

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

УПРАВЛЕНИЕ БОЛЬШИМИ СИСТЕМАМИ

***Выпуск 70
Ноябрь 2017***

**СБОРНИК
ТРУДОВ**

ISSN 1819-2467

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

Москва – 2017

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

**УПРАВЛЕНИЕ
БОЛЬШИМИ
СИСТЕМАМИ**

СБОРНИК ТРУДОВ

Выпуск 70

Москва – 2017

КООРДИНАЦИОННЫЙ СОВЕТ

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

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

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

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

РЕГИОНАЛЬНЫЕ РЕДАКЦИОННЫЕ СОВЕТЫ

Арзамас – д-р физ.-мат. наук проф. Пакшин П.В. **Волгоград** – д-ра физ.-мат. наук: проф. Воронин А.А., проф. Лосев А.Г. (ВолГУ); **Воронеж** – д-р техн. наук, проф. Баркалов С.А., д-р физ.-мат. наук, проф. Головинский П.А. (ВГАСУ), д-р техн. наук, проф. Подвальный С.Л. (ВГТУ); **Иркутск** – академик РАН Бычков И.В., д-р физ.-мат. наук, проф. Лакеев А.В. (ИДСТУ СО РАН); **Казань** – д-р физ.-мат. наук, проф. Маликов А.И., д-р техн. наук, проф. Сиразетдинов Р.Т. (КГТУ-КАИ); **Липецк** – д-ра техн. наук: проф. Погодаев А.К., Сараев П.В. (ЛГТУ); **Самара** – д-ра экон. наук: проф. Богатырев В.Д., проф. Гераськин М.И., д-р техн. наук, проф. Засканов В.Г. (СГАУ); **Петрозаводск** – д-р физ.-мат. наук, проф. Мазалов В.В., д-р техн. наук, доц. Печников А.А. (ИПМИ КарНЦ РАН); **Санкт-Петербург** – д-р физ.-мат. наук: проф. Петросян Л.А. (СПбГУ), д-р техн. наук проф. Фуртат И.Б. (ИПМ РАН); **Старый Оскол** – д-р техн. наук, проф. Еременко Ю.И. (СТИ).

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

Адрес в интернете: ubs.mtas.ru.

СОДЕРЖАНИЕ

Системный анализ

Бочаров П.С., Горяшко А.П.

*О субоптимальных решениях
антагонистических игр разбиений.....* 6

Бакаев А.М.

*Определение сложности задач
для зрительно-пространственной памяти
и пропускной способности человека-оператора* 25

Сетевые модели в управлении

Мелентьев В.А.

*О топологической отказоустойчивости
масштабируемых вычислительных систем* 58

Управление в социально-экономических системах

Ростова Е.П., Гераськин М.И.

*Оптимальная функция издержек предотвращения
промышленных рисков фирмы.....* 87

***Управление в медико-биологических
и экологических системах***

Лосев А.Г., Левшинский В.В.

*Интеллектуальный анализ термометрических
данных в диагностике молочных желез 113*

Управление подвижными объектами и навигация

Кузнецов А.В.

*Организация строя агентов с помощью
клеточного автомата 136*

СУБОПТИМАЛЬНЫЕ РЕШЕНИЯ АНТАГОНИСТИЧЕСКИХ ИГР РАЗБИЕНИЙ¹

Бочаров П. С.²

(ООО Вили Разработка, Москва)

Горяшко А. П.³

(Московский технологический институт, Москва)

Представлены экспериментальные результаты изучения субоптимальных решений в смешанных стратегиях антагонистических игр разбиений. Показано, что множество чистых стратегий ε -оптимального алгоритма обладает специальной структурой распределения ресурсов и эта структура может быть легко воспроизведена на базе ранее изученных свойств разбиений. Результаты подтверждены численным сравнением с поведением ε -оптимального алгоритма.

Ключевые слова: разбиение целого числа, композиция целого числа, теория игр, игры Блотто, игры Лотто, ε -оптимальное решение игры, цена игры, собственный ресурс разбиения, уравновешенность разбиения.

1. Введение

Изучение антагонистических игр размещения ресурсов, начало которому было положено классической работой Э. Бореля [4], в последние пару десятилетий велось достаточно интенсивно. Причиной явилось, в частности, то, что несмотря на

¹ Авторы признательны А.С. Немировскому за многочисленные консультации и помощь в проведении экспериментов.

² Павел Сергеевич Бочаров, технический директор (pavel@wheely.com).

³ Александр Петрович Горяшко, доктор технических наук, профессор (petrovich4you@gmail.com).

крайне простое описание этого класса антагонистических игр (называемых обычно играми Блотто и Лотто с добавлением различных воинских чинов), они ставят перед исследователями серьезные проблемы. Прежде всего в связи с астрономически большим числом возможных стратегий игры даже при относительно небольших значениях исходных параметров: целого числа возможных полей (m) и целого числа размещаемых по этим полям ресурсов (n). Из работ по теории разбиений (см., например, [2]) следует, что игры такого класса могут быть рассмотрены на языке комбинаторных алгоритмов как игры разбиений и игры композиций.

Ранее было установлено [8], что оптимальное решение игр Блотто возможно только в смешанных стратегиях⁴, а в [12] для Нэш-равновесия найдена функция распределения ресурсов (как для случая равных, так и для случая неравных ресурсов игроков)⁵. В частности, для Нэш-равновесия и случая, когда для обоих игроков A и B ресурс n распределяется по m полям, одномерная функция распределения ресурсов для обоих игроков выглядит следующим образом: для каждого поля $j \in 1, \dots, m$ $F_j(x) = xm/2n$, где $x \in [0, 2n/m]$. Очевидно, указанный способ задания support-множества⁶ мало обозрим.

