние, «пройденное» пятном, по-разному, но при обсуждении приходили к одному общему мнению. С помощью анализа результатов этого эксперимента автор предложил модель распространения как социальных норм, так и заблуждений в обществе.

Последующие эксперименты базировались на более реальных ситуациях [6]. Так, на этапе опознания испытуемым показывали «подозреваемого» сначала индивидуально, а затем в составе других подозреваемых. Им давалась одна секунда, чтобы опознать его, что делает задачу достаточно трудной. Одной группе сообщали, что их общее мнение чрезвычайно важно и будет учтено для изменения процедур следственного процесса. Другой группе было сказано, что это просто тест. В первой группе конформность проявили 51 % ее членов, тогда как во второй — всего 35 %.

В работе [31] была предложена гипотеза о том, что отклонения (fads) и тренды поведения в обществе формируются в результате рационального выбора отдельных индивидуумов на основе информации, полученной ими от окружающих. Информационные потоки (informational cascades — информационные каскады, лавины) очень быстро

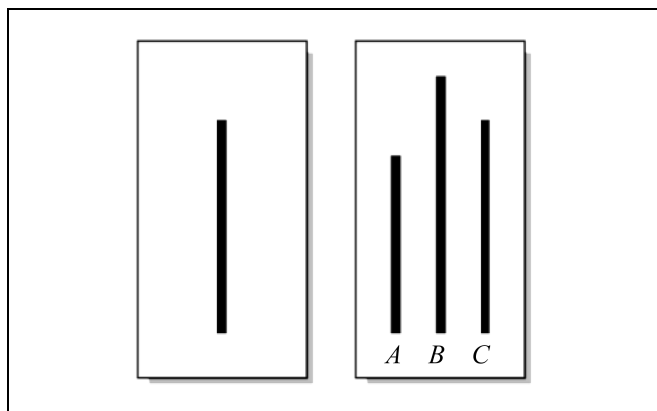


Рис. 1. Опыт С. Эша, иллюстрирующий конформность

формируются, так как люди преимущественно игнорируют свои внутренние сигналы и соглашаются с тем, что делают окружающие. Эти потоки являются преходящими (fragile) потому, что люди постепенно осознают, что они основываются на ограниченной информации. В круг рассматриваемых явлений авторы работы [31] включили, в том числе, и явления моды. Предложенная ими математическая модель описывается далее.

Нормативное (институциональное) социальное влияние имеет место тогда, когда индивидум хочет понравиться или быть принятым в определенную социальную группу. Это приводит к так называемой публичной уступчивости (public compliance), когда индивидум может делать или говорить то, во что он не верит.

Первым психологом, кто исследовал нормативное социальное влияние в лабораторных условиях, был Саломон Эш [7, 8]. Он модифицировал эксперимент со световым пятном, описанный ранее. Первоначально каждому испытуемому предлагали простой визуальный тест, в котором ошибка маловероятна. Ему следовало указать, какому из отрезков — *A*, *B* или *C* в правой части — равен отрезок в левой части рис. 1.

Правильный ответ, очевидно, «*C*» и 97 % индивидуально испытуемых отвечали правильно. Затем испытуемых тестировали в составе группы. Каждому испытуемому сообщали, что остальные члены группы — такие же участники, как и он. На самом деле последние были подставными лицами и отвечали единодушно по заранее составленному плану экспериментатора: например не «*C*», а «*A*». Уровень конформизма оказался очень высоким — число неправильных ответов у испытуемых возрастало в 10 раз до 32 % (с 3 % в индивидуальных тестах).

Во всех описанных ранее экспериментах, иллюстрирующих конформное поведение, индивидум выбирает одно решение из ограниченного числа вариантов. При выборе он испытывает влияние со стороны окружающих, причем это влияние тем более сильное, чем большая доля окружающих выбрала это решение. Внутренние факторы каждого конкретного индивидума очень разнотиповы, но конформность как бы преодолевает все эти барьеры.

Интересно, что перечисленные психологические эксперименты стали хрестоматийными — они приводятся практически в любом учебнике по социальной психологии, однако проводились они достаточно давно (в середине прошлого века). Среди относительно новых исследований можно указать на работу [6]. На основании экспериментов, проведенных ее авторами, можно сделать вывод, что уровень конформности убывает по мере возрастания важности поставленной перед индивидумом задачи.

3. МАТЕМАТИЧЕСКИЕ МОДЕЛИ КОНФОРМНОГО ПОВЕДЕНИЯ

Ранние социологические и психологические работы (см., например, книги [3, 32]) содержали утверждения о том, что массовое поведение больших социальных групп регулируется неким общим фундаментальным подсознанием (или массовым сознанием). У К.Г. Юнга, например, этими функциями обладал так называемый «архетип». Но эта теория не нашла применения в построении моделей массового поведения, и сегодня общими задачами социальных наук являются видимые изменения в социальных явлениях, которые возникают без участия явных фундаментальных для них причин. Чтобы описать подобного рода явления, необходимо учитывать «мелкие» изменения в переносных поведения большого числа агентов. Эти микроизменения в конечном итоге приводят к макроэффектам в социальных явлениях. Модели на базе такого подхода называются *моделями социального взаимодействия* (Models of Social Interactions) (см., например, работы [33, 34]). В них поведение агента зависит, наряду с прочими факторами, от выбора других агентов во всей социальной группе или в ее части, составляющей «окружение» данного агента. Макроэффект возникает тогда, когда значительная доля агентов придерживается одного и того же выбора.

Модели социального взаимодействия нашли отражение в обширной литературе, отражающей теорию социального поля [35], динамику распро-

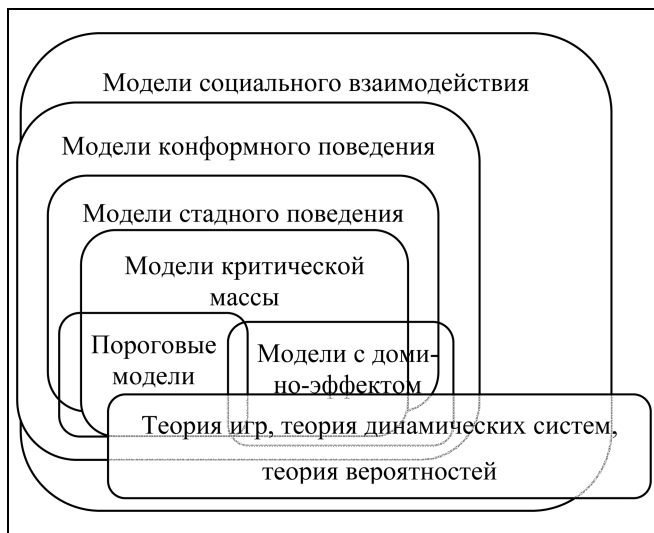


Рис. 2. Классификация моделей социального взаимодействия

странения сегрегации [9—13], участие в протестах [14], принятие религиозных предпочтений [36], бизнес-циклы [37], диффузия инноваций [38], экономический рост [39], изменение уровня преступности [40], принятие новых технологий [41], флуктуации цен на финансовых рынках [42, 43].

Схематично классификацию математических моделей социального взаимодействия можно представить в виде схемы — рис. 2 (при этом мы делаем акцент на «детализацию» вариантов конформного поведения).

Математический аппарат, применяемый для описания моделей социального поведения, достаточно разнообразен: теория вероятностей, дифференциальные уравнения, теория устойчивости, теория игр [44, 45] и др.

В рамках математических моделей социального взаимодействия [46, 47]) значительную долю составляют исследования конформного поведения. Работы по *моделям конформного поведения* [27, 48, 49] основываются на том, что индивидуальное поведение во многом мотивируется так называемыми социальными факторами, такими как желание престижа, уважения, популярности или желание быть принятым в различные социальные группы. Социологические, психологические и антропологические исследования подтверждают тот факт, что эти факторы широко распространены и приводят к конформности (см. § 1, 2).

Описание многих социальных процессов, в том числе моделей конформного поведения, с точки зрения системного анализа и когнитологии содержится в работе [50]. Вероятностные (неигровые)

модели конформного поведения рассматриваются в работах [51, 52]. В них учитываются два фактора — автономность агента и его зависимость от коллектива, и рассматриваются различные частные случаи соотношений между этими факторами. В работе [53] рассматривается динамическая (также неигровая) модель, и в ней кроме собственного мнения и «подражательного эффекта» (конформности) добавляется еще один фактор — внешний стандарт (или навязанная традиция).

Модели *подражательного поведения* (имитации) — см. работы [25, 54, 55] — строят, исходя из того, что агенты получают информацию, наблюдая за действиями остальных, и поэтому намереваются имитировать тех, кто, по общему мнению, лучше информирован или более эффективен.

В работе [26] строится модель «стадного поведения» (herd behavior) или «тупого поведения» (nerd behavior) в образовательном процессе. Так, студент может прикладывать или не прикладывать усилия в процессе учебы в зависимости от того, что он думает о том, как ведут себя его соседи. Соответственно, это приводит к нескольким возможным положениям равновесия, которые характеризуются различным уровнем усилий, прилагаемых студентами в процессе учебы. В работе [41], в которой исследуется модель конформного поведения для принятия новых технологий, делается вывод, что «стадное поведение» может привести к равновесию, которое не является общественно оптимальным. В работе [56] сравниваются эволюционный и теоретико-игровой подходы к исследованию конформного поведения в обучении. Оказывается, что эволюционный подход приводит к равновесию, которое совпадает с равновесием Нэша, *доминирующим по риску*. К уменьшению количества возможных множественных равновесий может привести учет в модели социальных норм, как это сделано в работе [57].

Одно из направлений исследования конформности составляют *модели критической массы* [58—61] (см. вложенные области на рис. 2). Эти модели характеризуются признаками:

- агенты осуществляют дискретный (или бинарный) выбор;
- агенты гомогенны в своих предпочтениях, т. е. их поведение можно описать одной целевой функцией;
- функция полезности агента возрастает с увеличением доли других агентов (его окружения), сделавших такой же выбор.

Такие термины, как «критическая масса», «эпидемии» и «инфекции», происходят от моделей, развитых в физике и эпидемиологии и во мно-

гом сходных между собой. Например, SIR-модель (Susceptibility/Infection/Recovery — восприимчивость/инфекция/выздоровление) в математической эпидемиологии восходит к работе [57] (см. также обзор применения этого класса моделей в социальных системах в работе [62]). В SIR-модели рассматривается популяция, в которой индивидум может заразиться инфекцией при контакте с другими. Весь процесс заболевания делится на три стадии — восприимчивость, заражение инфекцией и выход из состояния болезни по причине выздоровления или смерти. Один из важных вопросов, на который необходимо найти ответ, будет ли болезнь распространяться до тех пор, пока все восприимчивые индивидумы популяции не переболеют ею, или она остановится, не затронув всех членов популяции. Подверженность болезни, представляющая собой аналог индивидуального порога в моделях [14], считается одинаковой для всех индивидумов популяции. После выздоровления индивидум приобретает иммунитет к болезни («Все, что нас не убивает, делает нас сильнее» [2]). В качестве критической массы здесь выступает плотность населения. Если плотность населения больше некоторого критического значения, то болезнь захватывает всю восприимчивую часть популяции. Если же плотность меньше этого значения, то эпидемия затухает.

Как уже отмечалось, основополагающими работами по моделям критической массы в социальной науке являются работы Т. Шеллинга [9—13] и М. Грановеттера [14]. За исследования, отраженные в работе [13], Шеллинг получил Нобелевскую премию по экономике в 2005 г. Книга входит в сотню «самых влиятельных» научных работ второй половины XX в.

В работах Шеллинга [9, 11], содержащих анализ расовой сегрегации, развиваются несколько моделей критической массы. В этих моделях рассматривается принятие семьей решения по выбору района проживания в зависимости от предпочтений по отношению к своему окружению (своим соседям).

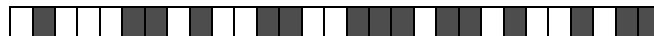
Модель 1 в работе [9] называется *моделью пространственного соседства* (spatial proximity model).

Состав модели. Рассматривается поведение двух групп агентов, которые различаются по одному из признаков: раса, пол, возраст, доход, язык, религия, цвет кожи и т. п. Агенты, в зависимости от личных предпочтений, могут жить в окружении агентов противоположной группы или перемещаться в то место, где агенты своей группы представлены в большей пропорции. Признак агента изображается графически белым или чер-

ным квадратом, и агенты распределяются на прямой (или на плоскости в двумерном варианте модели). В начальный момент времени агенты распределены на прямой по определенному закону, как это показано в табл. 1.

Таблица 1

Начальное распределение агентов в модели пространственного соседства



Функционирование. Агент может перемещаться по прямой (или на плоскости), выбирая место, удовлетворяющее критерию наличия минимальной доли соседей с таким же признаком. Агенты осуществляют выбор последовательно (порядок ходов определяется заданной первоначальной нумерацией агентов). При покидании агентом своего места, его соседи смыкаются, при появлении на новом месте, соседи раздвигаются.

Основные параметры. Выделяются пять факторов, от которых зависит эволюция системы:

- количество других агентов в окружении данного агента; окружение определяется как заранее заданное количество агентов по обе стороны от него;
- минимальная доля соседей с таким же признаком (цветом) в окружении;
- общее соотношение «белых» и «черных» агентов;
- правила перемещения;
- начальное распределение агентов вдоль прямой (или на плоскости).

Основные результаты. Модель исследуется путем имитационного моделирования. Шеллинг отмечает, что в рамках модели с минимальным процентом соседей с таким же признаком, равным 5/9, существует равновесие, которое может быть графически изображено в виде «кластеров». Таких равновесий много, и они представляют собой структуру с регулярным переменным составом в 2,5 и более агентов с одинаковым признаком.

Если уменьшить размер значимого для агента окружения соседями, то структура равновесия остается такой же — регулярные кластеры из определенного числа агентов, как это показано в табл. 2.

Таблица 2

Конечное распределение агентов в модели пространственного соседства



При изменении общего соотношения «белых» и «черных» агентов в равновесии, вместо чередующихся кластеров возникает так называемая сегрегация, т. е. агенты с цветом, составляющим меньшинство, образуют либо малое число кластеров, либо один большой кластер. При ограничении расстояния, на которое могут агенты перемещаться (изменение правила перемещения), число кластеров возрастает, а их размер, соответственно, убывает. Для распределения агентов на плоскости общие эффекты остаются теми же, но с более сложной структурой кластеров.

Модель 2 называется *моделью ограниченного окружения* (bounded-neighborhood model). В отличие от предыдущей модели, в ней окружение имеет другой смысл. Здесь оно не привязано к конкретному месту, а скорее к сообществу, в котором агенту комфортно находиться. Опять-таки, рассматривая случай «белых» и «черных» агентов, каждого агента интересует доля «своих», находящихся в его окружении. В этой модели под окружением понимается не непосредственное число соседей, а значительно большее число агентов (100—200) в некоторой области. Это позволяет заменить дискретное описание континуальным через функции распределения.

Состав модели. Рассматривается поведение двух групп агентов, которые различаются по одному из признаков, аналогичных модели пространственного соседства. Агент определенной группы (например, «белый») может находиться в окружении, где доля агентов противоположной группы («черных») не превышает его верхнего *уровня толерантности* — характерного для этого агента порога.

Функционирование. Если доля агентов противоположной группы не превышает порога данного агента, то последний продолжает находиться в этом окружении, иначе он его покидает. При благоприятных для себя обстоятельствах агент может вернуться в это окружение обратно. Таким образом, агент имеет только двоичный выбор, в отличие от предыдущей модели, где стратегия представляла собой перемещение вдоль прямой или на плоскости.

Основные параметры. Объект исследования — функции распределения *порогов толерантности* агентов разных типов. В зависимости от начальных условий и вида кривых распределения исследуется динамика процесса перехода данного окружения в состояние равновесия, при котором плотность агентов определенного типа не изменяется.

Общее знание (common knowledge [63]) состоит в том, что агенты обладают информацией только о

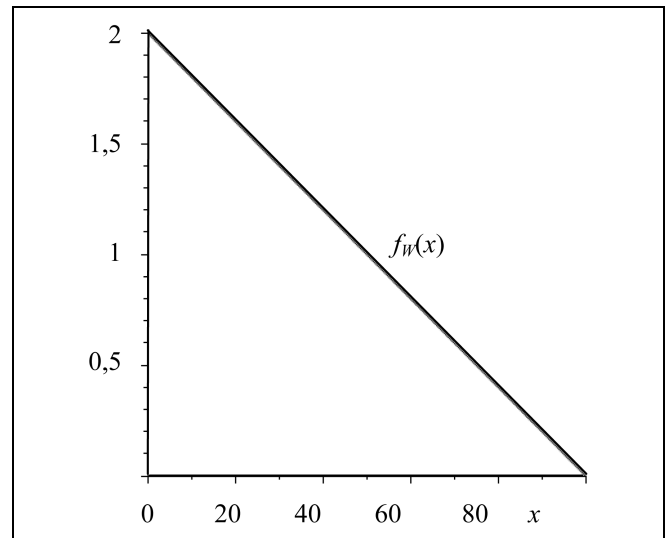


Рис. 3. Функция толерантности «белых» агентов

пропорции двух групп в своем окружении. Динамика процесса зависит также от относительной скорости реакции агентов разных групп на изменение общего знания.

Основные результаты. Простейшим случаем является линейная функция зависимости порога толерантности «белых» агентов f_W (см. рис. 2). Обозначим количество «белых» агентов через N_W , а максимальное значение порога толерантности отношения «черных» к «белым» агентам в своем окружении — через $R_{B/W}$. Количество «белых» агентов x , пороги которых не превышают величины $f_W(x)$, удовлетворяет линейному соотношению

$$f_W(x) = -\frac{R_{B/W}}{N_W}x + R_{B/W}.$$

Тогда величина $F_W(x) = x f_W(x)$ показывает количество «черных» агентов, которых «терпят» x «белых» агентов. График этой зависимости для $N_W = 100$ и $R_{B/W} = 2$ показан на рис. 3, где по оси абсцисс отложено количество «белых», а по оси ординат — «черных» агентов.

Для «черных» агентов введем такие же обозначения N_B , $R_{B/W}$, а функцию $f_B(x)$ будем считать также линейной. Построив на том же графике (рис. 4) функцию $F_B(x)$ для 50 «черных» агентов, где x изменяется по вертикальной оси от 0 до $N_B = 50$ и $R_{W/B} = R_{B/W} = 2$, получим точку S пересечения двух графиков (рис. 4). Эта точка и определяет равновесные пропорции между «черными» и «белыми» агентами в данном окружении.

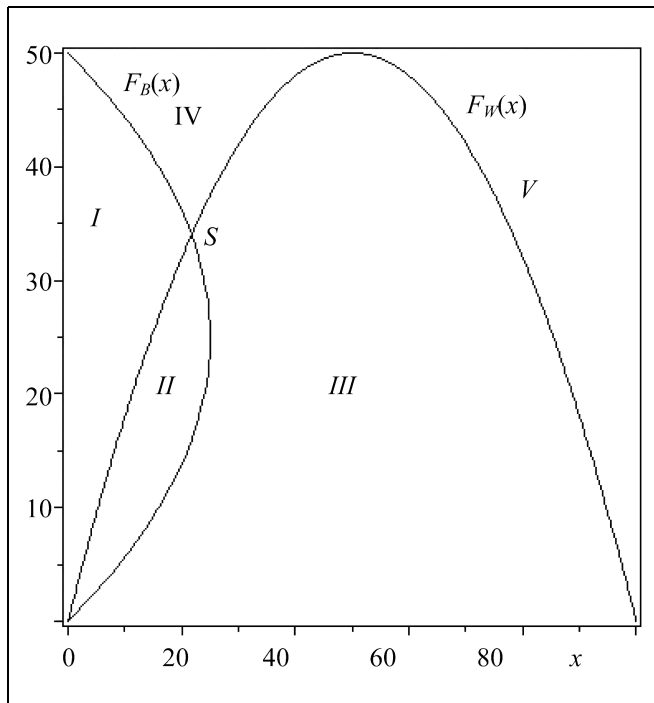


Рис. 4. Равновесие между «черными» и «белыми» агентами

Из областей *II* и *IV* (см. рис. 4) доли агентов сходятся со временем к *S*. Но положение равновесия *S* неустойчиво, так как при смещении из этой точки в области *I* и *III* возникают новые положения равновесия (0; 50) и (100; 0) соответственно. Последние два положения равновесия, т. е. когда во всем окружении присутствуют только «черные» или «белые» агенты, являются устойчивыми.

При других соотношениях между общим количеством «белых» и «черных» агентов могут появляться дополнительные устойчивые положения равновесия. Если имеют место внешние ограничения, а именно запрет на проживание в данном окружении «белых», «черных» или ограничение общего количества агентов, то дополнительные устойчивые положения равновесия также возникают.

Шеллинг показал, что в рассматриваемой модели существует несколько положений равновесия, в которые система «опрокидывается» («tipping») при превышении плотности агентов определенного типа некоторого критического значения.

В книге [12] Шеллинг развил и обобщил модель ограниченного окружения на другие примеры, такие как: коллективное поведение (посещение факультетских семинаров, волейбольных матчей, неосторожный переход улицы, присоединение к аплодисментам), ухудшение качества окружения в

районе проживания, сегрегация, решение суда без предварительного разбирательства, потеря доверия банкам, голосование, политические революции, дорожные знаки и переход на летнее время.

Модель 3 (пороговая модель М. Грановеттера). В работе [14] Грановеттер рассмотрел модели критической массы с несколько иной точки зрения. Его подход основывается на социологических исследованиях генезиса погромов и революций. В отличие от Шеллинга, Грановеттер исследовал устойчивость равновесия в поведении одного типа агентов.

Работа [14] стала отправной точкой для дальнейшей классификации порогового социального поведения (Threshold Models) [15, 64], а также для приложений этой модели к внедрению новых технологий [16], потребительскому спросу [17], сегрегации [18] и проблемам коллективного выбора [19].

Состав модели. Социальная группа состоит из *n* агентов. Любой агент может либо «действовать», либо «бездействовать». Для агента каждый из выборов действия или бездействия имеет свои положительные и отрицательные стороны. Агенты являются рациональными, а именно, исходя из своих целей и предпочтений, а также понимания ситуации, они действуют так, чтобы максимизировать свой выигрыш.

Основные параметры. Будем считать, что *выигрыш* агента зависит, с одной стороны, от действий (или бездействия) окружающих его агентов, т. е. от *социального фактора* — *давления группы*. С другой стороны — от его индивидуальных предпочтений, т. е. *индивидуального фактора* — *автономности* агента. В работе [14] такого рода поведение, зависящее от социального и индивидуального факторов, описывается в виде так называемой *модели порогового коллективного поведения*. Пусть индивидуальный фактор описывается числом *q*, принадлежащим отрезку [0; 1] (порогом), а социальный фактор *r* — долей действующих агентов.

Функционирование. Агент сравнивает значение своего индивидуального фактора (порога) со значением социального фактора — доли действующих агентов и принимает соответствующее решение — действовать или бездействовать: если $r < \theta$, то агент бездействует, если $r \geq \theta$, то агент действует. Обозначим функцию распределения порогов через $F_\theta(\cdot)$: $[0; 1] \rightarrow [0; 1]$, а долю действующих в момент времени *t* агентов через $r(t)$. По определению функции распределения, доля агентов с порогами, не превышающими $r(t)$, равна $F_\theta(r(t))$. Значит, в следующий момент времени будет действовать



именно эта доля агентов, т. е. справедливо рекуррентное соотношение:

$$r(t+1) = F_{\theta}(r(t)).$$

Положение равновесия характеризуется условием $r(t+1) = r(t)$, т. е. $r = F_{\theta}(r)$. Это положение равновесия может быть устойчивым или неустойчивым. Устойчивым оно будет, если график функции распределения $F_{\theta}(\cdot)$ пересекает биссектрису первого квадранта сверху. Если график функции распределения $F_{\theta}(\cdot)$ пересекает биссектрису снизу, то равновесие будет неустойчивым.

Основные результаты. Хотя формулировка функционирования модели включает доли действующих агентов и, соответственно, относительные значения порогов, в основных результатах Грановеттер использует абсолютные величины. Автор исследовал нормальную функцию распределения порогов для 100 агентов со средним значением 25:

$$F(x, \sigma) = \frac{100}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(y-25)^2}{2\sigma^2}} dy.$$

Зависимость точки равновесия от стандартного отклонения σ получилась разрывной. В точке $\sigma_c \approx 12,2$ существует «скачок» равновесия — с 6 агентов до 100. Этот разрыв объясняется тем, что при значении стандартного отклонения $\sigma_c \approx 12,2$ график функции распределения не пересекает биссектрису, а касается ее в точке $r = 6$ (рис. 5). Небольшое изменение значения стандартного отклонения переводит точку равновесия из $r = 6$ в $r = 100$. Таким образом, модель Грановеттера объясняет ситуации, когда две толпы, близкие по распределению порогов, могут вести себя совершенно по-разному.

Грановеттер почти не рассматривал случаи, когда некоторые из агентов являются друзьями, приводя только общие рассуждения [14] на эту тему. Социоматрица (дружеские связи между агентами) существенно изменяет поведение агентов, и тем самым намечается путь расширения области использования модели, который потом реализовал Чуэ [15, 65]. Другое ограничение модели состоит в том, что в ней не учитываются пространственные и временные ограничения для взаимодействий агентов.

Несмотря на эти ограничения, пороговая модель Грановеттера породила обширное направление исследований в области социального поведения. Она легко вписывается в класс моделей критической массы: в ней «критической массой»

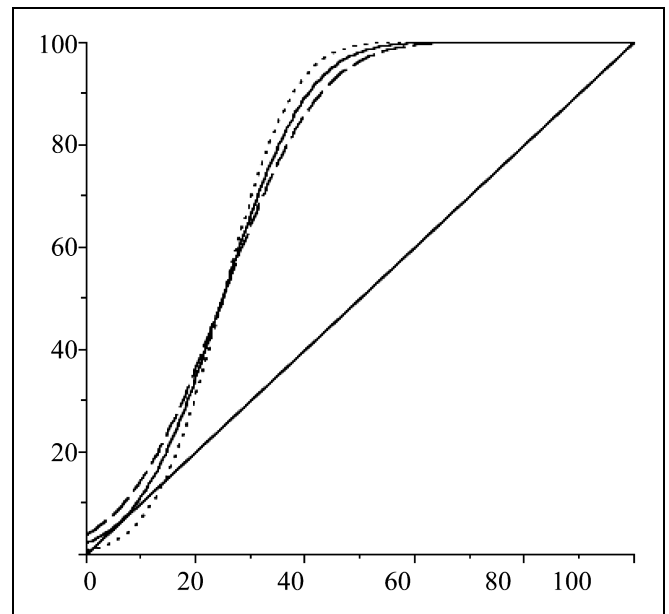


Рис. 5. Равновесия для различных значений стандартных отклонений порогов: ---- $\sigma = 10$; — $\sigma = 12,2$; - · - $\sigma = 14$; — x

будет, например, указанное критическое стандартное отклонение σ_c .

Содержательные интерпретации. Приведем соответствующие примеры порогового поведения — см. табл. 3.

Как видно из приведенных примеров, социальный фактор носит количественный характер (доля или число принявших решение о действии), тогда как индивидуальный фактор имеет различную (в том числе психологическую) природу.

В последующие годы модели 1–3 получили широкое развитие, которое продолжается и по настоящее время. Последователи попытались более подробно изучить социальный и индивидуальный факторы, наполняя их параметрами с содержательными интерпретациями. В результате применимость моделей значительно расширилась по сравнению с табл. 3, что будет проиллюстрировано в следующей части обзора.

ЗАКЛЮЧЕНИЕ

В первой части обзора рассмотрено, как широко распространенное явление конформного поведения рассматривалось различными науками. Оно не прошло мимо внимания философов, психологов и культурологов. Ими было показано, что в индивиде борются противодействующие психологические проявления поведения — следование действиям окружающей его социальной группе и

Примеры применимости модели порогового поведения [14]

Ситуация/действие	Факторы	
	Социальный r	Индивидуальный θ
Поведение при погромах и забастовках/принять участие	Кол-во забастовщиков	Боязнь неприятностей от властей
Распространение слухов/стать распространителем	Кол-во знакомых, сказавших о новости, разделяющих критическое мнение и т. п.	Уровень доверия к чужому мнению
Голосование за кандидатов/проголосовать "за"	Кол-во проголосовавших "за"	Подверженность влиянию других. Нежелание потерять голос
Выбор профессии	Кол-во выпускников, выбравших данную профессию	Готовность к предстоящей конкуренции на рынке труда. Индивидуальные предпочтения
Уход с просмотра фильма	Кол-во уже ушедших с просмотра	Вежливость по отношению к создателям фильма (даже если их в зале нет!) или перед сидящими рядом, которых нужно побеспокоить
Эмиграция/эмигрировать	Поток уезжающих граждан из данной страны	Склонность к перемене привычной жизни. Требования к уровню комфорта, свободе и т. п.

подчинение своему индивидуальному видению ситуации. Приведено несколько психологических экспериментов, которые иллюстрируют инстинктивную глубину конформного поведения. Рассмотрены математические модели Шеллинга и Грановеттера, которые дали начало описанию этого явления, нашедшего продолжение в последующих моделях критической массы и пороговых моделях.

В следующей части обзора будут рассмотрены более поздние математические модели и, в соответствии с их классификацией, приведены области их приложений.

ЛИТЕРАТУРА

1. Кант И. Антропология с прагматической точки зрения: Собр. соч. в 8 т. — М.: Чоро, 1994. — Т. 7.
2. Ницше Ф. Веселая наука. — СПб.: Азбука-классика, 2006.
3. Юнг К.Г. Психологические типы. — СПб.: Азбука, 2001.
4. Элиаде М. Миф о вечном возвращении. — М.: Ладомир, 2000.
5. Sherif M. The psychology of social norms. — N.-Y.: Harper Collins, 1936.
6. Baron R.S., Vandello J.A., & Brunsman B. The forgotten variable in conformity research: Impact of task importance on social influence // Journal of Personality and Social Psychology. — 1996. — Vol. 71. — P. 915–927.
7. Asch S.E. Opinions and social pressure // Scientific American. — 1955. — Vol. 193, N 5. — P. 31–35.
8. Asch S.E. Studies of independence and conformity: A minority of one against a unanimous majority // Psychological Monographs. — 1956. — Vol. 70, N 9 (whole no. 416). — P. 1–70.
9. Schelling T. A process of residential segregation: Neighborhood tipping, in Racial Discrimination / in: A. Pascal (Ed.), Economic Life, Lexington, MA: Lexington Books, 1972. — P. 157–184.
10. Schelling T. Hockey Helmets, Concealed Weapons, and Daylight Saving: A Study of Binary Choices with Externalities // The Journal of Conflict Resolution. — 1973. — Vol. 17, N 3. — P. 381–428.
11. Schelling T. Dynamic models of segregation // Journal of Mathematical Sociology. — 1971. — Vol. 1. — P. 143–186.
12. Schelling T. Micromotives and Macrobehavior. — N.-Y.: WW Norton & Co, 1978.
13. Schelling T. The Strategy of Conflict. — Oxford: Oxford Univ. Press, 1960.
14. Granovetter M. Threshold Models of Collective Behavior // The American Journal of Sociology. — 1987. — Vol. 83, N 6. — P. 1420–1443.
15. Chwe M. Structure and Strategy in Collective Action // American Journal of Sociology. — 1999. — Vol. 105, N 1. — P. 128–157.
16. Granovetter M., Soong R. Threshold models of diffusion and collective behavior // Journal of Mathematical Sociology. — 1983. — N 9. — P. 165–179.
17. Granovetter M., Soong R. Threshold models of interpersonal effects in consumer demand // Journal of Economic Behavior and Organization. — 1986. — N 7.
18. Granovetter M., Soong R. Threshold models of diversity: Chinese restaurants, residential segregation, and the spiral of silence // Sociological Methodology. — 1988. — Vol. 18. — P. 69–104.
19. Macy M. Chains of cooperation: threshold effects in collective action // American Sociological Review. — 1991. — Vol. 56, N 6. — P. 730–747.
20. Килошниченко М.И. Психология моды: теоретический и прикладной аспекты. — СПб.: СПбГУТ, 2001.
21. Орлова Л.В. Азбука моды. — М.: Просвещение, 1989.
22. Brock W., Durlauf S. A Multinomial Choice Model of Neighborhood Effects // American Economic Review. — 2002. — Vol. 92, N 2. — P. 298–303.



23. *Ницше Ф.* Генеалогия морали. — СПб.: Азбука-классика, 2010.
24. *Ницше Ф.* По ту сторону добра и зла: Сочинения. — М.: Эксмо; Харьков: Фолио, 2006.
25. *Axelrod R., Tesfatsion L.* A Guide for Newcomers to Agent-Based Modelling in the Social Sciences // *Handbook of Computational Economics*. — 2006. — Vol. 2.
26. *Banerjee A., Besley T.* Peer Group Externalities and Learning Incentives: A Theory of Nerd Behavior, John M. Olin Discussion Paper no. 68. Princeton. — N.-J.: Princeton Univ., December 1990.
27. *Akerlof G.A.* The Market for «Lemons»: Quality Uncertainty and the Market Mechanism // *The Quarterly Journal of Economics*. — 1970. — Vol. 84, N 3. — P. 488–500.
28. *Keynes J.M.* The General Theory of Employment, Interest and Money. — London: Macmillan, 1936.
29. *Aronson E., Wilson T.D., & Aker A.M.* Social Psychology. Upper Saddle River. — N.-J.: Pearson Prentice Hall, 2007.
30. *Kelman H.* Compliance, identification, and internalization: Three processes of attitude change // *Journal of Conflict Resolution*. — 1958. — Vol. II, N 1. — P. 51–60.
31. *Bikhchandani S., Hirshleifer D., Welch I.* A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades // *Journal of Political Economy*. — 1992. — Vol. 100, iss. 5. — P. 992–1026.
32. *Лебон Г.* Психология народов и масс. — М.: Макет, 1995.
33. *Macy M., Willer R.* From Factors to Actors: Computational Sociology and Agent-Based Modeling // *Annual Review of Sociology*. — 2002. — Vol. 28. — P. 143–166.
34. *Squazzoni F.* The Micro-Macro Link in Social Simulation // *Sociologica*. — 2008. — N 1.
35. *Levin K.* Field theory in social science. — N.-Y.: Harper, 1951. — 422 p.
36. *Becker G. and Murphy K.M.* Social Markets: Market Behavior in a Social Environment. — Cambridge: Belknap-Harvard University Press, 2001.
37. *Cooper J., John A.* Coordination failures in Keynesian models // *Quarterly Journal of Economics*. — 1988. — Vol. 103, N 3. — P. 441–464.
38. *Deffuant G., Huet S., Amblard F.* An Individual-Based Model of Innovation Diffusion Mixing Social Value and Individual Benefit // *American Journal of Sociology*. — 2005. — Vol. 110, N 4. — P. 1041–1069.
39. *Durlauf S.N.* A Framework for the Study of Individual Behavior and Social Interactions // *Sociological Methodology*. — 2001. — Vol. 31, iss. 1. — P. 47–87.
40. *Glaeser E., Sacerdote B., Scheinkman J.* Crime and social interactions // *Quarterly Journal of Economics*. — 1996. — Vol. CXI. — P. 507–548.
41. *Katz M., and Shapiro C.* Technology Adoption in the Presence of Network Externalities // *Journal of Political Economy*. — 1986. — Vol. 94. — P. 822–841.
42. *Follmer H., Horst U., and Kirman A.* Equilibria in financial markets with heterogeneous agents: A probabilistic perspective // *Journal of Mathematical Economics*. — 2004. — Vol. 41 (1–2). — P. 123–155.
43. *Horst U.* Financial price fluctuations in a stock market model with many interacting agents // *Econ. Theory*. — 2005. — Vol. 25, N 4. — P. 917–932.
44. *Губко М.В., Новиков Д.А.* Теория игр в управлении организационными системами. — М.: СИНТЕГ, 2002. — 148 с.
45. *Myerson R.* Game theory: analysis of conflict. — Cambridge—Massachusetts—London: Harvard University Press, 2001.
46. *Jackson M.* Social and Economic Networks. — Princeton: Princeton University Press, 2008.
47. *Glaeser E., Scheinkman J.* Non-Market Interactions. — Harvard Institute of Economic Research Working Papers number 1914, 2002.
48. *Bernheim D.* A Theory of Conformity // *Journal of Political Economy*. — 1994. — Vol. 102, N 5.
49. *Jones S.* The Economics of Conformism. — Oxford: Blackwell, 1984.
50. *Плотинский Ю.М.* Модели социальных процессов: Учебное пособие для высших учебных заведений. — Изд. 2-е, перераб. и доп. — М.: Логос, 2001. — 296 с.
51. *Краснощев П.С.* Некоторые результаты математического моделирования одного механизма коллективного поведения // *Социология*: 4М. — 1994. — № 3–4. — С. 65–83.
52. *Краснощев П.С.* Простейшая математическая модель поведения. Психология конформизма // *Математическое моделирование*. — 1998. — Т. 10, № 7. — С. 76–92.
53. *Гаврилец Ю.Н., Фомина Ю.В.* Моделирование динамики социальной установки (на примере отношения студентов к рекламе на телевидении) // *Социология*. — 2002. — № 15. — С. 110–134.
54. *Benito J.M., Hernandez P.* Schelling's Dynamic Models of Segregation: A cellular automata approach (2004). — URL: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.87.1841> (дата обращения 8.11.2013).
55. *Conlisk J.* Costly Optimizers versus Cheap Imitators // *Journal of Econ. Behavior and Organization*. — 1980. — Vol. 1, iss. 3. — P. 275–293.
56. *Kandori M., Mailath G., Rob R.* Learning, Mutation, and Long Run Equilibria in Games // *Econometrica*. — 1993. — Vol. 61, N 1. — P. 29–56.
57. *Kempe D., Kleinberg J., Tardos E.* Influential Nodes in a Diffusion Model for Social Networks. Automata, Languages and Programming In Automata // *Languages and Programming*. — 2005. — Vol. 3580. — P. 1127–1138.
58. *Oliver P.* Formal Models of Collective Action // *Annual Review of Sociology*. — 1993. — Vol. 19. — P. 271–300.
59. *Oliver P., Marwell G., Teixeira R.* A Theory of the Critical Mass. I. Interdependence, Group Heterogeneity, and the Production of Collective Action // *The American Journal of Sociology*. — 1985. — Vol. 91, N 3. — P. 522–556.
60. *Oliver P., Marwell G.* The Paradox of Group Size in Collective Action: A Theory of the Critical Mass. II // *American Sociological Review*. — 1988. — Vol. 53, N 1. — P. 1–8.
61. *Oliver P., Marwell G., Pahl R.* Social Networks and Collective Action: A Theory of the Critical Mass. III // *The American Journal of Sociology*. — 1988. — Vol. 94, N 3. — P. 502–534.
62. *Васин А.А.* Эволюционные и повторяющиеся игры: Учебное пособие. — М.: Российская экономическая школа, 2005. — 74 с.
63. *Новиков Д.А., Чхартишвили А.Г.* Рефлексия и управление: математические модели. — М.: Физматлит, 2012. — 420 с.
64. *Xue J.* Collective Behavior with Endogenous Thresholds. Faculty of Economics / University of Cambridge in its series Cambridge Working Papers in Economics number 0613. — 2006.
65. *Chwe M.* Communication and Coordination in Social Networks // *Review of Economic Studies*. — 2000. — Vol. 67, N 1. — P. 1–16.

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

Бреер Владимир Валентинович — бизнес-аналитик, ЗАО «АВИАХЭЛП ГРУПП», г. Москва, breer@live.ru.

ГАРАНТИРОВАННЫЕ РИСКИ И ИСХОДЫ В «ИГРЕ С ПРИРОДОЙ»

В.И. Жуковский, Н.Г. Солдатова

Рассмотрена однокритериальная задача при неопределенности, причем о последней известны только границы изменения, а какие-либо вероятностные характеристики по тем или иным причинам отсутствуют. Предложено два понятия гарантированного одновременно по исходам и рискам решения. Первый базируется на способе принятия решения в двухуровневой иерархической игре двух лиц, второй — на определении векторного максимина. Найден явный вид решения при линейно-квадратичном виде критерия.

Ключевые слова: стратегия, неопределенность, критерий, минимаксное сожаление, максимин, векторный оптимум.

ВВЕДЕНИЕ

В настоящей статье рассматривается однокритериальная задача $\Gamma^{(1)} = \langle X, Y, f(x, y) \rangle$ в условиях риска и неопределенности (игра с природой). В рамках этой задачи лицо, принимающее решение (ЛПР), выбирает свою стратегию $x \in X \subseteq R^n$ так, чтобы достичь возможно большего исхода (значения скалярного критерия $f(x, y)$), ориентируясь на реализацию любой чистой неопределенности $y \in Y \subseteq R^m$. Предполагается, что о неопределенностях ЛПР известны лишь границы изменения и отсутствуют какие-либо вероятностные характеристики. Такая модель $\Gamma^{(1)}$ возникает, например, на рынке сбыта, где продавец действует с учетом импорта (или конкуренции), добываясь как можно большей прибыли.

Заметим, что подробные обзоры различных видов неопределенности (неполноты и (или) неточности информации об условиях реализации выбранной стратегии) можно найти, например, в книгах [1, с. 106–114; 2, с. 20–32, и др.].

Наличие неопределенностей приводит к множественности исходов $f(x, Y) = \{f(x, y) | \forall y \in Y\}$, «порожденных» каждой конкретной стратегией $x \in X$. «Сужается» $f(x, Y)$ за счет рисков. Однако, как считает известный специалист в области оптимизации Т.К. Сиразетдинов, «строгого математического определения риска в настоящее время не

существует» [3, с. 31]. В книге [4, с. 15] приводится целая серия различных понятий риска. Все из них, кроме приводимого далее, требуют статистических данных. Однако зачастую у исследователя операций (ИО) просто отсутствует возможность описать «поведение» неопределенностей статистическими методами. Как раз этого случая будем в дальнейшем придерживаться.

Итак, приведем определение: «Риск — это возможность отклонения каких-либо величин от их желаемых значений».

Отметим, что именно такому понятию риска отвечают общепринятые многочисленные микроэкономические риски, вид которых приведен в работе [5, с. 40–50].

Численно оценивается риск значением функции сожаления

$$\Phi(x, y) = \max_{z \in X} f(z, y) - f(x, y), \quad (1)$$

предложенной Леонардом Сэвиджем [6] в 1951 г. Лауреат Нобелевской премии по экономике Милтон Фридман сказал о Сэвидже, что тот «... был одним из немногих встреченных мной людей, о которых я, не задумываясь, могу сказать — гений». Предложенный в работе [6] принцип минимаксного сожаления, сводящийся к построению пары (x^S, Φ^S) согласно

$$\min_{x \in X} \max_{y \in Y} \Phi(x, y) = \max_{y \in Y} \Phi(x^S, y) = \Phi^S,$$



активно используется для решения задачи $\Gamma^{(1)}$ наравне с принципом максимина (гарантированного результата по Вальду [7]), сводящемуся к нахождению пары (x^g, f^g) такой, что

$$\max_{x \in X} \min_{y \in Y} f(x, y) = \min_{y \in Y} f(x^g, y) = f^g.$$

Сама функция сожаления $\Phi(x, y)$ в отечественной и мировой литературе получила название «функция риска по Сэвиджу». Именно функцию риска $\Phi(x, y)$ и привлекаем в настоящей статье для оценки гарантированного риска.

Каким бывает отношение людей к риску? В ряде книг по финансовой экономике [1, с. 103; 5, с. 5; 8, с. 343] выделено три группы субъектов в зависимости от отношения их к риску:

- противники риска — рискофобы (люди, боящиеся риска и отвергающие его);
- рисконейтралы (люди, нейтрально относящиеся к риску);
- любители риска — рискофилы.

В экономике считается, что большинство людей относятся к противникам риска. На вопрос о том, как фактор неопределенности влияет на поведение людей, экономист обычно отвечает: «Люди не любят рисковать и готовы заплатить деньги за то, чтобы избежать бремени риска» [5, с. 6].

Однако возникают ситуации, когда риск просто необходим. Люди прошлого выходили в море, что часто было связано с риском для жизни. Существует даже латинская пословица: «Плывать по морю необходимо, жить — не очень». Так любители риска относятся и к альпинизму, авиации, экстремальным ситуациям. Более того, предпринимательство и риск — понятия неразделимые. В экономической практике принято, что некоторая доля риска является необходимым условием увеличения дохода. Зачастую возникают ситуации, когда без риска вообще обойтись невозможно (например, в чрезвычайных ситуациях).

Наконец, значительное большинство относится к рисконейтралам. Они будут пускаться пусть даже и в рискованные ситуации, но в том только случае, если доход будет выглядеть достаточно привлекательным и одновременно, чтобы возможно меньше нужно было бы рисковать.

В соответствии с приведенной градацией, работы [9, 10] и глава 3 из книги [11] посвящены исследованию бескоалиционных игр с позиции противников риска; те же игры, но с позиции любителей риска — в работе [12]; взгляд рисконейтрала на принятие решений в $\Gamma^{(1)}$ — в этой статье. Принятый здесь подход предполагается в дальнейшем распространить на бескоалиционные и кооперативные игры при неопределенности. Именно с

этой целью в статье привлечены некоторые положения теории «игр с приоритетом в действиях у управляющего центра, получивших название иерархических игр Гермейера» [13, с. 8].

Итак, цели настоящей работы:

- формализовать гарантированное решение задачи $\Gamma^{(1)}$ с *одновременным* учетом исходов и рисков (предложено два варианта понятия);
- установить существование при ограничениях, «привычных» для математического программирования;
- найти явный вид гарантированных решений в случае линейно-квадратичного варианта критерия $f(x, y)$ при ограниченных чистых неопределенностях.

1. ИНФОРМИРОВАННЫЕ НЕОПРЕДЕЛЕННОСТИ

Иерархическая игра представляет собой «математическую модель конфликтной ситуации при фиксированной последовательности ходов и обменом информацией участников» [14, с. 477]. Активное развитие теории иерархических игр в России началось со второй половины прошлого века и возглавлялось Юрием Борисовичем Гермейером ([13—17] и др.), продолжается его учениками. В игре двух лиц «такие игры описывают взаимодействие между верхним (ведущим) и нижним (ведомым) уровнями управления» [17, с. 103], именно, задают порядок ходов игроков, т. е. очередность выбора стратегий и (возможно) сообщение о таком выборе партнеру.

Основной момент в иерархических играх заключается в выборе класса используемых стратегий, зависящем от имеющейся у игроков информации. В теории иерархических игр Гермейера сформулировано точное математическое определение информационного расширения игры [14, с. 479; 15, с. 49—51], которое, в частном случае, приводит к использованию в задаче $\Gamma^{(1)}$ наряду с чистыми неопределенностями $y \in Y$ так называемых, «информированных неопределенностей» — m -вектор-функций $y(x): X \rightarrow Y$. Именно такие стратегии применялись в работе [18, с. 353] при изучении детерминированного варианта минимаксной антагонистической позиционной игры, в которой игроки наделены различными информационными возможностями. Такие возможности определяют соответствующие виды стратегий, что приводит, в свою очередь, к различным видам иерархических игр (Γ_1 , Γ_2 и т. д.) [13, 15, 17].

Наконец, в теории иерархических игр принято выделять *оперирующую сторону* — ЛПР, несущего полную ответственность за результаты, и *исследо-*

вателя операций — консультанта, который готовит аргументированные варианты решений.

При рассмотрении задачи $\Gamma^{(1)}$ будем считать, что один игрок (у нас ИО) ограничен только чистыми стратегиями $x \in X$, другой же может использовать «любую мыслимую информацию» [18, с. 353]. В частности, он может знать стратегию x (информационная дискриминация ИО) и формировать неопределенность в виде функции $y(x): X \rightarrow Y$. В этом случае критерий в задаче $\Gamma^{(1)}$ определяется скалярной функцией $f(x, y(x))$, а исходом будет (при выборе ИО конкретной стратегии $x^* \in X$) значение $f(x^*, y(x^*))$. Такие функции $y(\cdot) \in Y^X$ (множеству m -вектор-функций $y(x)$, определенных на X со значениями в Y) в теории дифференциальных игр иногда называют *контрстратегиями*, а задача вида $\Gamma^{(1)}$, где в качестве неопределенностей используются контрстратегии $y(x)$, названа в работе [18, с. 354] *минимаксной игрой*. Повторим, что такие задачи возникают при информационной дискриминации ИО и дополнительной информированности игрока, «ведущего» формированием неопределенностей. Заметим также, что далее будем применять подмножество Y^X , именно множество $C(X, Y)$ всех покомпонентно непрерывных на X m -вектор-функций $y(x): X \rightarrow Y$.

Итак, в статье используется два вида неопределенностей: чистые $y \in Y$ и информированные $y(\cdot) \in Y^X$.

Приведем два результата из теории исследования операций, касающиеся «информированных неопределенностей».

Лемма 1. Если в $\Gamma^{(1)} = \langle X, Y, f(x, y) \rangle$ множества X, Y суть компакты, а $f(x, y)$ непрерывна на $X \times Y$, то:

а) функция максимума (минимума) $\max_{x \in X} f(x, y)$ (соответственно, $\min_{y \in Y} f(x, y)$) непрерывна на Y (соответственно на X);

б) если дополнительно Y — выпукло и $f(x, y)$ строго выпукла по $y \in Y$ при каждом $x \in X$, то существует единственная непрерывная функция $y(\cdot) \in C(X, Y)$ такая, что

$$\min_{y \in Y} f(x, y) = f(x, y(x)) \quad \forall x \in X. \quad \blacklozenge$$

Напомним, что $f(x, y)$ строго выпукла по $y \in Y$ при каждом $x \in X$, если

$$f(x, \lambda y^{(1)} + (1 - \lambda)y^{(2)}) < \lambda f(x, y^{(1)}) + (1 - \lambda)f(x, y^{(2)})$$

при любых постоянных $\lambda \in (0, 1)$ и всяких $y^{(j)} \in Y$, $j = 1, 2$, $y^{(1)} \neq y^{(2)}$.

Утверждение *a* — известный факт, имеющийся во многих учебных книгах, например, [19, с. 146], а справедливость утверждения *б* указана, например, в книге [20, с. 54].

Замечание 1. Из леммы утверждения *a* следует непрерывность на $X \times Y$ функции риска (1) (если только в задаче $\Gamma^{(1)}$ множества X, Y суть компакты, а $f(x, y)$ непрерывна на $X \times Y$).

2. ДВУХКРИТЕРИАЛЬНАЯ ЗАДАЧА, СООТВЕТСТВУЮЩАЯ ЗАДАЧЕ $\Gamma^{(1)}$

Однокритериальной задаче

$$\Gamma^{(1)} = \langle X, Y, f(x, y) \rangle \quad (2)$$

поставим в соответствие двухкритериальную при неопределенности

$$\Gamma = \langle X, Y, F(x, y) \rangle, \quad (3)$$

где двухкомпонентная вектор-функция

$$F(x, y) = (F_1(x, y), F_2(x, y)), \quad F_1(x, y) = f(x, y), \quad F_2(x, y) = -\Phi(x, y). \quad (4)$$

В задаче (3) ИО выбором стратегии $x \in X$ стремится к возможно *большим* значениям обоих исходов $F_i(x, y)$, $i = 1, 2$, одновременно (именно для такого однообразия в выражении (4) функция риска по Сэвиджу фигурирует со знаком «минус»). При этом ИО учитывает возможную реализацию любой неопределенности $y \in Y$ (или $y(\cdot) \in Y^X$).

Приведем ряд сведений из теории многокритериальных задач вида $\Gamma^{(2)} = \langle X, F[x] \rangle$, где $x \in X$ (множеству стратегий x у ИО), вектор критериев $F[x] = (F_1[x], F_2[x])$ определен на X . Используем для двух векторов $F^{(j)} = (F_1^{(j)}, F_2^{(j)})$, $j = 1, 2$, отношения *строгого порядка*:

$$F^{(1)} < F^{(2)} \Leftrightarrow (F_i^{(1)} < F_i^{(2)}, i = 1, 2),$$

$$F^{(1)} \not< F^{(2)} \Leftrightarrow \neg (F^{(1)} < F^{(2)}),$$

и *нестрогого порядка*:

$$F^{(1)} = F^{(2)} \Leftrightarrow (F_i^{(1)} = F_i^{(2)}, i = 1, 2),$$

$$F^{(1)} \neq F^{(2)} \Leftrightarrow \neg (F^{(1)} = F^{(2)}),$$

$$F^{(1)} \geq F^{(2)} \Leftrightarrow (F_i^{(1)} \geq F_i^{(2)}, i = 1, 2),$$

$$F^{(1)} \not\geq F^{(2)} \Leftrightarrow \neg (F^{(1)} \geq F^{(2)} \wedge F^{(1)} \neq F^{(2)}).$$



Перейдем к формализации двух максимальных векторных оптимумов:

1) стратегия $x^S \in X$ называется *максимальной по Слейтеру* в $\Gamma^{(2)} = \langle X, F[x] \rangle$, если

$$F[x^S] \not\leq F[x] \quad \forall x \in X,$$

вектор $F[x^S]$ является *максимумом по Слейтеру* в $\Gamma^{(2)}$ (что эквивалентно: для каждой стратегии $x \in X$ найдется хотя бы один номер $j(x) = j \in \{1, 2\}$ такой, что $F_j(x) \leq F_j(x^S)$);

2) стратегия $x^P \in X$ называется *максимальной по Парето* в $\Gamma^{(2)}$, если

$$F[x^P] \not\leq F[x] \quad \forall x \in X,$$

а вектор $F[x^P] \in R^2$ есть *максимум по Парето* для $\Gamma^{(2)}$ (что эквивалентно любому из двух определений — для каждого $x \in X$:

а) либо $F[x^P] = F[x]$, либо $\exists j(x) = j \in \{1, 2\}$ такой, что $F_j[x] < F_j[x^P]$;

б) несовместна система неравенств $F_i[x] \geq F_i[x^P]$, $i = 1, 2$, из которых, по крайней мере, одно строгое).

Обозначим множество $x^S(x^P)$ через X^S (соответственно, X^P). Согласно определениям $X^P \subseteq X^S$, но они могут не совпадать. Факт максимальности в $\Gamma^{(2)}$ по Слейтеру (по Парето) обозначаем

$$F[x^S] = \text{MAX}_{x \in X}^S F[x] \quad (F[x^P] = \text{MAX}_{x \in X}^P F[x]).$$

Будем использовать и множества $F[X^S] = \{F[x] \mid x \in X^S\}$, $F[X^P] = \{F[x] \mid x \in X^P\}$. В теории многокритериальных задач [21, с. 158] установлены следующие факты:

Лемма 2 [21, с. 158]. Если в $\Gamma^{(2)} = \langle X, F[x] \rangle$ множество X есть непустой компакт, а компоненты вектора $F[x]$ непрерывны, то множество $X^P \neq \emptyset$ и стратегия $x^P \in X$, найденная из

$$\max_{x \in X} (\alpha_1 F_1[x] + \alpha_2 F_2[x]) = \alpha_1 F_1[x^P] + \alpha_2 F_2[x^P]$$

при каких-либо $\alpha_i = \text{const} > 0$, $i = 1, 2$, *максимальна по Парето* в $\Gamma^{(2)}$, множество X^P *внутренне P-устойчиво*, т. е. $\forall x^{(j)} \in X^P$ имеет место $F[x^{(1)}] \not\leq F[x^{(2)}]$, а также *внешне P-устойчиво*, т. е. $\forall x \in X$ и $x \notin X^P$ существует стратегия $x^P \in X^P$ такая, что $F[x] \leq F[x^P]$.

3. СТРОГО ГАРАНТИРОВАННОЕ ПО ИСХОДАМ И РИСКАМ РЕШЕНИЕ

3.1. Интерпретация максимина «с позиции» двухуровневой иерархической игры двух лиц

Максиминное решение (x^g, f^g) задачи (2) определяется цепочкой равенств

$$\max_{x \in X} \min_{y \in Y} f(x, y) = \min_{y \in Y} f(x^g, y) = f^g. \quad (5)$$

Используя «информированные неопределенности», выражение (5) можно представить как последовательное «действие» двух операций: *внутреннего минимума* — для игрока нижнего уровня — построение $y(x): X \rightarrow Y$ такого, что

$$\min_{y \in Y} f(x, y) = f(x, y(x)) = f[x] \quad \forall x \in X; \quad (6)$$

предполагая, что вектор-функция $y(x)$ единственна, переходим к операции *внешнего максимума* (для игрока верхнего уровня иерархии)

$$\max_{x \in X} f(x, y(x)) = f(x^g, y(x^g)) = f^g. \quad (7)$$

Тогда в иерархической двухуровневой игре с одним игроком на каждом уровне:

первый ход за ИО — игроком верхнего уровня: он передает на нижний уровень «свои» возможные стратегии $x \in X$;

второй ход за игроком нижнего уровня — он аналитически конструирует $y(x)$ согласно (6) и, если $y(x)$ единственно, передает $y(x)$ на верхний уровень;

третий ход за игроком верхнего уровня — он находит пару (x^g, f^g) согласно (7).

Приведенное «трехходовое понятие» укладывается *полностью* в определение гарантированного результата первого (ведущего) игрока в игре $\Gamma^{(1)}$ (по Гермейеру), если в работе [17, с. 104] заменить функцию выигрыша ведомого на $-f(x, y)$. Нетрудно видеть также, что находясь в рамках игры $\Gamma^{(1)}$, ведущий игрок, если знает *правило поведения ведомого*, может сам вычислить реакцию ведомого и сразу реализовать третий ход. Еще раз подчеркнем, что аналог и модификацию такого «трехходового понятия» удобно применять к построению гарантированного решения с учетом исходов и рисков для бескоалиционного и кооперативного вариантов конфликта.

Замечание 2. Максиминное решение определяется парой (x^g, f^g) по двум причинам:

а) каждой стратегии $x \in X$ (в результате операции *внутреннего минимума* (6)) ставится в соответствие гарантия $f[x]$, ибо

$$f[x] \leq f(x, y) \quad \forall y \in Y$$

(так как исход $f[x]$ «обеспечивает себе» ИО при любых $y \in Y$ благодаря применению стратегии x);

б) из таких гарантий ЛПР выбирает наибольшую (максимальную), ибо

$$f^g = f[x^g] \geq f[x] \quad \forall x \in X.$$

Итак, ЛПР предлагается применить в задаче (2) стратегию x^g , тем самым «обеспечивая себе» наибольшую (максимальную) гарантию $f[x^g] = f(x^g, y(x^g)) \leq f(x^g, y) \quad \forall y \in Y$. Этот же прием применим при формализации сильно гарантированного по исходам и рискам решения (СГР) задачи (3), (4).

3.2. Формализация

Определение 1. Пару $(x^P, F^g = F[x^P]) \in X \times R^2$ назовем *сильно гарантированным по исходам и рискам решением* однокритериальной задачи (2), если в задаче (3):

1°) существует для каждого $x \in X$ и $i = 1, 2$ «своя» гарантия

$$F_i[x] = \min_{y \in Y} F_i(x, y) \leq F_i(x, y) \quad \forall y \in Y;$$

2°) стратегия x^P максимальна по Парето ($x^P \in X^P$) в двухкритериальной «задаче гарантий» $\langle X, F[x] \rangle$.

Замечание 3. Отметим, что:

— в случае компактности X , Y , выпуклости Y , непрерывности $f(x, y)$ на $X \times Y$ и ее строгой выпуклости по y при каждом $x \in X$ такое гарантированное решение существует (что сразу следует из леммы 1, п. а и леммы 2), причем при использовании ИО стратегии $x \in X$ значение $F_1[x]$ будет гарантией по исходам ($F_1[x] \leq F_1(x, y) \quad \forall y \in Y$), а $F_2[x]$ — гарантией по рискам, ибо $F_2[x] \leq F_2(x, y) \quad \forall y \in Y$;

— п. 1° определения 1 связывает с каждой стратегией $x \in X$ векторную гарантию $F[x] \leq F(x, y) \quad \forall y \in Y$ (аналог операции внутреннего минимума в определении максимина);

— п. 2° предлагает лицу, принимающему решение, применять максимальную (по Парето) гарантию $F[x^P]$ (аналог операции внешнего максимума), ибо при $x \in X$ и $x \neq x^P$ увеличение одной из гарантий $F_j[x] > F_j[x^P]$ неизбежно влечет уменьшение другой $F_k[x] < F_k[x^P]$, $k \neq j \in \{1, 2\}$;

— вследствие внешней и внутренней Р-устойчивости множества X^P , выбор лицом, принимаю-

щим решение, стратегии $x^P \in X^P$ «обеспечивает» ему «самую большую» — неулучшаемую векторную гарантию $F[x^P]$ (конечно, в рамках максимальнойности по Парето);

— гарантия $F_1[x]$ ограничивает исходы $f(x, y)$ снизу, ибо $f(x, y) \geq F_1[x] \quad \forall y \in Y$, а риски $\Phi(x, y)$ — сверху, так как $\Phi(x, y) \leq -F_2[x] \quad \forall y \in Y$. Этот подход полностью соответствует желаниям рисконейтрала (см. Введение) увеличить исход и одновременно уменьшить риск.

3.3. «Иерархическая» интерпретация определения 1

Как и в п. 3.1, рассматриваем двухуровневую игру с одним игроком на каждом уровне (рис. 1).

Порядок построения СГР

Первый ход за игроком верхнего уровня: он, также как в понятии максимина, посылает на нижний уровень свои возможные стратегии $x \in X$.

Второй ход за игроком нижнего уровня; он аналитически конструирует две функции $F_i[x]$, $i = 1, 2$, согласно

$$F_i[x] = \min_{y \in Y} F_i(x, y) \quad \forall x \in X,$$

строая тем самым для каждой стратегии $x \in X$ векторную гарантию $F[x] = (F_1[x], F_2[x])$, и отправляет векторную гарантию $F[x]$ на верхний уровень иерархии.

Третий ход снова за игроком верхнего уровня: он находит максимальную по Парето стратегию x^P в двухкритериальной «задаче гарантий» $\langle X, F[x] \rangle$ и строит соответствующий вектор $F^g = (F_1[x^P], F_2[x^P])$. Тогда пара (x^P, F^g) и образует строго гарантированное по исходам и рискам решение задачи (2). «Двухкритериальный смысл» такого решения см. в замечании 3.

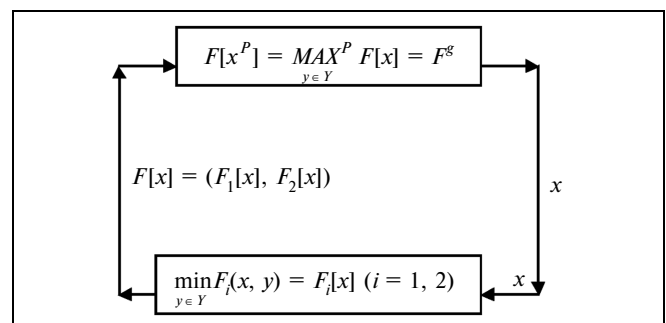


Рис. 1. Порядок построения СГР



3.4. Существование

Существование (x^P, F^g) устанавливает

Утверждение 1. Если в задаче (2) множества X и Y компактны (замкнуты и ограничены в соответствующих евклидовых пространствах), а критерий $f(x, y)$ непрерывен на $X \times Y$, то в однокритериальной задаче (2) существует сильно гарантированное по исходам и рискам решение.

Доказательство. Из непрерывности $f(x, y)$ и $\Phi(x, y) = \max_{z \in X} f(z, y) - f(x, y)$ на $X \times Y$, а также компактности X, Y и леммы 1, п. а следует существование непрерывных на X функций $F_i[x] = \min_{y \in Y} F_i(x, y)$, $i = 1, 2$. Поэтому линейная свертка $\Psi(x) = \alpha_1 F_1[x] + \alpha_2 F_2[x]$ при, например, $\alpha_1 = \alpha_2 = 1$ также непрерывна на X . По теореме Вейерштрасса тогда найдется стратегия $x^P \in X$, для которой $\max_{x \in X} \Psi(x) = \Psi(x^P)$. По лемме 2 такая стратегия x^P максимальна по Парето в задаче (3) (безотносительно к конкретному виду $F_i[x]$, $i = 1, 2$). Затем по стратегии $x^P \in X$ построим вектор $F^g = (F_1[x^P], F_2[x^P]) \in R^2$. Пара (x^P, F^g) как раз и образует сильно гарантированное по исходам и рискам решение задачи (2).

Замечание 4. Для построения явного вида сильно гарантированного решения (x^P, F^g) определение 1 «диктует» следующие этапы.

Этап 1. Найти $\min_{y \in Y} f(x, y) = F_1[x] \quad \forall x \in X$.

Этап 2. Найти $\min_{y \in Y} [-\Phi(x, y)] = -\max_{y \in Y} \Phi(x, y) = F_2[x] \quad \forall x \in X$.

Этап 3. Определить максимизатор $x^P = \arg \max_{x \in X} (F_1[x] + F_2[x])$ и построить $F_i[x^P]$, $i = 1, 2$.

Тогда пара $(x^P, F^g = (F_1[x^P], F_2[x^P]))$ является сильно гарантированным по исходам и рискам решением задачи (2).

3.5. Линейно-квадратичный вариант задачи с ограниченной скалярной неопределенностью

Рассматриваем однокритериальную задачу

$$\Gamma_{\text{л}} = \langle X = R^n, Y = \{y \in R | y_1 \leq y \leq y_2\}, f(x, y) \rangle,$$

в которой линейно-квадратичный критерий

$$f(x, y) = x'Ax + 2x'by + Cy^2 + 2a'x + 2cy, \quad (8)$$

где $(n \times n)$ -матрица A постоянна и симметрична, также постоянны n -векторы-столбцы a и b ; C и c — числа. Далее $A < 0$ означает, что квадратичная форма $x'Ax$ определено отрицательна, а штрих сверху отвечает операции транспонирования (например,

x' — вектор-строка); E_n — единичная $n \times n$ -матрица, 0_n — нулевой n -вектор, $\|\cdot\|$ — евклидова норма. Предполагаем (только для сокращения громоздких записей), что $y_2 = -y_1 = y^0 > 0$.

Условие $X = R^n$ не позволяет воспользоваться утверждением 1 для решения вопроса существования СГР в задаче $\Gamma_{\text{л}}$; поэтому найдем ограничения на постоянные параметры (в критерии (8)), достаточные для существования, и при этом построим явный вид сильно гарантированного по исходам и рискам решения.

Будем следовать этапам 1—3 из замечания 4.

Этап 1 (построение функции $F_1[x] = \min_{y \in [-y^0, y^0]} f(x, y)$, определенной на R^n).

Лемма 3. Если $C > 0$, то глобальный минимум (при каждом $x \in X$)

$$\begin{aligned} \Psi(x) &= \min_{y \in R} f(x, y) = \\ &= x' \left[A - \frac{\|b\|^2}{C} E_n \right] x + 2x' \left(a - b \frac{c}{C} \right) - \frac{c^2}{C}. \end{aligned} \quad (9)$$

Доказательство. Функция (8) строго выпукла по y для каждого $x \in R^n$, ибо $\frac{\partial^2 f}{\partial y^2} = 2C > 0$.

Построим теперь «информированную неопределенность» $y(x): R^n \rightarrow R$ такую, что

$$\min_{y \in R} f(x, y) = f(x, y(x)) \quad \forall x \in R^n.$$

Достаточные условия выполнения этого тождества имеют вид

$$\left. \frac{\partial f(x, y)}{\partial y} \right|_{y=y(x)} = 2x'b + 2Cy(x) + 2c = 0 \quad \forall x \in R^n,$$

$$\frac{\partial^2 f(x, y)}{\partial y^2} = 2C > 0.$$

Отсюда следуют при $\forall x \in R^n$, что

$$y(x) = -\frac{1}{C}(x'b + c)$$

и

$$2x'by + Cy^2(x) + 2cy = -Cy^2(x).$$

Поэтому

$$\min_{y \in R} f(x, y(x)) = x'Ax + 2a'x - \frac{1}{C}(x'b + c)(b'x + c)$$

и, наконец, перемножая, получаем справедливость выражения (9).

Лемма 4. Если $C > 0$, то для $x \in R^n$, удовлетворяющих ограничению

$$-Cy^0 - c \leq b'x \leq Cy^0 - c, \quad (10)$$

минимум функции $f(x, y)$ из выражения (9) совпадает с минимумом функции $f(x, y)$ на отрезке $[-y^0, y^0]$.

Доказательство. Для каждого $x \in R^n$ функция $f(x, y)$ из выражения (8) строго выпукла по y (ибо $C > 0$). Поэтому при $\forall x \in R^n$ частные производные

$$\begin{aligned} \left. \frac{\partial f(x, y)}{\partial y} \right|_{y=-y^0} &= 2x'b - 2Cy^0 + 2c < 0, \\ \left. \frac{\partial f(x, y)}{\partial y} \right|_{y=y^0} &= 2x'b + 2Cy^0 + 2c > 0 \end{aligned} \quad (11)$$

(т. е. функция $f(x, y)$ для каждого $x \in X$ убывает при возрастании $y \leq -y^0$ и возрастает при возрастании $y \geq y^0$).

Наконец, оба строгих неравенства (11) справедливы для тех и только тех $x \in R^n$, при которых выполнена цепочка неравенств (10).

Утверждение 2. При $C > 0$ искомая функция

$$\begin{aligned} F_1[x] &= \min_{y \in [-y^0, y^0]} f(x, y) = \\ &= \begin{cases} f(x, -y^0) & \text{при } b'x \geq Cy^0 - c, \\ \Psi(x) & \text{при } b'x \in (-Cy^0 - c, Cy^0 - c), \\ f(x, y^0) & \text{при } b'x \leq -Cy^0 - c, \end{cases} \end{aligned} \quad (12)$$

где скалярная функция (см. (9))

$$\Psi(x) = x' \left[A - E_n \frac{\|b\|^2}{C} \right] x + 2x' \left(a - b \frac{c}{C} \right) - \frac{c^2}{C}. \quad \blacklozenge$$

Доказательство следует из леммы 4.

Замечание 5. Согласно выражению (12) функция $F_1[x]$ непрерывна на R^n . $\blacklozenge$

Этап 2 (построение функции $F_2[x] = \min_{y \in [-y^0, y^0]} [-\Phi(x, y)]$). Прежде всего, укажем явный

вид функции риска $\Phi(x, y) = \max_{z \in R^n} f(z, y) - f(x, y)$.

Лемма 5. Если в критерии (8) постоянная $(n \times n)$ -матрица $A < 0$, то

$$\begin{aligned} \Phi(x, y) &= -x'Ax - 2x'by - b'A^{-1}by^2 - \\ &\quad - 2a'x - 2a'A^{-1}by - a'A^{-1}a = \\ &= -(x'A + b'y + a')A^{-1}(Ax + by + a). \quad \blacklozenge \end{aligned} \quad (13)$$

Доказательство этого факта см. в работе [2, с. 97–98].

Лемма 6. Если в задаче (2) матрица $A < 0$ и n -вектор $b \neq 0_n$, то функция $\Phi(x, y)$ строго выпукла по y и

$$\begin{aligned} F_2[x] &= \min_{-y^0 \leq y \leq y^0} [-\Phi(x, y)] = \\ &= -\max\{\Phi(x, -y^0), \Phi(x, y^0)\}. \end{aligned}$$

Доказательство. Справедлива цепочка импликаций

$$(A < 0) \Rightarrow (A^{-1} < 0) \Rightarrow (b'A^{-1}b < 0),$$

так как $b \neq 0_n$, а из выражения (13) получаем

$$\frac{\partial^2 [-\Phi(x, y)]}{\partial y^2} = b'A^{-1}b < 0.$$

Поэтому при каждом $x \in R^n$ и функция риска $\Phi(x, y)$ строго выпукла по $y \in [-y^0, y^0]$. $\blacklozenge$

Максимум строго выпуклой по y функции $\Phi(x, y)$ достигается на ее граничных точках. Поэтому

$$\begin{aligned} \min_{-y^0 \leq y \leq y^0} [-\Phi(x, y)] &= -\max_{-y^0 \leq y \leq y^0} \Phi(x, y) = \\ &= -\max\{\Phi(x, -y^0), \Phi(x, y^0)\}. \end{aligned} \quad (14)$$

Здесь

$$\begin{aligned} \Phi(x, y_l) &= -x'Ax - 2x'by_l - b'A^{-1}b(y_l)^2 - \\ &\quad - 2a'x - 2a'A^{-1}by_l - a'A^{-1}a, \quad l = 1, 2, \end{aligned}$$

где $y_1 = -y^0$ и $y_2 = y^0$.

Лемма 7. При $A < 0$, $b \neq 0_n$

$$\Phi(x, -y^0) - \Phi(x, y^0) = 4y^0(b'x + b'A^{-1}a). \quad \blacklozenge \quad (15)$$

Доказательство см. в работе [2, с. 99].

Утверждение 3. Если в критерии (8) матрица $A < 0$ и n -вектор $b \neq 0_n$, то

$$\begin{aligned} F_2[x] &= \min_{-y^0 \leq y \leq y^0} [-\Phi(x, y)] = \\ &= \begin{cases} -\Phi(x, -y^0) & \text{при } b'x > -\alpha, \\ -\Phi(x, y^0) & \text{при } b'x \leq -\alpha, \end{cases} = \\ &= \begin{cases} x'Ax + 2x'(-by^0 + a) + (-b'y^0 + a')A^{-1}(-by^0 + a) \\ \text{при } b'x > -\alpha, \\ x'Ax + 2x'(by^0 + a) + (b'y^0 + a')A^{-1}(by^0 + a) \\ \text{при } b'x \leq -\alpha, \end{cases} \end{aligned} \quad (16)$$

здесь $\alpha = b'A^{-1}a$.

Доказательство. Если в формуле (15) будет $b'x > -\alpha$, то $\Phi(x, -y^0) > \Phi(x, y^0) \Leftrightarrow -\Phi(x, -y^0) < -\Phi(x, y^0)$, и тогда из выражения (14) следует справедливость первой строки в выражении (16); аналогично устанавливается вторая строка.



Этап 3 (построение максимальной по Парето стратегии $x^P = \arg \max_{x \in R^n} (F_1[x] + F_2[x])$). Далее всюду до окончания п. 3.5, не оговаривая особо, будем считать выполненными

Условие 1. В задаче (2) постоянная симметричная $(n \times n)$ -матрица $A < 0$, постоянный n -вектор $b \neq 0_n$, число $C > 0$. ♦

Здесь выделяем три случая в зависимости от принадлежности точки $-\alpha$ одному из трех интервалов $(-\infty, -Cy^0 - c]$, $(-Cy^0 - c, Cy^0 - c)$, $[Cy^0 - c, +\infty)$. В каждом из этих случаев значения $b'x$ могут принадлежать одному из шести интервалов, на которые могут разбить множество R точки $-\alpha$, $-Cy^0 - c$ и $Cy^0 - c$.

Случай 1: $(-\alpha \in (-\infty, -Cy^0 - c])$

Утверждение 4. Пусть для задачи (2) выполняются условия 1, тогда при $-\infty < b'x \leq -\alpha \leq (-Cy^0 - c)$ гарантированное по исходам и рискам решение (x^P, f^P, Φ^P) имеет вид

$$\begin{aligned} x^P &= -A^{-1}(by^0 + a), \\ f^P &= f(x^P, y^0) = [-b'A^{-1}b + C](y^0)^2 + \\ &\quad + 2(-a'A^{-1}b + c)y^0 - a'A^{-1}a, \\ \Phi^P &= \Phi(x^P, y^0) = 0. \end{aligned}$$

Доказательство. В случае $b'x \leq -\alpha \leq (-Cy^0 - c)$, согласно утверждениям 2 и 3, найдем максимизатор x^P :

$$\begin{aligned} \max_{x \in R^n} (F_1[x] + F_2[x]) &= \max_{x \in R^n} [f(x, y^0) - \Phi(x, y^0)] = \\ &= f(x^P, y^0) - \Phi(x^P, y^0). \end{aligned} \quad (17)$$

Исходя из вида функции $f(x, y^0)$ в формуле (8) и $\Phi(x, y^0)$ из выражения (16), стратегия x^P удовлетворяет равенству (17), если

$$\max_{x \in R^n} \varphi_1(x) = \varphi_1(x^P),$$

где $\varphi_1(x) = 2x'Ax + 4x'(by^0 + a)$. Достаточные условия при этом будут иметь вид

$$\begin{aligned} \text{grad}_x \varphi_1(x) \Big|_{x^P} &= \frac{\partial \varphi_1(x)}{\partial x} \Big|_{x^P} = 4Ax^P + 4(by^0 + a) = 0_n, \\ \frac{\partial^2 \varphi_1(x)}{\partial x^2} &= 4A < 0. \end{aligned}$$

Из первого равенства получаем

$$x^P = -A^{-1}(by^0 + a),$$

откуда

$$\begin{aligned} f^P &= f(x^P, y^0) = F_1[x^P] = [-b'A^{-1}b + C](y^0)^2 + \\ &\quad + 2(-a'A^{-1}b + c)y^0 - a'A^{-1}a, \end{aligned}$$

$$\begin{aligned} \Phi^P &= \Phi(x^P, y^0) = \\ &= -((x^P)'A + b'y^0 + a')A^{-1}(Ax^P + by^0 + a) = 0. \end{aligned}$$

Утверждение 5. Если выполнены условия 1 и $-\alpha \leq b'x \leq (-Cy^0 - c)$, то гарантированное по исходам и рискам решение (x^P, f^P, Φ^P) принимает вид

$$\begin{aligned} x^P &= -A^{-1}a, \\ f^P &= C(y^0)^2 + 2(-a'A^{-1}b + c)y^0 - a'A^{-1}a, \quad (18) \\ \Phi^P &= -b'A^{-1}b(y^0)^2. \end{aligned}$$

Доказательство. По утверждениям 2 и 3 стратегия x^P является максимизатором $F_1[x] + F_2[x]$, т. е.

$$x^P = \arg \max_{x \in R^n} (f(x, y^0) - \Phi(x, -y^0)).$$

Согласно выражениям (12) и (16) тогда

$$\max_{x \in R^n} \varphi_2(x) = \varphi_2(x^P),$$

где $\varphi_2(x) = 2x'Ax + 4x'a$. Из достаточных условий

$$\begin{aligned} \frac{\partial \varphi_2(x)}{\partial x} \Big|_{x^P} &= 4Ax^P + 4a = 0_n, \\ \frac{\partial^2 \varphi_2(x)}{\partial x^2} &= 4A < 0 \end{aligned}$$

находим стратегию $x^P = -A^{-1}a$, подставляя которую в выражение (8), найдем

$$f^P = f(x^P, y^0) = (x^P)'Ax^P + 2(x^P)'(by^0 + a) + y^0(Cy^0 + 2c).$$

Аналогично для

$$\begin{aligned} \Phi^P &= -((x^P)'A - b'y^0 + a')A^{-1}(Ax^P - by^0 + a) = \\ &= -b'A^{-1}b(y^0)^2 > 0. \end{aligned}$$

Аналогично рассматриваются и остальные возможные случаи (результаты сведены в таблицу; предполагается, что в выражении (8) матрица $A < 0$, вектор $b \neq 0_n$, число $C > 0$).

Замечание 6. Почему тройка (x^P, f^P, Φ^P) является строго гарантированным по исходам и рискам решением задачи (2)? Напомним, что в задаче (2) ЛПР «выступает» как «рисконейтрал» (см. Введение): он выбирает стратегию $x \in X$ так, чтобы:

— его исход $f(x, y)$ стал возможно *большим* и одновременно соответствующий риск по Сэвиджу $\Phi(x, y)$ как можно *меньшим*, ориентируясь при этом на реализацию любой чистой неопределен-

ности $y \in Y$. Выбирая стратегию $x \in X$, ЛПР гарантирует себе исход, не меньший

$$F_1[x] = f[x] = \min_{-y^0 \leq y \leq y^0} f(x, y) \leq f(x, y) \quad \forall y \in Y,$$

и одновременно риск, не больший

$$-F_2[x] = \Phi[x] = \max_{-y^0 \leq y \leq y^0} \Phi(x, y) \geq \Phi(x, y) \quad \forall y \in Y;$$

— считая эти гарантии $(f[x], \Phi[x])$ «равноправными», ЛПР предпочитает использовать стратегию x^P , максимальную по Парето в двухкритериальной задаче $\langle X, \{f[x], -\Phi[x]\} \rangle$ (максимальную в $\langle X, f[x] - \Phi[x] \rangle$).

Из этой паретовости следует, что, прежде всего, при переходе ЛПР к стратегии $x \neq x^P$ увеличение гарантированного исхода $f[x] > f[x^P]$ неизбежно влечет увеличение гарантированного риска $\Phi[x] > \Phi[x^P]$, далее, уменьшение гарантированного риска $\Phi[x] < \Phi[x^P]$ приводит к уменьшению га-

рантированного исхода $f[x] < f[x^P]$. Таким образом, переход от двухкритериальной задачи при неопределенности

$$\langle X, Y, \{f(x, y), -\Phi(x, y)\} \rangle$$

к двухкритериальной «задаче гарантий без неопределенностей»

$$\langle X, \{ \min_{y \in Y} f(x, y) = f[x], -\max_{y \in Y} \Phi(x, y) = -\Phi[x] \} \rangle$$

с последующим использованием векторного максимума (по Парето) позволяет «выйти» на явный вид гарантированного решения с позиции «риско-нейтрала».

Однако недостатком предпринятого здесь подхода выступает ориентация на «самые плохие» (маленькие) исходы $\min_{y \in Y} f(x, y) = f[x]$ и на «самые большие» риски $\max_{y \in Y} \Phi(x, y) = \Phi[x]$. Ослабить этот недочет позволяет другой способ выбора гарантированного по исходам и рискам решения, ориен-

Явный вид гарантированного по исходам и рискам решения (x^P, f^P, Φ^P)

Случай 1: $-\alpha \leq (-Cy^0 - c)$	Случай 1а: $-\infty \leq b'x \leq -\alpha$	$x^P = -A^{-1}(by^0 + a),$ $f^P = [C - b'A^{-1}b](y^0)^2 + 2(c - a'A^{-1}b)y^0 - a'A^{-1}a,$ $\Phi^P = 0$
	Случай 1б: $-\alpha < b'x \leq (-Cy^0 - c)$	$x^P = -A^{-1}a,$ $f^P = C(y^0)^2 + 2(c - a'A^{-1}b)y^0 - a'A^{-1}a,$ $\Phi^P = -b'A^{-1}b(y^0)^2 > 0$
Случай 2: $(-Cy^0 - c) \leq -\alpha \leq Cy^0 - c$	Случай 2а: $(-Cy^0 - c) < b'x \leq -\alpha$	$x^P = -\left[2A - \frac{\ b\ ^2}{C}E_n\right]^{-1} \left(2a + b\left(y^0 - \frac{c}{C}\right)\right),$ $f^P = (x^P)' \left[A - \frac{\ b\ ^2}{C}E_n\right] x^P + 2(x^P)' \left(a - b\frac{c}{C}\right) - \frac{c^2}{C},$ $\Phi^P = -((x^P)'A + b'y^0 + a')A^{-1}(Ax^P + by^0 + a)$
	Случай 2б: $-\alpha < b'x < Cy^0 - c$	$x^P = -\left[2A - \frac{\ b\ ^2}{C}E_n\right]^{-1} \left(2a - b\left(y^0 + \frac{c}{C}\right)\right),$ $f^P = (x^P)' \left[A - \frac{\ b\ ^2}{C}E_n\right] x^P + 2(x^P)' \left(a - b\frac{c}{C}\right) - \frac{c^2}{C},$ $\Phi^P = -((x^P)'A - b'y^0 + a')A^{-1}(Ax^P - by^0 + a) > 0$
Случай 3: $-\alpha \geq Cy^0 - c$	Случай 3а: $(Cy^0 - c) \leq b'x \leq -\alpha$	$x^P = -A^{-1}a,$ $f^P = C(y^0)^2 - 2(c - a'A^{-1}b)y^0 - a'A^{-1}a,$ $\Phi^P = -b'A^{-1}b(y^0)^2 > 0$
	Случай 3б: $-\alpha < b'x < \infty$	$x^P = A^{-1}(by^0 - a),$ $f^P = [C - b'A^{-1}b](y^0)^2 - 2(c - a'A^{-1}b)y^0 - a'A^{-1}a,$ $\Phi^P = 0$



тированный на векторный максимин [22], разработка которого начата в 1990 г. одним из авторов в книге [23], близкой задаче посвящено уже значительное число публикаций, из которых особо отметим [24, 25].

4. МАКСИМИН ПО СЛЕЙТЕРУ

4.1. Формализация векторного максимина

Рассматриваем двухкритериальную задачу при неопределенности

$$\Gamma = \langle X, Y, F(x, y) = (F_1(x, y), F_2(x, y)) \rangle, \quad (19)$$

ориентация здесь, в конце концов, на $F_1(x, y) = f(x, y)$, $F_2(x, y) = -\Phi(x, y)$, где $f(x, y)$ — критерий из задачи (2), а $\Phi(x, y)$ — функция риска по Сэвиджу (1). Так же, как в задаче (3), (4), в двухкритериальной задаче (19) стратегии ИО есть $x \in X \subseteq R^n$, чистые неопределенности $y \in Y \subseteq R^m$, наряду с чистыми будем использовать «информированные неопределенности» — m -мерные вектор-функции $y(x): X \rightarrow Y$, в этом случае определенный на $X \times Y$ векторный критерий

$$F(x, y) = (F_1(x, y(x)), F_2(x, y(x))) = F(x, y(x)) = F[x] = (F_1[x], F_2[x]).$$

Пусть в критерии (19) «заморожена» стратегия $x \in X$, т. е. получаем $\Gamma(x) = \langle x, Y, F(x, y) \rangle$. Используем бинарные отношения строгого порядка ($>$) и ($\nless$) из § 2.

Неопределенность $y_S(x)$ называется *минимальной по Слейтеру* в задаче $\Gamma(x)$, если $F(x, y_S(x)) \nless F(x, y) \forall y \in Y$; множество таких $y_S(x)$ обозначим $Y_S(x)$. Далее используем обозначение $F(x, y_S(x)) = \min_{y \in Y} F(x, y)$.

Определение 2. Стратегия $x^S \in X$ называется *максиминной по Слейтеру* в задаче (19), если существует «информированная неопределенность» $\hat{y}_S(x) \in Y_S(x) \forall x \in X$, для которой

$$F(x^S, \hat{y}_S(x^S)) \nless F(x, y_S(x))$$

при всех $x \in X$ и $y_S(x) \in Y_S(x)$. ♦

При этом n -вектор x^S называют *максиминной по Слейтеру* стратегией в задаче (19), сам вектор $F(x^S, \hat{y}_S(x^S))$ — *максимином по Слейтеру*, а пару $(x^S, F(x^S, \hat{y}_S(x^S)))$ назовем *максиминным по Слейтеру решением задачи* Γ .

Если воспользоваться обозначениями из § 2 и

$$\min_{y \in Y} F(x, y) = F(x, y_S(x)) = F[x] \quad \forall x \in X, \quad (20)$$

$$\max_{x \in X, y_S(x) \in Y_S(x)} F(x, y_S(x)) = F(x^S, \hat{y}_S(x^S)), \quad (21)$$

то аналогом понятия максимина из п. 3.1 будет

$$F(x^S, \hat{y}_S(x^S)) = \max_{x \in X} \min_{y = y_S(x) \in Y_S(x)} F(x, y) = \min_{y \in Y} F(x^S, y).$$

Операция (20) соответствует нахождению внутреннего минимума (из п. 3.1), а операция (21) — операции внешнего максимума.

Замечание 7. Вектор $F[x] = F(x, y_S(x))$ является *векторной гарантией* для каждой стратегии $x \in X$. Действительно, из выражения (20) имеем

$$F[x] = \min_{y \in Y} F(x, y),$$

и поэтому $F[x] \nless F(x, y) \forall y \in Y$, т. е. при использовании любой стратегии $x \in X$ все компоненты $F_i(x, y)$ вектора $F(x, y)$ не могут стать одновременно меньшими $F_i[x]$, $i = 1, 2$, ни при каких $y \in Y$. Таким образом, $F[x]$ ограничивает убывание $F(x, y)$ одновременно по двум компонентам. Однако для одних и тех же стратегий $x \in X$

$$\min_{y \in Y} F_i(x, y) \leq F_i(x, y) \quad \forall y \in Y,$$

и тогда

$$\min_{y \in Y} F_i(x, y) \leq F_i(x, y_S(x)) \quad \forall x \in X, \quad i = 1, 2.$$

В этом смысле гарантии в строго гарантированном по исходам и рискам решении (см. определение 1) покомпонентно не больше, чем векторные гарантии $F[x]$ из определения 2. Как раз этим обстоятельством и вызван термин «*сильно*» (гарантированное решение). ♦

Затем, согласно выражению (21), из всех таких векторных гарантий выбрана в качестве максиминного по Слейтеру решения наибольшая в «векторном смысле» (максимальная по Слейтеру) по отношению ко всем компонентам $F[x]$, ибо $F[x] \nless F[x^S]$.

4.2. Геометрическая интерпретация

Пусть x^S — максиминная по Слейтеру стратегия в двухкритериальной задаче (19), а $\hat{y}_S(x^S)$ — соответствующее значение неопределенности. Согласно определению минимума по Слейтеру каждому

$x \in X$ ставится в соответствие множество минимальных по Слейтеру значений $F(x, Y_S(x))$ векторного критерия $F(x, y)$, т. е.

$$F(x, y) \not\prec F(x, y_S(x)). \quad (22)$$

Таким образом,

$$F(x, y_S(x)) = \cup_{y \in Y_S(x)} \text{MIN}^S f(x, y).$$

Тогда из условия (22) следует, что при каждом $x \in X$ и $y_S(x) \in Y_S(x)$ точки множества $F(x, Y) = \{F(x, y) | y \in Y\}$ не могут попасть внутрь прямого угла $G(x, y_S(x))$ с вершиной в точке $F(x, y_S(x))$ (для всех $y_S(x) \in Y_S(x)$). На рис. 2 указаны три таких прямых угла $G(x, y_S^{(k)}(x))$, $k = 1, 2, 3$, ограничивающих изменение $F(x, Y)$ в юго-западном направлении. В частности, таким свойством обладают максиминная стратегия x^S и соответствующая неопределенность $\hat{y}_S(x^S)$ (рис. 3).

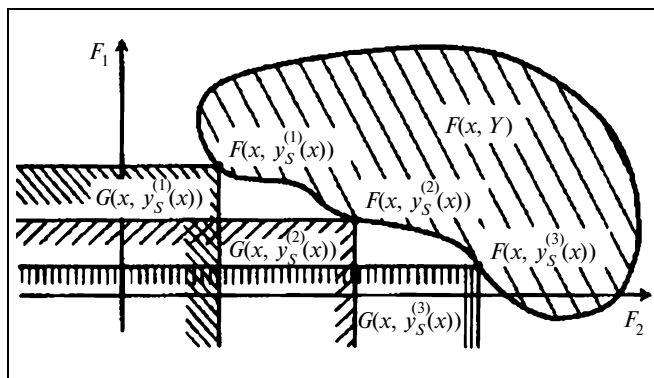


Рис. 2. Интерпретация множества $Y_S(x)$ при фиксированном x

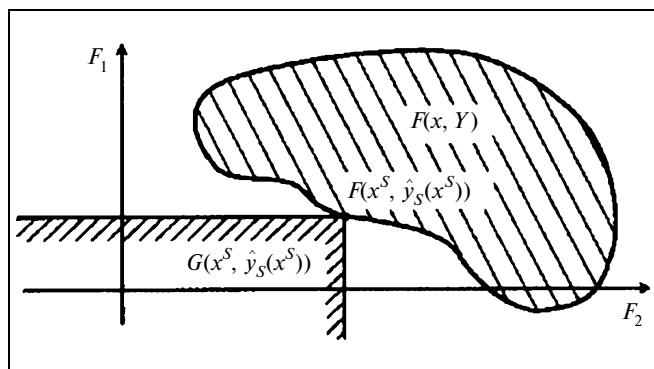


Рис. 3. Интерпретация эффекта векторной минимизации для максиминной стратегии x^S

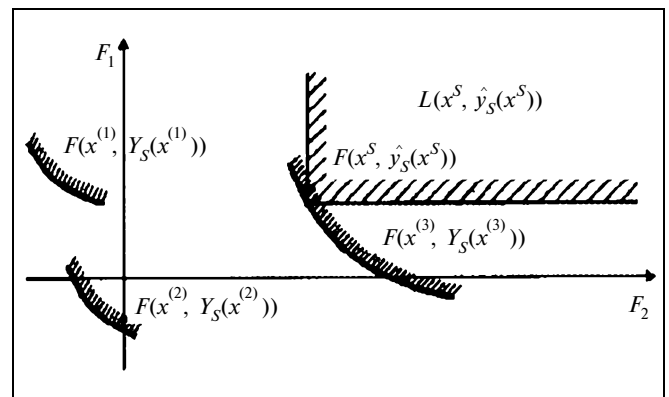


Рис. 4. Интерпретация эффекта векторной максимизации для x^S

С другой стороны, в условии (22) значение векторного критерия $F(x^S, \hat{y}_S(x^S))$ сравнивается с множествами $F(x, Y_S(x))$ при каждом фиксированном $x \in X$. Эти множества суть минимальные по Слейтеру значения векторного критерия $F(x, y)$ при данном x . Тогда условие (22) означает, что точки $F(x, Y_S(x))$ при всех $x \in X$ и $y_S(x) \in Y_S(x)$ не попадают внутрь прямого угла $L(x^S, \hat{y}_S(x^S))$ с вершиной в той же самой точке $F(x^S, \hat{y}_S(x^S))$, но направленного противоположно углам G . На рис. 4 указаны расположения трех множеств $F(x^{(k)}, Y_S(x^{(k)}))$, $k = 1, 2, 3$, относительно точки $F(x^S, \hat{y}_S(x^S))$.

Таким образом, максимин по Слейтеру $F(x^S, \hat{y}_S(x^S))$ ограничивает углом $L(x^S, \hat{y}_S(x^S))$ распространение (в северо-восточном направлении) минимальных по Слейтеру значений $F(x, Y_S(x)) \forall x \in X$.

4.3. Существование

Утверждение 6. Если в однокритериальной задаче при неопределенности $\Gamma^{(1)} = \langle X, Y, f(x, y) \rangle$ множества $X \in \text{comp} R^n$, $Y \in \text{comp} R^m$, а критерий $f(x, y)$ непрерывен на $X \times Y$, то для двухкритериальной задачи при неопределенности

$$\Gamma^{(1)} = \langle X, Y, F(x, y) = (F_1(x, y) = f(x, y), F_2(x, y) = -\Phi(x, y)) \rangle$$

существует максиминное по Слейтеру решение $(x^S, F(x^S, \hat{y}_S(x^S))) \in X \times R^2$.

Доказательство. Функция риска по Сэвиджу $\Phi(x, y)$ непрерывна на $X \times Y$ (замечание 1). Тогда из компактности X, Y и непрерывности компонент $F(x, y)$



на $X \times Y$, а также из работы [23, с. 17–21] сразу следует существование максиминного по Слейтеру решения $(x^S, F(x^S, \hat{y}_S(x^S)))$ задачи $\Gamma^{(1)}$.

4.4. Задача с «разделенными» критериями

Согласно определению 2, при «разделенном» (по стратегиям и неопределенностям) векторном критерии $F(x, y) = \varphi(x) + \psi(y)$, $F, \varphi, \psi \in R^2$ задачу построения максиминного по Слейтеру решения $(x^S, F(x^S, \hat{y}_S(x^S)))$ можно свести к следующим четырём этапам.

Этап 1. Для двухкритериальной задачи

$$\Gamma_\varphi = \langle X, \varphi(x) \rangle$$

найти *все* множество X^S максимальных по Слейтеру стратегий x^S , т. е.

$$\varphi(x^S) \not\leq \varphi(x) \quad \forall x \in X, \quad x^S \in X^S,$$

затем построить множество $\varphi(X^S) = \{\varphi(x) | x \in X^S\}$ максимумов по Слейтеру для задачи Γ_φ .

Этап 2. Для двухкритериальной задачи $\Gamma_\psi = \langle Y, \psi(y) \rangle$ найти *все* множество Y_S минимальных по Слейтеру неопределенностей y_S и $\psi(Y_S) = \{\psi(y) | y \in Y_S\}$, т. е.

$$\psi(y) \not\leq \psi(y_S) \quad \forall y \in Y, \quad y_S \in Y_S.$$

Этап 3. Сложить множества $\varphi(X^S) + \psi(Y_S)$, например, по правилу $\varphi(X^S) + \psi(Y_S) = \{\psi(Y_S) + \varphi | \varphi \in \varphi(X^S)\}$.

Этап 4. Построить множество максимумов по Слейтеру $\mathcal{F}^S$ для $\varphi(X^S) + \psi(Y_S)$ и уже для точки $F^S \in \mathcal{F}^S$ найти пару $\varphi^S \in \varphi(X^S)$ и $\psi^S \in \psi(Y_S)$ таких, что $\varphi^S + \psi^S = F^S$. Затем определить по φ^S и ψ^S точки x^S и y_S соответственно: $\varphi^S = \varphi(x^S)$, $\psi^S = \psi(y_S)$. Тогда максиминным по Слейтеру решением будет $(x^S, \varphi(x^S) + \psi(y_S))$.

Пример. Пусть в задаче $\Gamma = \langle X, Y, F(x, y) \rangle$ двухкомпонентный критерий $F(x, y) = x + y$, множества $X = \{x = (x_1, x_2) | x_1^2 + x_2^2 \leq c^2\}$ (рис. 5), $Y = \{y = (y_1, y_2) | -c \leq y_i \leq c, i = 1, 2\}$, где $c = \text{const} > 0$ (рис. 6).

Далее следуем по указанным четырём этапам.

Этап 1. Построим множество X^S максимальных по Слейтеру стратегий x^S двухкритериальной задачи $\Gamma_\varphi = \langle X, \varphi(x) = x \rangle$. На рис. 5 множество $X^S = \varphi(X^S)$

обведено жирной чертой (дуга AB), равенство имеет место благодаря специальному виду критерия

$$\varphi(x) = (\varphi_1(x), \varphi_2(x)) = x = (x_1, x_2).$$

Этап 2. Найдем множество Y_S минимальных по Слейтеру неопределенностей в двухкритериальной задаче $\Gamma_\psi = \langle Y, \psi(y) = y \rangle$, множество $Y_S = \psi(Y_S)$ выделено жирной чертой на рис. 6 (ломаная CDG). Опять-таки равенство вызвано видом критерия $\psi(y) = y$.