1.1. ИГРЫ РАЗМЕЩЕНИЯ РЕСУРСОВ

Из теоремы Каратеодори [7] следует, что существует оптимальное решение в смешанных стратегиях для задач того класса, к которым принадлежит игра полковника Блотто, с числом чистых стратегий не больше $(n+1)t+1$. Таким образом, например,

⁴ При $n = t$ существует оптимальное решение в чистых стратегиях, а именно $(1, 1, \dots, 1)$. Однако уже при $n \geq 2t$ решений в чистых стратегиях нет.

⁵ В [12] использована теория копул для нахождения смешанного равновесия игры при общих условиях на ресурсы игроков.

⁶ В теории игр support-множествами принято называть множества стратегий с ненулевой вероятностью для ситуаций равновесия в смешанных стратегиях.

для задачи полковника Блотто $(120, 6)$ ⁷ существует оптимальное решение, которое содержит не более 727 чистых стратегий.

Хотя результат К. Каратеодори известен уже 105 лет, до недавних пор, насколько нам известно, даже не предпринималось попыток представить оптимальное решение матричной антагонистической игры с постоянной суммой, в котором число стратегий было близким к оценке, установленной теоремой Каратеодори.

Только недавно в [6] предложен метод декомпозиций для линейных минимизационных оракулов, который позволяет находить ε -оптимальные решения широкого класса матричных задач с так называемыми «хорошо организованными»⁸ матрицами. В частности, требуемой структурой обладает платежная матрица игры, названной в [6] *attacker vs defender* и являющейся по существу расширением матричной игры полковника Блотто (полям, в которых распределяются целые значения ресурсов, могут быть приписаны различные веса). Хотя платежная матрица таких игр даже при небольших значениях n и m содержит сотни миллионов столбцов и строк, предложенный метод гарантирует нахождение ε -оптимального решения за полиномиальное время по n и m . Следует отметить, что алгоритм позволяет находить решение с полиномиальными оценками по n , m и $1/\varepsilon$ или по n , m , $\log 1/\varepsilon$, в зависимости от применяемого варианта алгоритма [6].

Например, программная реализация этого метода позволяет находить ε -оптимальное решение (при $\varepsilon \leq 0,02$) для игры Блотто $(120, 6)$ примерно за 20 минут машинного времени на среднем по производительности ноутбуке, причем количество стратегий в support-множествах в разы меньше верхней границы, установленной теоремой Каратеодори. При увеличении числа полей предлагаемый метод остается вполне доступным, хотя требуемое время

⁷ Здесь и далее в тексте используется обозначение (n, m) , где n – суммарное количество ресурсов, которое игрокам необходимо разместить по m участкам.

⁸ В [6] хорошо организованной называется такая $K \times L$ матрица B , что при любом $x \in \mathbb{R}^K$ можно легко найти максимальное и минимальное значение скалярного произведения столбцов матрицы $B[x]$.

решения может увеличиваться в разы.

Возможность численного анализа ε -оптимальных решений, состоящих не более чем из нескольких сотен чистых стратегий, позволяет понять некоторые особенности структуры стратегий, входящих в support-множества класса дискретных игр полковника Блотто и полковника Лотто. А использование в виде эталона тех смешанных стратегий, которые предоставляет метод [6], позволяет оценить качество субоптимальных решений, полученных случайным выбором из подмножества разбиений с заранее выбранными параметрами уравновешенности и собственного ресурса [1]. Таким образом, центр тяжести решения задачи нахождения цены антагонистической матричной игры с постоянной суммой в смешанных стратегиях для класса игр полковника Блотто и полковника Лотто может быть перенесен на поиск таких свойств разбиений, которые позволяют эффективно отобрать относительно малое число стратегий для включения в support-множества. (Под «эффективностью» здесь следует понимать нахождение цены игры с точностью порядка процента.)

1.2. ИГРА ЛОТТО И ИГРА РАЗБИЕНИЙ

Большинство теоретических и экспериментальных исследования игр размещения ресурсов относятся к играм композиций (играм полковника Блотто). И хотя при фиксированных значениях n и m число композиций (а значит, и размерность платежной матрицы) примерно в $m!$ раз больше, чем число разбиений, анализ структуры оптимальных решений в смешанных стратегиях достаточно прост, хотя и не претендует на конструктивность.

В [9] было введено понятие игры Лотто. В игре полковника Лотто m полей, на которых игроки распределяют ресурсы, для них неразличимые. Если в игре полковника Блотто $B(n, m)$ платежная функция $H_B(\alpha, \beta)$ для игроков $\alpha = (a_1, \dots, a_m)$ и $\beta = (b_1, \dots, b_m)$ вводится как

$$(1) \quad H_B(\alpha, \beta) = 1/m \sum_{i=1}^m \text{sign}(a_i - b_i),$$

то платежная функция $H_L(\alpha, \beta)$ для игры Лотто $L(n, m)$ вводится

в [9] как

$$(2) \quad H_L(\alpha, \beta) = 1/m^2 \sum_{i=1}^m \sum_{j=1}^m \text{sign}(a_i - b_j).$$

Платежную функцию $H_P(\alpha, \beta)$ игры разбиений $P(n, m)$ можно ввести как

$$(3) \quad \begin{aligned} H_P(\alpha, \beta) &= \frac{1}{|\Theta(\alpha)|} \sum_{\theta \in \Theta(\alpha)} H_B(\alpha, \beta) = \\ &= \frac{1}{m|\Theta(\alpha)|} \sum_{\theta \in \Theta(\alpha)} \sum_{i=1}^m \text{sign}(c_i - b_i), \end{aligned}$$

где $\Theta(\alpha)$ – множество всех перестановок α вида $\theta = (c_1, \dots, c_m)$.