Этап 3. Построим множество $X^S + Y_S$, добавив к каждой точке $y_S \in Y_S$ выделенную четверть окружности AB . Множество $X^S + Y_S$ заштриховано на рис. 7.

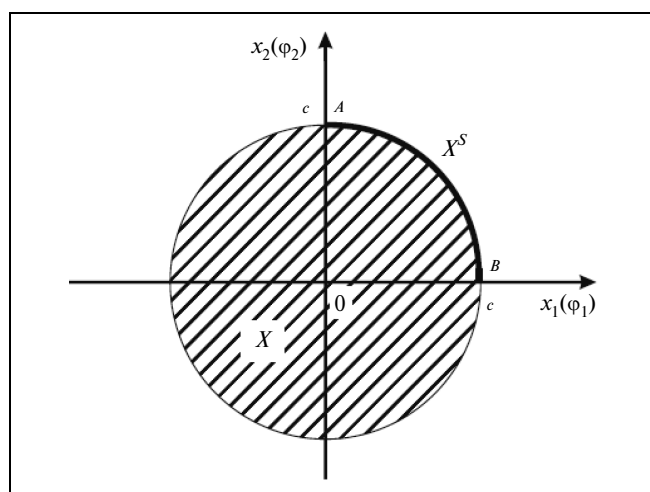


Рис. 5. Множество X^S

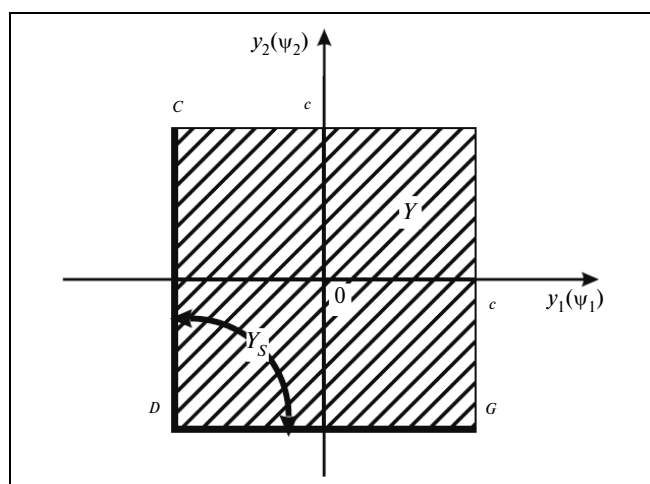


Рис. 6. Множество Y_S

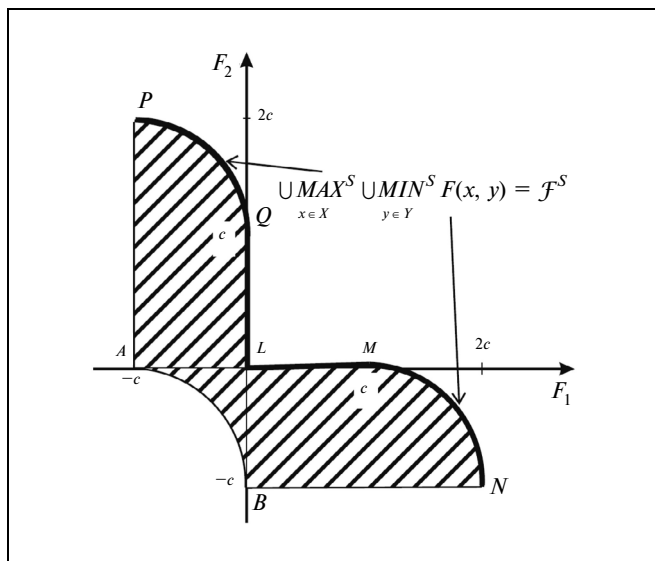


Рис. 7. Множество $X^S + Y_S$

Этап 4. Множество максимумов по Слейтеру для $X^S + Y_S$ образуют точки жирной линии $PQLMN$ (см. рис. 7). Например, точке M отвечает максиминное по Слейтеру решение $(x^S, F(x^S, y_S))$, где $x^S = (0, c)$, $y_S = (c, -c)$, $F(x^S, y_S) = x^S + y_S = (c, 0) = M$.

ЗАКЛЮЧЕНИЕ

В статье предложено два подхода к принятию решения в однокритериальной задаче при неопределенности, о неопределенностях известны лишь границы изменения. В обоих подходах ЛПР стремится не только увеличить исход, но и одновременно понизить риск. Первый подход базируется на подходящей модификации максимина и на принятии решения в иерархической двухуровневой игре, второй — на векторном максимине. Найден явный вид первого решения при линейно-квадратичном критерии.

Авторы благодарят рецензентов за замечания, большинство из которых учтено при подготовке рукописи к печати.

ЛИТЕРАТУРА

1. Черемных Ю.Н. Микроэкономика. Продвинутый уровень. — М.: ИНФРА, 2008. — 843 с.
2. Жуковский В.И. Риски при конфликтных ситуациях. — М.: URSS, Ленанд, 2011. — 328 с.

3. Сиразетдинов Т. К., Сиразетдинов Р.Т. Проблема риска и его моделирование // Проблемы человеческого риска. — 2007. — № 1. — С. 31—43.
4. Шахов В.В. Введение в страхование. Экономический аспект. — М.: Финансы и статистика, 1994. — 188 с.
5. Цветкова Е.В., Арлюкова И.О. Риски в экономической деятельности. — СПб.: ИВЭСЭП, 2002. — 64 с.
6. Savage L.Y. The theory of statistical decision // J. American Statistic Association. — 1951. — N 46. — P. 55—67.
7. Wald A. Contribution to the theory of statistical estimation and testing hypothesis // Annals Math. Statist. — 1939. — Vol. 10. — P. 299—326.
8. Фишер С., Дорнбуш Р., Шмалензи Р. Экономика. — М.: Дело, 1998. — 829 с.
9. Жуковский В.И. Введение в дифференциальные игры при неопределенности. — М.: МНИИПУ, 1997. — 459 с.
10. Zhukovskiy V.I. Lyapunov Functions in Differential Games. — London; N.-Y.: Taylor & Francis, 2003. — 282 с.
11. Жуковский В.И., Чикрий А.А. Линейно-квадратичные дифференциальные игры. — Киев: Наукова думка. — 1994. — 320 с.
12. Жуковский В.И., Жуковская Л.В. Риск в многокритериальных и конфликтных системах при неопределенности. — М.: Эдиториал УРСС. — 2004. — 272 с.
13. Гермейер Ю.Б. Игры с противоположными интересами. — М.: Наука, 1978. — 328 с.
14. Ватель И.А., Ерешко Ф.И. Игра с иерархической структурой // Математическая энциклопедия. — М., 1979. — Т. 2.
15. Кукушкин Н.С., Морозов В.В. Теория неантагонистических игр. — М.: МГУ, 1984. — 104 с.
16. Ватель И.А., Ерешко Ф.И. Математика конфликта и сотрудничества. — М.: Знание, 1974. — С. 478—481.
17. Морозов В.В. Основы теории игр. — М.: Изд. отдел фак-та ВМиК МГУ, 2002. — 150 с.
18. Красовский Н.Н., Субботин А.И. Позиционные дифференциальные игры. — М.: Наука, 1974. — 456 с.
19. Дмитрук А.В. Выпуклый анализ. Элементарный вводный курс. — М.: ВМиК МГУ, 2012. — 172 с.
20. Морозов В.В., Сухарев А.Г., Федоров В.В. Исследование операций в задачах и упражнениях. — М.: Высшая школа, 1968. — 286 с.
21. Подиновский В.В., Ногин В.Д. Парето-оптимальные решения многокритериальных задач. — М.: Наука, 1982. — 254 с.
22. Zhukovskiy V.I., Salukvadze M.E. The Vector-Valued Maximin. — N.-Y.: Academic Press, 1994. — 282 p.
23. Жуковский В.И., Молоствов В.С. Многокритериальная оптимизация систем в условиях неполной информации. — М.: МНИИПУ, 1990. — 112 с.
24. Поспелова И.И. Классификация задач векторной оптимизации с неопределенными факторами // Журнал выч. математики и мат. физики. — 2000. — Т. 40, № 6. — С. 860—876.
25. Новикова Н.М., Малащенко Ю.Е. Модели неопределенности в многопользовательских сетях. — М.: Эдиториал УРСС, 1999. — 160 с.

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

Жуковский Владислав Иосифович — д-р физ.-мат. наук, профессор, Московский государственный университет им. М.В. Ломоносова, zhkvlad@yandex.ru,

Солдатова Наталья Геннадьевна — ст. преподаватель, Московский государственный областной гуманитарный институт, г. Орехово-Зуево, solnata@pochta.ru.

ИНТЕЛЛЕКТНАЯ СИСТЕМА УПРАВЛЕНИЯ РАЗВИТИЕМ ТРАНСПОРТНО-ЛОГИСТИЧЕСКОЙ ИНФРАСТРУКТУРОЙ РЕГИОНА¹

И.В. Бычков, А.Л. Казаков, А.А. Лемперт, Д.С. Бухаров, А.Б. Столбов

Описана интеллектуальная система, реализованная на основе онтологий и формализованных знаний экспертов, математических моделей и численных методов. Разработан комплексный подход к исследованию транспортно-логистических систем, основанный на предложенной авторами концепции многоуровневого моделирования. Приведены краткие описания построенных математических моделей и интегрированных в систему программных модулей. Для организации взаимодействия между модулями и при обработке неполных статистических данных применены современные средства интеллектуализации. Описаны некоторые из решенных с помощью разработанной системы модельные и прикладные задачи.

Ключевые слова: интеллектуальная система, продукционные правила, база знаний, онтология, интеграция, многоуровневое моделирование, транспорт, логистика.

ВВЕДЕНИЕ

Транспортно-логистическая инфраструктура определяется как комплекс взаимосвязанных элементов, обеспечивающих движение материальных потоков [1]. В число ее ключевых функций входят повышение уровня социально-экономического развития региона благодаря эффективному функционированию различных видов транспорта и формированию сбалансированной транспортной сети, удовлетворение потребности экономических субъектов в транспортных и логистических объектах и совершенствование работы складской системы.

Как отмечают многие специалисты [2], в последние годы назрела необходимость комплексного исследования проблем транспорта и логистики, а именно рассмотрения процессов и объектов транспортно-логистической инфраструктуры с точки зрения их взаимодействия как сложной системы.

Для учета сложной иерархической структуры транспортных систем и особенностей транспортных процессов необходимы адекватные математические модели и методы и обработка разнородной

и сложно организованной информации. В зависимости от целей конкретного исследования могут применяться дискретные или непрерывные, детерминированные или стохастические описания. К настоящему времени разработано большое число моделей и методов, предназначенных для решения конкретных задач транспорта и логистики, среди которых можно отметить динамические модели на транспорте, применение теории и методов дискретной оптимизации и целочисленного программирования для оптимального размещения объектов инфраструктуры [2, 3], методы двухуровневой оптимизации [4], вероятностный подход к моделированию транспортных систем [5] и др.

Разработано и реализовано достаточно много программных систем, автоматизирующих решение различных задач, возникающих при исследованиях в области транспортной логистики. Так, в рамках общего исследования проблем территориального развития регионов России Отделом территориальных систем Института экономики и организации промышленного производства СО РАН разработан программно-модельный комплекс RTA. Основная задача комплекса состоит в оценке изменения показателей транспортной доступности при освоении новых регионов и оценке различных сценариев развития транспортной инфраструкту-

¹ Работа выполнена при частичной финансовой поддержке РФФИ (проекты №№ 12-07-13116-ОФИ-РЖД, 14-07-00222).

ры. В состав комплекса входят различные средства визуализации эффекта от прокладки железнодорожных путей и автодорог. По заказу Минтранса РФ консалтинговая компания «Геограком» разработала одноименную систему (на базе геоинформационной системы (ГИС) «Панорама») [6], предназначенную для диагностики «узких мест» транспортной системы и ее мониторинга, построения основных стратегий и сценариев развития дорожного хозяйства при разных налоговых схемах, оценки внешней окупаемости сетевых проектов. Система способна генерировать альтернативные инвестиционные программы. Система OLSIM позволяет визуализировать загруженность дорожной сети на основе информации, поступающей с детекторов движения, а также прогнозировать развитие дорожной ситуации на период до 60 мин с помощью имитационной модели, базирующейся на идеологии клеточных автоматов [7]. Система GAMMA (геометрические алгоритмы моделирования) предназначена для моделирования дорожного движения, основываясь на моделях сплошной среды и разбивая ее на ячейки с целью уменьшения вычислительных затрат путем применения массового параллелизма. В качестве модели среды используется сжимаемая вязкая жидкость [8]. В Институте динамики систем и теории управления СО РАН [9] выполнен анализ региональной информационно-телекоммуникационной инфраструктуры (на примере Иркутской области), предложен ряд ГИС и информационно-аналитических систем для органов государственной власти и местного самоуправления, включая ГИС «Мониторинг социально-экономического положения», в которой имеется раздел «Транспорт» (железнодорожный, воздушный, автомобильный, речной). Также существует ряд систем, программный инструментарий которых предназначен для обработки информации, поступающей в режиме реального времени на центральный сервер с различных датчиков, установленных на автомобильных дорогах: «Project Prometheus», «Project AMTICS and RACS», EMME/2, MVA, HDM-IV.

Отметим, что все упомянутые системы ориентированы на решение отдельных задач и не позволяют проводить комплексные исследования, направленные на решение проблем развития транспортно-логистической инфраструктуры региона как целостной системы. В частности, отсутствует возможность совместного использования моделей различных (макро-, мезо- и микро-) уровней.

В данной статье представлены результаты, позволяющие решить указанную проблему, описан комплексный подход к многоуровневому моделированию транспортных систем, разработана онтологическая модель предметной области и система представления обработки знаний экспертами в ви-

де продукции. На их основе создана интеллектуальная система управления развитием транспортно-логистической инфраструктуры региона (ИСУ).

1. МЕТОДОЛОГИЯ ИССЛЕДОВАНИЯ ТРАНСПОРТНО-ЛОГИСТИЧЕСКОЙ ИНФРАСТРУКТУРЫ И ЕЕ РЕАЛИЗАЦИЯ

Комплексное исследование транспортно-логистической инфраструктуры осуществляется на трех уровнях: на верхнем уровне разрабатывается качественный прогноз развития отраслей на основе экономических и эколого-экономических моделей; на среднем уровне с учетом прогноза, полученного на верхнем уровне, определяется качественный и количественный состав необходимых инфраструктурных объектов и производится их размещение на территории региона; на нижнем уровне моделируется работа отдельных объектов и анализируются экологические последствия полученного варианта, делаются выводы и уточняются модели верхнего и среднего уровней и формируются дополнительные ограничения (рис. 1).

Программным ядром ИСУ служит экспертная система (ЭС), позволяющая исследователям в интерактивном режиме выбирать методы и средства, подходящие для решения конкретной задачи развития инфраструктуры региона. Для используемых в рамках ИСУ моделей, методов и программных средств, реализующих необходимые вычисления, созданы формализованные описания, представленные в базе знаний фреймвого типа.

Для удобства разработки и поддержки в актуальном состоянии база знаний ЭС разбита на три части:

— МПО — онтологические модели предметных областей; здесь задаются понятия (отрасль, реги-

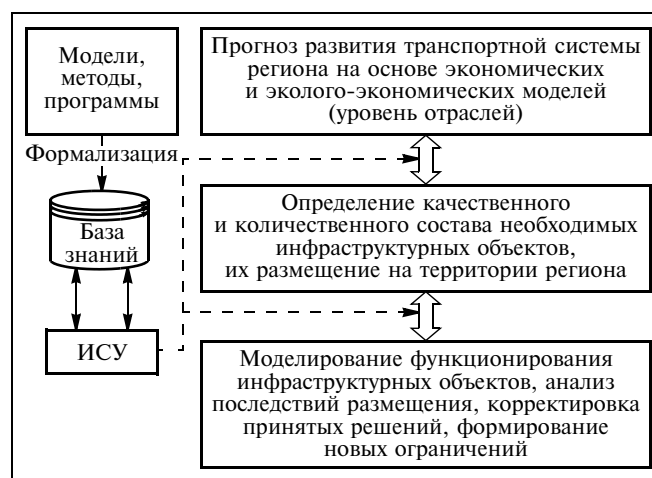


Рис. 1. Методология исследования транспортно-логистической инфраструктуры



Рис. 2. Типы ресурсов транспортно-логистической инфраструктуры

он, запасы, ресурсы, концентрация загрязняющих веществ и др.);

— МУМ — логическое описание взаимодействия уровней моделирования, оно включает такие термины, как модель, метод, прогноз развития отрасли, инфраструктурный объект, экологическое состояние территории, территориальное ограничение, схема корректировки параметров модели;

— МПМ — спецификация программных модулей и их взаимодействия; примерами используемых в ней понятий служат модель, метод, входные и выходные данные, настраиваемые параметры модуля, расчетная функция, файл данных и др.

Онтология МПО представляет собой оригинальную онтологию транспортно-логистической инфраструктуры региона. В ней используются существующие онтологии [10, 11], применяемые для описания различных частных проблем, возникающих в исследуемой предметной области. На рис. 2 представлена одна из ее основных частей — онтология типов ресурсов транспортно-логистической инфраструктуры.

Логическое описание взаимодействия уровней моделирования используется для представления связей между представлением инфраструктуры исследуемого объекта и поставленных пользователем задач, описанных в предметных терминах, с процедурами их решения. Эти связи формализованы в виде набора продукционных правил, используемых ЭС.

Спецификации программных модулей представляются в виде фреймов ЭС. Спецификация модуля включает в себя названия моделей и мето-

дов, реализованных в модуле; задачи (в терминах МУМ), которые он может решать; описания входных и выходных данных; настраиваемые параметры, режимы работы и т. п.

Примеры решаемых ИСУ задач: определение «узких мест» в транспортной системе региона, формирование рекомендаций по прокладке новых и изменению существующих маршрутов, определение местоположения терминалов и их количественного состава в условиях реализации различных сценариев развития региона. Возможные варианты изменений в транспортно-логистической инфраструктуре могут быть исследованы на уровне региональных многоотраслевых моделей.

Фрагмент онтологии логического описания взаимодействия уровней моделирования представлен на рис. 3.

2. АРХИТЕКТУРА ИСУ

В ИСУ вычислительный процесс планируется на основе интеграции модулей по данным: результаты работы каждого модуля служат исходными данными для другого модуля. В процессе передачи данных между модулями осуществляется преобразование их форматов. Для организации вычислительного процесса каждый включаемый в ИСУ модуль сопровождается спецификацией. В качестве основы представления данных, передаваемых между модулями, выбран формат XML, так как он поддерживается стандартными технологиями его обработки (трансляции, верификации, порождения и т. п.).

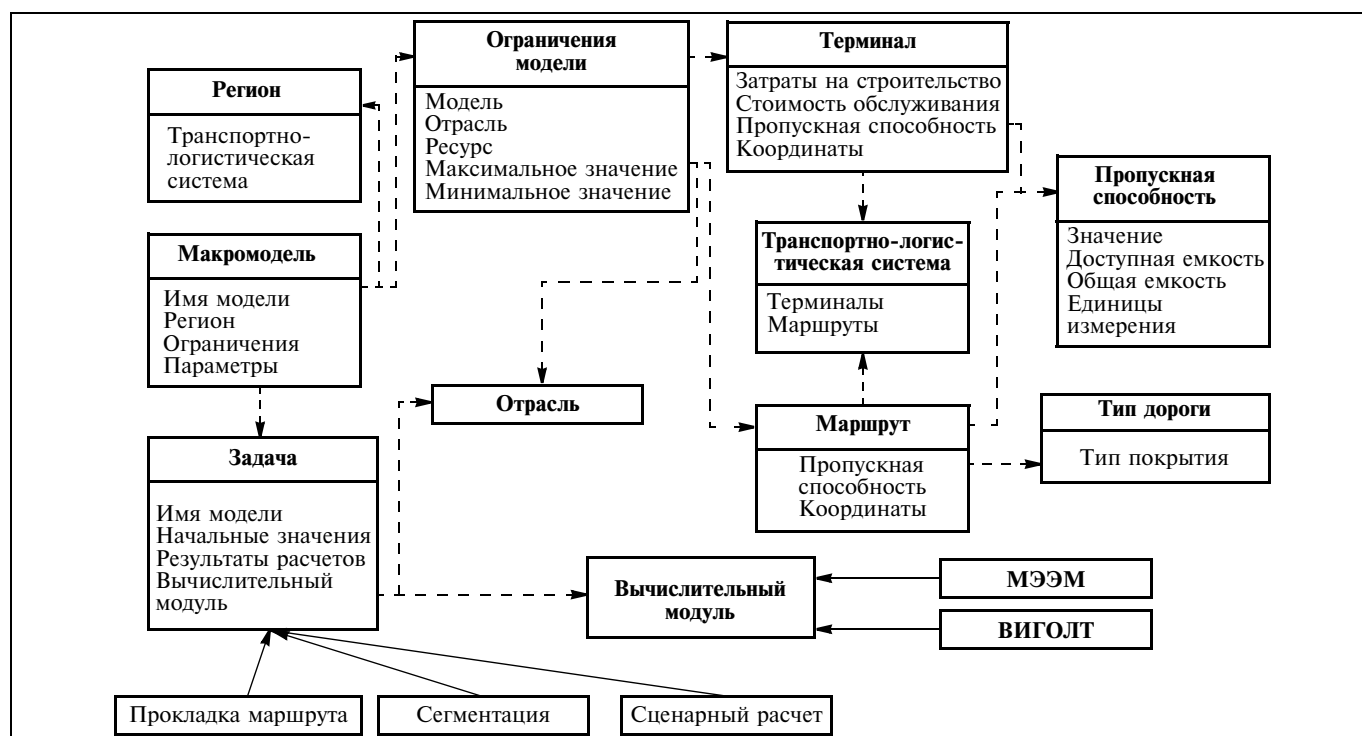


Рис. 3. Пример онтологии логического описания взаимодействия уровней моделирования: МЭЭМ, ВИГОЛТ — вычислительные модули (см. далее)

Проектирование и реализация ИСУ выполнены с помощью современных технологий и средств разработки программного обеспечения. Онтология предметной области разрабатывается в программной системе Protege [12]. База знаний ИСУ представлена в виде комбинации продукционных правил и фреймов. Обработка базы знаний осуществляется средствами механизма логического вывода системы CLIPS [13]. Продукционные правила сгруппированы по разделам, ЭС обрабатывает эти разделы независимо, осуществляя взаимодействие между ними через общее информационное пространство [14].

Функционирование ИСУ описывается последовательностью операций, обозначенных на рис. 4 цифрами: 1 — осуществляется выбор исследуемого региона и формулировка задач моделирования в терминах МПО и МУМ, запрос на вычисление от клиентского программного обеспечения к управляющему модулю ИСУ; 2 — запускается ЭС и выполняется логический вывод интерпретации задачи моделирования в терминах последовательности действий (плана), которую необходимо выполнить, чтобы получить решение задачи; 3 — в процессе вывода выполняется обмен служебной информацией между ЭС и УМ и вызов необходимых программных процедур; 4 — выбор вычислительных модулей, реализующих каждый этап синтезированного плана, при этом логический вывод

останавливается и текущее состояние сохраняется в рабочей памяти; на основе значений фреймов МПМ формируются входные данные для вычислительных модулей в формате xml; 5 — запускается полученная последовательность модулей, данные передаются между модулями в виде xml-файлов; 6 — передача результатов вычисления в управляющий модуль, который на их основе добавляет в БЗ новые факты в формате соответствующих фреймов МПМ и МУМ; 7 — запуск ЭС с обновленной базой знаний и передача результатов решения на анализ и выработку конкретных рекомендаций; 8 — формируется результат комплексного анализа в терминах МПО; 9 — результаты решений передаются от управляющего модуля к клиентскому программному обеспечению в терминах МПО.

Для иллюстрации возможностей применения разработанной ИСУ для комплексного изучения и управления транспортно-логистической инфраструктурой региона выбраны определенные вычислительные модули. В качестве средства моделирования на первом уровне исследования используется программный комплекс МЭЭМ (медико-эколого-экономическое моделирование) [15], предназначенный для анализа и прогнозирования экономического развития региона с учетом экологических и медицинских факторов, таких как загрязнение окружающей среды и его влияние на здоровье населения. В качестве математического аппарата взят

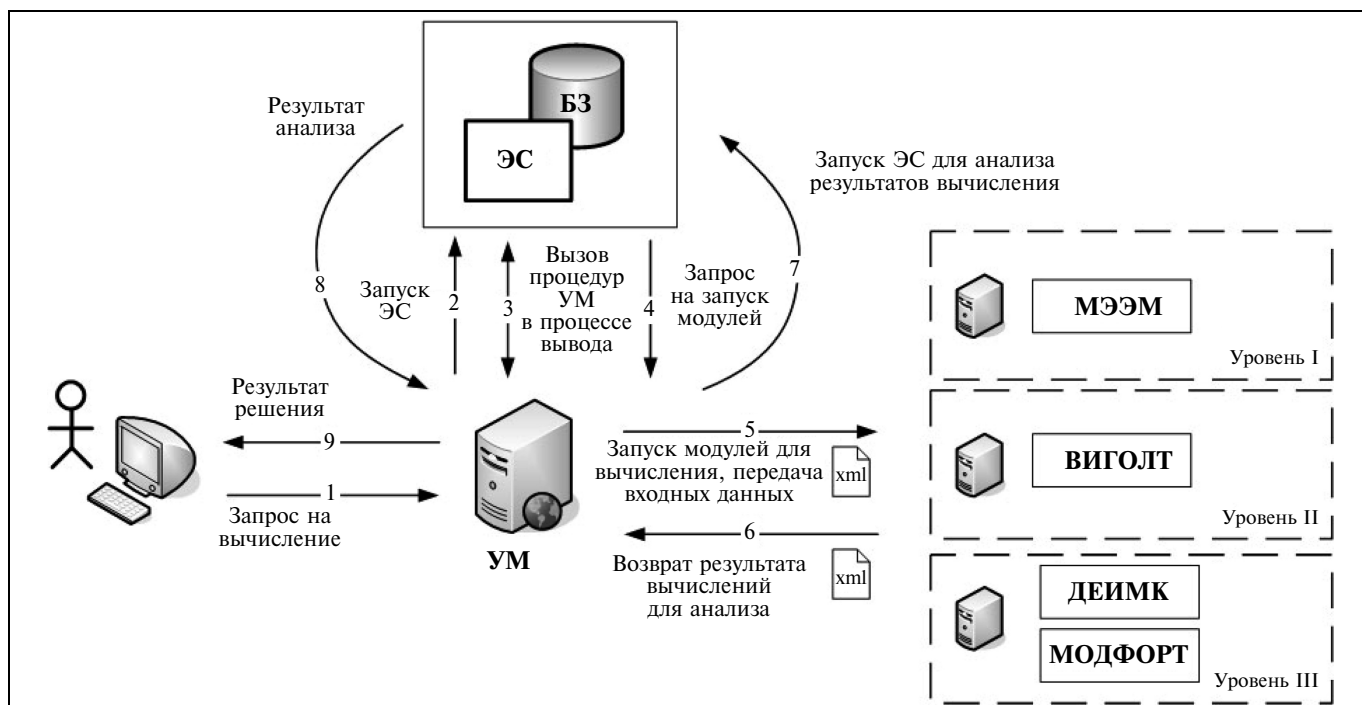


Рис. 4. Архитектура ИСУ: БЗ — база знаний; УМ — управляющий модуль; ДЕИМК, МОДФОРТ — вычислительные модули (см. далее)

ты динамические балансовые модели типа «Регион», предложена процедура исследования модели с применением магистральных решений.

Для второго уровня применяется система ВИГОЛТ (вариационное исчисление, геометрическая оптика, логистика, транспорт) [16, 17–20], предназначенная для решения задач построения оптимальных маршрутов прокладки коммуникаций, оптимального размещения логистических объектов и сегментации зон обслуживания. Применяется математический аппарат бесконечномерной оптимизации (вариационного исчисления) и оригинальные численные методы, в основе которых лежит аналогия между минимизацией интегрального функционала и распространением света в оптически неоднородной среде.

Для третьего уровня применяются программные комплексы МОДФОРТ (ModForT: Modeling, Forecasting, Transport) ДЕИМК (DEIMC: Differential Equations Inventory Management Computing) [21]. Первый предназначен для моделирования и прогнозирования работы грузовых терминалов железнодорожного транспорта. Предполагается, что грузы доставляются железнодорожным транспортом, а вывозятся автомобильным, причем входящий и исходящий транспортные потоки могут носить случайный или детерминированный характер, обслуженные заявки (принятые грузы) покидают терминал не напрямую, а через склад. В качестве математической модели применяется немарковская система массового обслуживания,

которая исследуется методами статистического моделирования (Монте-Карло). Более подробно данная модель описана далее. Программный комплекс ДЕИМК применяется для моделирования работы системы управления запасами (поставками), в которой поставки предполагаются непрерывными; учитывается наличие временного промежутка между принятием управленческого решения и его практической реализацией. В качестве математического аппарата использованы дифференциальные уравнения с отклоняющимся аргументом, при решении которых применяется авторский алгоритм, основанный на классическом методе шагов [22–24].

Рассмотрим пример спецификации модуля ВИГОЛТ. Он предназначен для решения ряда задач: прокладка оптимального маршрута (модель оптимальной организации коммуникаций); сегментация зон обслуживания (модель сегментации логистических зон обслуживания); размещение терминалов (модель оптимального размещения логистических объектов, модель сегментации логистических зон обслуживания); определение оптимального количества терминалов (модель оптимального размещения логистических объектов, модель сегментации логистических зон обслуживания); прокладка кратчайшей сети дорог (модель оптимальной организации коммуникаций).

Представим описание входных данных, используемых модулем ВИГОЛТ при решении каждой транспортно-логистической задачи.

— Прокладка оптимального маршрута: (x, y, p) , где p — некоторый исследуемый показатель (например, численность/плотность населения в точке (x, y) , стоимость прокладки маршрута, стоимость строительства инфраструктурного объекта); $A(x, y)$ — начальная точка искомого маршрута; $B(x, y)$ — конечная точка искомого маршрута.

— Сегментация зон обслуживания: (x, y, p) ; $A_i(x, y)$ — координаты расположения логистического объекта, $i = 1, \dots, k$, где k — число логистических объектов.

— Размещение терминалов: (x, y, p) ; $B_j(x, y)$ — координаты расположения потребителя, $j = 1, \dots, s$, где s — число потребителей.

— Прокладка кратчайшей сети дорог: (x, y, p) ; $A_i(x, y)$ — координаты расположения логистического объекта, $i = 1, \dots, k$, где k — число логистических объектов.

— Определение оптимального количества терминалов: (x, y, p) ; $B_j(x, y)$ — координаты расположения потребителя, $j = 1, \dots, s$, где s — число потребителей; C_i — стоимость открытия i -го терминала, $i = 1, \dots, k$.

Для того чтобы ЭС имела возможность оперировать описанными вычислительными модулями, они должны быть соответствующим образом описаны в терминах МПМ. Необходимо указать как технические данные (место хранения исполняемого файла, папки для загрузки исходных данных для вычислений и т. п.), так и содержательную информацию (список решаемых задач, поддерживаемых методов и моделей в терминах МУМ). После того, как закончен подготовительный этап по описанию программных модулей, начинается непосредственный процесс исследования транспортно-логистической инфраструктуры региона.

На *первом* этапе работы производится настройка ИСУ для конкретного региона: на основе предложенной онтологии предметной области необходимо ввести данные о состоянии региона и его транспортно-логистической инфраструктуры. Тем самым производится заполнение МПО.

Второй этап, согласно методологии исследования, состоит в постановке задач для макромоделей, где транспорт рассматривается как одна или несколько отраслей во взаимодействии с другими секторами экономики региона. На этом этапе у исследователя существует возможность многовариантного сценарного анализа, в ходе которого могут быть выявлены ситуации, когда экономическое развитие ограничивается возможностями транспортной отрасли. В этом случае ИСУ на третьем этапе должна предложить рекомендации по изменению транспортно-логистической инфраструктуры.

В зависимости от ситуации, экспертной системой пользователю будут предложены такие мероприятия, как прокладка новых и изменения существующих маршрутов, определение местоположения терминалов и их численности и т. п. Реализация рекомендаций основана на использовании моделей и методов, относящихся к среднему уровню комплексного исследования (см. рис. 1). Какие именно действия предпринять для реализации рекомендаций и с помощью каких средств эти действия осуществить, определяет ЭС. Основным источником информации для принятия решений ЭС на этом, *третьем*, этапе служит МУМ, но также организован диалог с пользователем, позволяющий получать дополнительную информацию для ЭС и влиять на процесс исследования. Результат третьего этапа — новая транспортно-логистическая инфраструктура, соответствующая некоторому заданному сценарию развития региона.

На *четвертом* этапе анализируются отдельные объекты транспортно-логистической инфраструктуры. Подробный пример моделирования работы такого объекта — грузового железнодорожного терминала — рассмотрен в § 3.

Для иллюстрации работы ЭС рассмотрим более подробно переход между вторым и третьим этапами исследования и приведем примеры продукционных правил, применяемых для анализа результатов первого уровня и последующего формулирования задач для второго уровня. Применение описанных далее правил реализует рекомендацию «если существуют ограничения по транспортной отрасли, то необходимо увеличить пропускную способность транспортной системы». Экспертная система предлагает несколько вариантов действий для выполнения этой рекомендации, формирует соответствующие постановки задач, определяет программные модули, способные решить эти задачи, устанавливает исходные данные и запускает модули на вычисление.

Первая группа правил в рассматриваемом примере применяется для анализа сценарных расчетов, проведенных на основе эколого-экономической модели с помощью программного комплекса МЭЭМ. Экспертная система должна определить, существуют ли на уровне отраслей региона ограничения по транспортной отрасли, которые не позволяют в полной мере реализовать инвестиционные проекты, заложенные в рассматриваемый сценарий. Получение ограничений осуществляется путем запуска программной процедуры, реализованной в управляющем модуле. Эта процедура должна получить результаты расчетов от соответствующего вычислительного модуля (в данном случае МЭЭМ) и представить их в формате фрейма «Ограничения модели» (см. рис. 3). В результате этой операции база знаний ЭС будет обновлена и



активированы правила, отвечающие за поиск инфраструктурных объектов, изменение которых может привести к устранению выявленных ограничений. Например, для такого элемента транспортно-логистической системы, как «Маршрут» (см. рис. 2 и 3), могут применяться правила:

ЕСЛИ выявленные ограничения связаны с *автотранспортом*,

ТО выбрать для анализа объекты типа «*Маршрут*» в моделируемом регионе,

ЕСЛИ для анализа выбран объект типа «*Маршрут*»,

ТО использовать рекомендацию «*Увеличение пропускной способности*» объекта.

На следующем шаге исследования ЭС определяет действия для инфраструктурных объектов (в данном случае «*Маршрут*»), связанные с рекомендацией «*Увеличение пропускной способности*»:

ЕСЛИ поставлена задача увеличения пропускной способности маршрута,

ТО задать следующие варианты действий для соответствующего объекта: «*Улучшение типа дорожного покрытия*», «*Прокладка новой полосы существующей дороги*», «*Прокладка нового маршрута*».

После этого будут активированы правила, осуществляющие подбор конкретных вычислительных модулей, предназначенных для реализации выбранных действий:

ЕСЛИ для объекта инфраструктуры региона установлено действие «*Прокладка нового маршрута*» **И** существуют модули, предназначенные для решения задачи «*Размещение*»,

ТО сформировать список вычислительных модулей.

При этом в правилах могут вызваться дополнительные процедуры, реализующие интерактивное взаимодействие с пользователем. Для описанного правила на экран будет выведен список модулей и предложено пользователю сделать выбор.

Пусть пользователем из предложенных вариантов выбран модуль ВИГОЛТ, тогда ЭС должна сформировать исходные данные для решения поставленной задачи с помощью процедуры *SetModuleData*, указав в качестве параметров: имя модуля, имя инструкции, объект с данными. Запуск модуля осуществляется процедурой *LaunchModule* с указанием его имени. Вызываемые процедуры реализованы в управляющем модуле. В общем виде правило запуска модуля представлено в следующем образом:

ЕСЛИ для решения выбран модуль ВИГОЛТ **И** типом выбранного объекта является «*Маршрут*» **И** существует карта моделируемого региона **И** известны затраты на строительство объекта,

ТО *SetModuleData* (ВИГОЛТ, «Карта», данные о карте моделируемого региона), *SetModuleData* (ВИГОЛТ, «Затраты по пространственным коор-

динатам», данные о затратах на строительство объекта по пространственным координатам), *SetModuleData* (ВИГОЛТ, «Режим работы», «Размещение»), *SetModuleData* (ВИГОЛТ, «Объект», выбранный объект типа «Маршрут»), *LaunchModule* (ВИГОЛТ).

Таким образом, приведенный пример иллюстрирует реализацию упомянутой рекомендации «если существуют ограничения по транспортной отрасли, то необходимо увеличить пропускную способность транспортной системы».

3. МОДЕЛИРОВАНИЕ РАБОТЫ ГРУЗОВОГО ЖЕЛЕЗНОДОРОЖНОГО ТЕРМИНАЛА

В качестве примера построенной математической модели и численного метода ее исследования для решения задач третьего уровня рассмотрим математическую и имитационную модели функционирования грузового железнодорожного терминала.

Под грузовым терминалом обычно понимается комплекс сооружений, технологических и технических устройств, которые предназначены для обработки и хранения различных грузов. Предполагается, что доставка грузов осуществляется железнодорожным транспортом, а вывоз — автомобильным. Поскольку входящий и исходящий транспортные потоки носят, как правило, случайный характер, наиболее адекватное математическое описание работы грузового терминала можно выполнить с помощью многоуровневой стохастической модели на основе применения математического аппарата немарковских систем массового обслуживания (СМО).

Система массового обслуживания характеризуется, как известно, наличием потока заявок, имеющего случайный характер, а также обслуживающих заявки каналов, время нахождения заявок в которых также, как правило, является случайным. В случае, если все каналы заняты, поступившая заявка может быть направлена в очередь либо отклонена. Мы будем дополнительно предполагать, что на вход поступают сложные заявки, состоящие из нескольких простых (мультизаявки); обработанная заявка покидает систему не мгновенно, а через склад, что, вообще говоря, служит чертой, отличающей движение материальных потоков в СМО от схожих по свойствам информационных.

Таким образом предполагается, что заявки могут находиться в четырех разных фазах: во входящем потоке, в очереди, в каналах и в выходящем потоке через склад. Поэтому построенную модель может разделить на уровни:

- потока, где задаются параметры заявок;
- очереди (входа), где задаются параметры очереди;

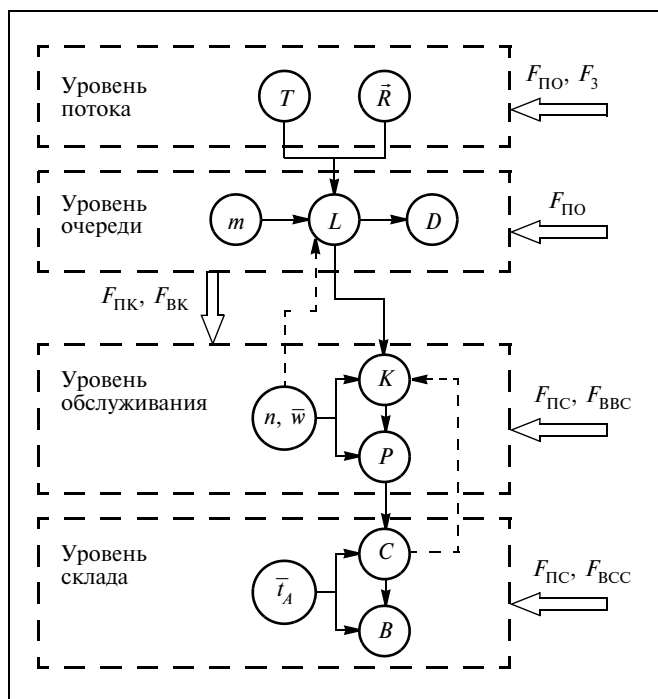


Рис. 5. Структура математической и управляющие функции имитационной модели

— канала (обслуживания), где задаются параметры обслуживания;

— выхода (склада), где задаются параметры склада и выходящего потока.

Поскольку аналитически исследовать подобного рода математические модели не представляется возможным, авторами была разработана и реализована имитационная модель, схема которой приведена на рис. 5.

В математической модели T — время поступления мультizaявки; $\vec{R} = (R_1, \dots, R_l)$ — вектор количества заявок в мультizaявке, l — количество типов заявок в мультizaявке; m — число мест в очереди; $L = L(t, T, \vec{R}, \vec{w}, m, n)$ — количество заявок в очереди; $D = D(t, L)$ — количество заявок, получивших отказ; n — число каналов обслуживания; $\vec{w} = (w_1, \dots, w_l)$ — вектор интенсивности обслуживания; $K = K(t, L, \vec{w}, C)$ — количество заявок в каналах (обслуживаемых); $P = P(t, K, \vec{w})$ — количество принятых заявок; $\vec{t}_A$ — средний интервал выгрузки; $C = C(t, P, \vec{t}_A, V)$ — состояние склада; $B = B(t, C, \vec{t}_A)$ — число выгруженных заявок.

В имитационной модели $F_{\text{во}}$ — функция генерирования времени поступления; F_3 — функция генерирования числа заявок в мультizaявке; $F_{\text{по}}$ — функция проверки состояния очереди; $F_{\text{ПК}}$ — функ-

ция проверки состояния каналов; $F_{\text{ВК}}$ — функция управления поступлением заявок в канал; $F_{\text{ПС}}$ — функция проверки состояния склада; $F_{\text{ВВС}}$ — функция генерирования времени обслуживания в каналах; $F_{\text{ВСС}}$ — функция генерирования времени выгрузки грузов со склада.

Имитационная модель, в частности, позволяет определять параметры эффективности работы грузового терминала: вероятность отказа, среднее время ожидания обслуживания, среднюю длину очереди, среднее число занятых каналов обслуживания, время простоя и др. Выполнена программная реализация имитационной модели в рамках программного комплекса МОДФОРТ.

4. ПРИКЛАДНЫЕ ЗАДАЧИ

С помощью построенных моделей и разработанных программных комплексов был решен ряд прикладных задач. Отметим две из них.

Выполнены численные расчеты для проектируемого участка железнодорожной высокоскоростной магистрали (ВСМ) между городами Иркутск и Красноярск. В качестве исходных данных использовалась информация о высоте над уровнем моря для каждой точки местности, о расположении населенных пунктов, болот, о речной сети и сети дорог. Как известно, при прокладке маршрута ВСМ большое значение имеют ограничения на кривизну маршрута (радиус кривизны — не менее 4 км) и перепад высот (не более 10 м на один километр) [25]. В необходимые допуски в основном удалось уложиться. Полученный маршрут ВСМ обладает большей гладкостью, чем маршрут существующей железной дороги. Тем не менее, на участке вблизи г. Тайшет имеется два «перехода» через горный хребет, в которых, по всей видимости, потребуется сооружение туннелей [26].

Выполнены расчеты для задачи о сегментации логистических зон обслуживания в регионе на примере Иркутской области. В случае, когда территория заселена равномерно, решение данной задачи не представляет особой сложности. Однако в случае неравномерной заселенности, как в большинстве регионов Сибири и Дальнего Востока, возникает противоречие между требованием равномоности логистических центров, что удобно с технологической точки зрения, если речь, например, идет о зонах утилизации (или, скажем, о предоставлении государственных услуг) и условием обеспечения равной доступности для граждан. Авторам представляется, что им удалось найти разумный компромисс между этими двумя требованиями. Более подробное описание методики можно найти в работе [18].



ЗАКЛЮЧЕНИЕ

В работе представлено описание архитектуры интеллектуальной системы управления, включающей в себя управляющий модуль, экспертную систему, базу знаний и разработанные авторами ранее вычислительные модули (МЭЭМ, ВИГОЛТ, МОДФОРТ, ДЕИМК). Разработана оригинальная онтология транспортно-логистической инфраструктуры региона, формализующая взаимодействие и взаимосвязь различных ресурсов (вид и тип транспорта, ресурса, терминалов обслуживания и др.). Представлен пример продукционных правил, составляющих основу базы знаний, посредством которой осуществляется принятие решений экспертной системой. Разработанная интеллектуальная система управления позволит частично заменить эксперта-человека в процессе исследований, сняв необходимость проведения рутинных вычислений, обеспечит оперативный анализ и интеграцию результатов решения задачи и принятие решений.

ЛИТЕРАТУРА

1. Юданов А.Ю. Логистика: учебник для вузов. — М.: Тандем, 2005. — 400 с.
2. Модели и методы теории логистики / Под ред. В.С. Лукинского. — СПб.: Питер, 2007. — 448 с.
3. Глухов В.В., Медников М.Д., Коробко С.Б. Математические методы и модели для менеджмента. — СПб.: Лань, 2005. — 528 с.
4. Strekalovsky A.S., Orlov A.V., Malyshev A.V. On computational search for optimistic solution in bilevel problems // Journal of Global Optimization. — 2010. — Vol. 48, N 1. — P. 159–172.
5. Казаков А.Л., Маслов А.М. Построение модели неравномерного транспортного потока на примере железнодорожной грузовой станции // Современные технологии. Системный анализ. Моделирование. — 2009. — № 3. — С. 27–32.
6. Bougromenko V. Defending local interests in project management: from team to decision making level // International Journal of Project Management. — 1999. — Vol. 17, N 2. — P. 131–135.
7. Schreckenberg M., Chrobok R.Y., Hafstein S.F., Pottmeier A. OLSIM — Traffic Forecast and Planning using Simulations // 17 Symposium «Simulationstechnik» ASIM 2003, Magdeburg, 16.09 bis 19.09.2003. — SCS—Europe, 2003. — P. 11–18.
8. Traffic simulation, reconstruction, and route planning. — URL: <http://gamma.cs.unc.edu/research/traffic/> (дата обращения: 21.11.2013).
9. Бычков И.В., Ружников Г.М., Хмельнов А.Е. и др. Интеграция информационно-аналитических ресурсов и обработка пространственных данных в задачах управления территориальным развитием. — Новосибирск: Изд-во СО РАН, 2011. — 369 с.
10. Smith S., Lassila O., Becker M. Configurable, mixed-initiative systems for planning and scheduling. In: Tate A. (ed): Advanced Planning Technology: Technological Achievements of the ARPA. — Menlo Park: AAAI Press, — 1996. — P. 235–241.
11. Becker M., Smith S. An Ontology for Multi-Modal Transportation Planning and Scheduling. Tech. report CMU—RI—TR—98—15. — The Robotics Institute, Carnegie Mellon University Pittsburgh, 1997. — 85 p.
12. Protege is a free, open source ontology editor and knowledge—base framework. — URL: <http://protege.stanford.edu/> (дата обращения 10.12.2013).
13. CLIPS: A Tool for Building Expert Systems. — URL: <http://clipsrules.sourceforge.net/> (дата обращения 10.12.2013).
14. Craig I. Blackboard Systems. — Norwood: Ablex Pub. Corp., 1995. — 241 p.
15. Павлов А.И., Столбов А.Б. Программный комплекс для поддержки моделирования медико-эколого-экономических систем // Программные продукты и системы. — 2011. — № 1. — С. 137–140.
16. Бухаров Д.С., Казаков А.Л. Программная система «ВИГОЛТ» для решения задач оптимизации, возникающих в транспортной логистике // Вычислительные методы и программирование. — 2012. — Раздел 2. — С. 65–74.
17. Казаков А.Л., Лемперт А.А. Об одном подходе к решению задач оптимизации, возникающих в транспортной логистике // Автоматика и телемеханика. — 2011. — № 7. — С. 50–57.
18. Казаков А.Л., Лемперт А.А., Бухаров Д.С. Об одном численном методе решения некоторых задач оптимизации, возникающих в транспортной логистике // Вестник ИРГТУ. — 2011. — Т. 53, № 6. — С. 6–12.
19. Казаков А.Л., Лемперт А.А., Бухаров Д.С. Математическая модель и программная система для решения задачи размещения логистических объектов // Управление большими системами. — 2013. — Вып. 41. — С. 270–284.
20. Казаков А.Л., Лемперт А.А., Бухаров Д.С. К вопросу о сегментации логистических зон для обслуживания непрерывно распределенных потребителей // Автоматика и телемеханика. — 2013. — № 6. — С. 87–100.
21. Фу Ф.Г., Казаков А.Л. Имитационное моделирование работы грузовых транспортных терминалов // Вестник ИРГТУ. — 2013. — № 9 (80). — С. 37–43.
22. Казаков А.Л., Лемперт А.А., Фунг Т.Б. Математическая модель управления запасами (поставками) с учетом запаздывания // Вестник ИРГТУ. — 2012. — № 4. — С. 131–137.
23. Лемперт А.А., Фунг Т.Б. Исследование устойчивости моделей управления запасами в непрерывной постановке // Вестник ВСГУТУ. — 2013. — № 4 (43). — С. 5–14.
24. Фунг Т.Б., Лемперт А.А. О моделировании процесса управления запасами с учетом запаздывания для двух видов товаров народного потребления // Вестник ИРГТУ. — 2013. — № 9 (80). — С. 43–52.
25. Журавская М.А., Казаков А.Л., Лемперт А.А. и др. О методе решения задачи оптимальной прокладки высокоскоростных железнодорожных магистралей с учетом региональных особенностей // Транспорт: наука, техника, управление. — 2012. — № 2. — С. 41–44.
26. Бухаров Д.С., Казаков А.Л. Трассировка на цифровой карте: математическая модель и численный метод // Транспорт Урала. — 2013. — № 3. — С. 7–12.

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

Бычков Игорь Вячеславович — академик РАН, директор, ☎ (3952) 42-71-00, ✉ bychkov@icc.ru,

Казаков Александр Леонидович — д-р физ.-мат. наук, гл. науч. сотрудник, ☎ (3952) 45-30-33, ✉ kazakov@icc.ru,

Лемперт Анна Ананьевна — канд. физ.-мат. наук, зав. лабораторией, ☎ (3952) 45-30-30, ✉ lempert@icc.ru,

Бухаров Дмитрий Сергеевич — программист, ☎ (3952) 45-30-30, ✉ bukharovds@gmail.com,

Столбов Александр Борисович — мл. науч. сотрудник, ☎ (3952) 45-30-30, ✉ stolboff@icc.ru,

Институт динамики систем и теории управления СО РАН, г. Иркутск.

ЧЕТЫРЕХЭТАПНАЯ МОЗГОВАЯ АТАКА

Ю.В. Сидельников

Путем анализа многих модификаций мозговой атаки выявлены их недостатки и предложена четырехэтапная мозговая атака, которая устраняет некоторые из них. Отмечено, что данная модификация может повысить эффективность мозговой атаки и содействовать исследователю в получении новой, в том числе прогнозной, информации. Для этой цели усилен вклад подготовительного этапа в общую эффективность метода и добавлен четвертый этап, в рамках которого предусмотрена возможность переформулирования постановки задачи на основании информации, полученной на предыдущих этапах.

Ключевые слова: экспертные оценки, модификация, мозговая атака, идея, участник.

ВВЕДЕНИЕ

Прежде чем описать новую модификацию процедуры классической мозговой атаки и обосновать сформулированную цель настоящей статьи:

— рассмотрим суть метода мозговой атаки, общую для любой его модификации, укажем, на чем он основывается, а также опишем его основные характеристики;

— дадим краткую историческую справку о классической мозговой атаке;

— перечислим модификации мозговой атаки;

— укажем недостатки метода.

Суть метода мозговой атаки. В любой из модификаций мозговой атаки используется то обстоятельство, что одни люди хорошо генерируют идеи, но плохо справляются с их анализом, а другие больше склонны к критическому анализу идей, чем к их генерированию [1, 2].

Мозговая атака (мозговой штурм¹), как инструмент, содействует выходу новых идей из подсознания участников группы генераторов. Этот метод был создан в те годы, когда в теории Фрейда видели универсальный ключ ко многим проблемам в социологии, психологии и искусстве и, по сути, он базировался на этой теории. Кроме того, мозговая атака возникла как попытка устранить одну из наиболее серьезных помех творческому мышлению — боязнь критики выдвигаемых идей. В целях устранения этой помехи в процессе применения метода предлагается выдвижение и анализ лю-

бых идей (в том числе самых фантастических, явно ошибочных, шуточных), так как они могут стимулировать появление более ценных идей. Этот коллективный метод решения задач предоставляет возможность развивать идеи друг друга.

На наш взгляд, мозговую атаку нельзя назвать экспертизой в строгом смысле этого понятия, так как участие в ней людей на этапе генерации идей, не являющихся специалистами по рассматриваемому вопросу, не только допускается, но даже поощряется.

Укажем на основные характеристики метода, которые определяют (идентифицируют) его принадлежность к совокупности модификаций мозговой атаки.

- Разделение по времени и по группам участников процессов генерации и критического анализа. Таким образом, сначала одна группа участников, получив задачу, только выдвигает идеи, другая же группа, но уже экспертов — анализирует полученные предложения. Как минимум, это двухэтапная процедура решения задачи.
- Дискурсивная² хаотизация перебора как реальный механизм получения новой информации от экспертов [3, 4].
- Управляемая дискуссия в малых группах участников и экспертов.

Историческая справка. Процедура мозговой атаки была предложена в середине прошлого столетия американским психологом Алексом Ф. Осборном

¹ Оба эти словосочетания — варианты перевода английского термина «brainstorming», однако иногда они употребляются в разных сферах и выполняют разные функции.

² О дискурсивности хаотизации перебора говорится с целью подчеркнуть, что хаотизируется познавательный процесс, протекающий в рамках уже имеющихся представлений и смыслов, означенных, оформленных, опредмеченных полностью или частично.



Модификации мозговой атаки

Модификация	Год	Первоисточник
Процедура синектики	1961	[6, 7]
Конференция идей, как метод на базе стандартной мозговой атаки	1970	[8, 9]
Обратный штурм	1981	[17, 18]
Массовая мозговая атака	1984	[11]
Визуальный мозговой штурм	1985	[12]
Деструктивно-конструктивный штурм	1985	[10]
Двухстадийный штурм (мозговая атака с двойной инверсией)	1991	[13]
Электронный мозговой штурм	1992	[14, 15]
Мозговая осада	1995	[16]
Метод "I-G-I"	1997	[19]
Индивидуальный штурм	2003	[20]
Поэтапный штурм	2005	[21]
Мозговой штурм со звездным импульсом (звездный мозговой штурм)	2011	[22]

(Alex Faickney Osborn). Как метод, мозговая атака стала известна широкому кругу специалистов с выходом в 1942 и 1953 гг. его книг [1, 2]. Хотя, конечно, базовым механизмом³ получения нового, лежащим в основе данного метода, и составными элементами метода пользовались гораздо раньше. Перечислим примеры, приведенные в обзоре [5]. Такой составной элемент мозговой атаки, как двухэтапная процедура решения задачи, использовался достаточно давно. Об этом писал древнеримский историк Тацит, исследовавший быт германцев. Еще одним примером того же подхода служит способ решения сложных задач, применявшийся в Древней Вифинии в 700—600 гг. до нашей эры. Существуют и современные варианты мифов о применении метода мозговой атаки в истории научно-технических инноваций, и связывают их с поиском защиты от торпедных атак.

В последние десятилетия исследователями предложено более тридцати модификаций мозговой атаки. Основные из них представлены в таблице.

Недостатки метода, присущие любой модификации мозговой атаки. По мнению А.Б. Соколова, «в литературе отсутствует полноценное методическое описание мозгового штурма, позволяющее обычному человеку использовать его для решения творческих задач. Помимо призывов «мыслить оригинально» и «отказаться от стереотипов» в литературных источниках не дается рекомендаций касательно того, как избавиться от этих самых стереотипов ...» [23].

³ Понятие базового механизмом получения нового впервые было введено в работе [4].

- Велика роль ведущего — от него зависит половина успеха [24].
- Бывают затруднения при определении авторства хороших идей [24].
- Неэффективность при решении изобретательских задач высших уровней⁴ [25].
- «Бестолковость» поисков, возведенная мозговой атакой в принцип. Четкое управление мышлением исключается. (Алгоритм целенаправленного движения к оригинальному решению отсутствует). И, хотя эти методы действительно помогают преодолевать инерцию: мысль сдвигается с «мертвой точки», разгоняется... и часто проскакивает то место, где следовало бы остановиться и развить перспективную мысль. (В ходе тестовых экспериментов десятки раз наблюдалась такая картина: один участник штурма высказывает идею, ведущую в перспективном направлении, другой подхватывает эту мысль и развивает ее; кажется, до выхода на финишную прямую остается несколько шагов, но в этот момент кто то выдвигает совершенно иную идею, «перспективная» цепь обрывается, и группа снова оказывается на исходных позициях) [24, 25].
- Существенные временные затраты. Внешне все выглядит эффектно — задача решается за один день. Но выигрыш этот в значительной мере кажущийся, поскольку задачу решает многочисленный коллектив участников. Таким образом, оказывается, что мозговой штурм (учитывая время на предварительную подготовку) требует достаточно большого количества человеко-часов [24, 25].
- В связи с тем, при мозговой атаке поощряется генерирование любых идей, даже фантастических, зачастую его участники уходят от реальной проблемы. В потоке разнообразных предложений бывает порой довольно трудно найти рациональные и продуктивные идеи. Кроме того, метод не гарантирует тщательную разработку предлагаемой идеи» [17, с. 209].
- Из-за высокой степени вовлеченности участников совещания ответственность за конечный результат несут все, и если идеи есть у всех, затраты времени на их обсуждение возрастают [26].
- При слабой обученности персонала сотрудничеству и командной работе, участники совещания могут быть не удовлетворены эффективностью своей деятельности. Кроме того, многие участники могут настаивать на своем авторстве обсуждаемых идей и предпочитают быть лиде-

⁴ Первоначальный текст данного пункта обоснованной критики мозговой атаки был рассмотрен Г.С. Альтшуллером в 1973 г. в его книге «Алгоритм изобретения» (2-е изд., испр., М.: Московский рабочий, 1973. — 296 с.).

рами творческого процесса за счет тех, кто менее развит и подготовлен [27].

- Недостаточно развитая способность осуществлять дистинкцию вызывает трудности по выбору из большого числа наработанных идей только тех, которые будут реально способствовать решению проблемы или задачи и, следовательно, их можно перевести в конкретные действия [26].
- Непригодность для решения сложных проблем и трудных задач. (Например, такой: «Свободен ли солдат от этических норм?». Это дискуссионный вопрос, ведь солдат принял присягу выполнять любые приказы. А если приказ безнравственный?) [24].
- Не имеет критериев сила решений [24].
- Не существует эффективной методики, которая позволила бы определить уровень качества проведения мозговой атаки⁵.
- Нет критериев оценки полученных идей⁶.
- В ходе мозговой атаки явная критика запрещена, но она почти неизбежно заменяется скрытой критикой в форме выдвижения новых предложений, пресекающих развитие других идей [25].

Цель настоящей работы состоит в повышении потенциальной эффективности метода мозговой атаки, путем создания его новой модификации и устранения первых трех из перечисленных недостатков.

ПРЕДЛАГАЕМАЯ МОДИФИКАЦИЯ КЛАССИЧЕСКОЙ МОЗГОВОЙ АТАКИ — ЧЕТЫРЕХЭТАПНАЯ МОЗГОВАЯ АТАКА

Этап 1, подготовительный. Его проводят заказчик и (или) ЛПР — лицо, принимающее решение, а также подобранные ими ведущие: выделенный эксперт (ВЭ) и специалист по процедуре (СП), который берет на себя обязанности фасилитатора⁷. При решении прогнозных задач такой специалист обычно называется прогнозистом, а при решении задач, связанных с экспертизой — экспертологом⁸.

Обязанности заказчика и (или) ЛПР

1. Поставить единственную задачу исследования, которую необходимо:

- 1.1) по возможности четко сформулировать и записать (в общих понятиях);

⁵ Замечено О.Д. Григоренко.

⁶ Этот недостаток был указан И.Л. Викентьевым в его работе «Система «ТРИЗ-ШАНС», 2000 г.».

⁷ Фасилитатор обеспечивает успешную групповую коммуникацию и соблюдение правил и регламента мозговой атаки.

⁸ Понятие экспертолога введено в работе [28].

- 1.2) представить в форме, наиболее удобной для участников [9];
- 1.3) в случае сложной задачи задать ее постановку слегка размыто (при этом посылка 1.1 не отвергает посылку 1.3). Данное положение для модификаций мозговой атаки впервые было рассмотрено в работе [7], а для модификаций метода фокальных объектов — в статье [29].

2. Решить организационные вопросы, включая определение времени и места проведения работы [9], подготовку помещения и оборудования (флип-чарт или просто доска), а также обеспечить:

- наличие техники фиксации мнений участников групп генераторов и аналитиков;
- транспаранты с ключевыми требованиями к участникам;
- запрет на присутствие посторонних лиц, включая начальство;
- комфортные условия работы участников и ведущих, например, обеспечить напитками (кофе, вода, чай).

3. Обеспечить финансирование деятельности всех участников.

4. Определиться с проблемой авторства и приоритета идей и решений, найденных предлагаемым методом. Эта проблема бывает довольно сложной. Обычно порядок установления авторства должен быть оговорен заранее. Имеются, например, варианты:

- установление авторства за лицами, сделавшими первый шаг в формировании новой идеи;
- установление авторства за лицами, сделавшими заключительный шаг в формировании новой идеи;
- признание всех участников авторами на одинаковых правах⁹ (далее мы будем придерживаться этого варианта).

5. Обеспечить возможность стимулирования участников групп генераторов и аналитиков, ведущих, а также самих групп как команд на основе теоретических и прикладных разработок психологии малых групп [30, 31].

6. Определиться с общей продолжительностью всех этапов (заседаний).

7. Определиться с требованиями к двум ведущим и критериями по их отбору.

Поясним данную задачу. Роль ведущих при проведении мозговой атаки многозначна, трудна и чрезвычайно важна. Конечно, требования к ведущим вытекают из их обязанностей. Эффективный отбор ведущих базируется на перечне их обязан-

⁹ Подобный вариант затрудняет решение проблемы индивидуальной мотивации участников.



ностей. Кроме того, на наш взгляд, критерии по отбору ведущих мозговой атаки определяются спецификой задачи, которую должны решать участники процедур с помощью ведущих.

Однако ряд априорных требований к ведущим можно указать и без учета специфики решаемой задачи. Так, например, А. Осборн предлагал выбирать ВЭ среди лиц, обладающих высокой творческой активностью в сочетании с доброжелательностью по отношению к идеям, высказанным другими людьми. Кроме того, такой ведущий должен органично сочетать в себе качества генератора и аналитика идей. Важно, чтобы он обладал быстрой реакцией, богатством ассоциативных связей, легкостью генерирования идей в сочетании с хорошими аналитическими способностями и чувством юмора [2].

8. В соответствии с обоснованно выбранными критериями отобрать двух ведущих процедуры.

9. Озвучить и реализовать процедуру стимулирования ведущих.

Обязанности СП и стоящие перед ним задачи

1. На данном этапе СП должен определить, имеет ли смысл применить рассматриваемую модификацию мозговой атаки как инструмент решения задачи, поставленной ЛПР и (или) заказчиком. Для этого он, например, может выяснить априорную оценку сложности рассматриваемой задачи [3].

1.1. Если СП сумел обосновать решение о неэффективности применения данной модификации мозговой атаки, то он должен передать соответствующее заключение заказчику и ЛПР и закончить процедуру решения задачи или предложить другой метод решения.

1.2. В случае, если СП принял положительное решение о возможности использования этой модификации мозговой атаки для решения задачи, которая была поставлена заказчиком, то он должен решать следующие вспомогательные задачи.

2. Определиться с критериями, по которым отбирать генераторов идей. Можно указать на такие критерии: «хорошие творческие способности; большую скорость мыслительных процессов; легкость включения в новые ситуации; гибкость (по-видимому, имеется в виду гибкость мышления — Ю.С.), способность быстро переключать внимание с одного аспекта на другой. Кроме того, умение отойти от привычных установок и психологических «тормозов», ограничений, что позволит им расширить область возможностей» [32].

3. Определиться с критериями, по которым отбирать аналитиков. Поясним эту задачу. К участникам этапа анализа идей (аналитикам) предъявляется очень широкий набор требований. «Важней-

шим требованием при отборе в эту группу является наличие у участников группы аналитиков творческих способностей. Естественно, что они должны быть интеллектуалами, обладать логическим, упорядоченным мышлением. Кроме того, важно, чтобы аналитики не относились ревниво к чужим идеям и отличались терпимостью к новым подходам. Эти люди должны обладать чувством повышенной ответственности за свое дело. Они, несомненно, должны быть оптимистами, но их оптимизм основывается на предположении, что лучшая идея — это та, которая рассматривается в данный момент» [1, 2]. Другие важные свойства, необходимые аналитику: выдержка, умение распределять свои силы на длительный срок. Ведь если генератора идей можно сравнить со спринтером, вкладывающим все свои силы в короткий рывок, то аналитик — это, несомненно, стайер.

4. Определиться с критериями, по которым формировать группы генераторов идей и аналитиков по структуре и составу. Существует точка зрения некоторых авторов, что в группу генераторов идей должны входить люди различных специальностей [33].

5. Определиться со способами формирования групп генераторов идей и аналитиков. Можно указать на такие способы: назначения; взаимные рекомендации («снежный ком»); последовательные рекомендации; выдвижение коллективами научных подразделений; документационный; тестирование [24]. Отметим, что при описании экспертного метода¹⁰ в научной литературе обычно способ (способы) формирования экспертной группы либо совсем не указан, либо уже зафиксирован, но это не всегда эффективно.

6. Определиться с числом участников малых групп генераторов и аналитиков и согласовать эти предложения с ЛПР и (или) заказчиком.

7. Определиться с пределами времени выступлений каждого из участников, а также количеству их выступлений на каждом заседании с тем, чтобы уложиться в общую продолжительность заседаний всех этапов, которая была задана заказчиком и (или) ЛПР).

8. Обеспечить процедурное отличие мозговой атаки на этапе генерации от совещания. Например, не ставить своей задачей получение мнения каждого из участников (это как раз совещание), а инициировать лавинообразное высказывание как можно большего числа необычных, оригинальных мнений или идей.

¹⁰ Имеются в виду лишь те методы, в которых предполагается работа группы участников.

Обязанности ВЭ, вытекающие из задач, поставленных перед ним ЛПР и (или) заказчиком

1. Подготовить предложения по персональному составу групп генераторов и аналитиков, принять их совместно с прогнозистом.

2. Разослать письменные приглашения участникам этих групп с указанием цели мозговой атаки и разъяснением роли каждого из них в решении поставленной задачи. (Очень важно, чтобы еще до начала все в группе были согласны с правилами проведения массовой атаки и представляли себе рамки задания [1, 2]). Кроме того, предложить участникам командное авторство идей и пояснить трудности относительно персонального авторства этих идей.

3. Подготовить краткую ориентирующую участников информацию об объекте (процессе) исследования, а также дополнительные материалы и информацию, касающиеся поставленной задачи (причины неудач аналогичных исследовательских попыток, существующие ограничения различного рода и др.). В ряде случаев бывает полезно подготовить для участников перечень основных направлений, по которым могут быть улучшен (изменен) тот или иной процесс, работа или продукт.

4. Предварительно проработать рассматриваемые вопросы [9].

Этап 2. Генерация идей. Его проводят двое ведущих: ВЭ и СП, кроме того, принимают участие ЛПР и (или) заказчик и участники из группы генераторов идей. Он реализуется лишь на одном заседании.

Если число выступлений любого участника из группы генераторов идей формально не ограничивать, то время каждого из его выступлений должно быть не более нескольких минут, а общая продолжительность заседания — не более часа.

На данном этапе осуществляется «принудительный» опрос экспертов, а затем инициируется лавинообразное высказывание мнений или идей. Обычно предлагают два пути организации этого этапа: «упорядоченный», когда каждый член группы высказывает идеи в порядке очередности, и «неупорядоченный», когда члены группы высказывают идеи по мере их возникновения. Нам представляется, что сначала члены группы должны высказываться в порядке очередности, но, как только процесс генерации идей начал действовать, члены группы должны высказывать идеи по мере их возникновения, даже если это будет и одновременно.

Обязанности СП и вспомогательные задачи, стоящие перед ним

1. Сформировать (совместно с ВЭ) персональный состав группы генераторов на основе предложений по ее составу, полученных на подгото-

вительном этапе¹¹, и утвердить его у ЛПР или заказчика¹². Число участников этой группы должно определяться, исходя из требований психологии малых групп [30, 31]. Обычно число участников от 5 до 15.

2. Ознакомить участников группы генераторов с процедурой проведения мозговой атаки, требованиями к ее участникам и их правами:

- каждый из участников должен на время проведения этого этапа забыть, что он начальник или подчиненны в рамках данной группы;
- каждый из участников должен считать поставленную на рассмотрение задачу актуальной;
- при генерировании идей необходимо свободно высказывать любые идеи, в том числе явно ошибочные, шуточные, фантастические (идеи выдвигаются без обоснования);
- критика предлагаемых идей и предложений, даже самых абсурдных, не допускается (не только словесная, но и невербальная — в виде, например, скептических улыбок и др.);
- участники мозговой атаки равноправны;
- разрешаются высказывание только новых идей и выступления в развитие предыдущего предложения.

3. После ознакомления участников группы с правилами проведения этапа и перед началом генерирования идей можно провести совместно с ВЭ имитацию мозговой атаки, которая некоторыми из авторов называется «разогревом» [23]. Если группа генераторов не имеет опыта участия в мозговых атаках, то «разогрев» необходим, чтобы мыслительный процесс был максимально эффективен. Можно предложить нейтральную, простую тему, которая не связана с основной задачей, или решить специально подготовленную пробную задачу (например, предложить рекламу для выдуманных проектов). В начале «разогрева» можно пересказать успешные яркие варианты применения мозговой атаки или посмотреть оригинальные видеоклипы.

4. Создать условия для снятия давления на участников их предыдущего профессионального и жизненного опыта. (Снятие давления предыдущего опыта повышает чувствительность к очень слабым ассоциациям, на основе которых зачастую и рождаются неожиданные, оригинальные, а иногда — шокирующие творческие находки).

¹¹ К сожалению, иногда отбор генераторов идей происходит чисто случайно, и в группу может попасть неисправимый критик.

¹² Это необходимо для повышения уровня доверия ЛПР к результатам, полученным в результате применения данной модификации мозговой атаки.



5. Создать и поддерживать непринужденную творческую и доброжелательную обстановку, способствующую смелому предложению любых идей, и доброжелательные отношения между участниками мозговой атаки. Процедура должна проходить в атмосфере раскованности и взаимного доверия.

6. Провести «принудительный» опрос участников, а затем инициировать лавинообразное высказывание мнений или идей, чтобы количество предложенных идей для решения проблемы было как можно больше. Для этого:

- дать возможность высказаться каждому и следить за тем, чтобы все участники выступали достаточно активно;
- поощрять выдвигающих идеи участников, включая нереальные, фантастические или даже бредовые;
- поощрять сам процесс фантастического ассоциирования: чем более сумасшедшей кажется идея, тем лучше. Создать свободу ассоциаций и творческого воображения. Для этого необходимо «сбить» сознание генераторов с навязанных культурой рутинных путей ассоциирования и смыслообразования;
- поощрять участников на дальнейшую разработку чьей-либо идеи другими участниками заседания (идея, выдвинутая одним участником атаки, должна подхватываться и развиваться другими);
- запрещать участникам обосновывать выдвигаемые идеи (это связано с тем, что в ходе атаки должно быть получено как можно больше новых предложений, и поэтому тщательное обсуждение каждого предложения приведет к снижению продуктивности работы);
- следить за тем, чтобы участники не критиковали друг друга, а также исключать любые критические жесты, иронические улыбки и т. п. участников группы;
- предупреждать участника, выступившего с критикой идей, о возможности вывода его из состава группы генераторов при повторении критических замечаний.

7. Фиксировать весь ход заседания таким образом, чтобы это не тормозило лавинообразный процесс выдвижения новых идей¹³. Например, с помощью диктофона либо видеосъемки.

8. В конце этапа генерации мягко и аккуратно вывести участников из состояния «творческого экстаза» [23].

9. Лишь по окончании этапа генерации записать каждую из выдвинутых идей в протокольный

¹³ На этапе лавинообразного высказывания идей попытка записывать их «от руки» категорически вредна по причине высокой скорости выдвижения новых идей.

список (с тем, чтобы избежать неоднозначности понимания предлагаемых идей и решений).

Обязанности ВЭ и задачи, стоящие перед ним

1. Ознакомить участников группы генераторов с единственной задачей, которую необходимо решить, и исходной информацией об объекте обсуждения.

2. Ознакомить участников группы с процедурой их стимулирования и реализовывать ее.

3. Внимательно рассматривать все выдвинутые идеи, даже те, которые кажутся явно ошибочными или несерьезными.

4. Следить, чтобы каждая предложенная идея формулировалась в терминах, понятных большинству участников.

5. Предлагать участникам развивать появившуюся идею и таким образом способствовать осознанию ее членами всей группы.

6. Ориентировать мышление участников на решение поставленной задачи и направлять высказывания участников в соответствующее русло.

7. Замечать в процессе обсуждения задачи новые предложения участников и способствовать их развитию.

8. Следить, чтобы участники не углублялись в анализ частных отдельных предложений, не давая увязнуть в потоке идей.

9. Направлять развитие дискуссии в нужное русло, к достижению заданной цели, не сбиваясь на беседу, на соревнование в остроумии и т. п.

10. Не допускать досрочного, по мнению ведущего, окончания дискуссии.

На этапе генерации ведущие не имеют право:

- выступать с идеями, касающимися сути обсуждаемой задачи;
- выражать сомнение, скептицизм или, тем более, объявлять предложение участника бесполезным, неверным или не относящимся к делу;
- требовать от участников фактов и предложений по проблеме экспертизы;
- навязывать участникам дискуссии свое мнение, ориентировать их на определенный способ мышления.

Вспомогательные задачи участников группы генерации идей:

- наладить доброжелательные отношения с другими участниками мозговой атаки;
- генерировать поток новых предложений;
- достаточно активно выступать, но не углубляться в частности;
- придерживаться принципов: есть идея — говорю, нет идеи — не молчу; чем более сумасшедшей кажется идея, тем лучше; поощряется заимствовать, как угодно комбинировать, а также видоизменять чужие идеи; высказывать любые мысли без боязни, что их признают плохими.

Этап 3. Творческий анализ высказанных идей. Этот этап проводят двое ведущих: ВЭ и СП, принимают участие ЛПР и (или) заказчик и участники из группы аналитиков. Он реализуется лишь на одном заседании. Идеи, полученные в результате обсуждения задачи группой генераторов, конкретизируются, развиваются и наиболее перспективные из них прорабатываются с тем, чтобы на их основе разработать альтернативные варианты решения.

Обязанности ведущих и вспомогательные задачи, стоящие перед ними

1. Сформировать персональный состав группы аналитиков на основе требований (критериев) к способам ее формирования, предложенных на подготовительном этапе, и утвердить его у ЛПР или заказчика¹⁴. Число ее участников должно определяться, исходя из требований психологии малых групп [30, 31].

2. Организовать работу группы аналитиков с целью поиска конструктивного решения задачи на основе списка идей, предложенных группой генераторов, согласно правилам (см. например, работы [1, 2, 24]):

- анализировать все идеи без исключения;
- найти место каждой идее в системе и найти систему под каждую идею;
- выявлять наиболее приемлемую творческую основу в каждой анализируемой идее (основное правило);
- не умножать сущностей без необходимости;
- искать принципиально новое видение проблемы;
- искать «жемчужину в навозе».

Задачи участников группы анализа идей

1. Рассмотреть идеи, полученные в результате обсуждения проблемы группой генераторов и дать ее глубинный анализ. Для этого:

- конкретизировать идеи;
- систематизировать и (или) классифицировать идеи по группам в соответствии с признаками, по которым их можно объединить. (Классификацию можно проводить так же, как в методе гирлянд случайностей и ассоциаций, разработанного Г.Я. Бушем [5]. Все идеи разбиваются на нерациональные (непригодные, плохие), полурациональные (привлекательные) и рациональные (хорошие). Нерациональные идеи отбрасываются, рациональные образуют ядро для выбора оптимального варианта, а полурациональные (которые чем-то привлекательны, но имеют видимые недостатки) снова анализи-

- руются, после чего включаются в список нерациональных или рациональных идей.);
- определиться, как оценивать идеи (например, по пятибалльной шкале);
- оценить значимость высказанных идей и предложений членов группы генераторов (при этом ведется тщательный поиск рационального зерна в каждой выдвинутой идее);
- выбрать перспективные идеи с учетом специфики творческой задачи, диагностики ситуации и анализа проблемы, а также возможных трудностей ее реализации;
- развить и наиболее перспективные из них проработать (например, оценить идеи по реализуемости).

2. Сформировать окончательный список практически реализуемых идей (заключительное мнение).

Сделаем несколько примечаний.

- Наилучшую идею можно выбрать, как в методе I-G-I (индивид — группа — индивид). Дж. Р. Росситер и Л. Перси предложили такую процедуру выбора наилучшей идеи: «... после группового обсуждения участники вновь переходят к самостоятельной работе и голосуют за ту идею, которую считают лучшей. Все имеют равное право голоса. В итоге выбираются несколько лучших идей. Далее происходит ранжирование от самой перспективной до наименее интересной. Идеи, получившие наивысшую общую оценку, передаются на реализацию и тестирование» [19].
- Аналитик в своей работе на данном этапе часто пользуется такими приемами, как обобщение и конкретизация.
- Для проведения глубинного анализа в ряде случаев группе аналитиков необходимо от часа до месяца. В обычном варианте мозговая атака длится 2 — 3 ч, большая часть времени приходится на последнюю стадию — выработку действенных управленческих решений.
- Результатом применения метода считается заключительное мнение всех участников процедуры.

Этап 4, дополнительный. Корректировка постановки задачи. Этот этап проводят двое ведущих: ВЭ и СП, принимают участие ЛПР и (или) заказчик.

Обязанности ведущих в начале этапа

Если первые три этапа не дали результата (задачу не удастся решить в приемлемые сроки или заказчик не устраивает уровень новизны), то ведущие должны:

- по возможности, разбить поставленную заказчиком сложную задачу на более простые и снова попытаться ее решить (в таком случае затраты времени на проведение штурма значительно возрастают);

¹⁴ См. сноску 12.



— попытаться внести такую корректировку в постановку задачи ЛПР, чтобы она удовлетворяла заказчика и задача могла быть решена с помощью того же метода.

Для корректировки постановки задачи можно, например, применять поэтапный штурм, т. е. такую модификацию мозговой атаки, когда последовательно штурмуются: постановка задачи, решение, доведение идеи до конструкции и ее реализации, причем с тем же самым составом участников [21].

Постановка задачи по ее корректировке близка к той, что была предложена Г.Я. Бушем в его методе гирлянд случайностей и ассоциаций. В рамках этого метода можно было частично корректировать постановку задачи и выбирать объект для совершенствования, чье название лишь синоним первоначально рассматриваемого объекта [5].

Близок к такой постановке и автор синектики [7]. В этой модификации мозговой атаки большое значение придается формулировке задачи. Автор полагает, что преждевременно четкая формулировка задачи приводит к затруднениям в поиске оригинальных решений. Обсуждение в группе часто начинается не с формулировки задачи, а с рассмотрения сущности принципа, лежащего в основе исследуемого процесса. Это дает возможность охватить в ходе обсуждения широкий круг общих проблем и использовать новые процессы и явления при выполнении функций, предусмотренных условиями задачи. Более того, в синектике решение задачи начинается с внесения изменений в ее постановку, предложенную заказчиком. Постепенно круг этих проблем (под влиянием вопросов руководителя) должен быть сужен. После формулировки уточненной постановки задачи участники группы, пользуясь некоторым набором приемов, предпринимая систематические попытки взглянуть на проблему с различных точек зрения и тем самым преодолеть «психологическую инерцию».

На наш взгляд, корректировать постановку заказчика следует не в начале процедуры, а лишь тогда, когда первые этапы оказались не результативными. Конечно, понятие «близкая» постановка задачи нужно попытаться формализовать, но при этом могут возникнуть трудности не только математического порядка.

Приведем два примера. Так, Г.Я. Буш, имея первоначальную задачу усовершенствовать стол, рассмотрел более общую задачу: усовершенствовать всю мебельную «линейку» и эффективно с ней справился [5]. Другой пример «близкой» постановки задачи: если нам нужно усовершенствовать кровать, то можно попытаться модифицировать надувной матрас, в том случае, если функции обоих объектов совпадают, а условия применения объекта, предназначенного для усовершенствования, другие — не в доме, а на природе. В данном

случае важно совпадение или близость функций применения обоих объектов.

Еще одним приемом, позволяющим эффективно использовать корректировку задачи, служит прямая аналогия. В этом случае, объект или процесс, рассматриваемый в первоначальной постановке задачи, сопоставляется с более или менее аналогичным (похожим) объектом или процессом из другой области техники или с объектом из живой природы. Примерами прямых аналогий служат: сердце и насос, мост и паутина, поток электронов и быстрый бег большой группы людей. И в случае такой задачи, как усовершенствование технических насосов, может быть полезна, при определенных условиях прямая аналогия с сердцем. Тем самым, делается попытка использования готовых решений из других областей техники или живой природы.

Обязанности ведущих после корректировки задачи

1. Определить, имеет ли смысл применить рассмотренную модификацию мозговой атаки при новой постановке задачи. Если смысла нет, процедура прекращается¹⁵. В противном случае процедура продолжается.

2. Согласовать новую постановку задачи с заказчиком и (или) ЛПР.

3. Используя материалы и результаты предварительного этапа, сразу приступить к этапу генерации идей с тем, чтобы закончить работу этапом творческого анализа новых высказанных идей.

ЗАКЛЮЧЕНИЕ

Перечислим основные результаты настоящей работы.

- Выявлены и описаны характеристики, определяющие принадлежность метода к совокупности модификаций мозговой атаки.
- Подготовительный этап мозговой атаки не только обозначен, как это делается, например, в работах [7, 24], но расширен и структурирован. Введены в рассмотрение такие субъекты, как заказчик и ЛПР и сформулированы стоящие перед ними задачи. Поэтому в предложенной модификации вклад этого этапа в потенциальную эффективность метода сопоставим с вкладом этапов генерации и аналитики. Задачи по определению требований к двум ведущим мозговой атаки и критериям по их отбору не только поставлены, но и предложены варианты их решения. Поставлена, как обязательная, задача

¹⁵ Конечно, можно попытаться решить задачу в новой постановке с помощью другого экспертного метода или модели.

предварительного определения эффективности применения мозговой атаки для решения задачи, поставленной перед ЛПР и (или) заказчиком.

- Рассмотрены критерии, по которым следует отбирать генераторов идей, специалистов, формировать группы генераторов идей и группы специалистов.
- В рамках второго и третьего этапов расширены и подробно прописаны обязанности ведущих.
- Добавлен дополнительный четвертый этап корректировки постановки задачи, на котором предусмотрена возможность переформулирования постановки задачи, если первые три этапа закончились не эффективно.

Разработанная модификация позволяет повысить потенциальную эффективность метода мозговой атаки.

В дальнейших исследованиях в области совершенствования мозговой атаки следует более глубоко учесть психологию коллективного мышления именно малых групп, будь то группы генераторов или аналитиков, и на этой основе, например, разработать более эффективное стимулирование, как этих групп, так и отдельных участников в их составе. Необходимо систематизировать все выявленные недостатки и достоинства разнообразных модификаций мозговой атаки, эффективно формализовать понятие «близкой» постановки задачи. Для решения вспомогательных задач можно попытаться применить многокритериальный анализ.

ЛИТЕРАТУРА

1. *Osborn A.F.* How to think up. — N.-Y. — London: McGraw-Hill Comp., 1942. — 38 p.
2. *Osborn A.F.* Applied imagination: principles and procedures of creative thinking. — N.-Y.: Charles Scribner's Sons, 1953. — 417 p.
3. *Сидельников Ю.В., Салтыков С.А.* Процедура отбора прогностических экспертных методов на основе установления соответствия между задачей и методом // Экономические стратегии. — 2008. — № 7. — С. 102–109.
4. *Сидельников Ю.В., Салтыков С.А.* Процедура отбора наиболее приемлемых разновидностей экспертных методов // Управление большими системами. — 2010. — Вып. 30 (15). — С. 35–66.
5. *Буш Г.Я.* Методы технического творчества. — Рига: Лиесма, 1972. — 73 с.
6. *Кудрявцев А.В.* Методы интуитивного поиска технических решений (методы анализа проблем и поиска решений в технике). — М.: Речной транспорт, 1991. — 112 с.
7. *Gordon William J.J.* Synectics: The development of creative capacity. — N.-Y.: Harper & Row, 1961. — 180 p.
8. *Гильде В., Штарке К.Д.* Нужны идеи: пер. с нем. — М.: Мир, 1973. — 120 с.
9. *Gilde Werner und Claus-Dieter Starke.* Ideen muss man haben. — Leipzig: Urania-Verl., 1970. — 158 p.
10. *Griggs R.E.* A Storm of Ideas, reported in Training, 1985, 22, 66 (November).
11. *Doherty B.T.* Thinkathon 2. — Educational Impressions, 1984. — 98 p.
12. *Roland Jon.* Questorming: An Outline of the Method, 1985. — URL: <http://pyntan.com/vri/questorm.htm> (дата обращения: 3.12.2013).
13. *Gallupe B.R., Bastianutti L.M., Cooper W.H.* Unblocking brainstorming // J. Appl. Psych. — 1991. — Vol. 76, N 1. — P. 137–142.
14. *Gallupe R.B., Dennis A.R., Cooper W.H., et al.* Electronic Brainstorming and Group Size // Academy of Management Journal. — 1992. — Vol. 35, N 2. — P. 350–369.
15. *Gandl S.* Neubau Modul: Electronic and Analogue Patterns. — Die Gestalten Verlag, 2007. — 383 p.
16. *Furnham A., Yazdanpanah T.* Personality differences and group versus individual brainstorming // Personality and Individual Differences. — 1995. — Vol. 19, N 1. — P. 73–80.
17. *Панфилова А.П.* Мозговые штурмы в коллективном принятии решений. — СПб.: Питер, 2005. — 320 с.
18. *Rawlinson J.G.* Creative Thinking and Brainstorming. — Westmead, Farnborough: Gower Publishing Company Limited, 1981. — 129 p.
19. *Rossiter J.R., Percy L.* Advertising Communications & Promotion Management. — 2-nd ed. — The McGraw-Hill Comp. Inc., 1997. — 351 p.
20. *Paulus P.B., Brown V.R.* Enhancing Ideational Creativity in Groups. — In: Paulus P.B., Nijstad B.A. Group Creativity: Innovation through Collaboration. — Oxford, UK: Oxford University Press, 2003. — P. 110–136.
21. *Toubia O.* Idea Generation, Creativity, and Incentives. — URL: <http://mitsloan.mit.edu/vc/Ideationpaper022805.pdf> (дата обращения: 3.12.2013).
22. *Beebe Sarah Miller, Pherson Randolph H.* Cases in Intelligence Analysis: Structured Analytic Techniques in Action. — Washington, DC: CQ Press College, 2011. — 241 p.
23. URL: <http://vikent.ru/sokolov/> (дата обращения: 3.12.2013).
24. *Сидельников Ю.В.* Системный анализ технологии экспертного прогнозирования. — М.: Изд-во МАИ-ПРИНТ, 2007. — 348 с.
25. *Альтшуллер Г.С.* Найти идею. — 3-е изд., доп. — Петрозаводск: Скандинавия, 2003. — 225 с.
26. *Bouchard T.J.* Whatever happened to brainstorming // J. of Creative Behavior, Buffalo (N.Y.). — 1971. — Vol. 5, N 3. — P. 182–189.
27. *Rickards T., Preeqman D.* Procedures for managers in idea — deficient situations. An examination of brainstorming approaches // J. of Management Studies, Oxford. — 1978. — Vol. 15, N 1. — P. 43–55.
28. *Сидельников Ю.В.* Экспертология — новая научная дисциплина // Автоматика и телемеханика. — 2000. — № 2. — С. 107–126.
29. *Сидельников Ю.В.* Модификация метода фокальных объектов: новые возможности в творчестве // Заводская лаборатория. — 2012. — Т. 78, № 1. — С. 93–102.
30. *Сингер Б.* Азбука создания победоносной Бизнес-Команды. — Минск: Попурри, 2007. — 176 с.
31. *Robbins Stephen P.* Organizational behavior. — Sixth edition. — Prentice Hall International, Inc., 1993. — 336 p.
32. *Глухов В.В.* Менеджмент. — 3-е изд. — М.: Питер, 2008. — 608 с.
33. *Lewis A.C., Sadosky T.L., Coraolly T.* The effectiveness of group brainstorming in engineering problem solving // IEEE Trans. on Engineering Management. — 1975. — Vol. KM-22, N 3. — P. 119–124.

Статья представлена к публикации членом редколлегии В.Н. Новосельцевым.

Сидельников Юрий Валентинович — д-р техн. наук, гл. науч. сотрудник, Институт проблем управления им. В.А. Трапезникова РАН, г. Москва, ☎ (495) 334-88-70, ✉ sidelnikov@mail.ru.



ПОСТРОЕНИЕ МОДЕЛЕЙ МАССОВОЙ ОЦЕНКИ ОБЪЕКТОВ НЕДВИЖИМОСТИ С УЧЕТОМ ПРОСТРАНСТВЕННОЙ ЗАВИСИМОСТИ

А.В. Беляева, Е.А. Гребенюк

Предложен подход к построению массовой оценки объектов недвижимости, учитывающий основные факторы местоположения объектов, влияющие на цену: расстояние до центров влияния и близость расстояний между объектами. Построены модели, реализующие этот подход, и исследованы их свойства. Показано, что совместный учет обоих факторов значительно улучшает качество модели.

Ключевые слова: пространственная корреляция, тест Морана, тест множителей Лагранжа, пространственная авторегрессионная модель, центр влияния.

ВВЕДЕНИЕ

Компьютерные методы массовой оценки объектов недвижимости активно развиваются по пути учета местоположения объектов. Среди множества факторов, влияющих на цену недвижимости (площадь квартиры, число комнат, категория дома, площадь участка и пр.) один из главных — местоположение объектов. В разработанной Дж.К. Эккертом [1] методологии, в соответствии с которой в настоящее время оценивается стоимость объектов недвижимости во многих странах, местоположение учитывается посредством выделения в пространстве «центров влияния», по мере приближения к которым цена недвижимости возрастает (центры позитивного влияния) или убывает (центры негативного влияния). При оценивании стоимости объекта расстояние до центров влияния вводится как экзогенная переменная в (линейную либо нелинейную) регрессионную модель.

Другое активно развивающееся направление учета местоположения заключается в применении пространственных корреляционных моделей [2, 3]. В их основе лежит гипотеза, что характеристики близко расположенных объектов модели связаны между собой, причем эта связь ослабевает с увеличением расстояния между объектами. В моделях с пространственной корреляцией учитывается расстояние между объектами, но не отражаются такие

воздействия на цену, как наличие поблизости загрязняющих окружающую среду объектов, шумных магистралей, кладбищ и пр., а также лесопарковых зон, центров коммуникаций и пр. Центр влияния не полностью определяет влияние местоположения объекта на цену, поскольку на моделируемом участке может присутствовать ряд центров влияния, расположенных в разных точках и различно влияющих на объект, и не все из них могут быть идентифицированы. Поэтому учет местоположения только путем выделения центров влияния может оказаться не полным.

В настоящей статье предложен подход, позволяющий совокупно оценивать влияние обоих факторов местоположения объекта: малое различие цен близких между собой объектов и зависимость цен от расстояния от объектов до центров влияния. Показано, что предлагаемый подход позволяет улучшить качество оценивания объектов. Структура моделей совпадает со структурой пространственных регрессионных моделей, рассмотренных в работах [2, 4]. Для методологии построения этих моделей характерны свои особенности как при выборе конкретной модели оценивания, так и при диагностике построенных моделей и выборе наиболее подходящего варианта. Рассмотрены основные этапы построения моделей, спецификации и выбора варианта, приведены результаты моделирования.

1. ПОСТРОЕНИЕ РЕГРЕССИОННЫХ МОДЕЛЕЙ МАССОВОЙ ОЦЕНКИ НЕДВИЖИМОСТИ С УЧЕТОМ ФАКТОРА МЕСТОПОЛОЖЕНИЯ ОБЪЕКТА

1.1. Постановка задачи

Имеется набор объектов недвижимости $Z = (z_1, \dots, z_n)$, представляющий собой выборку из объектов жилого массива (города, района, поселка). Каждый объект определяется вектором параметров: ценой, площадью, числом комнат, категорией дома, площадью участка и пр., а также координатами. Требуется построить регрессионную модель, позволяющую выполнять компьютеризированную массовую оценку цен объектов с учетом фактора местоположения объекта.

Здесь рассматриваются линейные регрессионные модели. Поскольку выбор независимых регрессоров до некоторой степени ограничен имеющейся в распоряжении статистика информацией, то повысить эффективность модели при построении массовой оценки можно путем оптимального учета информации о зависимости цены от расположения объекта в пространстве. Поэтому в качестве критериев качества модели выбираются:

- среднеквадратичная ошибка оценивания (сумма квадратов отклонений модельных значений цен от фактических);

- эффективность учета местоположения объекта, определяемая сравнением моделей, построенных с учетом пространственного фактора, со стандартными регрессионными моделями.

Местоположение объекта будем учитывать двумя способами: учитывать влияние на цену расстояния до центров влияния и, в случае наличия пространственной зависимости между объектами, учитывать влияние на цену объекта характеристик близко расположенных объектов.

Центр влияния — это объект, не принадлежащий множеству оцениваемых объектов Z и оказывающий позитивное или негативное влияние на стоимость объектов оценивания. К таким центрам могут относиться крупный деловой район города, лесопарк, торговый центр, близость к которым повышает привлекательность объекта недвижимости (центр положительного влияния) или промышленный район с неблагоприятной экологической обстановкой (центр отрицательного влияния).

Идея, лежащая в основе учета пространственной зависимости, заключается в том, что при оценивании характеристик объекта учитываются значения этих же характеристик близко расположенных объектов. Для учета этой зависимости используются пространственные модели [2]. При их построении стандартные процедуры регрессионного

анализа имеют свою специфику. Выбор структуры модели, помимо определения набора регрессоров, включает в себя выбор типа модели (SAR, SEM, GSM) и выбор пространственных матриц. При оценивании пространственных моделей обычным методом наименьших квадратов (МНК) получают смещенные и несостоятельные оценки, поэтому для оценивания применяются методы максимального правдоподобия, инструментальной переменной (IV) и обобщенный метод моментов (GMM) [3, 5, 6].

Построение модели с учетом пространственного фактора осуществляется:

- определением центров влияния;
- проверкой наличия пространственной корреляции;
- оцениванием различных вариантов моделей и их диагностикой;
- сравнением вариантов и выбором модели.

1.2. Учет расстояний до центров влияния

Предлагаемая процедура учета расстояний до центров влияния несколько отличается от процедуры, используемой Эккертом [1], и состоит из следующих шагов.

1. Если известны координаты центра влияния, то вычисляется вектор расстояний от центра влияния до каждого из объектов множества Z .

2. Если эта информация отсутствует, то координаты центра влияния определяются по выборочным данным. Стандартным МНК строится регрессия цен объекта на множество независимых переменных (координаты не включаются). Анализируется зависимость остатков регрессии от координат: строится трехмерный график зависимости остатков регрессионной модели от координат x и y и контурный график уровней остатков res . На рис. 1 и 2 приведены примеры построения трех-

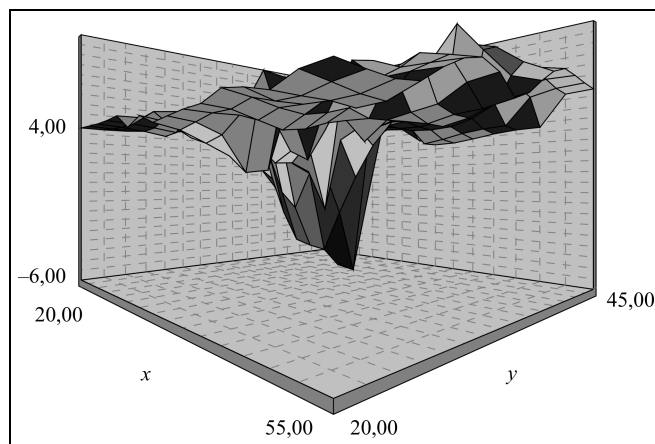


Рис. 1. Трехмерный график остатков

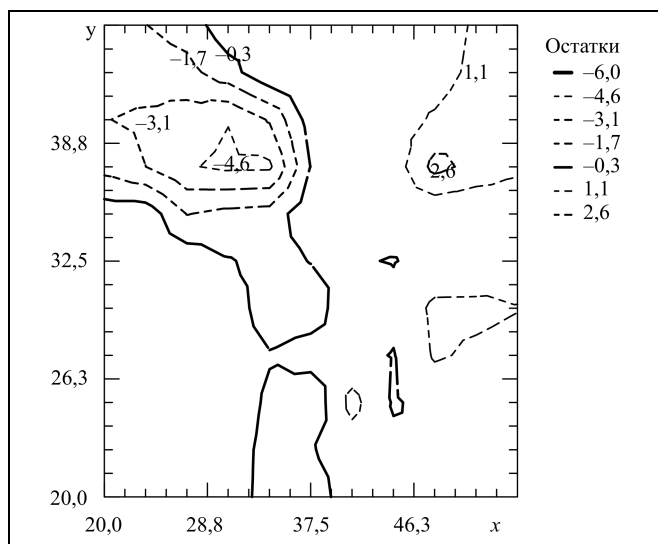


Рис. 2. Плоскость уровней остатков

мерного графика остатков и контурного графика его уровней¹. С помощью анализа пространственного графика и контурного графика выделяются участки, на которых остатки значительно отличаются от нуля. Координаты центров этих участков определяются как координаты центров влияния.

3. Для нахождения значений регрессора, учитывающего зависимость цен от расстояния до центров влияния, строится диаграмма рассеяния для переменных res и d , где res — остатки регрессионной модели: разность фактических цен объектов Y и цен Y^M , рассчитанных по регрессионной модели, не учитывающей пространственный фактор, d — вектор расстояний объектов до центров влияния. С помощью диаграммы выбирается вид функциональной зависимости $res = F = f(ad) + b$, где a и b — параметры функции, подбираемые методом наименьших квадратов. Вектор значений F добавляется в регрессионную модель в качестве регрессора. Если регрессор при оценивании модели оказывается значимым и не увеличивает величины критериев Акаике и Шварца, то он вводится в модель.