В [9] показано, что игры полковника Блотто $B(n, m)$ и полковника Лотто $L(n, m)$ имеют одну и ту же цену, когда все перестановки берутся с равными вероятностями. Это замечание остается справедливым и при определении цены игры разбиений (3).

Для игры разбиений точный ответ на вопрос «какая из пары заданных стратегий выигрывает» требует вычислительных ресурсов, экспоненциально растущих с ростом числа полей. Потому можно было ожидать, что общее решение игры разбиений окажется более сложным, нежели игры композиций. Например, в диапазоне (n, m) -разбиений от $(4, 2)$ до $(64, 8)$ существует алгоритм сложностью не более $const \cdot m^2$, который позволяет определить победителя с погрешностью, не превосходящей $0,02$.

Практический интерес к изучению игры Лотто связан с тем, что антагонистическая игра, результат которой не должен зависеть от *перестановок* по полям выбранного множества ресурсов, – случай столь же распространенный, как и ситуации с фиксированной нумерацией полей (игра полковника Блотто). В качестве примеров можно указать работы, в которых изучается проведение турниров [10] или аукционов (особенно в случаях приобретения набора предметов).

В то время как все известные нам методы решения общих случаев игр полковника Блотто и полковника Лотто носят характер доказательства «теорем существования», метод [6] примени-

тельно к игре полковника Блотто (точнее её расширенному варианту «attacker vs defender») позволяет получать приемлемые по затратам трудоемкости и точности численные решения – support-множества из небольшого числа чистых стратегий, которые обеспечивают точность нахождения цены игры в пределах 0,02. (Численно проверено для игр с параметрами $m \leq 10$, $n \leq 120$, для которых мощность support-множеств оказалась в разы ниже границы Каратеодори).

Однако метод [6], впрочем, как и любые другие вычислительные методы оптимизации, не позволяет априори выяснить, какие из чистых стратегий войдут в искомое support-множество. Целью настоящей работы является формулировка способа нахождения «достаточно хорошего» support-множества (близкого по параметрам к ε -оптимальному решению [6]), без решения соответствующей оптимизационной проблемы.

Дальнейшее изложение построено следующим образом. В разделе 2 рассмотрены численные результаты, полученные при синтезе ε -оптимальных решений игр полковника Блотто $B(120, 6)$ и $B(100, 10)$ методом [6]. В разделе 3 продемонстрированы результаты синтеза support-множеств для различных игр разбиений, полученные при выборе чистых стратегий из определенного класса разбиений. В заключении обсуждаются возможности предлагаемых субоптимальных методов и способы уточнения полученных результатов.

2. Анализ ε -оптимальных решений

Существовавшие до сих пор теоретические методы нахождения оптимальных решений игр типа игры полковника Блотто и полковника Лотто фактически не уделяли внимания конкретной структуре стратегий, входящих в support-множества. Оставались также неисследованными вопросы, относящиеся к зависимости «эффективности» чистых стратегий в игре разбиений от величины отношения n/m (за исключением асимптотических оценок в [13], которые вообще говоря никак не связаны с игровыми постановками).

Экспериментальные исследования (например, [3]) вообще не претендовали на исследование оптимальных решений, а лишь анализировали эвристические приемы «стратегических рассуждений», которые, по предположению, позволяли выбирать «конкурентоспособные» стратегии. Вопрос о том, насколько объективно хороши результаты подобных «стратегических рассуждений» оставался открытым.

В экспериментальной части нашей работы исследовалось четыре класса игр разбиений: (36, 6), (49, 7), (120, 6) и (100, 10). Во всех классах $n = m^2$. Далее, чтобы не загромождать изложение, приведены результаты только для классов (120, 6) и (100, 10).

Вначале для игры attacker vs defender были получены ε -оптимальные решения в смешанных стратегиях. Количество чистых стратегий в support-множествах таких решений не превосходило 230, а гарантированная погрешность цены игры составляла около 2%. (В приложении 5.2 приведены первые десять смешанных стратегий для обоих игроков в порядке убывания их вероятностей в support-множестве для игр $B(120, 6)$ и $B(100, 10)$.)

Представив все чистые стратегии ε -оптимальных решений в виде разбиений, упорядоченных по убыванию, можно заметить, что ресурс, размещенный в первой половине разбиения, часто *втрое* больше ресурса второй половины. Естественно попытаться выяснить, является ли неравномерность размещения ресурсов в ε -оптимальных support-множествах фактором, способствующим попаданию разбиения в ε -оптимальное решение.

Пусть множество $\Psi_{n,m}$ содержит некоторое число (n, m) -разбиений вида $\alpha = (a_1, \dots, a_m)$, где $a_j \geq a_{j+1}$, $j = 1, \dots, m - 1$ (везде далее m – четное).

Для любого разбиения $\alpha \in \Psi_{n,m}$ вычислим значение функции $Q(\alpha)$:

$$(4) \quad Q(\alpha) = \frac{S_1(\alpha) - S_2(\alpha)}{n} = \frac{2S_1(\alpha)}{n} - 1,$$

где $S_1(\alpha) = \sum_{j=1}^{m/2} a_j$, $S_2(\alpha) = \sum_{j=m/2+1}^m a_j$.

Для любого разбиения $\alpha \in \Psi_{n,m}$ справедливы границы

$$(5) \quad 0 \leq Q(\alpha) \leq 1,$$

где $Q(\alpha) = 0$ для разбиения $\alpha = (\frac{n}{m}, \frac{n}{m}, \dots, \frac{n}{m})$ и $Q(\alpha) = 1$ для разбиения $(n, 0, \dots, 0)$.

При достаточно больших значениях параметров n и m для значений $a_i, i = 1, \dots, m$, справедливо $a_i \approx n/m \ln(m/i)$ [13]. Пусть $\Psi_{n,m}^*$ – множество (n, m) -разбиений, для которых справедлива эта оценка. Тогда для любого $\alpha \in \Psi_{n,m}^*$

$$(6) \quad S_1(\alpha \in \Psi_{n,m}^*) = \sum_{i=1}^{m/2} \frac{n}{m} \ln \frac{n}{m} = \frac{n}{m} \ln \prod_{i=1}^{m/2} \frac{m}{i} = \frac{n}{m} \ln \frac{m^{\frac{m}{2}}}{(\frac{m}{2})!}.$$

Используя приближение Стирлинга, из (6) получаем

$$(7) \quad S_1(\alpha \in \Psi_{n,m}^*) \approx \frac{n}{m} \left(\frac{m}{2} \ln 2e - \frac{1}{2} \ln \pi m \right) = \frac{n}{2} \left(\ln 2 + 1 - \frac{\ln \pi m}{m} \right).$$

Из (4) и (7) следует

$$(8) \quad Q(\alpha \in \Psi_{n,m}^*) \approx \ln 2 - \frac{\ln \pi m}{m}.$$

Таким образом, при достаточно больших значениях n и m доля ресурсов, приходящихся на первую половину почти всех (n, m) -разбиений, близка к $\ln 2$.

На рис. 1 показано распределение величины $Q(\alpha \in \Omega_{100,10})$, где $\Omega_{100,10}$ – это множество всех $(100, 10)$ -разбиений. Легко заметить, что максимум распределения приходится на величину, близкую к $\ln 2 \approx 0,693$.

Как и следовало ожидать, при относительно небольших значениях n и m максимум распределения близок к $\ln 2$. В таблице 1 приведены значения этого максимума для различных величин отношения m^2/n .

В таблице 1 последний столбец – свидетельство того, что в изученных случаях при $n = m^2$ отклонение максимума распределения от $\ln 2$ близко к нулю, но заметно увеличивается, если $n < m^2$, т.е. случай $n = m^2$ дает оценку, практически совпадающую с асимптотической.

Отметим еще одну особенность, присущую всем разбиениям, входящим во множество рассмотренных в экспериментах

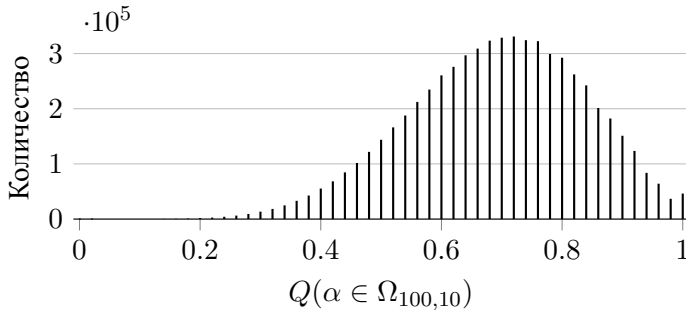


Рис. 1. Число $(100, 10)$ -разбиений в зависимости от величины $Q(\alpha \in \Omega_{100,10})$

ε -оптимальных решений. Она состоит в том, что среднее значение собственного ресурса (см. [1]) этих стратегий близко к максимальному, т.е. чаще всего в этих стратегиях мало повторяющихся значений (разбиение вполне упорядочено).

Таблица 1. Значение максимума распределения $Q(\alpha \in \Omega_{100,10})$ в зависимости от n, m

n	m	m^2/n	$M[Q(\alpha \in \Omega_{100,10})]$	$ M - \ln 2 $
90	16	2,8	0,764	0,071
100	10	1	0,691	0,002
120	6	0,3	0,632	0,061
120	20	3,3	0,780	0,087
144	12	1	0,698	0,005

Пусть $PR_{n,m} \subset \Omega_{n,m}$ – множество вполне упорядоченных по убыванию разбиений, т.е. всех тех разбиений $\alpha = (a_1, \dots, a_m)$, для которых значение собственного ресурса максимально, причем n/m – целое, n, m – четные и $a_m > 0$. Справедливо⁹

Утверждение 1. Для всех $\alpha \in PR_{n,m}$ выполняются следующие условия:

- 1) $\max \sum_{i=1}^{m/2} a_i \leq n - (\frac{m}{2} + 1)(\frac{m}{4})$;
- 2) $\min \sum_{i=1}^{m/2} a_i \geq \frac{n}{2} + m(m - 2)/8$.

⁹Доказательство утверждения приведено в приложении 5.1.

Пример 1. Для множества разбиений $(120, 6)$ единственными разбиениями с минимальным значением второй половины будет $(39, 38, 37, 3, 2, 1)$ и с максимальным – $(23, 22, 21, 19, 18, 17)$.

Таким образом, для разбиений из класса $PR_{n,m}$ величину функционала $Q(PR_{n,m})$ можно оценить как

$$(9) \quad \frac{2m}{n} \leq Q(PR_{n,m}) \leq 1 - \frac{m^2}{4n} - \frac{m}{2n}.$$

В случае когда $n = m^2$, из (9) получаем

$$(10) \quad 1/8 \leq Q(PR_{n,m}) \leq 2/3,$$

т.е. диапазон возможных значений функционала $Q(PR_{n,m})$ становится существенно меньше, чем в общем случае (5).