Предложенная процедура позволяет учитывать фактор центров влияния при оценивании объектов, не входящих в обучаемую выборку для построения модели.

1.3. Описание класса пространственных моделей

Если местоположение не имеет значения для рассматриваемой группы объектов, то их цена мо-

жет быть описана стандартной регрессионной моделью:

$$Y = X\beta + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 I_n), \quad (1)$$

где Y — n -мерный вектор цен объектов X — $(n \times k)$ -матрица объясняющих переменных, β — k -мерный вектор коэффициентов, ε — n -мерный вектор возмущений, σ^2 — дисперсия возмущений, n — число оцениваемых объектов.

Если местоположение объекта важно для анализа цены объекта, то для описания этой зависимости применяют различные способы «взвешивания» цен: $Y = \rho WY$, где Y — n -мерный вектор характеристик объектов ρ — константа, W — $(n \times n)$ -матрица пространственной зависимости: элемент w_{ij} принимает не нулевое значение, если объекты i и j близкие в смысле выбранной метрики, и нулевые — в противном случае. Для улучшения свойств оценок используют умножение элементов матрицы на различные весовые коэффициенты с последующей стандартизацией. Стандартизация матрицы выполняется таким образом, чтобы сумма элементов в строке равнялась единице. Будем рассматривать следующие типы пространственных моделей [2].

- Модель с пространственной корреляцией в объясняемой переменной (spatial autoregressive — SAR-модель):

$$Y = \rho WY + X\beta + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 I_n), \quad (2)$$

где ρ — коэффициент пространственной авторегрессии, W — пространственная матрица.

- Модель с пространственной автокорреляцией в возмущениях (spatial error model — SEM-модель):

$$Y = X\beta + u, \quad u = \lambda Wu + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 I_n), \quad (3)$$

где u — n -мерный вектор возмущений, λ — коэффициент пространственной корреляции возмущений, описывающий пространственную корреляцию в ошибках модели.

- Общая модель, включающая в себя пространственные корреляции в объясняемых переменных и в возмущениях (general spatial model — GSM-модель):

$$Y = \rho W_1 Y + X\beta + u, \quad u = \lambda W_2 u + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 I_n). \quad (4)$$

Пространственные матрицы W_1 и W_2 могут как различаться, так и быть равны между собой, причем в последнем случае могут возникнуть проблемы при оценивании модели.

При построении пространственных матриц, используемых в моделях (2) — (4), учитывается принцип: близко расположенные объекты имеют близ-

¹ Графики построены с помощью пакета NCSS (Number Cruncher Statistical System).

кие характеристики. Алгоритм построения матрицы W .

1. Элемент w_{ij} пространственной матрицы W полагается равным единице, если объекты i и j являются близкими в смысле выбранной метрики, и нулю в противном случае.

Один из наиболее распространенных алгоритмов построения пространственных матриц заключается в определении близких объектов на основе триангуляции Делоне. Триангуляция набора точек на плоскости состоит в соединении всех точек непересекающимися отрезками таким образом, чтобы новых отрезков уже нельзя было добавить без пересечения с имеющимися. Если в каждой точке плоскости (X, Y) задана координата Z , то триангуляция Делоне обеспечивает наилучшую аппроксимацию поверхности (X, Y, Z) . Объекты являются близкими к объекту i , если они соединены с ним отрезками в результате триангуляции Делоне.

Другой способ формирования пространственных матриц состоит в выделении для каждой из точек плоскости (X, Y) ее ближайших в смысле евклидовой метрики точек.

2. Значения расстояний между близкими точками учитываются заменой единиц в пространственной матрице весами: чем ближе расположены точки i и j , тем больше вес соответствующего элемента матрицы. Примеры назначения весов как функций расстояния d_{ij} :

$$w_{ij} = \begin{cases} d_{ij}^{-\alpha}, & \text{если } w_{ij} \neq 0, \\ 0, & \text{если } w_{ij} = 0, \end{cases}$$

или

$$w_{ij} = \begin{cases} \exp(-\alpha d_{ij}), & \text{если } w_{ij} \neq 0, \alpha > 0, \\ 0, & \text{если } w_{ij} = 0. \end{cases}$$

3. После выбора весов матрицу стандартизируют таким образом, чтобы сумма весов в каждой строке равнялась единице.

1.4. Проверка наличия пространственной автокорреляции

Для обоснования перехода от стандартной регрессионной модели к моделям с пространственной корреляцией и определения их вида разработано большое число тестов [2, 7, 8].

Тест Морана — один из широко применяемых. Он предназначен для проверки наличия пространственной зависимости, определяемой матрицей W . Статистика теста рассчитывается по формуле

$$I = (N/S) \frac{(Y - \bar{Y}) W (Y - \bar{Y})}{(Y - \bar{Y})(Y - \bar{Y})},$$

где $\bar{Y}$ — среднее значение Y , N — размерность вектора Y , W — матрица пространственной зависимости, $S = \sum_i \sum_j w_{ij}$ — сумма элементов матрицы W .

Если матрица W стандартизирована таким образом, что сумма элементов каждой строки равняется единице, то статистика принимает вид:

$$I = \frac{(Y - \bar{Y}) W (Y - \bar{Y})}{(Y - \bar{Y})(Y - \bar{Y})}.$$

В работе [3] обобщена статистика Морана и разработан тест для проверки пространственной корреляции для линейной регрессионной модели (2). Показано, что статистика

$$z = \frac{I - E(I)}{\sqrt{\text{Var}(I)}}$$

при выполнении нулевой гипотезы H_0 : «пространственная корреляция отсутствует» имеет стандартное нормальное распределение. Значения статистики зависят от пространственной матрицы W : при различных моделях пространственных матриц тест Морана может давать противоречивые результаты. Этот тест не определяет характер пространственной зависимости, но лишь указывает на ее существование.

Тест множителей Лагранжа (МЛ) [7, 8] при исследовании на наличие пространственной автокорреляции в данных не требует оценки пространственной модели, достаточно оценить модель в условиях нулевой гипотезы H_0 : «данные описываются регрессионной моделью (1)». Такие тесты могут также различать пространственную корреляцию в объясняемой переменной и пространственную корреляцию в ошибке модели. Статистика

$$LM_{err} = (ne^T W e / e^T e)^2 (\text{tr}(W^T W + W^2))^{-1}, \quad (5)$$

где e — вектор ошибок модели (1), используется для проверки гипотезы H_0 против альтернативной гипотезы H_1 : «данные описываются регрессионной моделью (3)».

Статистика МЛ-теста для проверки пространственной корреляции в объясняющей переменной имеет вид:

$$LM_{lag} = (ne^T W e / e^T e)^2 (n(WXb)^T M(WXb) / (e^T e) + \text{tr}(W^T W + W^2))^{-1}, \quad (6)$$

где b — МНК-оценка β в модели (1), $M = I - X(X^T X)^{-1} X^T$. Эта статистика проверяет гипотезу H_0 против альтернативной гипотезы H_1 : «данные описываются регрессионной моделью (2)». Обе статистики (5) и (6) асимптотически распределены



как $\xi^2(1)$. Если обе статистики значимы, то возникает ситуация неопределенности, для разрешения которой в работе [7] предложено выбирать ту модель, для которой соответствующая статистика имеет более высокий уровень значимости. Однако эти рекомендации работают только в том случае, если уровни значимости статистик существенно различаются.

1.5. Оценка качества и сравнение вариантов

Одним из основных вопросов остается оценка качества построенной модели. В силу структуры пространственных моделей и применяемых методов оценивания остатки могут не иметь нулевого математического ожидания, следовательно, разложения дисперсии на дисперсию, объясненную регрессией, и дисперсию остатков не существует, поэтому критерий R^2 теряет смысл. Для оценивания пространственных моделей выберем набор критериев:

- критерии Акаике и Шварца — оценивают среднее значение ошибки модели с учетом штрафов за каждый включаемый регрессор;
- квадрат коэффициента корреляции между наблюдаемыми и предсказанными по модели значениями (Corr2);
- максимум функции правдоподобия (МП);
- оценка смещения среднего остатков относительно нуля;
- оценка приведенной суммы квадратов остатков (Std).

2. ОПИСАНИЕ ПРОЦЕССА ПОСТРОЕНИЯ МОДЕЛИ ПО ЭКСПЕРИМЕНТАЛЬНЫМ ДАННЫМ

Для построения модели по экспериментальным данным рассмотрим пример Анселина [2], в котором определяется влияние пространственного фактора, доходов домохозяйств и стоимости жилья на уровень криминала (число ограблений на тысячу домохозяйств). Этот пример рассматривался многими исследователями для иллюстрации полученных результатов. Все вычисления проводились с помощью стандартных программ пакета MATLAB и набора программ, предназначенного для оценивания пространственных моделей и включающего в себя алгоритмы оценивания и построения пространственных матриц [4].

Определение центров влияния. Уравнение регрессии уровня криминала на размер доходов и стоимость жилья имеет вид:

$$Y = 68,619 - 1,597x_1 - 0,274x_2 + \varepsilon, \quad (7)$$

где Y — уровень криминала, x_1 — доходности домохозяйств, x_2 — стоимость жилья, ε — ошибка

модели. Анализируется 49 административных единиц, каждая из которых, помимо указанных переменных, определяется парой географических координат. В результате построения трехмерного графика отношений фактических значений цены к модельным в пространстве координат и контурного графика уровней отношений, определим координаты центра влияния: $x = 38,29$; $y = 30,35$. Построим диаграмму рассеяния переменных $d = \{d_i\}$, $i = 1, \dots, n$, где d_i — расстояние от i -го объекта до центра влияния, и res — остатков регрессии (7) и определим вид зависимости res от d : $res_i = \exp(-d_i)$.

Построим регрессионную модель, добавив в нее в качестве регрессора вектор $x_3 = \{\exp(-d_i)\}$, $i = 1, \dots, n$. Уравнение регрессии уровня криминала на размер доходов, стоимость жилья и расстояние до центра влияния имеет вид:

$$Y = 65,76 - 1,61x_1 - 0,229x_2 + 35,92x_3. \quad (8)$$

Характеристики моделей (7) и (8) представлены в табл. 1, где AIC — критерий Акаике, SIC — критерий Шварца, Corr2 — квадрат коэффициента корреляции между наблюдаемыми и предсказанными по модели значениями, Std — среднеквадратичное отклонение остатков. Проверка по критерию Жака—Бера указывает, что остатки модели (7) соответствуют нормальному распределению: вероятность ошибиться, отвергнув нулевую гипотезу, $p_{value} = 0,253$; гипотеза о нормальности остатков модели (8) отвергается, $p_{value} = 0,016$. По результатам, представленным в табл. 1, значения критериев в регрессии, построенной с учетом расстояния до центров, выше, чем в обычной. Однако более высокие значения критериев Акаике и Шварца указывают на то, что эти улучшения слишком незначительны по сравнению с платой за улучшение — добавлением новой переменной.

Выбор структуры пространственных матриц. При построении пространственных моделей будем пользоваться матрицами смежности различной структуры: W — матрица, построенная на основе триангуляции Делоне; W_2 и W_4 — матрицы с двумя и четырьмя ближайшими соседями, соответственно. Весовые коэффициенты матриц W , W_2 и W_4 полагаем равными единице и проводим стандартизацию матриц по строкам.

Таблица 1

Сравнение качества моделей

Критерии	R^2	AIC	SIC	Corr2	Std
Модель (7)	0,533	4,996	5,154	0,552	11,194
Модель (8)	0,628	5,059	5,278	0,651	9,885

Проверка наличия пространственной корреляции. Проверим наличие пространственной корреляции, применив тест Морана. Результаты представлены в табл. 2.

Таблица 2

Проверка наличия пространственной корреляции тестом Морана

Модель	Матрица модели	W	W_2	W_4
(7)	z p	4,0149 $5,8341 \cdot 10^{-5}$	3,1125 0,0019	4,7344 $2,1972 \cdot 10^{-6}$
(8)	z p	3,2838 0,001	2,3372 0,0194	4,2225 $2,4165 \cdot 10^{-5}$

где z — значения модифицированной статистики Морана [3], p — уровень значимости. Результаты применения теста Морана указывают на наличие пространственной корреляции (включение информации о центрах влияния, практически, не меняет результатов теста), но тип корреляции не определен.

Для определения типа пространственной корреляции проведем тесты множителей Лагранжа. Проверим наличие пространственной корреляции в остатках регрессионных моделей (7) и (8), используя статистику (5); результаты приведены в табл. 3, где L_{meff} — статистика, определяемая выражением (5).

Таблица 3

Проверка наличия пространственной корреляции в остатках тестом МЛ

Модель	Матрица модели	W	W_2	W_4
(7)	L_{meff} p	11,073 0,0009	7,506 0,0061	15,903 $6,67 \cdot 10^{-5}$
(8)	L_{meff} p	6,468 0,011	3,596 0,058	11,736 $6,132 \cdot 10^{-4}$

Тесты показывают наличие пространственной корреляции в остатках модели, однако добавление нового регрессора несколько повышает вероятность ошибки отвержения нулевой гипотезы об отсутствии корреляции в остатках.

Проверим наличие пространственной корреляции в лагах регрессионной модели с использованием статистики (6); результаты приведены в табл. 4, где L_{mlag} — статистика, определяемая выражением (6).

Таблица 4

Проверка наличия пространственной корреляции в лагах тестом МЛ

Модель	Матрица модели	W	W_2	W_4
(7)	L_{mlag} p	12,858 0,0003	13,816 0,0002	18,513 $1,69 \cdot 10^{-5}$
(8)	L_{mlag} p	10,128 0,0015	10,659 0,0011	14,96 $1,098 \cdot 10^{-4}$

Тесты показывают наличие пространственной корреляции в лагах, добавление нового регрессора мало влияет на статистику и вероятность ошибки.

По результатам тестов пространственная корреляция присутствует как в остатках, так и в лагах модели, причем включение информации о расстояниях до центра влияния не устраняет и, практически, не снижает уровень пространственной корреляции. Таким образом, по результатам тестов на наличие пространственной зависимости вид пространственной корреляции однозначно определить не удается.

Оценивание различных вариантов моделей и их диагностика. В силу неопределенности информации о типе пространственной корреляции для выбора наилучшей модели проверим все возможные варианты и построим пространственные модели (2)—(4). Оценки качества построенных SAR-моделей, учитывающих пространственную корреляцию в объясняемой переменной, приведены в табл. 5.

В первых трех столбцах таблицы приведены характеристики пространственных моделей, построенных без учета центров влияния, в следующих трех столбцах — моделей с добавленным регрессором x_3 , в первой строке таблицы указана пространственная матрица, используемая для построения модели. Как видно из данных, представленных в таблице, добавление информации о центрах влияния позволяет улучшить качество модели по всем критериям. Из всех построенных SAR-моделей наилучшими характеристиками обладает модель в 4-м столбце (с матрицей W).

Оценим качество построенных SEM-моделей, учитывающих пространственную корреляцию в остатках. Результаты приведены в табл. 6.

Как и для SAR-модели, в первых трех столбцах таблицы приведены характеристики пространственных моделей, построенных без учета центров влияния, в следующих трех столбцах — моделей с добавленным регрессором x_3 . Добавление нового регрессора практически не улучшает качество модели. Из всего набора построенных моделей опять же модель с матрицей W и с дополнительным регрессором обладает лучшими характеристиками, но она уступает модели с SAR-структурой.



Рассмотрим общую пространственную модель, предполагающую наличие пространственной корреляции как в лагах, так и в остатках, для различных вариантов моделей, различающиеся между собой комбинацией пространственных матриц. Так же, как и в предыдущих случаях, сравним два варианта: с учетом (табл. 7) и без учета информации до центров влияния (табл. 8).

Оптимальная модель — с дополнительным регрессором x_3 и комбинацией матриц $W - W_2$ (5-й столбец табл. 8). Ей незначительно уступает модель с SAR-структурой. Таким образом, по резуль-

татам экспериментальной проверки наилучшими характеристиками обладает GSM-модель с учетом расстояний до центра влияния.

Сравнение вариантов и выбор модели. В целом значения критериев Corr2, Std, МП в моделях с дополнительным регрессором x_3 выше, чем в моделях с добавленным регрессором, при этом, значения критериев Акаике и Шварца не увеличиваются, что говорит об эффективности использования дополнительного фактора.

Для проверки значимости эффекта учета расстояний до центра влияния используем диспер-

Таблица 5

Оценки качества SAR-моделей

Критерий	Матрица модели					
	W	W_2	W_4	W	W_2	W_4
Акаике	4,621	4,733	4,732	4,518	4,61	4,554
Шварца	4,776	4,888	4,887	4,711	4,8	4,747
Corr2	0,7303	0,6994	0,6988	0,7658	0,7439	0,7461
МП	-160,05	-163,15	-162,30	-155,98	-158,45	-157,66
Std	8,705	9,196	9,198	8,102	8,475	8,435

Таблица 6

Оценки качества SEM-моделей

Критерий	Матрица модели					
	W	W_2	W_4	W	W_2	W_4
Акаике	4,577	4,684	4,493	4,663	4,77	4,578
Шварца	4,731	4,838	4,648	4,856	4,96	4,771
Corr2	0,7297	0,6868	0,7274	0,7555	0,7173	0,7667
МП	-163,21	-165,49	-162,31	-159,75	-166,88	-157,88
Std	8,99	9,521	8,86	8,4697	8,995	8,167

Таблица 7

Оценки качества GSM-моделей без учета информации о расстоянии до центров влияния

Критерий	Матрица модели					
	$W_2 - W_4$	$W_4 - W_2$	$W - W_4$	$W_4 - W$	$W - W_2$	$W_2 - W$
Акаике	4,755	4,539	4,679	4,806	4,647	4,812
Шварц	4,948	4,694	4,872	5,000	4,84	5,005
Corr2	0,7196	0,7174	0,7382	0,7029	0,7459	0,7070
МП	-161,452	-161,179	-159,185	-161,866	-158,96	-162,514
Std	8,914	8,918	8,578	9,161	8,458	9,155

Таблица 8

Оценки качества GSM-моделей с учетом информации о расстоянии до центров влияния

Критерий	Матрица модели					
	$W_2 - W_4$	$W_4 - W_2$	$W - W_4$	$W_4 - W$	$W - W_2$	$W_2 - W$
Акаике	4,675	4,681	4,598	4,727	4,596	4,74
Шварц	4,906	4,913	4,829	4,958	4,827	4,972
Corr2	0,7613	0,7583	0,7765	0,7474	0,7782	0,7436
МП	-156,974	-156,8445	-154,936	-157,429	-155,17	-158,476
Std	8,203	8,234	7,917	8,418	7,892	8,48

сионный анализ. Будем рассматривать значения каждого из критериев Corr2, Std, МП (признаков) в моделях, построенных без дополнительного регрессора с соответствующими значениями в моделях с дополнительным регрессором. Каждый из признаков имеет $p = 2$ уровней и q наблюдений на каждом уровне. ($q = 3$ для SAR- и SEM-моделей и $q = 6$ для GSM модели). Для каждого из признаков вычислим факторную (порождаемую воздействием фактора) и остаточную (обусловленную случайными причинами) дисперсии и проверим значимость различия дисперсий по критерию Фишера—Снедекора. Если различие между этими дисперсиями значимо, то фактор оказывает существенное влияние на признак, следовательно, средние наблюдаемых значений на каждом уровне различаются значимо.

Значения критерия Фишера—Снедекора и критические значения приведены в табл. 9.

Таблица 9

Значения критерия Фишера – Снедекора

Критерий	Модель		
	SAR	SEM	GSM
Corr2	11,49	2,431	18,436
МП	14,43	0,557	26,96
Std	11,84	3,382	18,722
Акаике	7,644	1,19	0,545
Шварц	4,654	2,53	0,2
Критические значения	7,71	7,71	4,94

Критические значения критерия Фишера — Снедекора с уровнем значимости 0,05 равны: 7,71 (для SAR- и SEM-моделей), 4,94 (для GSM-модели). Из результатов, приведенных в табл. 9, следует, что увеличение значений критериев Corr2, Std и МП значимо в случае SAR- и GSM-моделей, а значения критериев Акаике и Шварца не увеличиваются, что свидетельствует об улучшении модели.

ЗАКЛЮЧЕНИЕ

Предложен подход к построению компьютерной массовой оценки объектов недвижимости, позволяющий учесть совместное влияние двух пространственных факторов: расстояния до центров влияния (объектов, не являющихся объектами выборки, близость к которым увеличивает или уменьшает стоимость оцениваемых объектов) и взаимного расположения объектов друг относительно друга. Предложенный подход был реализован в виде пространственной авторегрессионной модели, в которой близость объектов друг к другу описывается пространственной матрицей, построенной по координатам объектов. Для учета расстояний до центров влияния в модель включается дополнитель-

ный регрессор, значения которого вычисляются в виде функции от координат объекта и остатков (разностей между фактическими и модельными значениями) от линейной регрессионной модели, построенной без учета пространственных факторов.

Методика реализации предложенного подхода продемонстрирована на примере анализа пространственных данных. Эксперимент показал значимое улучшение качества оценки при введении в модель двух факторов учета пространственной зависимости, что показывает эффективность предложенного подхода.

Дальнейшее развитие подхода, учитывающего различные виды пространственной зависимости, заключается в исследовании и разработке моделей, позволяющих учитывать неравномерность влияния пространственного фактора: выделение локальных участков, на которых резко изменяется направление его воздействия.

Поскольку процедура построения моделей формализована на всех этапах и реализована в виде компьютерной программы, то предложенные модели и методика могут быть применены для построения компьютерной массовой оценки совокупности объектов недвижимости, расположенных на обширной территории, в целях налогообложения.

ЛИТЕРАТУРА

1. Eckert J.K., O'Connor P.M., Chamberlain C. Computer-Assisted Real Estate Appraisal. A California Savings and Loan Case Study // The Appraisal Journal. — 1993. — October. — P. 524—532.
2. Anselin L. Spatial Econometrics: Methods and Models. — Dordrecht: Kluwer Academic, 1988. — P. 284.
3. Cliff A., Ord J.K. Spatial Processes: Model, and Application. — London: Pion, 1981. — P. 327.
4. LeSage James P. Spatial Econometrics. — 1999. — URL: www.rri.wvu.edu/WebBook/LeSage/spatial/wbook.pdf (дата обращения 6.11.2013).
5. Kelejian H.H., Prucha I. A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Mode 1 // International Economic Review. — 1999. — Vol. 40, iss. 2. — P. 509—533.
6. Kelejian H.H., Prucha I.R., Yuzefovich E. Instrumental variable estimation of a spatial autoregressive model with autoregressive disturbances: large and small sample results // Advances in Econometrics. — 2004. — Vol. 18. — P. 163—198.
7. Anselin L., Raymond J.G., Florax M. Small Sample Properties of Tests for Spatial Dependence in Regression Models: Some Further Results., New Directions in Spatial Econometrics (L. Anselin and Raymond J.G.M. Florax, eds). — Berlin: Springer, 1995. — P. 21—74.
8. Anselin L., Rey S. Properties of Tests for Spatial Dependence in Linear Regression Models // Geographical Analysis. — 1991. — Vol. 23, N 2. — P. 112—131.

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

Беляева Анна Валерьевна — аспирант, ✉ BelyaevaAV@gmail.com, Гребенюк Елена Алексеевна — д-р техн. наук, вед. науч. сотрудник, ☎ (495) 334-46-40, ✉ lngrebenuk@rambler.ru, Институт проблем управления им. В.А. Трапезникова РАН, г. Москва.



МОДЕЛИРОВАНИЕ КОРРУПЦИИ В ТРЕХУРОВНЕВЫХ СИСТЕМАХ УПРАВЛЕНИЯ¹

Г.А. Угольников, А.Б. Усов

Рассмотрены динамические модели коррупции в трехуровневых системах управления. Даны определения равновесий в динамических играх трех лиц с учетом требований устойчивого развития в условиях коррупции, предложены алгоритмы их построения. Исследованы модели экономической коррупции в системах контроля качества поверхностных вод, основанные на определенных информационных регламентах, приведены примеры численных расчетов.

Ключевые слова: коррупция, динамические игры, иерархическая система управления, принуждение, побуждение, условия устойчивого развития.

ВВЕДЕНИЕ

Пионерской работой по математическому моделированию коррупции считается статья С. Роз-Аккерман [1], представляющую собой адаптацию идей Г. Беккера относительно моделирования абстрактного рода преступлений и методики выбора соответствующих наказаний [2].

В дальнейшем коррупция как отдельная ветвь математического моделирования социально-экономических процессов изучалась по нескольким основным направлениям:

— математические модели, направленные на изучение коррупции как социальной практики внутри отдельной организации (внутренняя коррупция) и между организациями (внешняя коррупция);

— моделирование коррупции в общественно-политической жизни (в основном, коррупционных схем, присутствующих в избирательном процессе);

— исследование таких явлений, как множественность равновесных коррупционных состояний, цикличность возникновения и др.

Модели коррупции в иерархических системах описаны, например, в работах [3–8]. В России наиболее активно математическим моделированием коррупции занимаются А.А. Васин [10, 11], М.И. Левин [12], В.М. Полтерович [13]. Обзор моделей коррупции приведен в книгах [14–16].

В рамках авторского подхода управления устойчивым развитием построены и исследованы различные модели коррупции в иерархических системах и предложен ряд принципов ее моделирования [16–23], в том числе следующие.

- Базовой схемой моделирования служит иерархическая система «принципал — супервайзер — агент — объект» в различных модификациях и ее теоретико-игровое исследование, причем в динамических моделях состояние объекта описывается явно. Коррупции подвержен средний уровень управления (супервайзер), верхний уровень (принципал) считается некоррумпируемым и выполняет функции борьбы с коррупцией.
- Предполагаются известными определенные требования устойчивого развития управляемой системы (объекта). В динамических моделях они формулируются в терминах состояния объекта. Если требования устойчивого развития выполняются, то задача принципала считается решенной даже при наличии коррупции.
- Пары «принципал — супервайзер» и «супервайзер — агент» состоят в отношениях «ведущий — ведомый». Ведущий игрок (принципал или супервайзер соответственно) для достижения своих целей использует методы принуждения (преимущественно административно-законодательные воздействия) и побуждения (преимущественно экономические воздействия); при математической формализации принуждение означает воздействие ведущего на множество допустимых стратегий ведомого (без обратной связи), а побуждение — на функцию выигрыша ведомого (с обратной связью).

¹ Работа выполнена при финансовой поддержке РФФИ (проект № 12-01-00017) и Южного федерального университета.

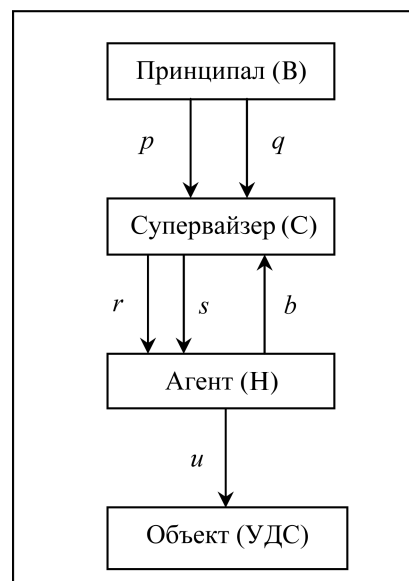
- Различаются административная коррупция, при которой за взятку ослабляются административные требования, и экономическая коррупция, при которой взятка позволяет ослабить экономические требования верхнего уровня управления. При моделировании административная коррупция означает принуждение агента супервайзером с обратной связью по размеру взятки, а экономическая коррупция — побуждение агента супервайзером с дополнительной обратной связью по размеру взятки.
- Коррупция представляет собой угрозу устойчивому развитию объекта, поскольку взяточнику выгодно в обмен на взятку ослаблять требования устойчивого развития. С другой стороны, коррупция есть специфическая форма обратной связи в иерархических системах управления, в силу которой управляющие воздействия становятся функциями размера взятки.

При моделировании коррупции в динамических иерархических системах нами используются результаты, полученные А.Ф. Кононенко с соавторами [24–28].

1. МЕТОДЫ УПРАВЛЕНИЯ И ИНФОРМАЦИОННЫЕ РЕГЛАМЕНТЫ

Рассмотрим трехуровневую систему управления, включающую в себя субъекты управления верхнего (принципал), среднего (супервайзер) и нижнего (агент) уровней, а также управляемую динамическую систему (УДС) — см. рисунок. Для простоты ограничимся случаем одного субъекта на каждом из уровней (В, С и Н соответственно). Основная цель субъекта В состоит в выполнении требований устойчивого развития (УР), которые формализуются как принадлежность траектории УДС определенной области пространства состояний. Субъект В может иметь дополнительные собственные интересы. Субъект С следит за выполнением требований УР от имени субъекта В, но может пренебрегать ими в обмен на взятку от субъекта Н. На УДС воздействует только субъект Н, который преследует исключительно свои частные интересы, исходя из которых может предлагать взятку субъекту С. Каждый из субъектов В и С при воздействии на нижестоящего субъекта (С и Н соответственно) может применять методы принуждения (С — compulsion) или побуждения (Н — impulsion). Подробное математическое описание этих методов как решений иерархической игры представлено в книге [17].

Таким образом, возникают четыре информационных регламента для иерархической игры трех лиц:



Трехуровневая система управления

— В и С используют принуждение (регламент СС);

— В использует принуждение, С — побуждение (регламент СИ);

— В использует побуждение, С — принуждение (регламент ИС);

— В и С используют побуждение (регламент II).

При этом имеет место административная (управления принуждения субъекта С становятся функциями размера взятки) и экономическая (управления побуждения субъекта С становятся функциями размера взятки) коррупция. Конечно, возможны и другие схемы информационного взаимодействия.

Исходная теоретико-игровая модель имеет вид:

$$J_B(z(\cdot)) = \int_0^T e^{-\alpha t} g_B(z(t)) dt \rightarrow \max; \quad (1)$$

$$q(t) \in Q; \quad p(t) \in P; \quad (2)$$

$$J_C(z(\cdot)) = \int_0^T e^{-\alpha t} g_C(z(t)) dt \rightarrow \max; \quad (3)$$

$$s(t) \in S; \quad r(t) \in R; \quad (4)$$

$$J_H(z(\cdot)) = \int_0^T e^{-\alpha t} g_H(z(t)) dt \rightarrow \max; \quad (5)$$

$$u(t) \in U; \quad b(t) \in B; \quad (6)$$

$$\dot{x} = f(u(t), x(t)), \quad x(0) = x_0; \quad (7)$$

$$x(t) \in X^*; \quad t \geq 0. \quad (8)$$



Здесь $z(t) = (q(t), p(t), s(t), r(t), u(t), b(t), x(t))$; $q(t)$ и $p(t)$ — векторы управлений принуждения и побуждения для субъекта В; $s(t)$ и $r(t)$ — векторы управлений принуждения и побуждения для субъекта С; $u(t)$ — вектор управлений субъекта Н; $b(t)$ — вектор управлений субъекта Н, трактуемых как взятки; векторы q , s и u имеют одинаковую размерность; J_B и g_B , J_C и g_C , J_H и g_H — интегральные и текущие функции выигрыша игроков В, С, Н соответственно; T — период рассмотрения, который может быть как конечным, так и бесконечным (второй вариант более адекватен задаче управления УР); α — коэффициент дисконтирования; $x(t)$ — вектор состояния УДС; соотношения (7) описывают динамику УДС при заданных начальных условиях и управлениях ведомого; условия (8) выражают требования УР применительно к УДС.

Предполагается, что Q , P , S , R , U и B — компактные подмножества соответствующих метрических пространств; функции J_B , J_C и J_H непрерывны по своим аргументам, для функции f из соотношений (7) выполнены условия Каратеодори.

Математически задача управления УР в условиях коррупции представляет собой динамическую игру (1)–(7) с фазовыми ограничениями (8). Ее смысл — нахождение решения динамической игры (1)–(7), т. е. приемлемого компромисса между субъектами управления, удовлетворяющего требованиям (8) к состоянию УДС, в условиях, когда управления игрока С зависят от размера взятки, предлагаемой игроком Н. Будем далее в ряде случаев для простоты опускать зависимость от времени.

Пусть сначала оба субъекта В и С применяют принуждение. Управления побуждения не используются. Будем интерпретировать управление принуждения $s(t)$ игрока С как некоторую «квоту» и считать ее функцией размера взятки: $\tilde{s}(t) = s(b(t))$, т. е. квота может быть увеличена в обмен на взятку (административная коррупция). В общем случае административная коррупция означает изменение управления принуждения за взятку в пользу агента. При этом принципал должен обеспечить выполнение требований УР (8) посредством принуждения супервайзера.

Информационный регламент в игре трех лиц типа СС («принуждение — принуждение») в условиях административной коррупции имеет следующий вид.

- Субъект верхнего уровня В (принципал) выбирает управление принуждения (постоянную стратегию) $q \in Q$ и сообщает ее субъекту С.
- Обозначим $\tilde{S} = \{\tilde{s} = s(q, b): Q \times B \rightarrow S\}$. С учетом известного значения $q \in Q$ субъект среднего

уровня С (супервайзер) выбирает и сообщает субъекту Н стратегию принуждения с обратной связью по размеру взятки (механизм административной коррупции) $\tilde{s} = s(q, b)$ из своего множества оптимальных ответов $R_C^{CC}(q, u, b) = \text{Arg max}_{\tilde{s} \in \tilde{S}} J_C^{CC}(q, s(q, b), u, b, x)$, где траектория x

определяется стратегией u . Структура стратегии $\tilde{s} = s(q, b)$ обусловлена следующими соображениями. Параметрическая зависимость $s(q)$ определяется воздействием принципа на множество допустимых стратегий супервайзера при принуждении. Зависимость $s(b)$ — это дополнительная обратная связь по величине взятки, возникающая при наличии административной коррупции.

- Зная механизм $\tilde{s} \in R_C^{CC}(q, u, b)$, субъект нижнего уровня (агент) выбирает свою стратегию $(\tilde{u}, b)$ из множества оптимальных ответов

$$R_H^{CC}(\tilde{s}) = \text{Arg max}_{(\tilde{u}, b) \in \tilde{U} \times B} J_H^{CC}(u(s(q, b)), b, x),$$

где $\tilde{U} = \{\tilde{u} = u(s(q, b)): \tilde{S} \rightarrow U\}$, траектория x порождается стратегией $\tilde{u} = u(s(q, b))$.

Гарантированный выигрыш принципа

$$\gamma_B^{CC} = \sup_{q \in Q} \inf_{\tilde{s} \in R_C^{CC}(q, u, b)} \inf_{(\tilde{u}, b) \in R_H^{CC}(\tilde{s})} J_B^{CC}(\pi(q, \tilde{s}, \tilde{u}, b, x)),$$

где $\pi: Q \times \tilde{S} \times \tilde{U} \times B \times X \rightarrow Q \times S \times U \times B \times X$ — проекция [24]. Равновесие «принуждения — принуждения» с учетом требований УР в условиях административной коррупции есть

$$(q^*, \tilde{s}^*, \tilde{u}^*, b^*) \in BR^{CC}: J_B^{CC}(\pi(q^*, \tilde{s}^*, \tilde{u}^*, b^*, x)) = \gamma_B^{CC},$$

$$BR^{CC} = \{(q, \tilde{s}, \tilde{u}, b) \in Q \times \tilde{S} \times \tilde{U} \times B: s(q, b) \in R_C^{CC}(q, u, b), u(s(q, b)) \in R_H^{CC}(\tilde{s})\} \quad (9)$$

при условии (8), траектория x^* порождается управлением $u^*(s^*(q^*, b^*))$.

Алгоритм построения равновесия (9) заключается в следующем.

1. Решается параметрическая задача оптимального управления субъектом Н (5)–(7): ее решением является пара $(b^*(t), u^*(t) = u^*(s(q(t), b^*(t))))$, при этом оператор $u^*(t)$ порождает для каждой тройки $(q(t), s(t), b^*(t))$ единственную траекторию УДС $x^*(t)$.

2. Решается параметрическая задача оптимального управления субъекта С (3), (4) в силу соотношений (7) при $b(t) = b^*(t)$, $u(t) = u^*(t)$. Ее решением является оператор $s^*(t) = s^*(q(t), b(t))$.

3. Решается задача оптимального управления субъекта В (1), (2) в силу соотношений (7) с дополнительным условием (8) при $s(t) = s^*(q(t), b^*(t))$, $u(t) = u^*(s^*(q(t), b^*(t)))$, $b(t) = b^*(t)$.

Если решение этой задачи $q^*(t)$ существует, то ситуация $(q^*, s^*(q^*, b^*), u^*(s^*(q^*, b^*)), b^*)$ есть равновесие «принуждения — принуждения» с учетом требований УР в условиях административной коррупции (9).

Пусть теперь оба субъекта В и С используют побуждение. Управления принуждения не применяются. Будем интерпретировать управление побуждения $r(t)$ субъекта С как некий «налог» и считать его функцией размера взятки: $\tilde{r}(t) = r(b(t))$, т. е. налог может быть уменьшен в обмен на взятку (экономическая коррупция). Возможны и другие интерпретации экономической коррупции, например, $r(t)$ есть некоторые ассигнования, которые могут быть увеличены за взятку, и т. п. Таким образом, общий смысл экономической коррупции — изменение стратегии побуждения за взятку в пользу агента. При этом принципал должен обеспечить выполнение требований УР посредством побуждения супервайзера.

Информационный регламент в игре трех лиц типа II («побуждение — побуждение») в условиях экономической коррупции имеет следующий вид.

- Субъект В выбирает стратегию побуждения с обратной связью $\tilde{p} \in \tilde{P} = \{\tilde{p} : R \rightarrow P\}$ и сообщает ее субъекту С.

- Обозначим $\tilde{R} = \{\tilde{r} = r(\tilde{p}, u, b) : \tilde{P} \times U \times B \rightarrow R\}$.

С учетом известной стратегии $\tilde{p} \in \tilde{P}$ субъект С выбирает и сообщает субъекту Н стратегию побуждения с дополнительной обратной связью по размеру взятки (механизм экономической коррупции) $\tilde{r} = r(\tilde{p}, u, b)$ из своего множества оптимальных ответов $R_C^{\text{II}}(\tilde{p}, u, b) = \text{Argmax}_{\tilde{r} \in \tilde{R}} J_C^{\text{II}}(\pi(\tilde{p}, \tilde{r}, u, b, x))$, где траектория x порождается стратегией u .

Рассмотрим подробнее структуру стратегии $\tilde{r} = r(\tilde{p}, u, b)$. Параметрическая зависимость $r(\tilde{p})$ обусловлена тем, что в момент принятия решения субъект С знает стратегию $\tilde{p}$ субъекта В, воздействующую на функцию выигрыша субъекта С. Зависимость $r(u)$ — это исходный механизм побуждения с обратной связью по действию агента u . За-

висимость $r(b)$ — это дополнительная обратная связь по размеру взятки b , возникающая при наличии экономической коррупции.

- Зная механизм $\tilde{r} \in R_C^{\text{II}}(\tilde{p}, u, b)$, субъект Н выбирает свою стратегию $(\tilde{u}, b)$ из множества оптимальных ответов $R_H^{\text{II}}(\tilde{r}) = \text{Argmax}_{(\tilde{u}, b) \in \tilde{U} \times B} J_H^{\text{II}}(u(\tilde{r}),$

$b, x)$, где $\tilde{U} = \{\tilde{u} = u(r(\tilde{p}, u, b)) : \tilde{R} \rightarrow U\}$, траектория x порождается управлением $u(r(\tilde{p}, u, b))$. Гарантированный выигрыш принципала равен

$$\gamma_B^{\text{II}} = \sup_{\tilde{p} \in \tilde{P}} \inf_{\tilde{r} \in R_C^{\text{II}}(\tilde{p}, u, b)} \inf_{(\tilde{u}, b) \in R_H^{\text{II}}(\tilde{r})} J_B^{\text{II}}(\pi(\tilde{p}, \tilde{r}, \tilde{u}, b,$$

$x))$. Равновесие «побуждения — побуждения» с учетом требований УР в условиях экономической коррупции есть

$$(\tilde{p}^*, \tilde{r}^*, \tilde{u}^*, b^*) \in BR^{\text{II}} : J_B^{\text{II}}(\pi(\tilde{p}^*, \tilde{r}^*, \tilde{u}^*, b^*, x^*)) = \gamma_B^{\text{II}},$$

$$BR^{\text{II}} = \{(\tilde{p}, \tilde{r}, \tilde{u}, b) \in \tilde{P} \times \tilde{R} \times \tilde{U} \times B : \tilde{r} \in R_C^{\text{II}}(\tilde{p}, u, b), (\tilde{u}, b) \in R_H^{\text{II}}(\tilde{r})\} \quad (10)$$

при условии (8), траектория x^* порождается управлением $u^*(r^*(\tilde{p}^*, u^*, b^*))$.

Алгоритм построения равновесия (10) опирается на предложенную в работе [24, с. 39] процедуру.

1. Обозначим для краткости $y = (\tilde{p}, \tilde{r}, \tilde{u}, b, x)$. Вычислим максимальный гарантированный выигрыш субъекта Н при использовании субъектами В и С стратегий наказания:

$$L_H^{\text{II}} = \sup_{(\tilde{u}, b) \in \tilde{U} \times B} \inf_{\tilde{p} \in \tilde{P}} \inf_{\tilde{r} \in \tilde{R}} J_H^{\text{II}}(\pi(y)).$$

2. Определим множество $D_H^{\text{II}} = \{(\tilde{r}, \tilde{u}, b) : J_H^{\text{II}}(\pi(\tilde{p}^P, \tilde{r}, \tilde{u}, b, x)) \geq L_H^{\text{II}}, \text{ где } \tilde{p}^P \text{ — стратегия наказания субъекта С субъектом В, т. е. } J_C^{\text{II}}(\pi(\tilde{p}^P, \tilde{r}, \tilde{u}, b, x)) = \min_{\tilde{p} \in \tilde{P}} J_C^{\text{II}}(\pi(y))\}$.

3. Вычислим максимальный гарантированный выигрыш субъекта С при учете интересов субъекта Н и использовании субъектом В стратегии наказания: $L_C^{\text{II}} = \sup_{(\tilde{r}, \tilde{u}, b) \in D_H^{\text{II}}} J_C^{\text{II}}(\pi(\tilde{p}^P, \tilde{r}, \tilde{u}, b, x))$.

4. Введем множество $D_C^* = \{(\tilde{p}, \tilde{r}, \tilde{u}, b) : J_C^{\text{II}}(\pi(y)) \geq L_C^{\text{II}}, J_H^{\text{II}}(\pi(y)) \geq L_H^{\text{II}}, \forall t x(t) \in X^*\}$ учета интересов субъектов С и Н при выполнении требований УР в условиях экономической коррупции.



5. Тогда гарантированный выигрыш субъекта В $\gamma_B^{\Pi} = \sup_{(\tilde{p}, \tilde{r}, \tilde{u}, b) \in D_C^*} J_B^{\Pi}(\pi(y))$, откуда получаем равновесие (10). Конечно, реализация этого алгоритма возможна при условиях $D_H^{\Pi} \neq \emptyset$, $D_C^* \neq \emptyset$.

Наконец, рассмотрим информационный регламент типа CI, т. е. субъект В использует принуждение, а субъект С — побуждение. Таким образом, супервайзер реализует стратегию побуждения агента с дополнительной обратной связью по размеру взятки (экономическая коррупция), а принципал должен обеспечить выполнение требований УР посредством принуждения супервайзера.

- Субъект В выбирает управление (постоянную стратегию) $q \in Q$ и сообщает ее субъекту С.
- Субъект С выбирает и сообщает субъекту Н стратегию $\tilde{r} = r(q, u, b)$ из своего множества оптимальных ответов $R_C^{CI}(q, u, b) = \text{Arg max}_{\tilde{r} \in \tilde{R}} J(\pi(q, \tilde{r}, u, b, x))$, где траектория x порождается стратегией u .
- Субъект Н выбирает свою стратегию $(\tilde{u}, b)$ из множества оптимальных ответов $R_H^{CI}(\tilde{r}) = \text{Arg max}_{(\tilde{u}, b) \in \tilde{U} \times B} J_H^{CI}(u(\tilde{r}), b, x)$.

Гарантированный выигрыш принципала $\gamma_B^{CI} = \sup_{q \in Q} \inf_{\tilde{r} \in R_C^{CI}(q, u, b)} \inf_{(\tilde{u}, b) \in R_H^{CI}(\tilde{r})} J_B^{CI}(\pi(q, \tilde{r}, \tilde{u}, b, x))$.

Равновесие «принуждения — побуждения» с учетом требований УР при экономической коррупции:

$$(q^*, \tilde{r}^*, \tilde{u}^*, b^*) \in BR^{CI} : J_B^{CI}(\pi(q^*, \tilde{r}^*, \tilde{u}^*, b^*, x^*)) = \gamma_B^{CI},$$

$$BR^{CI} = \{(q, \tilde{r}, \tilde{u}, b) \in Q \times \tilde{R} \times \tilde{U} \times B : \tilde{r} \in R_C^{CI}(q, u, b), (\tilde{u}, b) \in R_H^{CI}(\tilde{r})\} \quad (11)$$

при условии (8), траектория x^* порождается управлением $u^*(r^*(q^*, u^*, b^*))$.

Алгоритм нахождения равновесия (11) заключается в следующем.

1. Вычислить максимальный гарантированный выигрыш субъекта Н при использовании субъектами В и С стратегий наказания: $L_H^{CI} =$

$$= \sup_{(\tilde{u}, b) \in \tilde{U} \times B} \inf_{q \in Q} \inf_{\tilde{r} \in \tilde{R}} J_H^{CI}(\pi(q, \tilde{r}, \tilde{u}, b, x)).$$

2. Найти множество $D_H^{CI} = \{(\tilde{r}, \tilde{u}, b) : J_H^{CI}(\pi(q, \tilde{r}, \tilde{u}, b, x)) \geq L_H^{CI}\}$ для каждого $q \in Q$.

3. Вычислить максимальный гарантированный выигрыш субъекта С $\gamma_C^{CI}(q) = \sup_{(\tilde{r}, \tilde{u}, b) \in D_H^{CI}} J_C^{CI}(\pi(q, \tilde{r}, \tilde{u}, b, x))$; обозначим через $\tilde{r}^*$, $(\tilde{u}^*, b^*)$ стратегии субъектов С и Н, реализующие (возможно, с ε -точностью) величину $\gamma_C^{CI}(q)$ для каждого $q \in Q$.

4. Решить задачу оптимального управления с фазовыми ограничениями субъекта В (1), (2), (7), (8) при $\tilde{r} = \tilde{r}^*$, $(\tilde{u}, b) = (\tilde{u}^*, b^*)$. Если ее решение q^* существует, то ситуация $(q^*, \tilde{r}^*(q), \tilde{u}^*(r(q, u, b), b^*))$ есть равновесие «принуждения — побуждения» с учетом требований УР в условиях экономической коррупции (11).

Информационный регламент «побуждения — принуждения» рассматривается аналогично.

В качестве примера рассмотрим использование информационного регламента «побуждение — побуждение» (II) в условиях экономической коррупции в задаче контроля качества поверхностных вод.

2. МОДЕЛИ ЭКОНОМИЧЕСКОЙ КОРРУПЦИИ В СИСТЕМАХ КОНТРОЛЯ КАЧЕСТВА ПОВЕРХНОСТНЫХ ВОД

Рассмотрим трехуровневую систему контроля качества поверхностных вод, включающую в себя субъекты верхнего (федеральный центр, ФЦ), среднего (местные органы управления, ОУ), нижнего (промышленные предприятия, ПП) уровней и УДС [21]. Таким образом, $(В, С, Н) = (ФЦ, ОУ, ПП)$. Будем считать, что на каждом уровне имеется единственный субъект.

Пусть на берегу реки (УДС) расположено ПП, сбрасывающее загрязняющие вещества (ЗВ) в реку вместе со сточными водами. Для простоты исследуем случай только одного вида ЗВ, например, азотосодержащих ЗВ. Предполагается, что взаимоотношения между элементами исследуемой динамической системы устроены следующим образом: ФЦ воздействует на ОУ, ОУ — на ПП, а ПП — только на УДС.

Промышленное предприятие стремится к максимизации получаемой в результате производства прибыли. В процессе производства в водоток (УДС) сбрасываются ЗВ. Органы управления определяют размер платы за сброс ПП загрязнений в водоток, который (размер) является функцией взятки, и стремятся к максимизации поступающих к ним от ПП средств. Федеральный центр должен поддерживать УДС в устойчивом состоянии. При по-

буждении ФЦ определяет, какая часть средств, полученных с ПП в виде платы за сброс загрязнений в водоток, остается у ОУ; при принуждении — минимально допустимую степень очистки сточных вод на ПП (если ОУ применяет принуждение по отношению к ПП) или минимально допустимый размер платы за сброс ЗВ (если ОУ применяет побуждение по отношению к ПП).