3. Синтез субоптимальных решений для игр разбиений: результаты экспериментов

Основной прием, использованный в экспериментальной части работы, состоял в следующем: вначале для заданного класса разбиений вычислялось ε -оптимальное решение задачи attacker vs defender. Затем рассматривалось «синтетическое» решение – набор чистых стратегий, выбранных из всего множества допустимых стратегий по определенным правилам. Проверка качества такого решения состояла в вычислении результатов игры этих двух наборов стратегий. Близость к нулю цены игры предполагалась свидетельством в пользу «близости» ε -оптимального и «синтетического» решений.

Назовем ε -оптимальное решение игроком A и «синтетическое» – игроком C . Эксперимент (игра между A и C) состоял из следующих шагов.

1. Зададим одни и те же параметры n, m для A и C и обозначим через N_A, N_C числа чистых стратегий A и C соответственно.

2. Из ε -оптимального support-множества $E_{n,m}$, упорядоченного по вероятностям, для игрока A выбираем первые N_A чистых стратегий.

3. Для игрока C выберем из множества разбиений с максимальным значением собственного ресурса $PR_{n,m}$ случайно и

равномерно подмножество мощности N_C с некоторым значением уравновешенности pb .

4. Для игроков, полученных на шаге 2 и 3, находим значение цены, решая матричную подигру на этих стратегиях с помощью программы выпуклой оптимизации с использованием библиотеки CVXPY [5].

5. Повторяем эксперимент I раз и вычисляем среднее значение цены игры.

Поскольку уже было обосновано решение выбирать чистые стратегии только среди разбиений с максимальным значением собственного ресурса $PR_{n,m}$, в экспериментах, помимо числа итераций I , остается только три свободных параметра: N , M и pb . Эти значения выбирались в процессе компьютерного моделирования. Так, например, были получены наиболее эффективные значения уравновешенности pb в диапазоне $[-0,85; 0]$ для множества разбиений $(120, 6)$ и pb в диапазоне $[-1,55; 0]$ для множества разбиений $(100, 10)$. В качестве примера на рис. 2 показаны результаты цены игры A и C для случая разбиений $n = 100$, $m = 10$ и $n = 120$, $m = 6$ в зависимости от числа стратегий (все результаты получены при значениях $I = 200$).

Число стратегий N_A в support-множестве для $(120, 6)$ было равно 100 (суммарная вероятность 0,8 для одного игрока и 0,76 для другого) и 200 (суммарная вероятность 0,99976 для одного и 0,99971 для другого) и в support-множестве для $(100, 10)$ – 100 (суммарная вероятность 0,849) и 194 (суммарная вероятность 1).

Таблица 2. Зависимость цены игры при различных значениях числа стратегий из ε -оптимального решения A и синтетического решения C

N_C	(120, 6)		(100, 10)	
	$N_A = 100$	$N_A = 200$	$N_A = 100$	$N_A = 194$
100	0,032	0,053	0,020	0,080
200	0,000	0,006	0,019	0,030
300	-0,003	-0,012	0,000	0,011
400	-0,028	-0,023	-0,001	0,004

Результаты проведенных экспериментов позволяют сделать следующий вывод.

При $N_C/N_A \geq 2$ цена игры A против C находится в диапазоне $[-0,03; 0,03]$, т.е. качество решения, которое обеспечивает «синтетический» игрок C в этих экспериментах, практически совпадает с качеством ε -оптимального решения.

Тем не менее утверждать, что игрок C «равносилен» игроку A , только на основании того, что они – предположим, даже в любом эксперименте – играют между собой вничью, было бы преждевременно. Можно допустить, что при встречах с другими игроками C и A будут показывать различные результаты. Поэтому был проведен цикл экспериментов, в которых эти два игрока (« ε -оптимальный» и «синтетический») – с фиксированными параметрами уравновешенности и собственной силы – играют по очереди с произвольным игроком T (тестовым).

Пусть игрок T – это множество чистых стратегий $T \subset \Omega_{n,m}$, которые случайно выбираются из $\Omega_{n,m}$ с определенными значениями собственного ресурса и уравновешенности. В проведенных экспериментах рассматривались четыре тестовых игрока: T_{200}^I , T_{400}^I , T_{100}^{II} и T_{200}^{II} (см. таблицу 3).

Таблица 3. Параметры тестовых игроков T

Тестовый игрок	Число стратегий	Собственный ресурс	Уравновешенность
T_{200}^I	200	Максимальный	$[-0,85; 0]$
T_{400}^I	400	Максимальный	$[-0,85; 0]$
T_{100}^{II}	100	[Макс./2, Макс.]	$[-0,85; 0]$
T_{200}^{II}	200	[Макс./2, Макс.]	$[-0,85; 0]$

В таблице 4 показаны значения цены игры для параметров $n = 120$, $m = 6$ игроков A и C с различными тестовыми игроками T (значение цены усреднялось по 200 экспериментам).

Полученные результаты позволяют сделать следующий вывод.

Наилучшие результаты синтетический игрок C показывает при числе стратегий, вдвое превышающих число стратегий ε -оптимального решения. В этом случае среднее значение абсолютной разницы в ценах игры этих игроков не превосходит 0,03, а средняя разница по всем экспериментам составляет 0,00518.