Орган управления преследует только собственные интересы и может быть заинтересован в получении взяток от ПП. За взятки ОУ занижает размер платы за сброс загрязнений в водоток. Чиновники ОУ взятки рассматривают как один из факторов (наряду с доходами от платы за загрязнение) в общем балансе их интересов. Фактически ОУ, вопреки декларируемым целям защиты государственных интересов, часто выступает как коммерческая структура. Поскольку благополучие ОУ зависит только от ФЦ, но не от качества жизни в регионе, то стандарты качества речных и сточных вод могут нарушаться в силу экономической коррупции. Федеральный центр должен создать условия, при которых поддержание системы в устойчивом состоянии даже в условиях коррупции будет экономически выгодно для ОУ. Добиться этого ФЦ может неединственным образом, поэтому, кроме того, он стремится к оптимизации индивидуального функционала. Если на конечном интервале времени дисконтирование не учитывается, то целевой функционал ФЦ

$$J_B(p(\cdot), q_1(\cdot), q_2(\cdot), r(\cdot), u(\cdot), b(\cdot)) = \int_0^T \{-C_B(a(t)) + (1 - p(t))r(b(t))a(t)\}dt \rightarrow \max, \quad (12)$$

где t — временная координата; $u(t)$ — доля загрязняющих веществ, удаляемых на ПП в процессе очистки сточных вод; $W(t)$ ($a(t) = W(t)(1 - u(t))$) — количество ЗВ, сбрасываемых в реку предприятием до (после) очистки сточной воды в момент времени t ; $C_B(a(t))$ — функция затрат центра на улучшение качества речной воды, зависящая от общего количества сброшенных в реку ЗВ ($a(t)$); T — момент времени, до которого ведется рассмотрение; $r(b(t))$ — функция платы за единицу сброшенных ЗВ, в условиях коррупции зависящая от размера взятки $b(t)$; $p(t)$ — доля поступающих от ПП средств, которая остается в распоряжении ОУ; $q_1(t)$ и $q_2(t)$ — управление ФЦ при принуждении, когда ОУ, в свою очередь, применяет принуждение и побуждение соответственно.

Слагаемое $C_B(a(t))$ в выражении (12) отражает затраты ФЦ на очистку речной воды; $a(t)(1 - p(t)r(b(t)))$ — поступления средств от ПП в виде платы за сброс ЗВ в речную систему. Максимум функционала (12) ищется по $p(t)$ при побуждении и по $q_1(t)$ или $q_2(t)$ — при принуждении. Таким образом, в целевом функционале (12) учтены затраты на централизованную очистку речной воды и поступления от ПП в виде платы за сброс загрязнений.

Орган управления стремится к максимизации средств, поступающих к нему от ПП в виде взяток и платы за сброс загрязнений с учетом взяток, за вычетом расходов на очистку речной воды. Целевой функционал ОУ

$$J_C(p(\cdot), r(\cdot), s(\cdot), u(\cdot), b(\cdot)) = \int_0^T \{-C_C(a(t)) + a(t)r(b(t))p(t) + b(t)a(t)\}dt \rightarrow \max. \quad (13)$$

Здесь $C_C(a(t))$ — функция затрат ОУ на улучшение качества речной воды; ОУ управляет размером платы за сброс загрязнений в водоток $r(b(t))$ при побуждении и минимально допустимой степенью очистки сточных вод на ПП $s(b(t))$ — при принуждении. Слагаемое $C_C(a(t))$ отражает затраты ОУ на очистку речной воды; $a(t)p(t)r(b(t))$ — поступления средств от ПП за сброс загрязнений в водоток; $a(t)b(t)$ — размер взятки за сброс в водоток загрязняющих веществ в количестве $a(t)$.

Цель ПП — максимизация своей прибыли в условиях коррупции, т. е.

$$J_H(r(\cdot), u(\cdot), b(\cdot)) = \int_0^T \{zR(\Phi(t)) - C_H(u(t))W(t) - r(b(t))a(t) - b(t)a(t)\}dt \rightarrow \max. \quad (14)$$

Здесь $C_H(u(t))$ — функции затрат ПП на очистку единицы сбрасываемых ЗВ; $\Phi(t)$ — производственные фонды; $R(\Phi(t))$ — производственная функция ПП; $z = \text{const}$ — цена единицы продукции. Максимум функционала (14) ищется по функциям $u(t)$ и $b(t)$.

Слагаемое $zR(\Phi(t))$ выражает доход ПП от реализации $R(\Phi(t))$ единиц произведенной на предприятиях продукции, $a(t)r(b(t))$ — плату за сброс ЗВ в водоток; $C_H(u(t))W(t)$ — затраты ПП на очистку сточной воды; $a(t)b(t)$ — размер взятки за сброс в водоток загрязняющих веществ в количестве $a(t)$. Далее рассмотрим случай $C_H(u(t)) = D \frac{u(t)}{1 - u(t)}$; $D = \text{const}$.



Оптимизационные задачи (12) — (14) решаются при ограничениях на управления

$$\begin{aligned} & \text{— ФЦ } \gamma_1 \leq p(t) \leq \gamma_2; \quad 0 \leq q_1(t) \leq 1 - \varepsilon; \\ & \quad 0 \leq q_2(t) \leq r_{\max}; \quad \forall t: t \in [0, T]; \end{aligned} \quad (15)$$

$$\begin{aligned} & \text{— ОУ } q_1(t) \leq s(t) \leq 1 - \varepsilon; \quad q_2(t) \leq r(b(t)) \leq r_{\max}; \\ & \quad \forall t: t \in [0, T]; \end{aligned} \quad (16)$$

$$\begin{aligned} & \text{— ПП } s(t) \leq u(t) \leq 1 - \varepsilon; \quad 0 \leq b(t) \leq b_{\max}; \\ & \quad \forall t: t \in [0, T]. \end{aligned} \quad (17)$$

Здесь значение величины ε определяется технологическими возможностями ПП по очистке сточных вод; $b_{\max}$ — максимально возможная взятка, приходящаяся на единицу сброшенных ЗВ; $r_{\max}$ — максимально допустимый размер платы за сброс ЗВ без коррупции; γ_1 и γ_2 — минимально допустимые доли средств, поступающих от ПП к ФЦ и ОУ; значения $r_{\max}$, $b_{\max}$, γ_1 и γ_2 заданы.

Динамика изменения производственных фондов ПП описывается обыкновенным дифференциальным уравнением

$$\frac{d\Phi}{dt} = -k\Phi + Y, \quad \Phi(0) = \Phi_0, \quad (18)$$

где $k = \text{const}$ — коэффициент амортизации производственных фондов; Y — инвестиции в производство, которые будем считать постоянной величиной; величина $\Phi_0 = \text{const}$ задана.

Пусть общее количество сбрасываемых ЗВ (до очистки) зависит от количества произведенной на ПП продукции линейно и производственные функции имеют вид

$$\begin{aligned} W(t) &= \beta R(\Phi), \quad R(\Phi) = \chi \Phi^{0.5}(t), \\ \chi, \beta &= \text{const}. \end{aligned} \quad (19)$$

В качестве характеристики речной воды возьмем концентрацию ЗВ $x(t)$, изменение которой со временем описывается обыкновенным дифференциальным уравнением

$$\frac{dx}{dt} = F(x(t), u(t), t), \quad x(0) = x_0, \quad (20)$$

где $F(x(t), u(t), t)$ — некоторая заданная функция.

Известны предельно допустимые концентрации ЗВ в водотоке $x_{\max}$

$$0 \leq x(t) \leq x_{\max}, \quad \forall t: t \in [0, T], \quad (21)$$

и ограничения на качество сточной воды

$$\frac{W(t)[1 - u(t)]}{Q^0(t)} \leq Q_{\max}, \quad \forall t: t \in [0, T], \quad (22)$$

где $Q^0(t)$ — расход воды на ПП в момент времени t ; значения $x_{\max}$ и $Q_{\max}$ заданы.

Рассмотрим вначале информационный регламент II. Будем считать, что $q_1(t) \equiv q_2(t) \equiv s(t) \equiv 0$, управлениями служат функции $p(t)$, $r(t)$, $u(t)$, t и $b(t)$. Введем обозначения для проекций $\pi(\tilde{r}) = r(p(r), u, b)$; $\pi(\tilde{u}) = u(r(p(r), u, b))$. Будем считать, что $\varepsilon^2 r_{\max} \leq D \leq r_{\max}$.

Решение задачи (12)–(22) строится в соответствии с описанным в § 1 алгоритмом в случае информационного регламента II в условиях экономической коррупции, а именно:

1) вычисляется величина $L_H^{\text{II}} = T(D - 2\sqrt{Dr_{\max}}) \times \int_0^T W(t)dt + \int_0^T zR(\Phi)dt$; стратегии наказания имеют вид $\tilde{p}^P = \tilde{p}(r_{\max}) = 0$; $\tilde{r}^P = \tilde{r}(\tilde{p}(r_{\max}), u^P, b^P) = r_{\max}$; $u^P = 1 - \sqrt{D/r_{\max}}$; $b^P = 0$;

2) определяется множество $D_H^{\text{II}} = \{(\tilde{r}, \tilde{u}, b) : - \int_0^T W(t) \left\{ \frac{D - \pi(\tilde{u})}{1 - \pi(\tilde{u})} + (1 - \pi(\tilde{u})(\pi(\tilde{r}) + b(t))) \right\} dt \leq T \times (D - 2\sqrt{Dr_{\max}}) \int_0^T W(t)dt \}$;

3) вычисляется величина

$$\begin{aligned} L_C^{\text{II}} &= \sup_{(\tilde{r}, \tilde{u}, b) \in D_H^{\text{II}}} \int_0^T \{-C_C(W(t)(1 - \pi(\tilde{u})) + \\ & \quad + (1 - \pi(\tilde{u}))W(t)b(t))\} dt; \end{aligned}$$

4) вводится множество

$$\begin{aligned} D_C^* &= \{(\tilde{p}, \tilde{r}, \tilde{u}, b) : \int_0^T \{-C_C((1 - \pi(\tilde{u}))W(t) + \\ & \quad + (1 - \pi(\tilde{u}))W(t)(\pi(\tilde{r})p(r(t)) + b(t)))\} dt \geq L_C^{\text{II}}, \\ & \quad - \int_0^T W(t) \left\{ \frac{D - \pi(\tilde{u})}{1 - \pi(\tilde{u})} + (1 - \pi(\tilde{u})(\pi(\tilde{r}) + b(t))) \right\} dt \leq \\ & \quad \leq T(D - 2\sqrt{Dr_{\max}}) \int_0^T W(t)dt, \end{aligned}$$

$$\forall t: 0 \leq x(t) \leq x_{\max}, \quad W(t)[1 - u(t)]/Q_0(t) \leq Q_{\max};$$

5) определяется величина

$$\begin{aligned} \gamma_B^{\text{II}} &= \sup_{(\tilde{p}, \tilde{r}, \tilde{u}, b) \in D_C^*} \int_0^T \{-C_B(W(t)(1 - \pi(\tilde{u})) + \\ & \quad + (1 - p(r(t)))W(t)\pi(\tilde{r})(1 - \pi(\tilde{u})))\} dt \end{aligned}$$

и находится равновесие «побуждения — побуждения».

Приведем примеры нахождения равновесий. Во всех примерах равновесие «побуждения — побуждения» находится численно методом прямого упорядоченного перебора с учетом гипотезы постоянства управлений на некоторых временных интервалах [19, 20]. Именно, считается, что субъекты меняют свои управления n раз в известные моменты времени, т. е.

$$m(t) = \begin{cases} m_1, & \text{если } 0 \leq t < t_1, \\ m_2, & \text{если } t_1 \leq t < t_2, \\ \dots & \dots \\ m_N, & \text{если } t_{N-1} \leq t < T, \end{cases}$$

где $m_i = \text{const}$; $t_i = i\Delta t$; $\Delta t = T/n$; $m(t)$ — управление одного из субъектов управления (одна из функций $p(t)$, $r(t)$, $u(t)$ и $b(t)$).

Пример 1. Пусть $F(x(t), u(t), t) = -\omega x + C_1 W(t)(1 - u(t))$; $C_B(a(t)) = C_2 a(t)$; $C_C(a(t)) = C_3 a(t)$; $C_1, C_2, C_3 = \text{const}$; $C_2 = 20$ у. е.; $C_3 = 30$ у. е.; $C_1 = 0,015$ (л · сут⁻¹); $n = 3$; $\omega = 0,002$ сут⁻¹; $z = 25$ сут⁻¹; $r_{\max} = 200$ у. е.; $b_{\max} = 70$ у. е.; $T = 540$ сут; $x_{\max} = 10$ мг/л; $Q_{\max} = 100$; $\Phi_0 = 10^{16}$ у. е.; $D = 18$; $x_0 = 6$ мг/л; $\chi = 2$; $\gamma_1 = 0,3$; $\gamma_2 = 0,8$; $\varepsilon = 0,1$; $Y = 0$ у. е.; $\beta = 0,3$ мг/(сут у. е.); $Q^0(t) = 0,6 W(t)/Q_{\max}$; $k = 0,001$ сут⁻¹ (у. е. — условная единица). Получим, что коррупция в системе есть и $L_H^{\text{II}} = -132\,507$ у. е.; $L_C^{\text{II}} = 285\,317$ у. е.; $\gamma_B^{\text{II}} = 105\,764$ у. е.; $J_C^{\text{II}} = 285\,882$ у. е.; $J_H^{\text{II}} = -132\,483$ у. е., где J_C^{II} , J_H^{II} — прибыль ОУ и ПП соответственно в равновесии «побуждения — побуждения».

Пример 2. В случае входных данных примера 1 и уменьшения затрат на очистку сточных вод на ПП ($D = 1$) коррупция в системе также присутствует, прибыль ПП растет, остальных субъектов уменьшается, так как на ПП выгодна высокая степень очистки сточных вод: $L_H^{\text{II}} = 397\,870$ у. е.; $L_C^{\text{II}} = 93\,911$ у. е.; $\gamma_B^{\text{II}} = 2166$ у. е.; $J_C^{\text{II}} = 94\,262$ у. е.; $J_H^{\text{II}} = 398\,121$ у. е.

Пример 3. В случае входных данных примера 1 и увеличения расходов на очистку речной воды со стороны ОУ ($C_3 = 200$ у. е.) ОУ терпит убытки, коррупция в системе есть: $L_H^{\text{II}} = -132\,507$ у. е.; $L_C^{\text{II}} = -269\,677$ у. е.; $\gamma_B^{\text{II}} = 175\,732$ у. е.; $J_C^{\text{II}} = -269\,156$ у. е.; $J_H^{\text{II}} = -132\,452$ у. е.

Пример 4. Равновесие II с учетом условий УР системы существует не всегда. Например, его не существует в случае входных данных примера 1 и роста на ПП затрат на очистку сточных вод ($D = 100$), увеличения максимально возможной величины удельной взятки ($b_{\max} = 500$ у. е.), уменьшения максимально допустимой

платы за сброс ЗВ ($r_{\max} = 10$ у. е.) или ухудшения экологического состояния системы ($x_0 = 9$ мг/л).

Пример 5. В случае входных данных примера 1 и уменьшения максимально возможной удельной взятки ($b_{\max} = 10$ у. е.) коррупция в системе отсутствует, прибыль ФЦ и ПП растет, ОУ — падает: $L_H^{\text{II}} = -132\,507$ у. е.; $L_C^{\text{II}} = -477\,12$ у. е.; $\gamma_B^{\text{II}} = 301\,948$ у. е.; $J_C^{\text{II}} = 68\,267$ у. е.; $J_H^{\text{II}} = -111\,052$ у. е. ♦

Рассмотрим информационный регламент СС. Считаем, что $q_2(t) \equiv 0$, а функции $p(t)$ и $r(t)$ заданы и не меняются, управлениями служат функции $q_1(t)$, $s(t)$, $u(t)$ и $b(t)$. Пусть $\forall t: t \in [0, T] \rightarrow D \geq \varepsilon^2 r(t)$.

Решение задачи (12)—(22) опять строится в соответствии с описанным в § 1 алгоритмом в случае информационного регламента СС, а именно:

1) если $1 - \sqrt{D/r(t) + b_{\max}} \leq s(t) < 1 - \varepsilon$, то $b^*(t) = 0$; $u^*(t) = s(t)$;

если $1 - \sqrt{D/r(t)} \leq s(t) < 1 - \sqrt{D/r(t) + b_{\max}}$, то $b^*(t) = b_{\max}$; $u^*(t) = 1 - \sqrt{D/r(t) + b_{\max}}$ или $(b^*(t) = 0$; $u^*(t) = s(t))$ в зависимости от входных данных модели;

если $s(t) < 1 - \sqrt{D/r(t)}$, то $(b^*(t) = b_{\max}$; $u^*(t) = 1 - \sqrt{D/r(t) + b_{\max}}$) или $(b^*(t) = 0$; $u^*(t) = 1 - \sqrt{D/r(t)})$.

2) решается параметрическая задача максимизации функционала (13), (16) с учетом уравнения (18) и функцией (19) при $b(t) = b^*(t)$, $u(t) = u^*(t)$; ее решением является оператор $s^*(t) = s^*(q_1(t), b(t))$;

3) решается задача оптимального управления субъекта В (12), (15), (18)—(22) при $s(t) = s^*(q_1(t), b^*(t))$; $u(t) = u^*(s^*(q_1(t), b^*(t)))$, $b(t) = b^*(t)$. Находится равновесие «принуждения — принуждения».

Равновесие ищется численно методом прямого упорядоченного перебора с учетом гипотезы постоянства управлений на некоторых временных интервалах [19, 20] в случае $p(t) = 0,4$; $r(t) = r_{\max}/2$. Заметим, что результаты существенно зависят от заданных функций $p(t)$ и $r(t)$.

Пример 6. В случае входных данных примера 1 доходы ФЦ и ПП выросли по сравнению с примером 1, ОУ — упали и $J_B^{\text{CC}} = 142\,658$ у. е.; $J_C^{\text{CC}} = 35\,664$ у. е.; $J_H^{\text{CC}} = 109\,371$ у. е., где J_B^{CC} , J_C^{CC} и J_H^{CC} — прибыль ФЦ, ОУ и ПП соответственно в равновесии «принуждения — принуждения».

Пример 7. В случае входных данных примера 2 доходы ФЦ и ПП опять выросли по сравнению с примером 2,



ОУ — упали, и $J_B^{CC} = 28\,531$ у. е.; $J_C^{CC} = 7132$ у. е.; $J_H^{CC} = 4\,588\,851$ у. е.

Пример 8. В случае входных данных примера 3 доходы ФЦ и ПП упали по сравнению с примером 3, ОУ — выросли, $J_B^{CC} = 28\,531$ у. е.; $J_C^{CC} = -114\,127$ у. е.; $J_H^{CC} = -632\,454$ у. е.

Пример 9. В случае входных данных примера 4 равновесие СС с учетом условий УР системы существует во всех указанных случаях (при $D = 100$, $b_{\max} = 500$ у. е., $x_0 = 9$ мг/л). Путем принуждения удастся добиться выполнения условий УР. ♦

В случае информационного регламента СИ будем считать, что $q_1(t) \equiv 0$; $s(t) \equiv 0$, а функция $p(t)$ задана, управлениями служат функции $q_2(t)$, $r(t)$, $u(t)$, $b(t)$. Введем обозначения для проекций $\pi(\tilde{r}) = r(q, u, b)$; $\pi(\tilde{u}) = u(q, u, b)$.

Решение задачи (12)—(22) строится в соответствии с описанным в § 1 алгоритмом в случае информационного регламента СИ в условиях экономической коррупции, а именно:

1) вычисляется величина $L_H^{CI} = -\frac{1-\varepsilon}{\varepsilon} \int_0^T W(t)dt - \varepsilon \int_0^T r(t)W(t)dt + \int_0^T zR(\Phi)dt$; $q_2^p = 1 - \varepsilon$; $\tilde{r}^p = \tilde{r}(q_2^p, u^p, b^p) = r_{\max}$; $u^p = 1 - \varepsilon$; $b^p = 0$;

2) определяется множество

$$D_H^{CI} = \{(\tilde{r}, \tilde{u}, b) : -\int_0^T W(t) \left\{ D \frac{\pi(\tilde{u})}{1 - \pi(\tilde{u})} + (1 - \pi(\tilde{u})(\pi(\tilde{r}) + b(t))) \right\} dt \leq \frac{1-\varepsilon}{\varepsilon} \int_0^T W(t)dt + \varepsilon \int_0^T r(t)W(t)dt;$$

3) вычисляется величина

$$\gamma_C^{CI}(q_2) = \sup_{(\tilde{r}, \tilde{u}, b) \in D_H^{CI}} \int_0^T \{-C_C(W(t)(1 - \pi(\tilde{u})) + p(t)W(t)\pi(\tilde{r})(1 - \pi(\tilde{u})))\} dt,$$

обозначим через $\tilde{r}^*$ и $(\tilde{u}^*, b^*)$ стратегии субъектов С и Н, реализующие величину $\gamma_C^{CI}(q_2)$;

4) решается задача оптимального управления с фазовыми ограничениями субъекта В (12), (15), (18)—(22) при $\tilde{r} = \tilde{r}^*$, $(\tilde{u}, b) = (\tilde{u}^*, b^*)$ и находится равновесие «принуждения — побуждения» с уче-

том требований УР в условиях экономической коррупции.

Равновесие СИ, как и в рассмотренных ранее случаях, ищется численно и существенно зависит от выбранной функции $p(t)$. Далее считается, что $p(t) \equiv 0,4$.

Пример 10. В случае входных данных примеров 1, 4 и 5 равновесия «принуждения — побуждения» с учетом условий УР системы не существует. Для ОУ и ПП выгодно нарушать условия УР, а у ФЦ недостаточно рычагов воздействия на них.

Пример 11. В случае входных данных примера 2 доходы ОУ и ПП выросли по сравнению с примером 2, ФЦ — упали, коррупция в системе есть, и $L_H^{CI} = 387\,556$ у. е.; $J_B^{CI} = 62\,665$ у. е.; $L_C^{CI} = 42\,002$ у. е.; $J_H^{CI} = 389\,882$ у. е., где J_B^{CI} , J_C^{CI} и J_H^{CI} — прибыль ФЦ, ОУ и ПП соответственно в равновесии «принуждения — побуждения».

Пример 12. В случае входных данных примера 3 доходы ПП упали по сравнению с примером 3, ФЦ и ОУ — выросли, коррупция в системе есть, $L_H^{CI} = -703\,783$ у. е.; $J_B^{CI} = 136\,309$ у. е.; $J_C^{CI} = -68\,154$ у. е.; $J_H^{CI} = -636\,750$ у. е. ♦

Таким образом, как в случае информационного регламента П, так и регламента СИ равновесие с учетом условий УР системы строится не всегда, т. е. ФЦ не удастся обеспечить выполнение условий УР из-за коррупции. В случае регламента СС равновесие существует для более широкого класса входных функций. Прибыль различных субъектов управления в равновесии системы существенно зависит от принятого информационного регламента.

ЗАКЛЮЧЕНИЕ

В настоящей работе даны определения и проведено исследование равновесий «принуждения и побуждения» с учетом требований устойчивого развития в условиях коррупции в трехуровневых системах управления. Методы принуждения и побуждения при иерархическом управлении [17] формализуются как информационные расширения игры двух лиц, в основном соответствующие регламентам Γ_{1t} и Γ_{2t} , описанным в книгах [24, 25]. Требования устойчивого развития учитываются дополнительно как фазовые ограничения в дифференциальной игре. Административная и экономическая коррупция трактуются как обратная связь управлений принуждения и побуждения по размеру взятки [18—23].

В играх трех лиц в аспекте методов управления возможны четыре варианта информационных регламентов: оба субъекта применяют один и тот же метод (принуждение или побуждение) либо они

применяют различные методы, при этом во всех случаях управления игрока среднего уровня дополняются административной или экономической коррупцией. Равновесие «принуждения — принуждения» с учетом требований устойчивого развития в условиях административной коррупции строится путем решения последовательности связанных параметрических задач оптимального управления с соответствующими фазовыми ограничениями. Для нахождения равновесия «побуждения — побуждения» с учетом требований устойчивого развития в условиях экономической коррупции использована предложенная в работе [24] процедура, основанная на построении взаимосвязанных, выгодных для игроков, программ. Равновесия «принуждения — побуждения» и «побуждения — принуждения» с учетом требований устойчивого развития в условиях коррупции ищутся с помощью комбинации этих подходов.

В качестве иллюстрации представлено применение указанных методов к задачам контроля качества поверхностных вод. Показано, что для рассмотренных примеров в случае регламента «принуждения — принуждения» принципалу всегда удастся обеспечить выполнение требований устойчивого развития даже при коррупции, а для других регламентов — только при определенных значениях модельных параметров. Таким образом, в рамках примеров административно-законодательные методы в борьбе с коррупцией более эффективны, чем экономические.

ЛИТЕРАТУРА

1. *Rose-Ackerman S.* The Economics of Corruption // *Journal of Political Economy*. — 1975. — N 4. — P. 187–203.
2. *Becker G.* Crime and Punishment: An Economic Approach // *Journal of Political Economy*. — 1968. — N 76. — P. 169–218.
3. *Bac M.* Corruption and supervision costs in hierarchies // *Journal of Comparative Economics*. — 1996. — N 2. — P. 99–118.
4. *Bag P.K.* Controlling corruption in hierarchies // *J. Comparative Economics*. — 1997. — Vol. 25, N 3. — P. 322–344.
5. *Mookherjee D., Png I.P.* Corruptible Law Enforcers; How Should They Be Compensated // *Economic Journal*. — 1995. — N 105. — P. 112–121.
6. *Hindriks J., Keen M., Muthoo A.* Corruption, Extortion and Evasion // *J. Public Economics*. 1999. — Vol. 74, N 3. — P. 395–430.
7. *Mishra A.* Hierarchies, incentives and collusion in a model of enforcement // *J. Economic Behavior and Organization*. — 2002. — Vol. 47. — P. 165–178.
8. *Lambert-Mogiliansky A., Majumdar M., Radner R.* Petty corruption: A game — theoretical approach // *Journal of Economic Theory*. — 2008. — Vol. 4. — P. 273–297.
9. *Drugov M.* Information and delay in an agency model // *The Rand J. Economics*. — 2010. — Vol. 41. — P. 598–615.
10. *Васин А.А., Картунова П.А., Уразов А.С.* Модели организации государственных инспекций и борьбы с коррупцией // *Математическое моделирование*. — 2010. — Т. 22, № 4. — С. 67–89.
11. *Васин А.А., Николаев П.В., Уразов А.С.* Механизмы подавления коррупции // *Журнал новой экономической ассоциации*. — 2011. — № 10. — С. 10–30.
12. *Левин М.И., Цирик М.Л.* Коррупция как объект математического моделирования // *Экономика и математические методы*. — 1998. — Т. 34, № 3. — С. 40–62.
13. *Полтерович В.М.* Факторы коррупции // *Экономика и математические методы*. — 1998. — Т. 34, № 3. — С. 30–39.
14. *Выборнов Р.А.* Модели и методы управления организационными системами с коррупционным поведением участников. — М.: ИПУ РАН, 2006.
15. *Левин М.И., Левина Е.А., Покатович Е.В.* Лекции по экономике коррупции. — М.: Изд. дом ВШЭ, 2011.
16. *Угольницкий Г., Денин К.* Математические модели коррупции. Теория и приложения. — LAP Lambert Academic Publishing, 2011. — 152 с.
17. *Угольницкий Г.А.* Иерархическое управление устойчивым развитием. — М.: Физматлит, 2010. — 336 с.
18. *Денин К.И., Угольницкий Г.А.* Теоретико-игровая модель коррупции в системах иерархического управления // *Известия РАН. Теория и системы управления*. — 2010. — № 1. — С. 192–198.
19. *Угольницкий Г.А., Усов А.Б.* Управление устойчивым развитием иерархических систем в условиях коррупции // *Проблемы управления*. — 2010. — № 6. — С. 19–26.
20. *Угольницкий Г.А., Усов А.Б.* Устойчивое развитие систем управления с учетом коррупции // *Математическая теория игр и ее приложения*. — 2010. — Т. 2, вып. 4. — С. 106–119.
21. *Угольницкий Г.А., Усов А.Б.* Статические модели коррупции в системах контроля качества водных ресурсов // *Проблемы управления*. — 2012. — № 4. — С. 38–44.
22. *Антоненко А.В., Угольницкий Г.А., Усов А.Б.* Статические модели борьбы с коррупцией в иерархических системах управления // *Известия РАН. Теория и системы управления*. — 2013. — № 4. — С. 165–176.
23. *Угольницкий Г.А., Усов А.Б.* Динамические иерархические игры двух лиц в программных стратегиях и их приложения // *Математическая теория игр и ее приложения*. — 2013. — Т. 5, вып. 2. — С. 82–104.
24. *Горелик В.А., Кононенко А.Ф.* Теоретико-игровые модели принятия решений в эколого-экономических системах. — М.: Радио и связь, 1982. — 144 с.
25. *Горелик В.А., Горелов М.А., Кононенко А.Ф.* Анализ конфликтных ситуаций в системах управления. — М.: Радио и связь, 1991.
26. *Горелов М.А., Кононенко А.Ф.* Динамические модели конфликтов. Язык моделирования // *Автоматика и телемеханика*. В печати.
27. *Горелов М.А., Кононенко А.Ф.* Динамические модели конфликтов. Равновесия // Там же. В печати.
28. *Горелов М.А., Кононенко А.Ф.* Динамические модели конфликтов. Иерархические игры // Там же. В печати.

Статья представлена к публикации членом редколлегии В.Н. Бурковым.

Угольницкий Геннадий Анатольевич — д-р физ.-мат. наук, зав. кафедрой, ☎ (863) 297-51-14, ✉ ougoln@mail.ru,

Усов Анатолий Борисович — д-р физ.-мат. наук, профессор, ✉ usov@math.rsu.ru,

Южный федеральный университет, г. Ростов-на-Дону.

ЭКСАФЛОПНЫЕ СУПЕР-ЭВМ: КОНТУРЫ АРХИТЕКТУРЫ¹

С.А. Степаненко, В.В. Южаков

Исследованы архитектурные аспекты вычислительных систем эксафлопной производительности. Получены оценки параметров вычислительной и коммуникационной сред. Показаны необходимость и возможности применения архитектурных средств масштабирования эффективности.

Ключевые слова: гибридная архитектура, средства масштабирования, эффективность, реконфигурируемая структура, минимизация, длительность обменов, топологическое резервирование.

ВВЕДЕНИЕ

Задача эффективного применения супер-ЭВМ актуальна в течение всей истории развития вычислительной техники. Это обусловлено как наличием сложнейших задач, для решения которых собственно и разрабатываются супер-ЭВМ, так и большими ресурсами, требуемыми для создания последних.

Достижение эффективности требует учета свойств архитектуры вычислительных систем в прикладных программах и реализации в архитектуре средств, позволяющих ускорить выполнение вычислений. На различных этапах эволюции вычислительной техники применялись различные архитектурные средства — от введения кэш-памяти до создания специализированных вычислителей, аппаратно реализующих алгоритмы [1].

В настоящей работе исследуются архитектурные аспекты, которые с большой вероятностью будут присущи супер-ЭВМ эксафлопной производительности, необходимость которой и возможности создания показаны, например, в публикациях [2, 3].

Эти аспекты обусловлены объективными факторами — энергопотреблением системы эксафлопной производительности и числом задействованных в ней процессорных ядер, определяющим степень параллелизма. Цель работы заключается в:

— обосновании необходимости применения гибридных архитектур для достижения эксафлопной производительности;

— качественной оценке параметров вычислительной и коммуникационной сред;

— описании архитектурных средств масштабирования эффективности, позволяющих на различных уровнях параллелизма учитывать особенности исполняемых процессов, что при прочих равных условиях позволяет уменьшить длительность вычислений и достигнуть практически приемлемых значений производительности и эффективности.

1. ЭТАПЫ ЭВОЛЮЦИИ АРХИТЕКТУРЫ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ

Этапы эволюции вычислительных систем согласно, например, работе [1] можно охарактеризовать применяемыми дисциплинами вычислений и архитектурами, реализующими эти дисциплины.

Для достижения производительности $1-10^9$ оп/с оказалось достаточно SISD-дисциплины (Single Instruction Single Data) и однопроцессорной архитектуры.

Достижение $10^{12}-10^{15}$ оп/с потребовало MIMD-дисциплины (Multiple Instructions Multiple Data) и мультипроцессорной архитектуры с разделенной памятью.

Достижение 10^{18} оп/с — эксафлопс — предполагает применение MIMD- и SIMD-дисциплин (Single Instruction Multiple Data) вычислений, реализуемых гибридными архитектурами. Процессорные элементы в них содержат универсальные про-

¹ Статья написана по материалам доклада на шестой международной конференции «Параллельные вычисления и задачи управления» (РАСО'2012), Москва, 24 — 26 октября 2012 г.

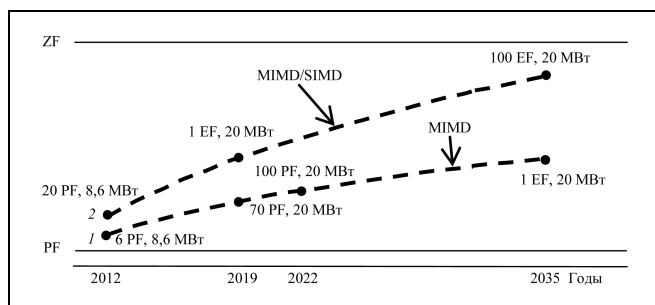


Рис. 1. Оценки производительности и потребляемой мощности: 1 — CPU-only «Titan»; 2 — CPU-Accelerated «Titan»; PF, EF и ZF — пета-, экса- и зеттафлопс соответственно

цессоры — MIMD-компонент и арифметические ускорители — SIMD-компонент.

Применение SIMD-компонентов позволяет гибридной системе достигнуть при определенных условиях производительности 10^{18} оп/с, потребляя 10–20 МВт (в тех же условиях для MIMD системы потребуется не менее 100 МВт); количество MIMD-ядер универсальных процессоров и SIMD-ядер ускорителей составит в системе соответственно $\sim 10^7$ и 10^8 шт.

Значения производительности и потребляемой мощности, требуемые для систем, реализующих MIMD-дисциплину и MIMD/SIMD-дисциплину, показаны на рис. 1.

Эффективное задействование гибридных архитектур требует разработки соответствующих вычислительных процессов и анализа их особенностей, в частности, выделения фрагментов, «быстро» исполняемых универсальными процессорами (MIMD-компонентом), и фрагментов, «быстро» исполняемых арифметическими ускорителями (SIMD-компонентом). В свою очередь это влечет необходимость применения нового прикладного и системного программного обеспечения.

Масштабность и трудоемкость создания качественно новых аппаратных и программных средств породили многочисленные исследовательские проекты, выполняемые в различных странах и направленные на освоение гибридных архитектур [2, 5].

Из результатов исследований, выполняемых в мире, следует, что эксафлопная производительность может быть достигнута в результате комплекса взаимозависимых работ, которые включают в себя разработку и создание:

- оптимальной архитектуры, позволяющей обеспечить эффективное исполнение приложений вычислительной системой, состоящей из $\sim 10^8$ ядер;
- аппаратных компонентов, удовлетворяющих конструктивным ограничениям и требованиям надежности;

— прикладного и системного программного обеспечения, реализующего управление ресурсами и надежное исполнение приложений на разных уровнях параллелизма;

— экспериментальных систем, позволяющих верифицировать проектные решения.

Удовлетворительным результатом этих работ, приемлемым для практики, будет создание вычислительной машины, имеющей пиковую производительность не менее 1 эксафлопс и соответствующую пропускную способность средств обмена информацией, энергопотребление 10–20 МВт, занимающую 100–200 стоек, оснащенную системным программным обеспечением, позволяющим эффективно распараллеливать приложения на $\sim 10^8$ процессов, а также соответствующим прикладным программным обеспечением, допускающим эффективное исполнение с указанным параллелизмом.

Оценим параметры компонентов и некоторые архитектурные средства, требуемые для достижения такого результата.

2. ПАРАМЕТРЫ АППАРАТНЫХ КОМПОНЕНТОВ

Ключевыми аппаратными компонентами служат:

— процессоры для научных расчетов, в качестве которых в ближайшей перспективе рассматриваются MIMD/SIMD-процессоры (MIMD — универсальная часть, SIMD — арифметические ускорители), называемые также гибридными; в более отдаленной — MIMD/SIMD/FPGA;

— система межпроцессорного обмена, включая средства реализации коммуникационной среды.

Оценим параметры вычислительной среды и коммуникационной среды, необходимые для достижения эксафлопной производительности.

2.1. Параметры и состав вычислительной среды

Вычислительный компонент эксафлопной машины (включающий в себя не только процессоры, но и память) должен обеспечить достижение эксафлопной производительности при «разумном» значении энергопотребления — 10–20 МВт и технологической надежности.

Первое может быть достигнуто совместным применением MIMD- и SIMD-компонентов. Вследствие сравнительно простой структуры, энергопотребление, конструктивные размеры и стоимость, приходящиеся на единицу производительности SIMD-компонентов, примерно в 10 раз меньше по сравнению с MIMD-компонентами.

Из приведенных в работах [6–8] данных следуют представленные в табл. 1 значения удельной производительности для MIMD- и SIMD-компо-



Таблица 1

Значения удельной производительности q , Гфлопс/Вт

Год	MIMD	SIMD
2014	2—4	24
2016	4—8	50
2018	10—15	100

нентов. В соответствии с этими значениями возможна разработка MIMD/SIMD-процессоров производительностью (значения до и после символа «/» означают соответственно производительность MIMD- и SIMD-компонентов):

— 500...1000 Гфлопс/4000...8000 Гфлопс в 2014 г. проектные нормы 22 нм;

— 1000...2000 Гфлопс/10 000...16 000 Гфлопс в 2017 г. проектные нормы 17 нм (заметим, что в планах ведущих производителей микросхем — 8 нм в 2017 г. [6]).

Потребляемая мощность процессора постоянна: 300—500 Вт.

Можно показать, что вычислительная среда пиковой производительностью 1000 Пфлопс, из которых 100 и 900 Пфлопс составляют производительность MIMD-компонента и SIMD-компонента соответственно, при указанных условиях будет потреблять около 19 МВт, из них 10 МВт приходится на MIMD-компонент и 9 МВт на SIMD-компонент.

В составе такой вычислительной среды понадобится задействовать $50 \cdot 10^3$ — $90 \cdot 10^3$ MIMD/SIMD-процессоров пиковой производительностью $(1000\text{—}2000)/(10\,000\text{—}16\,000)$ Гфлопс каждый.

Полагаем, что MIMD/SIMD-процессор содержит $(100\text{—}200)$ MIMD-ядер и $(1000\text{—}2000)$ SIMD-ядер. Общее количество MIMD/SIMD-ядер в системе составит $\sim 10^7/10^8$ шт.

2.2. Коммуникационная среда

Оценим параметры коммуникационной среды, требуемые для объединения указанного количества процессоров в единую систему определенной выше производительности.

2.2.1. Уровни параллелизма и структура соединений

Будем различать четыре уровня параллелизма: процессор, вычислительный блок, стойка и система. Их иерархия показана на рис. 2 (см. третью полосу обложки).

Полагаем, что в процессоре связь между MIMD- и SIMD-компонентами и образующими их ядрами осуществляется внутрипроцессорными средствами.

Структура гибридного процессорного элемента показана на рис. 3. Он содержит несколько MIMD/SIMD-процессоров и коммутатор K , через который осуществляется его взаимодействие с другими элементами.

В процессорном элементе задействованы каналы I—III уровней, реализующие соответственно связи между процессорными элементами в вычислительном блоке, в стойке и в системе.

Укажем на идентичность рассматриваемой структуры связей и структуры, примененной, например, в K компьютере [9] и в Cray XC [10].

2.2.2. Оценки параметров коммуникационной среды

Функционирование современных процессоров требует ~ 1500 внешних выводов на его корпусе. Полагаем, что это количество, определяемое механическими параметрами, не изменится. Чтобы уменьшить количество связей, реализуемых проводными соединениями, MIMD/SIMD-процессоры, задействованные в процессорном элементе, объединяют на общей «подложке» или в виде трехмерной сборки. Это позволяет микроэлектронными технологиями реализовать связи между процессорами, а также внешний интерфейс, через который осуществляется связь с системой межпроцессорного обмена. Примером внешнего интерфейса служит совокупность одновременно задействованных разъемов интерфейса PCI Express, Hypertransport или QPI. Возможны другие конструктивные элементы.

Для определенности в расчетах будем использовать характеристики процессорного элемента, производительность MIMD/SIMD-компонентов которого составляет 8 Тфлопс/64 Тфлопс; такая производительность в 2017 г. может быть достигнута объединением на одной подложке восьми MIMD/SIMD-процессоров производительностью 1 Тфлопс/8 Тфлопс.

В качестве каналов связи будем использовать каналы IB 12xHDR ($480 \cdot 480$) Гбит/с, параметры которых указаны в сообщении [11]; пропускная спо-

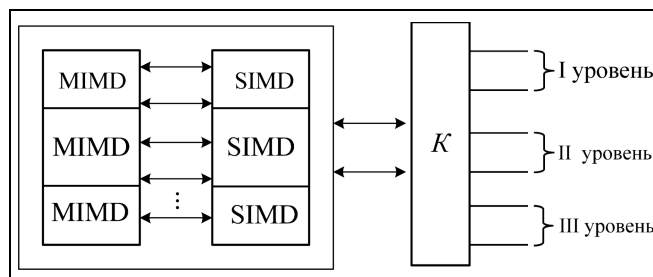


Рис. 3. Процессорный элемент

способность одного линка составляет $(40 + 40)$ Гбит/с = $(5 + 5)$ Гбайт/с.

Каждый канал содержит 12 линков, его пропускная способность составляет $v_k = (0,005 \cdot 12) = 0,06$ Тбайт/с.

Для достижения производительности 1 Тфлопс потребуется 2^{14} процессорных элементов.

Полагаем, что вся система содержит 128 стоек, в каждой стойке 8 блоков, в каждом блоке 16 процессорных элементов.

Каналы первого уровня применяются для объединения 16-ти процессорных элементов в вычислительный блок, достаточно длины $l_1 = 20$ см. Каналы второго уровня — для объединения вычислительных блоков в стойке, достаточно длины $l = 2$ м. Каналы третьего уровня объединяют стойки, достаточно длины $l = 20$ м.

Реализация каналов первого и второго уровней возможна путем применения многослойных печатных плат. Реализация каналов третьего уровня, по-видимому, невозможна без применения много-модовых оптических средств связи.

Оценим три варианта топологии среды: 3D тор, гиперкуб (H) и dragonfly (DF).

При расчете значений параметров среды полагаем, что выполняются условия:

— размерности среды 3D тор: $32 \times 32 \times 16$; причем, по координате z объединены 32 элемента (два блока) в топологию 1D тор; четыре 1D тора (восемь блоков) в стойке объединяются по координате x , восемь стоек образуют ряд по координате x ; 16 рядов по координате y , по 8 стоек в каждом, образуют систему; каждый процессорный элемент содержит 6 каналов;

— размерности среды H: $n_1 = 4$ (2^4 элементов в блоке), $n_2 = 3$ (2^3 блоков в стойке), $n_3 = 7$ (2^7 стоек в системе); каждый процессорный элемент содержит 14 каналов;

— размерности среды DF: 16 элементов (blades) объединены полносвязным графом в блок (chassis), в каждом элементе 15 каналов первого уровня; 8 блоков объединяются полносвязным графом в стойку (group), в каждом элементе 7 каналов второго уровня; стойки объединены полносвязным

графом в систему, каждая пара стоек соединена восемью каналами; у каждого элемента 8 каналов третьего уровня; в скобках указаны термины, используемые для этой среды в работе [10]; размерности рассматриваемой здесь среды и среды, описанной в той же работе [10], также близки; каждый процессорный элемент содержит 30 каналов.

Для каждой из рассматриваемых топологий — 3D тор, H и DF, в табл. 2 указаны значения C_i — количество связей среды i -уровня и L_i — суммарная длина этих связей. Символом D обозначено значение диаметра — наибольшего расстояния между процессорными элементами; γ — отношение суммарной пропускной способности каналов связи процессорного элемента к производительности его SIMD-компонента.

Заметим, что значения γ в табл. 2 существенно меньше $\gamma = 0,12$ для системы Cray XE6 [12] и $\gamma = 0,2$ для объявленного IBM проекта системы «Blue Waters» [13].

Приведенные в табл. 2 данные демонстрируют реальность создания системы эксафлопной производительности. Они иллюстрируют достоинства и недостатки рассмотренных топологий, влияние которых понадобится оценивать на этапе создания систем, исходя из достигнутого технологического уровня.

3. АРХИТЕКТУРНЫЕ СРЕДСТВА МАСШТАБИРОВАНИЯ ЭФФЕКТИВНОСТИ

Рассмотренные варианты вычислительной системы характеризуются:

— гибридной (неоднородной) структурой процессорных элементов;

— сложностью коммуникационной среды и, как следствие, сравнительно малым значением отношения пропускной способности каналов связи процессорного элемента к его производительности.

В этих условиях необходимы инструментальные средства, которые позволяют учитывать архитектурные особенности вычислительной системы и обеспечивают масштабирование эффективности на различных уровнях параллелизма.

Таблица 2

Значения параметров коммуникационных сред

Топология	Уровень I	Уровень II	Уровень III	D	γ
	C_1 , шт./ L_1 , км	C_2 , шт./ L_2 , км	C_3 , шт./ L_3 , км		
3D	$0,18 \cdot 10^6/36$	$0,16 \cdot 10^6/319$	$0,2 \cdot 10^6/3,9 \cdot 10^3$	40	0,005
H	$0,39 \cdot 10^6/78$	$0,29 \cdot 10^6/600$	$0,69 \cdot 10^6/14 \cdot 10^3$	14	0,013
DF	$1,5 \cdot 10^6/294$	$0,69 \cdot 10^6/1376$	$0,78 \cdot 10^6/15 \cdot 10^3$	5	0,028



Рассмотрим средства архитектурного масштабирования эффективности:

- реконфигурацию структуры гибридных процессорных элементов, посредством вариации количества MIMD-ядер и задействованных с ними SIMD-ядер, в соответствии с особенностями выполняемого вычислительного процесса для достижения наибольшей в заданных условиях производительности и эффективности;

- применение бесконфликтных множеств источников и приемников и/или декомпозиции вычислительного процесса на подпроцессы и размещение их по элементам среды в соответствии с особенностями элементов и топологией связей среды в целях минимизации длительностей обменов информацией между процессорными элементами;

- топологическое резервирование процессорных элементов, позволяющее при отказах элементов сохранять неизменными топологию среды и ее производительность на данном процессе, тем самым сохранять значение эффективности, достигнутое перечисленными средствами реконфигурации структуры и минимизации длительностей обменов.

3.1. Гибридные реконфигурируемые структуры

Значения ускорения вычислений гибридными системами и их эффективность зависят от особенностей решаемой задачи и параметров вычислительной среды.

К особенностям задачи, точнее, алгоритма ее решения, относятся длительности нераспараллеливаемых фрагментов, количество и тип операций обмена информацией, синхронность вычислительных процессов и т. п.

Для гибридных архитектур (в отличие от однородных) характерно, что вычислительный процесс распределяется в начале между MIMD- и SIMD-компонентами и лишь затем между процессорами, образующими эти компоненты.

Результирующее ускорение зависит от ускорений достигаемых на MIMD- и SIMD-компонентах и от размера «долей» вычислительного процесса, приходящихся на эти компоненты.

Варьируя производительностями MIMD- и SIMD-компонентов — в частности, количеством задействованных в них ядер, можно изменять длительности выполнения вычислительного процесса.

В работе [14] получены оценки длительности вычислений гибридными системами (процессорными элементами) в зависимости от соотношений между фрагментами вычислительного процесса и производительностью MIMD- и SIMD-компонентов, выполняющих эти фрагменты. Предложены критерии динамической реконфигурации структуры процессора, предусматривающие разделение

ядер MIMD- и SIMD-компонентов на определенные взаимодействующие подмножества, состав и производительность которых определяются в соответствии с параметрами исполняемого процесса.

Варьирование составом и производительностью MIMD- и SIMD-компонентов позволяет, исходя из определенных первичных свойств процесса, получить максимальное для заданных условий ускорение вычислений.

В частности, коэффициент ускорения вычислений гибридной системой, содержащей q ядер и один ускоритель, по сравнению с системой, содержащей одно ядро универсального процессора, имеет вид:

$$K_{q,1} = \frac{q}{\varphi + (1 - \varphi)q/\rho},$$

если задействованы одно ядро и q ускорителей, то

$$K_{1,q} = \frac{q}{\varphi q + (1 - \varphi)/\rho}, \text{ где } 0 \leq \varphi \leq 1 \text{ — доля вычис-}$$

лительного процесса, выполняемого универсальным процессором (доля MIMD-фрагмента), $\rho > 1$ — коэффициент ускорения по сравнению с универсальным ядром процессора, достигаемый применением ускорителя на SIMD-фрагменте.

В качестве иллюстрации приведем, согласно работе [15], пример вычислений значений потенциала Морзе по программе молекулярной динамики гибридной системой, содержащей четырехядерный процессор Intel Core i7920 и арифметические ускорители Nvidia Tesla C2050.

Длительность вычислений одним ядром значений потенциала Морзе для задачи размером $55 \times 55 \times 55$ периодов кристаллической решетки составила $T_1 = 22,96$ с. Этот вычислительный процесс можно разделить на два фрагмента: MIMD-фрагмент, выполняемый ядром в течение $T_M = 7,07$ с (следовательно, $\varphi = 7,7/22,96 \approx 0,31$) и SIMD-фрагмента, выполняемый ускорителем в течение $T_s = 2,8$ с (имеем $\rho = (1 - \varphi) T_1/T_s = 5,67$). Длительность выполнения этого процесса в режиме умножения (weak scaling) гибридной системой, содержащей одно ядро универсального процессора и четыре ускорителя, составляет $T_{1,4} = 30,3$ с, а длительность выполнения системой, содержащей четыре ядра и один ускоритель $T_{4,1} = 18,3$ с, т. е. система из четырех ядер и одного ускорителя на этом процессе в 1,65 раз быстрее системы из одного ядра и четырех ускорителей.

На рис. 4 (см. третью полосу обложки) для рассматриваемого вычислительного процесса по программе молекулярной динамики приведены графики: $K_{q,1}$ — значения ускорения вычислений по сравнению с одним ядром, достигаемые гибридной системой, содержащей q ядер и один ускоритель.

тель; $K_{1,q}$ — значения ускорения вычислений по сравнению с одним ядром, достигаемые гибридной системой, содержащей одно ядро и q ускорителей.

Из значений $K_{q,1}$ и $K_{1,q}$ следует, что этот вычислительный процесс быстрее выполняется системой с большим количеством ядер.

Другие подробности применения гибридных систем изложены в работах [14, 16]. Изложенный метод может быть применен на первом уровне параллелизма.

3.2. Минимизация длительностей обменов

Минимизация длительностей обменов достигается взаимной адаптацией вычислительного процесса и структуры связей между процессорными элементами с целью исключения конфликтов при выполнении обменов и уменьшения расстояний обменов. Возможности адаптации зависят как от топологии вычислительной среды, так и от свойств вычислительного процесса (явные схемы, регулярные связи и т. д.) С увеличением сложности вычислительной системы актуальность (и результативность) этих средств возрастает.

3.2.1. Бесконфликтные множества

Различные топологии мультипроцессорных сред накладывают различные принципиальные ограничения на количество свободных непересекающихся маршрутов, исключающих возникновение конфликтов в процессе передачи информации.

Пусть S и R соответственно множества источников и приемников такие, что для любой пары $a \in S$ и $b \in R$ существует свободный маршрут длины, не превышающей $l_{\max}$ при условии, что все остальные источники и приемники из множеств S и R также выполняют парные обмены ($l_{\max}$ — диаметр среды, расстояние достаточное для соединения любой пары «источник — приемник» из данной среды). Другими словами, S и R — такие множества источников и приемников, находящиеся на максимальном для данной среды расстоянии, при задействовании которых для любой пары «источник — приемник» существует и заранее известен свободный маршрут.

Таблица 3

Мощности и численность бесконфликтных множеств

Среда	3D тор	H^m	DF
C	$\omega^{2/3}$	$\omega/2 = 2^{m-1}$	$\omega/2$
G	$\omega^{2/3} 2^{\omega^{2/3}}$	$\left(\log_2 \omega \right) \log_2 \omega \cdot 2^{\frac{\omega}{\log_2 \omega} - 1}$	$2^{\omega-1}$
$l_{\max}$	$(3/2)\omega^{1/3}$	$m = 2^k - 1, k = 0, 1, 2$	2

Пусть $C = \|S\| = \|R\|$ — мощности этих множеств, G — количество таких множеств для сред с различной топологией. Количество процессорных элементов в среде обозначим ω . Оценки C и G для сред с различными топологиями приведены в работе [17].

Значения C , G и $l_{\max}$ для сред с топологиями 3D тор, H^m и DF указаны в табл. 3 (для среды DF оценки приведены лишь для двух первых уровней (chasis и group) [10]). Из них следует, что сравнительно хорошими коммуникационными возможностями характеризуются среды H^m и DF.

По определению бесконфликтных множеств, выполнение обменов между принадлежащими им источниками и приемниками свободно от конфликтов; потребуются затраты времени лишь на построение маршрута длины $l_{\max}$ и передачи информации по этому маршруту. Это позволяет в определенных ранее условиях сохранять практически неизменной эффективность среды при наращивании ее сложности.

Недостаток рассматриваемого метода заключается в сравнительно «малой» численности бесконфликтных множеств, по сравнению с общей численностью $2^{\omega-1}$ множеств источников и приемников, содержащих по $\omega/2$ элементов; исключение составляют лишь среда N — полный матричный коммутатор и первые два уровня среды DF; однако их аппаратная реализация очень сложна.

3.2.2. Декомпозиция и размещение процессов

Другим средством масштабирования эффективности минимизацией длительностей обменов служит размещение источников и приемников на минимальном расстоянии с целью достижения наименьшего значения длительности обменов.

Рассмотрим возможности, предоставляемые различными топологиями для реализации одного класса вычислительных процессов, относящихся к наиболее применимым (рис. 5, a — см. третью полосу обложки). Потребуем, чтобы «массовые» обмены выполнялись только между подпроцессами, расположенными на соседних процессорных элементах — непосредственно соединенных каналом. Тем самым попытаемся создать условия, при которых с увеличением числа процессорных элементов для обменов задействуются лишь соседние коммутаторы.

Средства минимизации длительностей обменов включают в себя:

— декомпозицию вычислительного процесса в соответствии с особенностями процессорных элементов;

— размещение полученных подпроцессов по элементам в соответствии с направлениями обменов между ними.

Возможности декомпозиции и размещения вычислительных процессов, предоставляемые топологией 3D тор, изложены, например, в работах [18, 19].

Покажем возможность применения других топологий, в частности, H^n , из которой можно получить 1D, 2D и 3D торы разных размерностей.

В гиперкуб размерности n , обозначаемый как H^n , помещаются (вкладываются) с сохранением физического соседства: 1D тор из 2^n процессов; 2D тор из $2^{n_1} \times 2^{n_2}$ процессов, где $n_1 + n_2 = n$; 3D тор из $2^{n_1} \times 2^{n_2} \times 2^{n_3}$ процессов, где $n_1 + n_2 + n_3 = n$.

В 3D тор можно помещать 3D, 2D и 1D торы меньших размерностей; в 2D тор можно помещать 2D и 1D торы меньших размерностей.

Потребуем, чтобы и для трехмерного, и для двумерного процесса обмена с соседями по каждому измерению обеспечивались одинаковыми связями. Тогда, в качестве процессорного элемента целесообразно использовать элемент, содержащий 2^m процессоров, где m — число, кратное 3 и 2.

Процессорный элемент, содержащий $2^6 = 64$ процессора, позволяет размещать на нем «квадраты» размерностью $2^3 \times 2^3$ и «кубики» $2^2 \times 2^2 \times 2^2$.

Для вычисления требуемого отображения процессов и исполняющих их элементов может применяться прикладная программа, результатом выполнения которой служит таблица соответствия. Эта таблица передается системным средствам, которые загружают процессы на соответствующие процессоры.

Описанные средства легко распространяются на топологию DF и позволяют обеспечить физическое соседство компонентов, образующих вычислительный процесс. Эти средства применимы в условиях современных аппаратных платформ — вычислительных модулей из нескольких, в частности, гибридных многоядерных процессоров на общей памяти.

В табл. 4 приведены значения производительности [20], достигнутые на тесте NPB LU [21]. В столбцах 1 — значения производительности, оп/с, для варианта размещения процессов в соответствии со структурой связей вычислительной

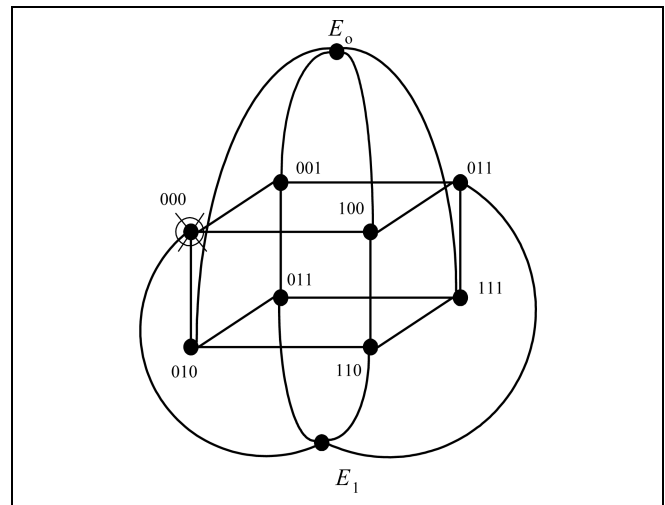


Рис. 6. Топологическое резервирование H^3 избыточными элементами

системы, приведенной на рис. 5, б (для вычисления соответствия использовался код Грэя), в столбцах 2 — значения производительности для последовательного размещения процессов, обычно реализуемого системным планировщиком. В частности, на тесте NPB LU система (класс C) из 512 процессорных ядер при оптимальном размещении процессов показала производительность в 1,72 раза большую по сравнению с достигаемой при «обычном» последовательном размещении.

Представленные в табл. 4 данные показывают, что эффект от применения декомпозиции и размещения подпроцессов возрастает с увеличением числа процессорных элементов (ПЭ) (ядер), задействованных в процессе вычисления. Этот эффект иллюстрируется рис. 6.

3.3. Средства отказоустойчивого масштабирования эффективности

Сбои и отказы отдельных элементов обусловлены как аппаратными, так и программными эффектами, характер и источник которых «некогда» выяснять в процессе счета, их надо исключать и изолировать.

Таблица 4

Значения производительности на тесте NPB LU

Число процессорных ядер, шт.	Класс В		Класс С		Класс D	
	1	2	1	2	1	2
128	88 950	73 478	97 072	95 398	83 421	83 008
256	164 537	108 826	177 953	152 613	223 547	221 760
512	—	—	283 926	164 283	409 697	406 182

Архитектурные средства обеспечения надежности (дополняющие технологические и схемотехнические достижения) должны не только устранять источники сбоев и отказов, но и сохранять эффекты масштабирования эффективности, достигаемые в результате применения средств, описанных ранее.

Масштабирование эффективности может быть достигнуто применением методов топологического резервирования [22, 23], позволяющих обеспечить в случае отказов и сбоев элементов неизменность топологии среды и ее производительности.

Могут быть применены два метода топологического резервирования. Их отличительными особенностями:

- сохранение топологии вычислительной среды, выполняющей вычислительный процесс (деградации в случае отказа не происходит);
- идентичность резервных и резервируемых элементов.

Первый метод [22] основан на введении избыточных процессорных элементов. Он иллюстрируется на рис. 6, где E_0 и E_1 — резервные элементы. Если, например, откажет элемент (000), он и его каналы связи заменяются элементом E_0 и его каналами.

Пусть $p = p(t)$ — вероятность безотказной работы процессорного элемента на интервале t . $P = p^\omega$ — вероятность безотказной работы среды из ω элементов на интервале t . Можно показать, что для d -кратного резервирования, когда на $\log \omega = n$ элементов вводится d резервных, вероятность безотказной работы среды на интервале t составит:

$$P = p^\omega + \binom{\omega}{1} p^{\omega-1} (1-p)^1 + \dots + \binom{\omega}{d} p^{\omega-d} (1-p)^d.$$

Второй метод [23] основан на избирательном резервировании части вычислительной среды. Процессорные элементы, занятые выполнением одного вычислительного процесса, резервируются в случае необходимости другими процессорными элементами этой же среды. Эти элементы изначально могут использоваться для выполнения других, менее «важных» процессов, которые при необходимости резервирования удаляются. Множество процессорных элементов, предоставляемое резервируемому процессу, имеет ту же топологию и мощность, что и исходное.

Второй метод иллюстрируется на рис. 7, где процесс, исполняемый элементами (0000, 0001, 0010, 0011), может быть перенесен на другую плоскость. Например, в случае отказа элемента (0000) про-

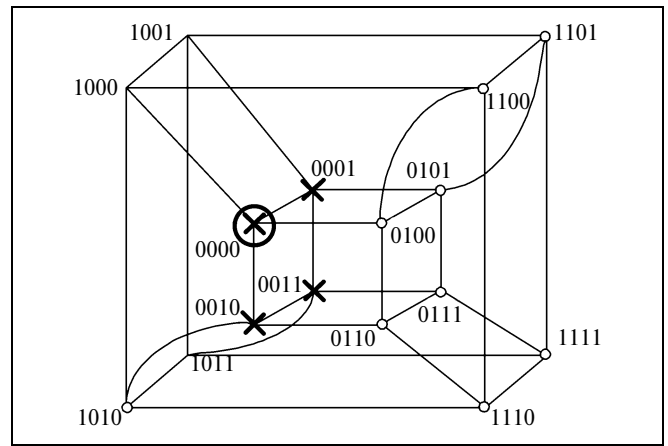


Рис. 7. Топологическое избирательное резервирование H^4 выделенными элементами

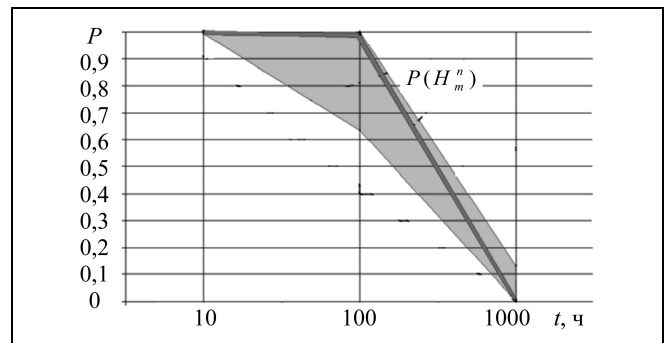


Рис. 8. Топологическое избирательное резервирование: вероятность выполнения процесса

цесс можно перенести на элементы (1000, 1001, 1010, 1011).

В общем виде вероятность $P(H_m^n)$ выполнения средой с топологией H^n вычислительного процесса, занимающего в ней подмножество процессорных элементов H^m , где $m < n$, запишется как

$$\sum_{i=0}^d \binom{\omega}{i} p^{\omega-i} (1-p)^i \leq P(H_m^n) \leq \sum_{i=0}^{\mu} \binom{\omega}{i} p^{\omega-i} (1-p)^i,$$

где $d = 2^{n-m}$, $\mu = \binom{n}{0} + \binom{n}{1} + \dots + \binom{n}{m}$.

На рис. 8 показаны значения вероятностей безотказного выполнения средой из 128 элементов процесса длительностью t , требующего половину элементов среды.



Процессорный элемент имеет длительность наработки на отказ (MTBF) 10^4 ч. Вероятность безотказного выполнения процесса длительностью $t = 100$ ч без применения топологического избирательного резервирования составляет $0,28 \leq P \leq 0,53$, с применением резервирования $0,64 \leq P \leq 0,99$, на более вероятное значение близко к 0,99 (показано утолщенной линией).

Оба метода могут быть применены для сред с различными топологиями. Разумеется, различные топологии обеспечивают различные оценки вероятностей выполнения вычислительного процесса.

Эффективность вычислительной системы зависит также от длительности записи и чтения контрольных точек [24, 25].

Избыточные элементы могут использоваться для хранения контрольных точек, для этого они располагаются аналогично резервным на минимальном расстоянии от элементов, непосредственно выполняющих вычисления.

Неизменность топологии среды и ее производительность исключают необходимость каких бы то ни было изменений исполняемых программ и процессов в случае отказов.

Реализация средств топологического резервирования на различных уровнях параллелизма применительно к процессору (резервирование ядер MIMD- и SIMD-компонентов), вычислительному блоку (резервирование процессорных элементов), стойке (резервирование блоков) и т. п. позволяет создавать среды с наперед заданными значениями вероятностей исполнения вычислительного процесса определенной длительности, занимающего заданное число элементов.

4. УРОВНИ ПАРАЛЛЕЛИЗМА И КОНТУРЫ АДАПТИРУЕМОЙ АРХИТЕКТУРЫ

Из изложенного следует, что перспективный вариант достижения эксафлопной производительности заключается в применении гибридных архитектур, реализующих различные дисциплины вычислений и допускающих реконфигурацию компонентов в соответствии с особенностями исполняемого процесса.

Эффективное применение гибридных архитектур предусматривает создание системного и прикладного программного обеспечения, позволяющего как можно «сильнее» задействовать возможности аппаратных средств, в частности, возможности их адаптации к особенностям исполняемой программы.

Реализуемая реконфигурация — средство создания архитектуры, динамически адаптируемой к особенностям исполняемого процесса на первом

уровне параллелизма — на уровне MIMD/SIMD-компонентов. Эти компоненты более общие, по сравнению с узкофункциональными арифметическими и логическими устройствами, обычно рассматриваемыми при построении реконфигурируемых систем. Их задействование в соответствии с особенностями исполняемого вычислительного процесса позволяет ускорить вычисления.

Задействование бесконфликтных множеств процессорных элементов, маршрутизация в соответствии с топологией среды и размещение процессов по процессорным элементам, минимизирующее длительности обменов, обеспечивают эффективное задействование коммуникационной среды на втором и третьем уровнях параллелизма в соответствии с особенностями исполняемой программы.

Топологическое резервирование на различных уровнях позволяет реализовывать отказоустойчивое масштабирование эффективности среды.

В свою очередь, в создаваемых программах должны быть учтены особенности и параметры вычислительной системы — наличие и состав MIMD- и SIMD-компонентов, структура и характер связей между элементами, модулями и стойками.

Представляется сомнительным, что без реализации перечисленных средств гибридную систему эксафлопной производительности удастся эффективно использовать на содержательных задачах.

Изложенное означает принципиально новый уровень взаимозависимости аппаратных и программных средств (именуемый в литературе «co-design»), реализуемый через специальный инструментарий, позволяющий максимально задействовать возможности аппаратуры и использовать алгоритмические особенности прикладных программ.

ЗАКЛЮЧЕНИЕ

Исследованы архитектурные особенности вычислительных систем, необходимые для достижения эксафлопной производительности.

Оценены параметры процессорной среды и коммуникационной среды.

Показана целесообразность применения архитектурных средств масштабирования эффективности, включающих в себя:

- реконфигурацию структуры гибридных процессорных элементов посредством вариации количества MIMD-ядер и задействованных с ними SIMD-ядер в соответствии с особенностями выполняемого вычислительного процесса в целях достижения наибольшей в заданных условиях производительности и эффективности;

- применение бесконфликтных множеств источников и приемников и/или декомпозиции вычислительного процесса на подпроцессы и раз-

мещение их по элементам среды в соответствии с особенностями элементов и топологией связей среды; этим достигается минимизация длительностей обменов информацией между процессорными элементами;

— топологическое резервирование элементов среды, позволяющее при отказах элементов сохранять неизменными топологию среды и ее производительность на данном процессе. Тем самым сохраняется значение эффективности, достигнутое перечисленными средствами реконфигурации структуры и средствами минимизации длительностей обменов, т. е. обеспечивается отказоустойчивое масштабирование эффективности.

В результате достигается динамическая адаптируемость архитектуры к особенностям исполняемой программы (при условии, что в самой программе учтены возможности архитектуры), что в свою очередь должно обеспечить эффективность применения экскафлопных супер-ЭВМ.

ЛИТЕРАТУРА

1. Цилькер Б.Я., Орлов С.А. Организация ЭВМ и систем. — СПб.: Питер, 2004.
2. Концепция по развитию технологии высокопроизводительных вычислений на базе супер-ЭВМ экскафлопного класса на 2012—2020 гг. — URL: <http://www.rosatom.ru/wps/wcm/connect/rosatom/rosatomsite/aboutcorporation/nauka/> (дата обращения 13.11.2013).
3. SC11 Keynote by Nvidia CEO Jen-Hsun Huang. — URL: <http://blogs.nvidia.com/2011/11/exascale-an-innovator%E2%80%99s-dilemma/> (дата обращения 13.11.2013).
4. Stevens R. and White A. A DOE Laboratory plan for providing exascale applications and technologies for critical DOE mission needs. — URL: http://computing.ornl.gov/workshops/scidac2010/presentations/r_stevens.pdf (дата обращения 16.12.2013).
5. International Exascale Software Project. — URL: www.exascale.org (дата обращения 13.11.2013).
6. URL: http://www.ecmwf.int/newsevents/meetings/workshops/2010/high_performance_computing_14th/presentations/barkai.pdf (дата обращения 13.11.2013).
7. SC'09 Exascale Panel. Steve Scott. Cray Chief Technology Officer. Exhibitor Forum, SC'09. — URL: http://www.exascale.org/mediawiki/images/c/c1/SC09_Exascale_panel_Scott.pdf (дата обращения 13.11.2013).
8. The Future of GPU Computing. — URL: http://www.nvidia.com/content/GTC/documents/SC09_Dally.pdf (дата обращения 13.11.2013).
9. Tomohiro Inoue. Fujitsu Limited. The 6D Mesh/Torus Interconnect of K Computer. — URL: <http://www.fujitsu.com/downloads/TC/sc10/interconnect-of-k-computer.pdf> (дата обращения 13.11.2013).
10. Alverson B., Froese E., Kaplan L. and Roweth D. Cray Inc. Cray XC Series Network. — URL: <http://www.cray.com/Assets/PDF/products/xc/CrayXC30Networking.pdf> (дата обращения 13.11.2013).
11. Infiniband Roadmap. — URL: http://www.infinibandta.org/content/pages.php?pg=technology_overview (дата обращения 13.11.2013).
12. IBM Blue Waters. — URL: <http://www.ncsa.illinois.edu/BlueWaters> (дата обращения 13.11.2013).
13. Cray Titan. — URL: <http://www.knoxnews.com/news/2011/mar/07/oak-ridge-lab-to-add-titanic-supercomputer/> (дата обращения 13.11.2013).
14. Степаненко С.А. Оценки ускорения вычислений гибридными системами // Пятая междунар. конф. «Параллельные вычисления и задачи управления» РАСО 2010»: плен. докл., 26—28 октября 2010 г., Москва / ИПУ РАН. — М., 2010. — С. 61—71.
15. Крючков И.А., Степаненко С.А., Рыбкин А.С. Реализация статической маршрутизации и оптимального размещения вычислительных процессов в мультипроцессорных средах // Молодежь в науке: сб. докл. Шестой науч.-техн. конф. — Саров, 2008. — С. 172—176.
16. Степаненко С.А. Топологическое резервирование мультипроцессорных сред выделенными элементами // Тр. РФЯЦ—ВНИИЭФ. — 2005. — № 10. — С. 50—60.
17. Применение арифметических ускорителей для расчета задач молекулярной динамики по программному комплексу МД / Б.Л. Воронин, А.М. Ерофеев, С.В. Копкин и др. // Вопросы атомной науки и техники. Сер. Математическое моделирование физических процессов. — 2009. — Вып. 2.
18. Степаненко С.А. Топологическое резервирование мультипроцессорных сред // Вопросы атомной науки и техники. Сер. Математическое моделирование физических процессов. — 2002. — Вып. 4. — С. 55—60.
19. NASA, NAS Parallel Benchmarks. — URL: <http://www.nas.nasa.gov/Resources/Software/npb.html> (дата обращения 13.11.2013).
20. Barrett B., Barrett R., Brandt J., et al. Report of Experiments and Evidence for ASC L2 Milestone 4467 — Demonstration of a Legacy Application's Path to Exascale; Sandia Report, SAND2012-1750, Printed March 2012.
21. Daly J.T. A higher order estimate of the optimum checkpoint interval for restart dumps. Los Alamos National Laboratory. M/S, Los Alamos, NM 87545, USA. 28 December 2004. — URL: <http://www.sciencedirect.com> (дата обращения 13.11.2013).
22. Hao Yu, I.-Hsin Chung, Jose Moreira. Topology Mapping for Blue Gene/L Supercomputer. SC2006 November 2006. — URL: <http://www.ibm.com> (дата обращения 13.11.2013).
23. Network Resiliency for Cray XETM Systems. — URL: <http://fs.hlr.de/projects/craydoc/docs/books/S-0032-3101/html-S-0032-3101/index.html> (дата обращения 13.11.2013).
24. Степаненко С.А. Коммуникационные параметры мультипроцессорных сред // Сб. докл. IX Междунар. семинара по супервычислениям и математическому моделированию. Саров, 3—7 октября 2006 г. — Саров, 2006. — С. 96.
25. Stepanenko S.A. Estimated speedups of hybrid reconfigurable systems // XIV International conference «Supercomputing and Mathematical Modeling» / RFNC — VNIIEF, Sarov, October 1—5, 2012. — Sarov, 2012.

Статья представлена к публикации членом редколлегии В.Г. Лебедевым.

Степаненко Сергей Александрович — д-р физ.-мат. наук, нач. отдела, ☎ (83130) 4-53-54, ✉ ssa@vniief.ru,

Южаков Василий Васильевич — зам. нач. отдела, ☎ (83130) 2-79-66, ✉ v.v.yuzhakov@vniief.ru,

Российский федеральный ядерный центр — Всероссийский научно-исследовательский институт экспериментальной физики, г. Саров.

ДИСКРЕТНО-СОБЫТИЙНЫЕ МОДЕЛИ ЖЕЛЕЗНОДОРОЖНОЙ СЕТИ¹

А.И. Потехин, С.А. Браништов, С.К. Кузнецов

Разработаны дискретно-событийные модели основных элементов участка железной дороги: перегона с разъездом, сегмента (как части перегона), секции сегмента (блок-участка), стрелки, а также модели движения поездов. Модели представлены в виде сетей Петри с ограничивающими дугами. Групповое управление моделями при параллельно-конвейерном движении поездов осуществляется специально разработанными управляющими компонентами (супервизорами), обеспечивающими требования безопасности движения.

Ключевые слова: дискретно-событийная модель, сети Петри, супервизор, железнодорожная сеть, участок, перегон, блок-участок.

ВВЕДЕНИЕ

Системы управления железнодорожным транспортом относятся к классу структурно сложных систем, которые при математическом описании не сводятся к последовательно-параллельным или древовидным структурам, а описываются структурами сетевого типа с циклами.

Современные мировые научные разработки в области железнодорожного транспорта направлены, в первую очередь, на решение вопросов безопасности движения. Сейчас безопасность обеспечивается системами сигнализации, обнаружения и блокирования аварийных ситуаций, контроля движения поездов. Ввиду недопустимости ошибок в таких системах для их проектирования и моделирования необходимо применять формальные методы.

Наличие автоматических блокировок в современных железнодорожных (ж/д) сетях позволяет каждый перегон между станциями представить в виде совокупности электрически изолированных блок-участков (секций) с соответствующими датчиками состояния и светофорами. Это обстоятельство позволяет перейти от традиционных непре-

рывных моделей компонентов движения поездов к дискретно-событийным моделям (ДС-моделям).

В случае нахождения на перегоне нескольких поездов возникает проблема обеспечения безопасного их перемещения. Принят критерий безопасного параллельно-конвейерного движения поездов: безопасность движения обеспечивается, если между любыми двумя соседними поездами, находящимися на перегоне, имеется, по крайней мере, один свободный блок-участок, а в ряде случаев два.

Поставить и решить задачу обеспечения безопасности предлагается с помощью теории дискретно-событийных систем (ДСС), которая сейчас активно развивается в ряде научных центров мира [1–3]. Основная задача ДСС: как по ДС-модели объекта и заданным ограничениям на его функционирование построить управляющий компонент (*супервизор*), обеспечивающий поведение объекта в соответствии с заданными ограничениями.

Таким образом, обеспечение безопасного движения на железнодорожном транспорте (обнаружение и блокирование аварийных ситуаций) состоит, прежде всего, в разработке ДС-моделей основных компонентов ж/д сети (перегона с разъездом, сегмента, блок-участка, стрелки, светофора) и, далее, в разработке супервизора, обеспечивающего заданный уровень безопасности при параллельном (одновременном) движении поездов по перегону.

Традиционно компоненты ДСС (ДС-модели объектов, ограничений и супервизоров) обычно

¹ Работа выполнена при частичной финансовой поддержке РФФИ (проект № 13-08-00973-а, проект офи-м-РЖД № 12-08-13179).

представляются конечными автоматами, что не приемлемо для транспортной системы с параллельным движением нескольких подвижных средств.

Для представления ДС-моделей основных компонентов ж/д сети вместо конечных автоматов целесообразно применить сети Петри: они обладают точной математической семантикой, что позволяет достаточно просто интерпретировать поведение ж/д сети. Парадигма параллельности сети Петри хорошо подходит к реальной параллельности ж/д сетей. Графическое представление моделей ж/д компонентов в сети Петри наглядно и удобно для анализа. Движение поездов также удобно моделировать в сети Петри: поезда моделируются метками (фишками), перегоны и блок-участки (секции) представляются позициями, управляемые переходы сети Петри реализуют условия движения поездов. Сети Петри в виде текущей разметки всегда точно определяют состояние ж/д сети, положение поездов, ж/д стрелок, светофоров. Поэтому ДС-модели ж/д компонентов в виде сетей Петри могут использоваться как основа для программной поддержки мониторинга, моделирования и анализа ж/д сети.

Однако моделирование и графическое представление реальной ж/д сети в виде сети Петри получается громоздким и сложным для анализа. Поэтому в работах [4–6] сети Петри применяются для моделирования небольших участков ж/д сети. В настоящей работе предлагается строить несколько иерархически зависимых уровней моделирования.

На верхнем уровне с помощью сети Петри строится модель перегона с разъездом ж/д сети по его структурной схеме. Основное назначение этого уровня — наблюдать за движением поездов по перегону и контролировать соответствие заданным маршрутам, рассчитывать временные параметры движения поездов и др.

На среднем уровне моделирования моделируется часть перегона — сегмент, состоящий из нескольких последовательно соединенных блок-участков (секций). На этом уровне проектируется управляющий компонент ДСС — супервизор в виде сети Петри, обеспечивающий заданную зону безопасности, состоящую из одной, двух или трех свободных секций между двумя соседними поездами, находящимися на перегоне.

На нижнем уровне моделирования определяется связь ДС-моделей блок-участков, ж/д стрелок, светофоров с соответствующими датчиками состояния и управляющими командами оператора, тем самым обеспечивается соответствие между состояниями моделей всех уровней и состояниями физических объектов.

Далее в работе показано применение ДС-моделей для расчета пропускной способности перегона.

Разработанные модели участка железной дороги, а также модели управляющих компонентов (супервизоров) на базе сетей Петри могут быть положены в основу построения системы моделирования параллельно-конвейерного движения поездов в режиме реального времени.

1. КРАТКОЕ ОПИСАНИЕ ОБЪЕКТА МОДЕЛИРОВАНИЯ

В общем виде ж/д сеть представляется совокупностью железнодорожных станций, соединенных между собой перегонами с разъездами (вставками).

На рис. 1 изображена топологическая схема участка, состоящего из двух станций A и B и однопутного перегона с одним разъездом. Через v_1 — v_4 обозначены сегменты перегона. Сегмент, как часть перегона, состоит из нескольких последовательно соединенных электрически изолированных друг от друга секций (блок-участков).

Нахождение поезда на i -й секции контролируется датчиком состояния секции d_i . Предполагаем, что в любой момент времени доступны сигналы всех секционных датчиков. В узлах схемы расположены стрелки c_1 и c_2 . С каждой стороны секции установлены светофоры. Для простоты изложения будем считать, что светофор может находиться в двух состояниях: красный или зеленый. В реальной ситуации светофор может находиться в трех или четырех состояниях.

Одно важное свойство перегона с разъездом, содержащим ж/д стрелки, (свойство корректности): поезд, находящийся, например, на секции сегмента v_2 физически не может переместиться на секцию сегмента v_3 , минуя секцию сегмента v_1 или v_4 . Таким образом, по топологической схеме перегона, содержащего ж/д стрелки, а в общем случае по произвольной схеме железнодорожной сети, невозможно определить корректные маршруты движения, применяя формальные методы теории графов.

Содержательное описание объекта закончим перечислением основных свойств перегона:

— каждый сегмент перегона (благодаря наличию в современных ж/д сетях автоматических

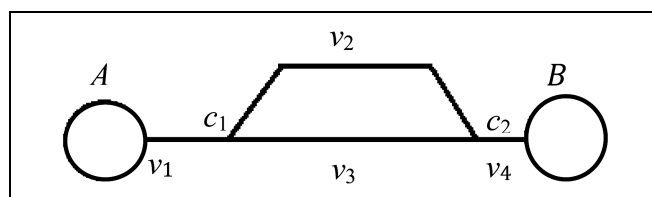


Рис. 1. Топологическая схема перегона с одним разъездом

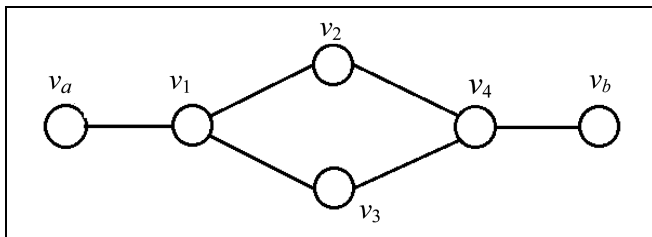


Рис. 2. Неориентированный граф перегона с одним разъездом

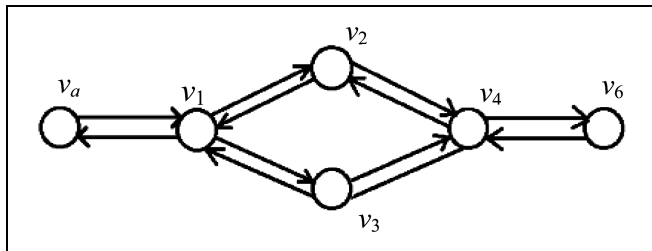


Рис. 3. Ориентированный граф перегона с одним разъездом

блокировок) состоит из нескольких электрически изолированных друг от друга секций (блок-участков);

- длина сегмента определяется числом секций;
- каждая секция длиннее самого длинного поезда и ее длина находится в строго фиксированном диапазоне;

- критерий безопасного параллельно-конвейерного движения поездов: безопасность движения обеспечивается, если между любыми двумя соседними поездами на перегоне имеется, по крайней мере, один свободный блок-участок.

Наличие автоматических блокировок позволяет перейти от непрерывных моделей компонентов движения поездов к ДС-моделям.

С целью построения графа перегона, по которому можно формально определить все возможные пути, маршруты, циклы и т. д., поступаем следующим образом.

Вначале строится неориентированный граф $G = (V, W)$ перегона с разъездом: каждому i -му сегменту (ребру топологической схемы перегона) сопоставляется вершина v_i графа G . Ребро $(v_i, v_j) \in W$, если в схеме возможно движение от i -го сегмента к j -му или наоборот. Станции A и B представляются вершинами v_a и v_b . На рис. 2 изображен граф перегона с одним разъездом, стрелки не показаны, будем считать, что они принадлежат сегментам v_1 и v_4 .

Для организации движения поездов в обе стороны по однопутному перегону заменим каждое

ребро графа G двумя разнонаправленными дугами. Таким образом, граф G преобразуется в ориентированный граф $G' = (V, W')$ (рис. 3).

2. ДИСКРЕТНО-СОБЫТИЙНАЯ МОДЕЛЬ ПЕРЕГОНА

Задачи обеспечения безопасности движения поездов аналогичны задачам, решаемым в теории ДСС. Основная из них состоит в том, как по ДС-модели объекта и заданным ограничениям на его функционирование построить управляющий компонент-супервизор, обеспечивающий поведение объекта в соответствии с заданными ограничениями.

Таким образом, обеспечение безопасного движения на железнодорожном транспорте (блокирование аварийных ситуаций) сводится к дискретно-событийному моделированию, включая парадигму супервизорного управления, основных компонентов ж/д сети: перегона с разъездом, сегмента, блок-участка, стрелки, светофора.

Напомним основные положения ДСС. Под ДСС понимается триплет (G, K, S) , где объект $G = \{G^1, G^2, \dots, G^n\}$ — совокупность автономных компонентов, K — спецификации (ограничения на поведение объекта), S — супервизор (управляющий компонент ДСС), обеспечивающий поведение объекта G в соответствие со спецификацией K .

Функционирование ДСС характеризуется множеством событий E и генерируемыми последовательностями из этих событий. Множество событий E состоит из множества управляющих событий E_1 (генерируются собственным контроллером объекта), множества ожидаемых событий E_2 (моделируют состояния (положения) исполнительных механизмов). Супервизор S блокирует те управляющие события, появление которых противоречит спецификации K .

Поведение ДСС рассматривается с самых общих позиций как поведение некоторого генератора (источника) последовательностей (строк) *событий* из конечного множества событий E . Событие $e \in E$ — это абстракция для множества фактов наблюдения «жизни» ДСС. События мгновенны, их появление спонтанно и происходят они в непредсказуемые моменты времени, поэтому все, что можно наблюдать, это их последовательности, которые и представляются *строками*. Примеры событий: факты изменения положения и состояния отдельных компонентов объекта; команды на объект; характеристика исправного или неисправного состояния объекта и т. п.

В теории ДСС для моделирования компонентов объекта и ограничений K применяются конечные

автоматы (КА). Каждый КА определяется собственным набором: $(Q, E, \delta, F, Q_m, q_0)$, где Q — множество состояний; E — множество событий; δ — функция переходов $Q \times E \rightarrow Q$; $F: Q \rightarrow 2^E$ — функция допустимых событий в каждом состоянии; Q_m — множество достижимых состояний; q_0 — начальное состояние.

Для построения ДС-модели объекта G в ДСС используют одну из двух операций над множеством КА компонентов: декартовое произведение или параллельную композицию, что естественно приводит к огромному числу состояний ДС-модели объекта G .

В работах [7, 8] А.А. Амбарцумян показал, как по ДС-моделям объекта G , представленным в виде сетей Петри, и заданным ограничениям K проектировать супервизор S , обеспечивающий поведение G в соответствии с ограничениями K , не применяя указанных операций.

В нашем случае объект представляет собой перегон с разъездом, он состоит из компонентов: сегментов, стрелок, светофоров, кроме того, сам перегон является элементом ж/д сети.

2.1. Дискретно-событийная модель перегона с разъездом

Построение ДС-моделей начнем с ДС-модели перегона (П) как элемента ж/д сети. Объект $\Pi = \{V, M, E, G\}$, где V — множество вершин графа G (сегментов), в каждой вершине (сегменте) находится m_i поездов, M — множество поездов, находящихся на перегоне, E — множество событий. Наступление события $e_{i,j} \in E$ разрешает движение поезда из сегмента v_i в сегмент v_j ; G — граф соединения сегментов перегона. Ограничения K содержательно определим следующим образом: движение поезда из сегмента v_i в сегмент v_j может быть разрешено только тогда, когда в сегменте v_j обеспечивается заданный интервал безопасности между поездами, иными словами, ограничения K задаются множеством запрещенных состояний каждого сегмента. Задача супервизора S состоит в выполнении ограничений K путем разрешения или блокирования $e_{i,j} \in E$. Нетрудно видеть, что автоматная модель перегона сложна.

Поэтому для представления ДС-моделей основных компонентов ж/д сети вместо конечных автоматов целесообразно применить сети Петри: поезда моделируются метками (фишками), перегон, блок-участки (секции) и т. д. представляются позициями, управляемые переходы сети Петри реализуют движение поездов. Сети Петри в виде текущей разметки всегда точно определяют состо-

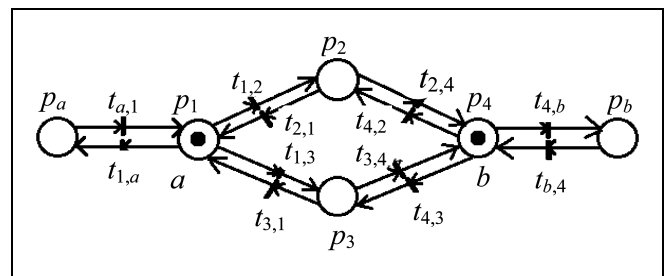


Рис. 4. Сеть Петри, моделирующая перегон с разъездом, $\mu = (0, a, 0, 0, b, 0)$

яние ж/д сети, положение поездов, ж/д стрелок, светофоров.

Дискретно-событийную модель перегона с разъездом представим сетью Петри $(P, T, Pre, Post)$ следующим образом:

— каждой вершине v_i (сегменту) графа G' (см. рис. 3) сопоставляется позиция p_i в графе сети Петри (рис. 4);

— дуга $(v_i, v_j) \in W'$ заменяется структурой $(p_i, t_{i,j}, p_j)$, где $t_{i,j}$ — управляемый переход, $pre(t_{i,j}) = p_i, post(t_{i,j}) = p_j$. Множества позиций и переходов — соответственно P и T (множества конечны, не пусты и в графической форме сети Петри изображаются двумя типами вершин — кружками и полочками). Состояние сети Петри определяется ее маркировкой $\mu = (\mu_1, \mu_2, \dots, \mu_n)$, где $n = |P|$, значение μ_i равно числу фишек в позиции p_i (числу поездов в i -м сегменте).

Движение поездов по перегону имитируется перемещением фишек по позициям при срабатывании управляемых переходов. Запуск перехода $t_{i,j}$ разрешается при наличии фишек в управляющих позициях (разрешающих и блокирующих). Управление переходами на рис. 4 не показано, далее этому будет уделено основное внимание.

Иногда требуется различать фишки, перемещающиеся по позициям сети Петри от A к B и от B к A . Например, фишку соответствующую поезду, следующему от A к B , можно обозначить как a , соответственно фишки, перемещающиеся от B к A , обозначить как b . Таким образом, получаем *раскрашенную* сеть Петри. На рис. 4 изображен граф базовой сети Петри перегона с разъездом с маркировкой $\mu = (0, a, 0, 0, b, 0)$.

Основное назначение этого уровня моделирования — наблюдать за движением поездов по перегону на соответствие заданным маршрутам, рассчитывать временные параметры движения поездов др.

2.2. Дискретно-событийная модель сегмента

Как уже отмечалось, сегмент перегона состоит из последовательно соединенных секций (блок-участков). Сеть Петри сегмента представляется как (P, T, μ) , где P — множество позиций (секций), T — множество переходов (событий, управляющих движением поездов), μ — маркировка сегмента: $\mu = (\mu_1, \mu_2, \dots, \mu_k)$, где $k = |P|$, $\mu_i \in (0, 1)$; $\mu_i = 1$ соответствует наличию фишки в позиции p_i .

Рассмотрим движение поездов по сегменту. Пусть без потери общности сегмент состоит из трех секций и представляется сетью Петри, изображенной на рис. 5, где позиции p_1, p_2, p_3 соответствуют секциям сегмента, $t_1 - t_4$ — управляемые переходы, $\mu = (1, 0, 0)$.

В позиции p_1 находится фишка, что означает нахождение поезда на первой секции. Позиции a_2, a_3, b_1, b_2 будем называть управляющими позициями сети, их состояния (наличие или отсутствие фишек) определяются значениями соответствующих логических переменных, с помощью которых осуществляется управление движением фишек по позициям. Значения управляющих переменных задаются системой организации маршрутов, маневров. Например, если требуется переместить поезд с первой секции на вторую секцию, то управляющей переменной a_2 присваивается единичное значение ($a_2 = 1$), а в позицию a_2 помещается фишка, создавая тем самым условия запуска перехода t_1 , в результате фишка перемещается из позиции p_1 в позицию p_2 . Если затем будет $b_1 = 1$, то фишка переместится из позиции p_2 в позицию p_1 . Комбинируя наборами управляющих переменных, можно осуществить желаемое движение фишек. Естественно, что не все наборы управляющих переменных допустимы, например, должно быть $b_1 a_3 = 0$, т. е. перемещение фишки из каждой позиции должно быть однозначным.

Помимо управляющих позиций, каждый переход ДС-модели сегмента дополнительно управляется (разрешается или блокируется) супервизором, обеспечивающим заданный уровень безопасности движения поездов, проектирование которого показано в п. 2.3. Кроме того, каждый переход ДС-модели сегментов перегона соответствует так называемому в теории сетей Петри непримитивному событию. Исследуем его более подробно.

Рассмотрим связь состояний ДС-моделей секций, сегментов, перегона в целом с состояниями соответствующих физических объектов. Покажем процесс перемещения фишки из одной позиции в другую на примере перехода фишки из позиции p_1

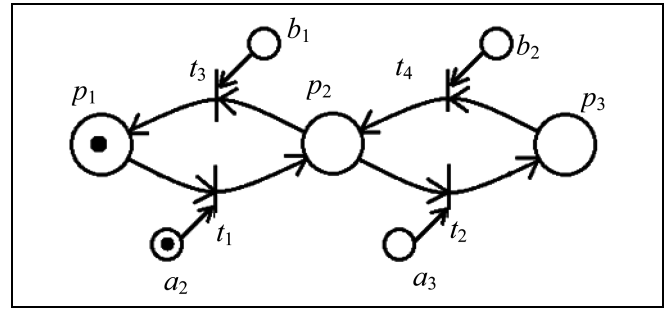


Рис. 5. Сеть Петри сегмента перегона, состоящего из трех секций

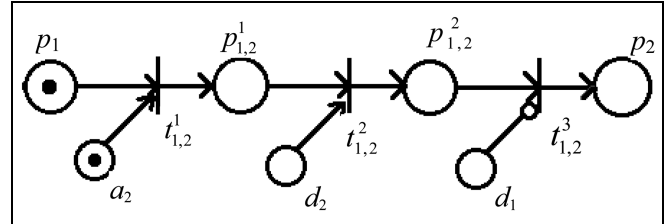


Рис. 6. Сеть Петри, моделирующая движение поездов по сегменту

в позицию p_2 (рис. 6). Срабатывание перехода $t_{1,2}^1$ ($a_2 = 1$) переводит фишку в позицию $p_{1,2}^1$ (начало движения поезда). После того, как голова поезда окажется на второй секции, срабатывает ее датчик состояния $d_2 = 1$, в результате чего в позицию d_2 помещается фишка. Тем самым создаются условия срабатывания перехода $t_{1,2}^2$, в результате фишка перемещается из позиции $p_{1,2}^1$ в позицию $p_{1,2}^2$ (поезд находится одновременно на обеих секциях).

Фишка будет оставаться в позиции $p_{1,2}^2$ до тех пор, пока конец поезда будет находиться на первой секции. После того как $d_1 = 0$ (первая секция свободная) фишка из позиции $p_{1,2}^2$ переместится в позицию p_2 (дуга $(d_1, t_{1,2}^3)$ является блокирующей — вместо стрелки эта дуга оканчивается кружком).

Время нахождения фишек в позициях $p_{1,2}^1$ и $p_{1,2}^2$ определяется сигналами от датчиков d_1 и d_2 , оно зависит от скорости движения поезда, длины поезда и секций.

Рассмотренный процесс движения поезда проиллюстрирован в таблице.

Светофор на входе второй секции со стороны первой секции обозначим как $C_{2,1}$, аналогично, светофор первой секции со стороны второй — $C_{1,2}$.

Пошаговое описание движения поезда

Шаг	a_2	d_1, d_2	t_1	$p_1, p_{1,2}^1, p_{1,2}^2, p_2$	Состояние секций		Состояние светофора	
					Первая	Вторая	$C_{1,2}$	$C_{2,1}$
1	0	1, 0	0	1, 0, 0, 0	Занята	Свободна	Красный	Зеленый
2	1	1, 0	$t_{1,2}^1$	0, 1, 0, 0				
3	0	1, 1	$t_{1,2}^2$	0, 0, 1, 0		Занята		Зеленый
4	0	0, 1	$t_{1,3}^3$	0, 0, 0, 1	Свободна			

Состояние светофора (красный/зеленый) в простейшем случае определяется состоянием секций (свободна/занята), т. е. сигналом соответствующего датчика. В таблице приведена смена состояний светофоров $C_{1,2}$ и $C_{2,1}$ при движении поезда из первой секции на вторую.

2.3. Обеспечение безопасности при параллельно-конвейерном движении поездов

Требование безопасности состоит в том, чтобы между двумя движущимися поездами в попутном или встречном направлении по одному пути была бы, по крайней мере, одна свободная секция. В сети Петри это требование означает, чтобы между позициями сегмента, содержащими фишки, была бы, по крайней мере, одна пустая позиция. Рассмотрим процесс конструирования супервизора безопасности.

На рис. 7 изображена сеть Петри сегмента, состоящего из пяти секций. Пусть в исходном состоянии сети позиции p_1 и p_3 содержат фишки, а позиция p_2 — пустая, тем самым в исходном состоянии сети выполняется требование безопасности. При $a_2 = 1$ срабатывает переход $t_{1,2}$, и фишка из по-

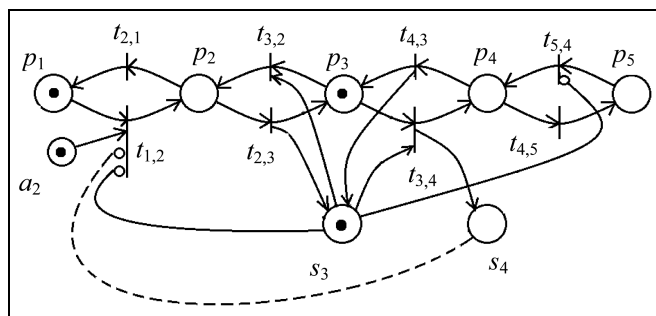


Рис. 7. Объединенная сеть Петри сегмента и супервизора безопасности

зиции p_1 перемещается в позицию p_2 , но при этом будет нарушено требование безопасности. Для блокирования срабатывания перехода $t_{1,2}$ позиция p_3 дополняется позицией s_3 супервизора и блокирующей дугой ($s_3, t_{1,2}$). Одновременно с этим блокируется симметричный ему переход $t_{5,4}$ с помощью дуги ($s_3, t_{5,4}$). Позиция s_3 дублирует позицию p_3 , т. е. наличие (отсутствие) фишки в позиции s_3 совпадает с наличием (отсутствием) фишки в позиции p_3 . Таким образом, каждая позиция p_i сети дополняется управляющей позицией s_i и соответствующими блокирующими дугами. Число управляющих позиций (позиций типа s) равно числу секций перегона.