Таблица 4. Усредненные результаты игры « ε -оптимального» игрока A и «синтетического» игрока C с тестовыми игроками T

T	Число стратегий		Цена игры			
	A	C	A против T	C против T	Разница	
T_{200}^I	100	100	0,020	0,052	0,030	
		200	0,059	0,054	0,005	
		400	0,084	0,047	0,027	
	200	100	0,015	0,070	-0,058	
		200	0,059	0,054	0,005	
		400	0,057	0,076	-0,018	
T_{400}^I	100	100	-0,034	0,009	-0,040	
		200	0,009	0,003	0,006	
		400	0,033	0,002	0,030	
	200	100	-0,028	0,026	-0,055	
		200	0,009	0,029	-0,020	
		400	0,034	0,031	0,003	
T_{100}^{II}	100	100	0,342	0,262	-0,080	
		200	0,336	0,318	-0,018	
	200	200	0,410	0,325	-0,080	
		400	0,362	0,364	0,002	
	T_{200}^{II}	100	100	0,188	0,139	-0,048
			200	0,177	0,176	-0,001
200		200	0,228	0,182	-0,046	
		400	0,212	0,215	0,002	

4. Заключение

Практическое использование моделей антагонистических матричных игр трудно осуществимо без наличия относительно простых методов получения решений в смешанных стратегиях небольшой мощности, пусть и за счет (заранее известной) потери в точности результата. Поэтому методы, предложенные в [6], представляются серьезным продвижением в этом направлении. В данной работе мы попытались сделать следующий шаг в направлении разработки простых субоптимальных методов решения антагонистических матричных игр. Несомненно, предлагаемый метод требует не только экспериментального подтверждения, но и теоретических доказательств его эффективности, что и является нашей ближайшей целью.

Безусловный интерес вызывает также получение приближенных оценок цены игры в случаях несимметричных ресурсов противников и сравнение этих результатов с известными.

Одна из интересных на наш взгляд возможностей применения рассмотренного метода – экспериментальное изучение игр размещения ресурсов в рамках поведенческой теории игр [3, 10]. Объяснение принципов принятия «стратегических решений» на основе психологических теорий типа «глубины рассуждений» [3] позволяет в основном отделить совсем тривиальные рассуждения лиц, принимающих решения, от чуть более сложных. Но когда дело доходит до поиска нетривиальных особенностей изучаемого объекта, оценка «глубины рассуждений» оказывается мало информативной. Необходимо оценивать *качественные* различия в содержании рассуждений.

Заметим, что участникам было бы логично рассматривать турнир игры полковника Блотто [3] именно как турнир *разбитий*, поскольку при большом числе участников, о возможных стратегиях которых нет никакой информации, естественно пытаться найти «сильное» решение для случая *ненумерованных* полей. Участник, пришедший к подобному выводу, сделал бы главный шаг к победе в турнире. Для этого ему было достаточно в

любом варианте, предложенном десятью победителями, исключить повторяющиеся значения полей с *большими значениями ресурсов*¹⁰. Например, вместо (32, 32, 31) (первое место в турнире) играть (33, 32, 30) и уверенно выиграть у всех разбиений первой десятки. К сожалению, такого рода наблюдения – если испытуемые не обладают некоторыми (врожденными или благоприобретенными) способностями – требуют специального обучения (желательно с поощрением), что исключено в условиях one-shot игры. Представляется важным и интересным понять, можно ли обучить людей принятию *нетривиальных* решений и, если «да», то как.

Методы, рассмотренные в работе, могут быть, в частности, использованы в таких исследованиях, как [11], когда необходимо вычисление равновесных решений в игре полковника Блотто, поскольку их использование могло бы помочь изучить случаи гораздо большей размерности.

5. Приложение

5.1. ДОКАЗАТЕЛЬСТВО УТВЕРЖДЕНИЯ 1

1. Оценим величину второй половины разбиения. Очевидно, вторая половина разбиения минимальна, когда является убывающей последовательностью натуральных чисел $(\frac{m}{2}, \frac{m}{2} - 1, \dots, 1)$. Отсюда несложно установить, что $\max \sum_{i=1}^{m/2} a_i = n - \frac{((m/2)+1)m}{4}$.

2. Для того чтобы оценить снизу величину первой половины разбиения, достаточно рассмотреть такие разбиения, в первой половине которых значение каждой части $a_i = \frac{n}{m} + \frac{m}{2} - i$ ($i = 1, 2, \dots, \frac{m}{2}$), т.е. минимально увеличивать каждую часть первой половины по сравнению с равномерным разбиением. В этом случае $\sum_{i=1}^{m/2} a_i = \frac{n}{m} \cdot \frac{m}{2} + \sum_{i=1}^{m/2} (i-1) = \frac{n}{2} + \frac{(m/2)(m/2-1)}{2}$. Отсюда получаем, что $\min \sum_{i=1}^{m/2} a_i \geq \frac{n}{2} + \frac{m(m-2)}{8}$.

¹⁰ Это замечание следует из того простого обстоятельства, что в игре разбиений любое подмножество полей, например, a_1, a_2, a_3 , выигрывает у подмножества полей b_1, b_2, b_3 с таким же суммарным ресурсом, если $a_1 = b_1 + 1, a_2 = b_2, a_3 = b_3 - 1$.