Совокупность управляющих позиций типа s и соответствующих блокирующих дуг образует так называемый супервизор безопасности, основные функции которого заключаются в синхронизации и обеспечении требований безопасности.

Требование безопасности можно ужесточить, а именно, потребовать, чтобы между активными позициями были бы две пустые позиции. Это достигается следующим образом. Рассмотрим перемещения фишки из позиции p_3 в позицию p_4 . После срабатывания перехода $t_{3,4}$ в позицию s_4 супервизора помещается фишка. Блокирующая дуга в переход $t_{1,2}$ от позиции s_4 (штриховая линия) обеспечивает его блокирование и тем самым между позициями p_1 и p_4 , содержащими фишки, будут две пустые позиции p_2 и p_3 .

Супервизор безопасности одновременно управляет переходами светофоров из состояния «красный» в состояние «зеленый» и наоборот. Рассмотрим управление светофором $C_{2,1}$. Представим модель светофора $C_{2,1}$ в виде конечного автомата $A_{2,1}$ с двумя состояниями (красный, зеленый). Функцию переходов, переводящую автомат $A_{2,1}$ из со-

стояния «красный» в состояние «зеленый» обозначим как f_1 . Нетрудно найти, что $f_1 = \bar{d}_2 \bar{s}_3$, т. е. функция f_1 принимает единичное значение при отсутствии поезда на второй секции и отсутствия фишки в позиции s_3 супервизора. Функцию переходов, переводящую автомат $A_{2,1}$ из состояния «зеленый» в состояние «красный» обозначим как $f_2 = d_2 \vee s_3$, т. е. функция f_2 принимает единичные значения при занятой второй секции или при наличии фишки в позиции s_3 супервизора.

2.4. Дискретно-событийная модель стрелки

На рис. 8 изображена ДС-модель стрелки c_1 перегона в виде сети Петри. Позиции $m_{1,1}$ и $m_{1,3}$ соответствуют устойчивым состояниям стрелки, причем фишка в позиции $m_{1,1}$ означает, что стрелка находится в состоянии, в котором соединены сегменты 1 и 3 перегона (см. рис. 4).

Определим это состояние как положение 1 стрелки. Аналогично: фишка в позиции $m_{1,3}$ — сегмент 1 соединен с сегментом 2 (положение 2 стрелки). Датчики положения стрелки — $d_{c,1}^1$ и $d_{c,1}^2$ и управляющие переменные $a_{1,2}$, $a_{1,3}$, $b_{1,2}$, $b_{3,1}$ определяют состояние соответствующих управляющих позиций сети Петри. Рассмотрим переход из положения 2 стрелки (фишка находится в устойчивой позиции $m_{1,3}$) в положение 1 стрелки.

При значениях управляющих переменных $a_{1,3} = 1$ или $b_{3,1} = 1$ фишка перемещается из позиции $m_{1,3}$ в позицию $m_{1,4}$, в которой на стрелку c_1 выдается

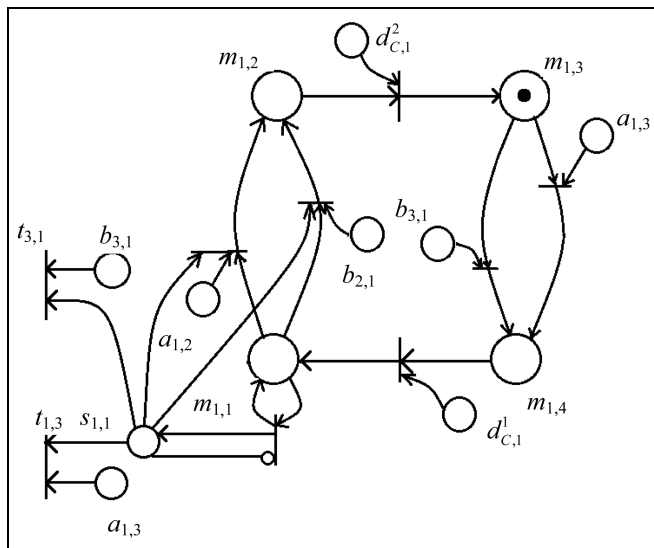


Рис. 8. Дискретно-событийная модель стрелки

управляющее воздействие. В результате изменения состояния стрелки будет $d_{c,1}^1 = 1$. В итоге фишка из позиции $m_{1,4}$ переместится в устойчивую позицию $m_{1,1}$. При этом управляющая позиция $s_{1,1}$ (позиция супервизора стрелки) и управляющие переменные $a_{1,3}$, $b_{3,1}$ ($a_{3,1}b_{3,1} = 0$) создают условия срабатывания переходов либо $t_{1,3}$, либо $t_{3,1}$ (см. рис. 4). Аналогично происходит переход стрелки из положения 1 в положение 2. В этом случае фишка перемещается из позиции $m_{1,1}$ в устойчивую позицию $m_{1,3}$, где аналогичным образом формируется управляющая позиция $s_{1,2}$ (на рис. 8 не показана), которая в зависимости от значений $a_{1,2}$ и $b_{2,1}$ ($a_{1,2}b_{2,1} = 0$) определяет условия срабатывания переходов $t_{1,2}$, $t_{2,1}$ (см. рис. 4).

Таким образом, в данном разделе показано как строить ДС-модели ж/д сети: сегмента, секции (блок-участка), светофора, стрелки, перегона с разъездом на основе аппарата сети Петри, а также показано, как строить блокирующие супервизоры, обеспечивающие безопасное групповое движение поездов.

3. ПРИМЕНЕНИЕ ДС-МОДЕЛЕЙ ЭЛЕМЕНТОВ ПЕРЕГОНА

3.1. Применение ДС-моделей элементов перегона для расчета пропускной способности

Рассмотрим применение ДС-моделей элементов перегона для определения времени движения поездов по перегону при их конвейерном движении.

Каждый перегон состоит из сегментов. Упрощенную сеть Петри сегмента (движение фишек только слева направо) представим в виде цепи последовательно соединенных k_i позиций, где k_i — число секций i -го сегмента.

Пример 1. Пусть сегмент содержит четыре секции (рис. 9).

Для обеспечения безопасности движения поездов введем ограничение на взаимное расположение фишек в сети Петри: между позициями, в которых находятся фишки, должна быть, по крайней мере, одна позиция, не содержащая фишку.

Такие маркировки сети Петри будем называть допустимыми. Пример допустимой маркировки: $\mu = (\mu_1, \mu_2, \mu_3, \mu_4) = (1, 0, 1, 0)$; пример недопустимой маркировки: $\mu = (1, 1, 0, 0)$. Допустимость маркировки обеспечивается блокирующим супервизором (см. п. 2.3).

Сеть Петри функционирует следующим образом. Переход t может послать фишку в позицию p_i всякий раз, когда это не противоречит принятому ограничению. В дереве достижимости сети Петри каждая последовательность маркировок определяется последовательностью

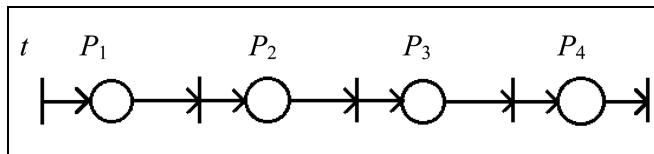


Рис. 9. Сеть Петри сегмента, состоящего из четырех секций

срабатывания переходов. Можно показать, что самая короткая последовательность допустимых маркировок для пакета из двух поездов в нашем примере имеет место при одновременном сдвиге фишек в сети Петри:

$$\begin{aligned} \mu_0 &= (0, 0, 0, 0), \quad \mu_1 = (1, 0, 0, 0), \quad \mu_2 = (0, 1, 0, 0), \\ \mu_3 &= (1, 0, 1, 0), \quad \mu_4 = (0, 1, 0, 1), \quad \mu_5 = (0, 0, 1, 0), \\ \mu_6 &= (0, 0, 0, 1), \quad \mu_0 = (0, 0, 0, 0). \quad \blacklozenge \end{aligned}$$

Длину последовательности маркировок, начинающейся и оканчивающейся начальной маркировкой μ_0 при прохождении пакетом поездов i -го сегмента, обозначим как N_i . Иными словами, величина N_i равна числу тактов прохождения пакетом поездов i -го сегмента.

Нетрудно видеть, что для прохождения одного поезда через k_i секций i -го сегмента потребуется $N_i = k_i + 1$ тактов. Для прохождения пакета из двух поездов (с одной секции безопасности между ними) потребуется $N_i = (k_i + 2) + 1$ тактов. Для прохождения n поездов в пакете потребуется не менее $N_i = (k_i + n) + (n - 1)$ тактов или

$$N_i = k_i + 2n - 1. \quad (1)$$

В общем случае, когда интервал безопасности равен r ($r = 1, 2, 3$), минимальное число тактов для прохождения пакета из n поездов через k секций определяется как

$$N = k_i + n + r(n - 1), \quad (2)$$

а время прохождения

$$N\tau = (k + n + r(n - 1))\tau, \quad (3)$$

где τ — среднее время прохождения поездом одной секции.

Если задан интервал времени T и число секций k и требуется определить максимальное число прошедших поездов, то из формулы (3) следует, что

$$n_{\max} = (T/\tau - k + r)/r. \quad (4)$$

Пример 2. Определим максимальную пропускную способность сегмента за сутки ($T = 1440$ мин), пусть $\tau = 1$ мин, $k = 10$, тогда при интервале безопасности $r = 1$ по формуле (4) получаем 67 поездов, при $r = 2$ получаем 45 поездов, при $r = 3$ имеем 34 поезда.

Таким образом, формула (4) определяет пропускную способность сегмента перегона. В любом маршруте сегмент, имеющий минимальную пропускную способность, определяет пропускную способность всего маршрута.

3.2. Варианты организации параллельного движения поездов по перегону с разъездом

Перегон с разъездом (см. рис. 3) содержит два маршрута от станции A до станции B :

$$M_A^1 = (A, 1, 3, 4, B), \quad M_A^2 = (A, 1, 2, 4, B),$$

где цифрами 1—4 обозначены сегменты.

Длины маршрутов соответственно

$$L_A^1 = k_1 + k_3 + k_4, \quad L_A^2 = k_1 + k_2 + k_4,$$

где k — число секций соответствующего сегмента.

От станции B до станции A имеются два маршрута

$$M_B^1 = (B, 4, 3, 1, A), \quad M_B^2 = (B, 4, 2, 1, A).$$

Существует большое число вариантов организации движения поездов между станциями A и B . Рассмотрим наиболее интересные. Пусть требуется отправить n_a поездов со станции A на станцию B , а со B на станцию A — n_b . Интервал безопасности равен 1.

Вариант 1. Рассмотрим случай, когда в начале поезда отправляются из A в B , затем — из B в A по основному маршруту M^1 . Для простоты анализа положим, что все секции имеют равные временные задержки. По формуле (1) определяем длину пути из A в B

$$N_{ab} = k_1 + k_3 + k_4 + 2n_a - 1.$$

Аналогично длина пути из B в A

$$N_{ba} = k_1 + k_3 + k_4 + 2n_b - 1,$$

тогда

$$\begin{aligned} N_{ab} + N_{ba} &= 2(k_1 + k_3 + k_4) + 2n_a + 2n_b - 2 = \\ &= 2(k_1 + k_3 + k_4 + n_a + n_b - 1). \end{aligned}$$

Время в пути $T = (N_{ab} + N_{ba})\tau$, где τ — среднее время прохождения поездом одной секции перегона.

Вариант 2. Пакет поездов размерности n_a , следующий от станции A по маршруту $M_A^2 = (A, 1, 2, 4, B)$, и пакет поездов размерности n_b , одновременно следующий от станции B по маршруту $M_B^1 = (B, 1, 3, 4, A)$ не должны мешать друг другу.



Необходимо определить величины n_a и n_b .

Величина n_a определяется как

$$2n_a = k_3 + k_4 - k_1 + 1.$$

Величину n_b определяем аналогично:

$$2n_b = k_1 + k_2 - k_4 + 1.$$

Пусть, например $k_1 = k_4$, тогда

$$2n_a = k_3 + 1 \text{ и } 2n_b = k_2 + 1,$$

отсюда

$$n_a = (k_3 + 1)/2 \text{ и } n_b = (k_2 + 1)/2.$$

ЗАКЛЮЧЕНИЕ

Сформулируем основные результаты настоящей работы.

- Построены ДС-модели компонентов перегона с разездом: сегмента, секции (блок-участка), стрелки, светофора, перегона с разездом на основе сети Петри.
- Показано, как строить блокирующие супервизоры, управляющие ДС-моделями элементов ж/д сети и обеспечивающие безопасное параллельно-конвейерное движение поездов. Разработаны супервизоры для компонентов перегона: сегмента, секции (блок-участка), стрелки, светофора, перегона с разездом. Модели элементов перегона и супервизоров в виде сетей Петри могут быть использованы в системе моделирования железнодорожной сети при параллельно-конвейерном движении поездов.
- Разработан метод оценки времени параллельно-конвейерного безопасного движения поездов по перегону, что позволит приблизиться к решению задачи определения оптимальных по времени маршрутов в произвольной железнодорожной сети.

Авторы признательны О.П. Кузнецову, конструктивная и доброжелательная критика которого способствовала улучшению статьи.

ЛИТЕРАТУРА

1. Ramadge J.G., Wonham R.W.M. The control of discrete-event systems // IEEE Trans. Automat. Control. — 1989. 77 (1). — P. 81—98.
2. Chalmers Golaszewski C. H., Ramadge P. J. Control of discrete-event processes with forced events // Proc. 28th Conf. Decision Control. Los Angeles. — 1987. — P. 247—251.
3. Cassandras C.G., Lafortune S. Introduction to discrete event systems // Springer Science + Business Media, LLC. USA. — 2008. — P. 781.
4. Mustafa Seçkin Durmuş, Mehmet Turan Söylemez. Automation Petri Net Based Railway Interlocking and Signalization Design, Istanbul Technical University. International Symposium on INnovations in Intelligent Systems and Applications, INISTA'09, Karadeniz Technical University, Trabzon, Turkey, 29 June—01 July, 2009.
5. Joakim Björk, Anders M. Hagalisletto, and Paal Enger. Large scale simulations of Railroad Nets, Precise Modeling and Analysis, Department of Informatics, University of Oslo, 2006.
6. Maged Dessouky, Ph.D, Fernando Ordonez, Ph.D, Robert Leachman, Ph.D, Pavankumar Mural. STRATEGIES FOR EFFECTIVE RAIL TRACK CAPACITY USAGE. Final Report METRANS Project January 16, 2010. University of Southern California. Automation Petri Net Based Railway Interlocking and Signalization Design.
7. Амбарцумян А.А. Моделирование и синтез супервизорного управления на сетях Петри для рассредоточенных объектов. Ч. 1. // Автоматика и телемеханика. — 2011. — № 8. — С. 151—169. Ч. 2. // Автоматика и телемеханика. — 2011. — № 9. — С. 173—189.
8. Амбарцумян А.А. Потехин А.И. Групповое управление в дискретно-событийных системах // Проблемы управления. — 2012. — № 5. — С. 46—53.

Статья представлена к публикации членом редколлегии О.П. Кузнецовым.

Потехин Анатолий Иванович — канд. техн. наук, вед. науч. сотрудник, ☎ (495) 334-90-29, ✉ apot@ipu.ru,

Браништов Сергей Александрович — канд. техн. наук, и. о. зав. лабораторией, ☎ (495) 334-93-70, ✉ branishtov@mail.ru,

Кузнецов Сергей Константинович — науч. сотрудник, ☎ (495) 334-90-29, ✉ skuznetsov@bk.ru,

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

ПРИМЕНЕНИЕ ТЕОРЕМЫ ЛАГРАНЖА О КОНЕЧНЫХ ПРИРАЩЕНИЯХ ДЛЯ РЕШЕНИЯ ПРОБЛЕМ УПРАВЛЕНИЯ ТРАНСПОРТНЫМИ СИСТЕМАМИ¹

С.Л. Блюмин, А.С. Сысоев

Рассмотрена постановка задачи оптимизации процесса функционирования регулируемого перекрестка. Исследована возможность применения теоремы Лагранжа о конечных приращениях к выработке управленческих решений по реорганизации работы регулируемого перекрестка. Приведен численный пример.

Ключевые слова: теорема Лагранжа о конечных приращениях, управление транспортными системами, транспортная задержка.

ВВЕДЕНИЕ

Проблемы, связанные с управлением транспортными потоками, становятся все острее, о чем свидетельствуют, в частности, транспортные заторы на пересечении городских магистралей (так называемые «пробки»).

Различают два направления исследований [1] этих проблем — управление системой связанных перекрестков (реализуется с помощью настройки планов координации и программы «зеленая волна») и управление отдельным регулируемым перекрестком, которое бывает полезно в часы пик для некоторых элементов транспортной сети мегаполиса.

Для оценки качества функционирования регулируемых перекрестков обычно принимается значение средней транспортной задержки регулирования, представляющей собой функцию одного или нескольких аргументов. Минимизация этой функции по соответствующим аргументам дает оптимальное значение длительности цикла и ее оптимальное распределение по фазам регулирования.

Представляет интерес задача исследования поведения этой функции при изменении ее аргумен-

тов. Ее решение может быть получено с помощью анализа функций на основе теоремы Лагранжа о промежуточной точке. Ранее подобный метод был описан в контексте экономического факторного анализа производственных функций [2]. Приведенные в настоящей статье рассуждения аналогичны рассуждениям, положенным в основу [2], с той лишь разницей, что оцениваются не производственные функции, а функции, описывающие параметры транспортных систем.

1. ЗАДАЧА ОПТИМИЗАЦИИ РЕГУЛИРУЕМОГО ПЕРЕКРЕСТКА

Задача оптимизации функционирования регулируемого перекрестка понимается как поиск оптимальных длительности цикла светофорного регулирования и ее распределения по фазам регулирования.

В строгой постановке: найти вектор $[T, g_1, \dots, g_m]$, такой что

$$D = \sum_{i=1}^m \left(k_f^i \frac{(T - g_i)^2 s_i}{T(s_i - v_i)} + \frac{T}{4} \left(\left(\frac{v_i T}{g_i s_i} - 1 \right) + \sqrt{\left(\frac{v_i T}{g_i s_i} - 1 \right)^2 + 14 \cdot 400 \frac{v_i T}{g_i s_i^2}} \right) \right) v_i \Bigg/ \sum_{i=1}^m v_i \rightarrow \min \quad (1)$$

¹ Работа выполнена при финансовой поддержке РФФИ (проект № 13-07-97519 p_центр_a).



при ограничениях

$$\sum_{i=1}^m g_i = T, \quad (2)$$

$$T_{\min} \leq T \leq T_{\max}, \quad (3)$$

$$(g_i)_{\min} \leq g_i \leq (g_i)_{\max}, \quad i = 1, \dots, m. \quad (4)$$

Здесь T — длительность цикла светофорного регулирования, с, g_i , $i = 1, \dots, m$, — длительность фаз светофорного регулирования, с, $D(T, g_1, \dots, g_m)$ — значение функции агрегированной задержки, с, k_f^i , $i = 1, \dots, m$, — коэффициент, определяющий тип прибытия транспортных средств (т. е. насколько плотным потоком транспортные средства прибывают к рассматриваемому перекрестку), v_p , $i = 1, \dots, m$, — интенсивность прибытия транспортных средств на i -е направление (приведенные автомобили в час), s_p , $i = 1, \dots, m$, — поток насыщения i -го направления, прив. авт./ч.

Учитывая ограничения (2) и (3), получим множество X , на котором минимизируется функция (1). Задача сводится к задаче условной минимизации с ограничением в виде равенства (2). Для перехода к задаче безусловной минимизации построим квадратичную штрафную функцию $\phi(g, T, \beta)$, которая будет стремительно увеличиваться за пределами множества X и обращаться в ноль внутри него.

Получаем задачу многомерной безусловной минимизации

$$\tilde{D} = D + \phi(g, T, \beta) \rightarrow \min. \quad (5)$$

Решение задачи (5) проведено численно на основе предложенного мультистартового параллельного алгоритма безусловной многомерной минимизации [3].

Численный пример решения задачи управления регулируемым перекрестком путем отыскания оптимальных длительности цикла регулирования и ее распределения по фазам регулирования минимизацией функции агрегированной задержки рассмотрен в § 4.

2. ФАКТОРНЫЙ АНАЛИЗ ФУНКЦИЙ, ОСНОВАННЫЙ НА ПРИМЕНЕНИИ ТЕОРЕМЫ ЛАГРАНЖА О ПРОМЕЖУТОЧНОЙ ТОЧКЕ

Описываемый подход пришел из экономики. Математическому понятию «значение функции» в экономическом анализе соответствует понятие

«значение результирующего показателя», а понятие «приращение переменной» — «приращение фактора, оказывающего на этот показатель влияние». Основная задача экономического факторного анализа состоит в оценке влияния факторов на результирующий показатель [2, 4]. Однако данный подход может применяться не только для анализа экономических процессов, но и многих других, в том числе процессов в технических системах.

Задача оценки влияния факторов опирается на исчисление конечных разностей, связываемое, в свою очередь, через теорему Лагранжа о промежуточной точке (конечных приращениях, средней точке) с дифференциальным исчислением [5]. В зависимости от вида представления изменения значения функции (показателя) и ее аргументов (факторов), выделяются аддитивное исчисление (исчисление конечных разностей, в качестве изменения значения величины принимают приращение или разность $\Delta x = x - x_0$), мультипликативное исчисление (исчисление конечных частных, в качестве изменения значения величины — индекс или частное $\dot{x} = x/x_0$, при $x_0 \neq 0$), а также их комбинации.

Из математического анализа известна модель, характеризующая зависимость приращения (отклонения, разности) значения функции (отклика, показателя) от приращения (отклонения, разности) аргумента (переменной, фактора). Это — теорема Лагранжа о средней (промежуточной) точке [6]:

$$\begin{aligned} \Delta f(x_1, \dots, x_n) &= \\ &= \sum_{i=1}^n \frac{\partial}{\partial x_i} \Delta x_i (x_1^0 + \alpha \Delta x_1, \dots, x_n^0 + \alpha \Delta x_n), \end{aligned} \quad (6)$$

где x_i^0 и Δx_i , $i = 1, \dots, n$, — начальные значения и приращения аргументов функции соответственно, $\alpha \in (0; 1)$ — параметр смещения промежуточной точки.

Ранее возникала идея получения альтернативных форм теоремы Лагранжа, в которой конечные разности (приращения) — аддитивные меры различия чисел — замещены конечными частными (индексами) — мультипликативными мерами различия [5, 7, 8].

Итак, теорема Лагранжа представляет собой основную модель аддитивного исчисления (6). Путем логарифмирования (экспоненцирования) аргумента (функции) возможно получение основной модели мультипликативного исчисления (7) и смешан-

ных моделей мультипликативно-аддитивного (8) и аддитивно-мультипликативного исчислений (9):

$$\Delta y = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(\xi) \Delta x_i, \quad (7)$$

$$iy = \prod_{i=1}^n (ix_i)^{\gamma_i}, \quad (8)$$

где $\gamma_i = \frac{\xi_i}{f(\xi_1, \dots, \xi_n)} \cdot \frac{\partial f}{\partial x_i}(\xi)$ — частные эластичности в промежуточной точке.

$$iy = \prod_{i=1}^n \left(\exp \left(\frac{1}{f(\xi_1, \dots, \xi_n)} \cdot \frac{\partial f}{\partial x_i}(\xi) \right) \right)^{\Delta x_i}, \quad (9)$$

$$\Delta y = \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i}(\xi) \cdot \xi_i \right) \ln ix_i. \quad (10)$$

Перспектива применения моделей для анализа функции транспортной задержки с целью выработки управляющих решений на примере модели (7) показана далее.

3. ФАКТОРНЫЙ АНАЛИЗ ОДНОГО КЛАССА РАЦИОНАЛЬНЫХ ФУНКЦИЙ

Рассмотрим класс рациональных функций специального вида:

$$f(x_1, \dots, x_n) = \left(a_0 \sum_{i=1}^n a_i x_i + \sum_{i=1}^n b_i x_i^2 \right) / \sum_{i=1}^n x_i, \quad (11)$$

где $x_i \geq 0$, $a_0, a_i, b_i = \text{const} \geq 0$.

Начальным значениям аргументов $x_i^0, i = 1, \dots, n$, отвечает начальное значение функции $f(x_1^0, \dots, x_n^0)$. Аргументы претерпевают изменения Δx_i , которые порождают новые значения аргументов $x_i^0 + \Delta x_i$ и функции $f(x_1^0 + \Delta x_1, \dots, x_n^0 + \Delta x_n)$.

Частные производные функции (11):

$$\frac{\partial f(x_1, \dots, x_n)}{\partial x_i} = (a_i + 2b_i x_i) / \sum_{i=1}^n x_i - \left(a_0 \sum_{i=1}^n a_i x_i + \sum_{i=1}^n b_i x_i^2 \right) / \left(\sum_{i=1}^n x_i \right)^2.$$

С одной стороны, приращение функции может быть вычислено как

$$f(x_1^0 + \Delta x_1, \dots, x_n^0 + \Delta x_n) - f(x_1^0, \dots, x_n^0), \quad (12)$$

а с другой стороны, как

$$\sum_{i=1}^n \frac{\partial}{\partial x_i} \cdot \Delta x_i (x_1^0 + \alpha \Delta x_1, \dots, x_n^0 + \alpha \Delta x_n). \quad (13)$$

Приравнявая выражения (12) и (13) и решая полученное уравнение относительно α , получаем:

$$\alpha_{1,2} = \left(- \sum_{i=1}^n x_i \pm \sqrt{\sum_{i=1}^n x_i \sum_{i=1}^n (x_i^0 - \Delta x_i)} \right) / \sum_{i=1}^n \Delta x_i.$$

В силу предположения о положительности значений аргументов и их приращений, только значение α_1 (знак «+» перед квадратным корнем) удовлетворяет условию теоремы Лагранжа о принадлежности параметра промежуточной точки промежутку (0; 1).

В качестве примера рассмотрим анализ функции вида (11) от двух аргументов (оценка полученного результата с позиций транспортных систем приводится далее).

Рассматривая функцию двух аргументов, приходим к структуре вида

$$f(x_1, x_2) = \frac{a_0 + a_1 x_1 + a_2 x_2 + b_1 x_1^2 + b_2 x_2^2}{x_1 + x_2}.$$

Начальным значениям аргументов x_1^0, x_2^0 отвечает начальное значение функции $f(x_1^0, x_2^0)$. Аргументы претерпевают изменения $\Delta x_1, \Delta x_2$, которые порождают новые значения аргументов $x_1^0 + \Delta x_1, x_2^0 + \Delta x_2$ и функции $f(x_1^0 + \Delta x_1, x_2^0 + \Delta x_2)$.

Частные производные:

$$\frac{\partial f(x_1, x_2)}{\partial x_1} = \frac{a_1 + 2b_1 x_1}{x_1 + x_2} - \frac{a_0 + a_1 x_1 + a_2 x_2 + b_1 x_1^2 + b_2 x_2^2}{(x_1 + x_2)^2},$$

$$\frac{\partial f(x_1, x_2)}{\partial x_2} = \frac{a_2 + 2b_2 x_2}{x_1 + x_2} - \frac{a_0 + a_1 x_1 + a_2 x_2 + b_1 x_1^2 + b_2 x_2^2}{(x_1 + x_2)^2}.$$

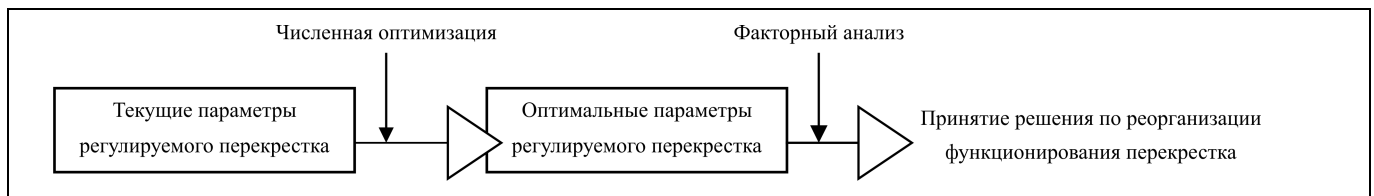


Рис. 1. Оптимизация и анализ функционирования регулируемого перекрестка

С одной стороны, приращение функции может быть вычислено как $f(x_1^0 + \Delta x_1, x_2^0 + \Delta x_2) - f(x_1^0, x_2^0)$, а с другой — как $\left(\frac{\partial}{\partial x_1} \cdot \Delta x_1 + \frac{\partial}{\partial x_2} \cdot \Delta x_2\right)(x_1^0 + \alpha \cdot \Delta x_1, x_2^0 + \alpha \cdot \Delta x_2)$.

Находя из равенства этих выражений, получаем (учитывая положительность аргументов и приращений):

$$\alpha = \frac{-(x_1^0 + x_2^0) + \sqrt{(x_1^0 + x_2^0)(x_1^0 + \Delta x_1 + x_2^0 + \Delta x_2)}}{\Delta x_1 + \Delta x_2}. \quad (14)$$

Перейдем к рассмотрению конкретных приложений.

4. ОБСУЖДЕНИЕ ПОЛУЧЕННЫХ РЕЗУЛЬТАТОВ. ЧИСЛЕННЫЙ ПРИМЕР

Процесс последовательного применения оптимизации и анализа системы представлен на рис. 1 в виде структурной схемы. Минимизируется функция транспортной задержки, для анализа применяется теорема Лагранжа о промежуточной точке.

На этапе численной оптимизации находится оптимальный набор $[T, g_1, \dots, g_m]$ — длительность цикла регулирования и ее распределение по фазам. Затем на этапе факторного анализа проводится (на основе статистической информации, собранной на данном объекте регулирования) предсказательный анализ изменения интенсивности движения в критических направлениях с целью выявления фазы, оказывающий наибольшее воздействие на изменение задержки на перекрестке. Особенность анализа состоит в том, что найденная фаза не обязана обладать наибольшим среди всех рассматриваемых в системе фаз отношением интенсивности к пропускной способности, v/c .

В качестве примера рассмотрим двухфазный регулируемый перекресток (рис. 2). Фаза 1 включает в себя направления движения 1, 2, 5 и 6; критическим в этой фазе является направление 2.

Фаза 2 объединяет направления 3, 4, 7 и 8; критическим в ней является направление 4.

Входные интенсивности в критических группах фаз, а также результаты оптимизации для рассматриваемого примера представлены в табл. 1, где также приведены результаты оптимизации функционирования перекрестка традиционным методом Вебстера. Близость отношений v/c , полученных с помощью стандартного метода Вебстера и предлагаемого метода, подтверждает, что оптимизация проведена корректно.

Покажем применение факторного анализа, основанного на использовании модели (7), для выявления проблемной фазы регулирования. Большой практический интерес такой подход приобретает при большом числе фаз регулирования. Но для простоты понимания ограничимся рассмотрением двухфазной системы.

Рассмотрим модификацию функции из задачи (1), предложенную в работе [9]:

$$D = k_f^i \frac{(T - g_i)^2 s_i}{T(s_i - v_i)} + \frac{T}{4} \left(\left(\frac{v_i T}{g_i s_i} - 1 \right) + \sqrt{\left(\frac{v_i T}{g_i s_i} - 1 \right)^2 + 14 \cdot 400 \frac{v_i T}{g_i^2 s_i^2}} \right). \quad (15)$$

Алгоритм получения средней задержки для регулируемого перекрестка приводит к модели вида

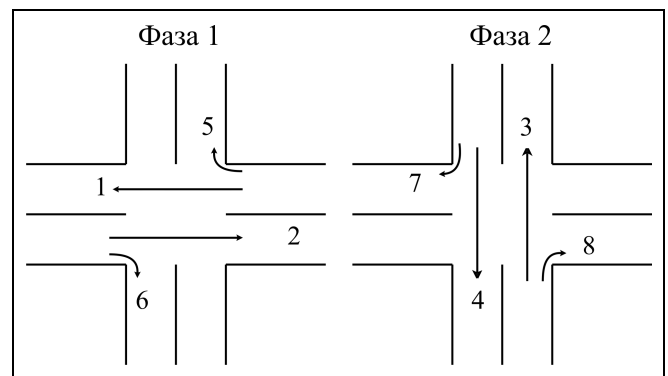


Рис. 2. Пример двухфазного регулируемого перекрестка

Таблица 1

Результаты оптимизации

Характеристика		Фаза 1	Фаза 2
Интенсивность критической группы, прив. авт./ч		180	1601
Поток насыщения, прив. авт./ч		1800	1800
Длительность фазы, с	до оптимизации	18	39
	после оптимизации	15,5	24,3
	после оптимизации методом Вебстера	7,9	13,8
Задержка в критической группе, с	до оптимизации	16,97	27,72
	после оптимизации	9,69	23,01
	после оптимизации методом Вебстера	6,08	21,56
Отношение v/c	до оптимизации	0,317	1,299
	после оптимизации	0,257	1,457
	после оптимизации методом Вебстера	0,275	1,399

$D = \sum_{i=1}^n d(v_i)v_i / \sum_{i=1}^n v_i$. Анализ этой функции, где в качестве каждого $d(v_i)$ выступает выражение (15), представляет собой вычислительно сложную задачу. Для упрощения анализа применим к функции D разложение в ряд Маклорена (до трех членов). В результате получим модель, структурно схожую с

выражением (11). Здесь $a_0 = 0$, $a_i = \frac{1}{2} k_f^i T \left(1 - \frac{g_i}{T}\right)^2$, $b_i = \frac{1}{2} \frac{k_f^i (T - g_i)^2}{T s_i} + \frac{1}{4} T \left(\frac{T}{g_i s_i} - \frac{T(g_i s_i - 7200)}{g_i^2 s_i^2} \right)$, $i = 1, \dots, m$.

Параметр средней точки для данной модели в случае двух аргументов, согласно формуле (14), находится как

$$\alpha = \frac{-(v_1 + v_2) + \sqrt{(v_1 + v_2)(v_1 + \Delta v_1 + v_2 + \Delta v_2)}}{\Delta v_1 + \Delta v_2}.$$

Для некоторых начальных значений интенсивностей и их приращений выполнен полный факторный эксперимент (значение потока насыщения для всех фаз одинаково и равно 1800 прив. авт./ч). Значения параметров средней точки и приращения транспортной задержки, а также процентные значения коэффициентов факторного влияния приведены в табл. 2.

Например, для ситуации, когда интенсивность движения критической группы первой фазы равна 1000 прив. авт./ч, а второй — 120 прив. авт./ч; приращения интенсивностей равны соответственно 75 и 5, задержка станет больше на 159,47 с, причем на 70 % такое приращение определяется увеличением интенсивности критической группы второй фазы.

Таблица 2

Результаты анализа функции задержки регулирования

v_1	v_2	Δv_1	Δv_2	$v_1 + \alpha v_1$	$v_2 + \alpha v_2$	ΔD , с	Влияние v_1 , %	Влияние v_2 , %
1000	70	15	5 25	1007,465 1007,431	72,488 82,385	2,485 197,180	57 48	43 52
		75	5 25	1036,824 1036,663	72,455 82,221	329,072 444,307	58 49	42 51
	120	15	5 25	1007,467 1007,434	122,489 132,390	2,313 287,127	29 26	71 74
		75	5 25	1036,853 1036,698	122,457 132,233	159,473 389,049	30 27	70 73
1500	70	15	5 25	1507,476 1507,453	72,492 82,421	2,271 100,154	83 71	17 29
		75	5 25	1537,034 1536,921	72,469 82,307	398,044 437,816	85 72	15 28
	120	15	5 25	1507,477 1507,454	122,492 132,424	2,494 218,672	44 40	56 60
		75	5 25	1537,048 1536,938	122,470 132,313	267,177 434,851	45 41	55 59



ЗАКЛЮЧЕНИЕ

Описанный метод может служить основанием для принятия управленческих решений по реорганизации существующего режима функционирования регулируемого перекрестка. Свою дальнейшую работу авторы связывают с исследованием возможности применения данного метода для анализа сети перекрестков.

ЛИТЕРАТУРА

1. Иносэ Х., Хамада Т. Управление дорожным движением. — М.: Транспорт, 1983. — 248 с.
2. Блюмин С.Л., Суханов В.Ф., Чеботарев С.В. Экономический факторный анализ: Монография. — Липецк: ЛЭГИ, 2004. — 148 с.
3. Сысоев А.С. Численный алгоритм оптимизации функции транспортной задержки на регулируемом перекрестке // Научно-технический вестник Поволжья. — 2012. — № 5. — С. 324—330.
4. Баканов М.И., Шеремет А.Д. Теория экономического анализа — М.: Финансы и статистика, 1997. — 416 с.

5. Блюмин С.Л. Исчисление конечных разностей и частных // Фундаментальные исследования. — 2008. — № 3. — С. 105.
6. Математический анализ: учеб. / Под ред. А.Н. Тихонова. — 3-е изд., перераб. и доп.: в 2 ч. Ч. 1. — М.: ТК Велби, Проспект, 2006. — 672 с.
7. Кац В.Г., Чен П. Квантовый анализ / Пер. с англ. — М.: МЦНМО, 2005. — 128 с.
8. Сысоев А.С. Альтернативные формы представления теоремы конечных приращений Лагранжа и их применение // Сб. материалов междунар. форума студенческой и учащейся молодежи «Первый шаг в науку — 2009». Центр студенческих научных инициатив при Совете молодых ученых НАН Беларуси. — Минск: Право и экономика, 2010. — Т. 2. — С. 524—527.
9. Canadian Capacity Guide for Signalized Intersections 2008. Institute of Transportation Engineers, District 7 — Canada. ISBN 978-1-933-45224-1.

Статья представлена к публикации членом редколлегии А.А. Дорофеевым.

Блюмин Семен Львович — д-р физ.-мат. наук, профессор,
☎ (4742) 32-80-50, ✉ sabl@lipetsk.ru,

Сысоев Антон Сергеевич — ассистент,
☎ (4742) 32-80-50, ✉ anton_syssoyev@mail.ru,

Липецкий государственный технический университет.



НАУКОМЕТРИЯ И ЭКСПЕРТИЗА В УПРАВЛЕНИИ НАУКОЙ

Специальный выпуск электронного научного журнала «Управление большими системами»

Выпуск посвящен применению наукометрических и экспертных методов к задачам оценки эффективности научно-исследовательской деятельности — как отдельных ученых, организаций, институтов, научных направлений, так и науки страны в целом. Эта тема приобрела особую актуальность в последние годы с ростом популярности формальных показателей, основанных на библиометрической информации. Потенциально эти показатели могут и должны учитываться на всех этапах процесса управления научно-исследовательской деятельностью.

Однако применение библиометрических (и, более общо, наукометрических) показателей в управлении наукой встречает множество трудностей, а также сильное противодействие, в первую очередь самих ученых, указывающих на невозможность количественного измерения значимости научного результата, неполноту и подверженность любого индекса манипулированию со стороны заинтересованных лиц.

Альтернатива состоит в расширении и совершенствовании тех или иных экспертных процедур для поддержки принятия решений по управлению наукой. Цель Спецвыпуска как раз и состоит в том, чтобы разобраться, что же в настоящее время предлагается этими двумя подходами к оценке эффективности научной деятельности, насколько обоснованно их противопоставление и как их следует сочетать для достижения наилучших результатов.

Главный итог, с которым согласны большинство авторов, можно сформулировать таким образом: только профессиональная экспертиза может дать всестороннюю объективную оценку научных результатов и заслуг; наукометрические же показатели служат инструментом поддержки принятия решений экспертами.

Полные тексты статей Спецвыпуска доступны онлайн по адресу:
http://ubs.mtas.ru/archive/index.php?SECTION_ID=685.

CONTENTS & ABSTRACTS

CONFORMAL BEHAVIOR MODELS.

PART 1. FROM PHILOSOPHY TO MATH MODELS. 2

Breer V.V.

The first part gives the review of conformal behavior studies by various sciences — philosophy, culturology, theoretical, practical, and social psychology. Attempt of classification of mathematical models of conformal behavior as a special case of social interactions models is carried out. Schelling and Granovetter models which gave rise to other mathematical models are considered in details.

Keywords: conformal behavior, social interaction, model of critical mass, threshold model of social interaction, social-physics model, crowd control.

GUARANTEED RISKS AND OUTCOMES

IN «GAME AGAINST NATURE» 14

Zhukovskiy V.I., Soldatova N.G.

One-criteria problem under uncertainty is considered. There are only the limits of change that are known about uncertainty. The two concepts of formalization of guaranteed (for outcomes and risks simultaneously) solution are offered. The first concept is based on the method of decision-making in two-level hierarchical game. The second concept is based on the vector-maximin definition. The explicit form of solution under linear-quadratic criterion is found.

Keywords: strategy, uncertainty, criterion, minimax regret and maximin, vector-optimum.

THE INTELLIGENT MANAGEMENT SYSTEM

OF DEVELOPMENT OF REGIONAL

TRANSPORT-LOGISTIC INFRASTRUCTURE 27

Bychkov I.V., Kazakov A.L., Lempert A.A., Bukharov D.S., Stolbov A.B.

The paper considers the intelligent system based on ontology, formalized expert knowledge, mathematical models, and numerical methods. The system focuses on decision support for development of regional transport-logistic infrastructure. Particularly, for examination of transport-logistic systems the developed complex approach is used with the author's conception of multilevel modeling. Some short descriptions of formalized mathematical models and integrated program units are presented. Modern intellectualization tools are used for the program units communication and processing of statistical data. Several model tasks and applied problems are solved with use of the intelligent system, and results of tasks solution are given.

Keywords: intelligent system, production rules, knowledge bases, ontology, integration, multilevel modeling, transport logistics.

FOUR-STAGE BRAINSTORM 36

Sidelnikov Yu.V.

The paper proposes the four-stage brainstorm, which removes the number of deficiencies of many other brainstorm modifications. The proposed modification allows to increase the brainstorm effectiveness by increasing the contribution of preparatory stage and adding the fourth stage, in which the possibility of problem definition reformulation, on the basis of information obtained at the previous stages, is provided. Furthermore, the responsibilities of leaders of the second and third stages are extended and thoroughly prescribed.

Keywords: expert estimations, brainstorm modification.

REAL ESTATE MASS APPRAISAL MODELS DESIGN

WITH THE SPATIAL DEPENDENCE. 45

Belyaeva A.V., Grebenuk E.A.

The paper proposes the new approach for real estate mass appraisal models design with location effect factors influencing the price: the distance to influence centers and the real estate listing neighborhood. The models are designed to implement the approach. The results showed that the joint consideration of both factors significantly improves the quality of the model.

Keywords: spatial correlation, Moran test, Lagrange multiplier tests, spatial autoregressive model, influence centers.

MODELING OF CORRUPTION IN THREE-LEVEL

CONTROL SYSTEMS 53

Ougolnitsky G.A., Usov A.B.

The dynamic models of corruption in three-level control systems of the consequent structure are considered. The definitions of equilibria in three-person dynamic games with respect to sustainable development requirements with corruption are given, algorithms of their construction are proposed. The models of economic corruption in the systems of surface water quality control based on corresponding informational patterns are considered, the examples of numerical calculations are given.

Keywords: corruption, dynamic games, hierarchical control system, compulsion, impulsion, conditions of sustainable development.

EXASCALE SUPERCOMPUTERS. ARCHITECTURAL

OUTLINES. 63

Stepanenko S.A., Yuzhakov V.V.

Architectural aspects of exascale computing systems are explored. Assessments of performance of computing systems and communication environment are made. Necessity and potential of architectural efficiency scaling solutions are demonstrated.

Keywords: hybrid architectures, architectural efficiency scaling solutions, hybrid reconfigurable structures, minimization of communication time, topological redundancy.

RAILWAY NETWORK DISCRETE-EVENT MODELS 73

Potekhin A.I., Branishtov S.A., Kuznetsov S.K.

The paper develops discrete-event models of railway network basic elements: track region with track for passing, track, track segment, track-section, turnout (router) and models of train movements. The elements models are presented as Petri Nets with inhibitor and enabling arcs. Group control for these models is implemented by special developed control components (supervisors) that meet traffic safety requirements.

Keywords: discrete event models, Petri nets, supervisor, railway net, railway region, station-to-station block, railway track.

APPLICATION OF LAGRANGE MEAN VALUE THEOREM

TO CONTROL OF TRANSPORT SYSTEMS 82

Blyumin S.L., Sysoev A.S.

The paper considers the definition of optimal functioning problem for a regulated city intersection as well as the applicability of Lagrange mean value theorem to elaboration of control decisions for this kind of intersection based on the factor analysis of transport delay. The numerical example is given.

Keywords: Lagrange mean value theorem, control of transport systems, transport delay.

Рисунки к статье С.А. Степаненко, В.В. Южакова
«ЭКСАФЛОПНЫЕ СУПЕР-ЭВМ: КОНТУРЫ АРХИТЕКТУРЫ»
 (см. с. 63–72)

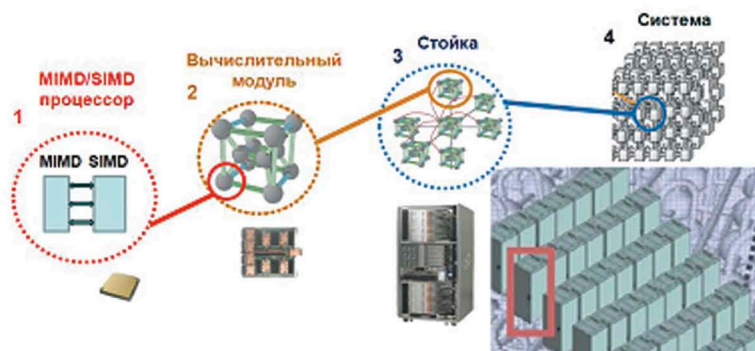


Рис. 2. Уровни параллелизма

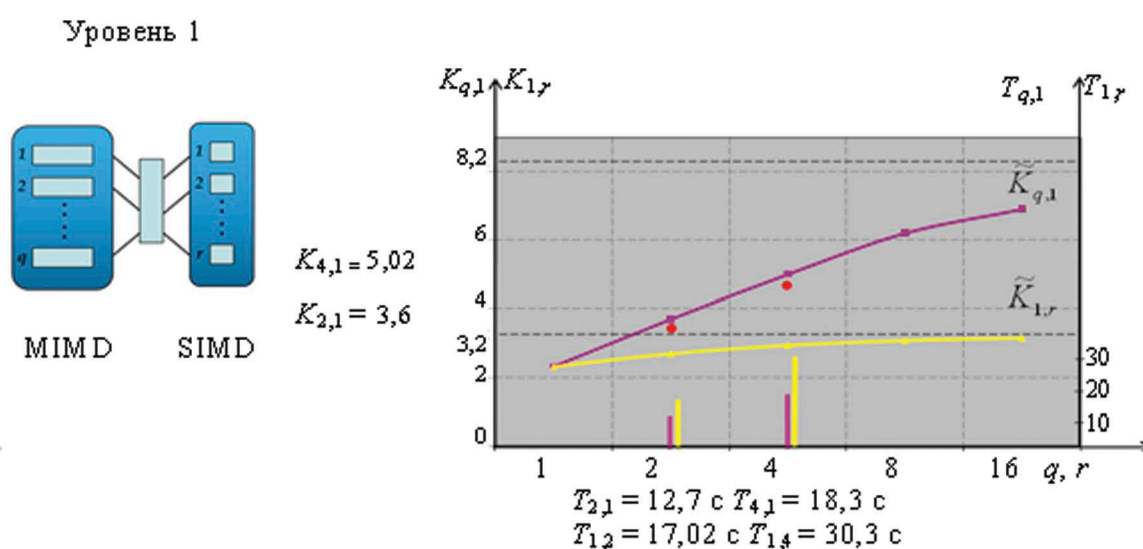


Рис. 4. Значения коэффициентов ускорения гибридными элементами различных конфигураций

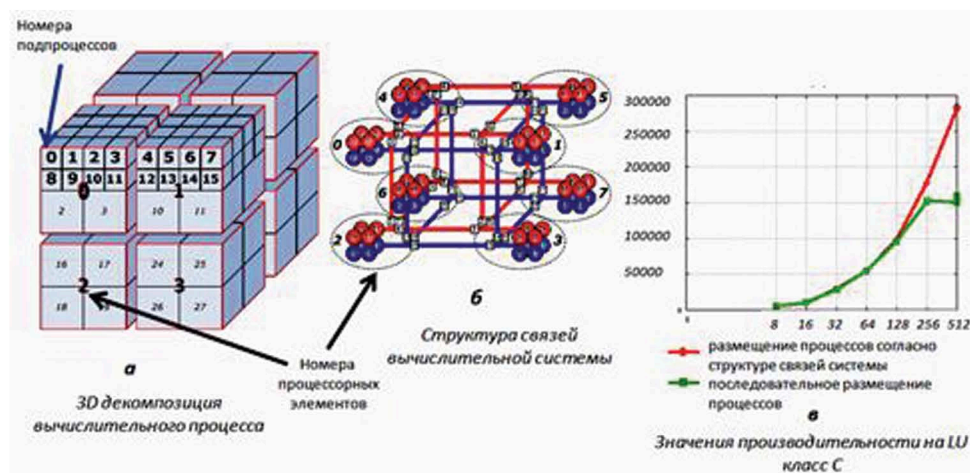


Рис. 5. Структура системы и значения производительности:
 а – 3D декомпозиция вычислительного процесса; б – структура связей вычислительной системы; в – значения производительности на LU, класс C