5.2. ε -ОПТИМАЛЬНЫЕ РЕШЕНИЯ ИГРЫ БЛОТТОТаблица 5. Игра Блотто $B(120, 6)$, A против D

#	A, 227 стратегий		D, 229 стратегий	
	Вероятн.	Разбиение	Вероятн.	Разбиение
1	0,0182	(22, 0, 39, 21, 29, 9)	0,0167	(2, 27, 20, 3, 39, 29)
2	0,0171	(9, 37, 20, 34, 7, 13)	0,0165	(30, 28, 34, 7, 11, 10)
3	0,0170	(5, 40, 25, 40, 5, 5)	0,0165	(13, 8, 23, 28, 15, 33)
4	0,0149	(36, 30, 7, 4, 18, 25)	0,0153	(23, 15, 17, 15, 24, 26)
5	0,0135	(37, 17, 32, 28, 3, 3)	0,0138	(14, 10, 39, 39, 13, 5)
6	0,0131	(14, 16, 6, 16, 37, 31)	0,0131	(5, 3, 24, 35, 21, 32)
7	0,0129	(25, 29, 5, 20, 19, 22)	0,0128	(6, 21, 40, 37, 3, 13)
8	0,0126	(13, 19, 39, 14, 15, 20)	0,0127	(8, 5, 14, 40, 34, 19)
9	0,0126	(40, 10, 8, 5, 39, 18)	0,0124	(34, 32, 5, 6, 40, 3)
10	0,0117	(17, 5, 16, 36, 40, 6)	0,0119	(35, 10, 25, 17, 12, 21)
...
Гарантированный результат				
-0,02587182			0,02040023	

Таблица 6. Игра Блотто $B(100, 10)$, A против D

#	A, 185 стратегий		D, 182 стратегий	
	Вероятн.	Разбиение	Вероятн.	Разбиение
1	0,0222	(5, 4, 12, 2, 15, 8, 15, 13, 11, 15)	0,0232	(8, 18, 20, 15, 7, 4, 9, 15, 1, 3)
2	0,0173	(3, 8, 8, 20, 10, 9, 9, 17, 4, 12)	0,0211	(9, 2, 15, 18, 20, 8, 2, 9, 13, 4)
3	0,0168	(16, 19, 14, 0, 14, 13, 3, 2, 0, 19)	0,0184	(7, 14, 18, 14, 3, 1, 12, 14, 10, 7)
4	0,0160	(0, 10, 19, 1, 1, 20, 10, 18, 1, 20)	0,0180	(17, 4, 4, 17, 10, 15, 6, 10, 15, 2)
5	0,0159	(20, 13, 7, 3, 2, 2, 13, 4, 18, 18)	0,0173	(15, 11, 3, 20, 10, 9, 4, 8, 7, 13)
6	0,0156	(15, 19, 11, 15, 13, 3, 7, 0, 15, 2)	0,0167	(1, 13, 7, 4, 17, 20, 8, 13, 12, 5)
7	0,0150	(1, 20, 2, 8, 12, 18, 8, 8, 8, 15)	0,0148	(3, 8, 16, 0, 20, 0, 18, 12, 8, 15)
8	0,0149	(9, 8, 3, 6, 13, 17, 19, 14, 5, 6)	0,0145	(10, 11, 17, 19, 0, 13, 18, 4, 0, 8)
9	0,0148	(7, 20, 7, 11, 5, 7, 0, 15, 11, 17)	0,0141	(13, 9, 6, 9, 11, 14, 17, 0, 2, 19)
10	0,0148	(9, 15, 5, 5, 20, 19, 11, 10, 3, 3)	0,0135	(19, 7, 17, 9, 18, 7, 0, 9, 3, 11)
...
Гарантированный результат				
-0,05591			0,06349	

Благодарности

Авторы признательны рецензентам за чрезвычайно полезные замечания и комментарии.

Литература

1. БОЧАРОВ П.С., ГОРЯШКО А.П. *О способах анализа игр разбиений* // Управление большими системами. – 2016. – №61. – С. 6–40.
2. КНУТ Д.Э. *Искусство программирования, том 4, А. Комбинаторные алгоритмы, часть 1*: Пер. с англ. – М.: ООО «И.Д. Вильямс», 2013.
3. ARAD A., RUBINSTEIN A. *Multi-dimensional iterative reasoning in action: The case of the Colonel Blotto game* // J. of Economic Behavior & Organization. – 2012. – No. 84. – P. 571–585.
4. BOREL E. *La theorie du jeu et les equations integrales a noyau symetrique* // Comptes Rendus de l'Academie. – 1921. – No. 173. – P. 1304–1308. English: Transl. by L.J. Savage *The theory of play and integral equations with skew symmetric kernels* // Econometrica. – 1953. – No. 21. – P. 97–100.
5. DIAMOND S., BOYD S. *CVXPY: A Python-Embedded Modeling Language for Convex Optimization* // J. of Machine Learning Research. – 2016. – Vol. 17(83). – P. 1–5.
6. COX B., JUDITSKY A., NEMIROVSKI A. *Decomposition Techniques for Bilinear Saddle Point Problems and Variational Inequalities with Affine Monotone Operators* // J. of Optimization Theory and Applications. – 2015. – P. 1–34.
7. ECKHOFF J. *Helly, Radon, and Carathéodory type theorems* // Handbook of convex geometry. – 1993. – P. 389–448.
8. GROSS O., WAGNER R. *A Continuous Colonel Blotto Game* // RAND Corporation. – 1950.
9. HART S. *Discrete Colonel Blotto and General Lotto Games* // International J. of Game Theory. – 2008. – No. 36. – P. 441–460.
10. MOLDOVANU B., SELA A. *Contest Architecture* // J. of Economic Theory. – 2006. – No. 126(1). – P. 70–96.

11. MONTERO M., POSSAJENNIKOV A., SEFTON M., TUROCY T. *Majoritarian Contests with Asymmetric Battlefields: An Experiment* // Centre for decision Research & Experimental Economics Discussion. – Paper No. 2013-12.
12. ROBERSON B. *The Colonel Blotto Game* // Econ. Theory. – 2006. – No. 29. – P. 1–24.
13. VERSHIK A.M., YAKUBOVICH Y. *The Limit Shape and Fluctuations of Random Partitions of Naturals with Fixed Number of Summands* // Mosc. Math. – 2001. – Vol. 1(3) – P. 457–468.

SUBOPTIMAL SOLUTIONS OF ANTAGONISTIC PARTITION GAMES

Pavel Bocharov, Wheely, Moscow, CTO (pavel@wheely.com).

Alexander Goryashko, Moscow Technological Institute, Moscow, Doctor of Science, professor (petrovich4you@gmail.com).

Abstract: The paper examines (n, m) -partition games in order to develop tractable method of suboptimal decision of the resource allocation games such as colonel Blotto game or colonel Lotto game. The main goal is to develop tractable method for building suboptimal solution in mixed strategies for these games without solving the relevant optimization problem. The foundation of proposed method lies in the specific combinatorial properties of the (n, m) -partition games. It turned out that if for all game strategies the values of the balance and peculiar resource have the values from specific range it could be sufficient to get of suboptimal decisions of the games mentioned above. The proposed methods are based on both the analytical and numerical results, analytical partitions properties and numerical simulation results. The numerical simulation for the partition games $(120, 6)$ and $(100, 10)$ demonstrated that one could design of the ε -optimal decision where $\varepsilon \leq 0.02$. The support set of these decisions contain no more two hundred pure strategies and decisions complexity equals $\text{const} \cdot m^2$ for considered games. Results of the numerical simulation provide reasons to suppose that our approach is quite competitive with ε -optimal solution. The simplicity of our suboptimal solution method could be advantage in the behavioral game theory.

Keywords: integer partition, composition on integer, game theory, Blotto games, Lotto games, ε -optimal solution, value of game, peculiar resource, partition balance.

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

Поступила в редакцию 27.02.2017.

Дата опубликования 30.11.2017.

УДК 004.7-053.88:340.130.53
ББК 32.81

ОПРЕДЕЛЕНИЕ СЛОЖНОСТИ ЗАДАЧ ДЛЯ ЗРИТЕЛЬНО-ПРОСТРАНСТВЕННОЙ ПАМЯТИ И ПРОПУСКНОЙ СПОСОБНОСТИ ЧЕЛОВЕКА-ОПЕРАТОРА¹

Бакаев М. А.²

*(Новосибирский государственный технический
университет, Новосибирск)*

Разумникова О. М.³

*(Новосибирский государственный технический
университет, НИИ физиологии и фундаментальной
медицины, Новосибирск)*

Исследуется эффективность выполнения задач, возникающих в ходе человеко-машинного взаимодействия и требующих задействования зрительно-пространственной памяти. По аналогии с известным законом Фиттса и исходя из принципов восприятия гештальт-психологии предложена формулировка индекса сложности для таких задач, включающая в себя количество запоминаемых информационных единиц (фигур, посчитанных по правилам окрестности фон Неймана 1-го порядка), размер словаря отображаемых объектов и размер рабочего поля. На основе индекса сложности запоминания вычислена пропускная способность для участников специально проведённого экспериментального исследования.

¹ Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 16-37-00184 мол_а.

² Максим Александрович Бакаев, кандидат технических наук (maxis81@gmail.com).

³ Ольга Михайловна Разумникова, доктор биологических наук, доцент (razum@physiol.ru).

Ключевые слова: человеко-машинное взаимодействие, информационная сложность, когнитивная психология, гештальт-принципы, графический интерфейс пользователя.

1. Введение

С 1980-х годов наиболее популярным средством взаимодействия оператора с системами управления остаётся графический интерфейс, который, как правило, предусматривает прямую манипуляцию объектами и использование меню; задействует зрение (в меньшей степени слух) и моторную реакцию рук; широко использует графические элементы; позволяет применять такие устройства как монитор, мышь, сенсорный экран, клавиатура и др.; имеет слабо выраженную интеллектуальность и направлен на прямое взаимодействие с человеком. Разновидностью графического интерфейса является веб-интерфейс, основной особенностью которого является то, что он работает в специальной программе – браузере – функционирующей на компьютере пользователя, в то время как данные и логика веб-приложения размещаются на удалённом веб-сервере. Для современных информационных систем и АСУ всё более характерен пользовательский интерфейс, разработанный именно на веб-платформе.

Выражение «Пользователь всегда прав» стало необсуждаемым правилом для любого разработчика человеко-машинных интерфейсов и уже более десятка лет считается очевидным и не допускающим сомнений, а проектировщики интерфейсов и промышленные дизайнеры прилагают все усилия для создания у пользователя положительных впечатлений. Тем не менее, в задачах управления на первый план выходит достижение общей эффективности человеко-машинной системы, а не удовлетворенность пользователя. С 1970-х гг. в ходе активной разработки и внедрения АСУ в ряде ведущих стран мира специально обученный оператор или аналитик рассматривался как составная часть человеко-машинной системы управления, получая, обрабатывая и выдавая информацию так же, как любой другой её компонент. При этом отмечалось, что человек является неотъ-

емлемой частью такой системы, выполняя функции интегрирующего звена, придавая ей способность адаптироваться, возможность саморазвития [2, т. 1]. Тем не менее, при всех своих сильных сторонах человек определённо уступает машине в способностях по восприятию больших объемов информации, и отсутствие учёта данного аспекта ведёт к дорогостоящим ошибкам, возникающим по вине человека, особенно в системах управления [4].

Человек, как обработчик информации, может вводить (получать) и выводить информацию, причём для обоих процессов могут быть введены меры скорости (R_{in} и R_{out}) как производные от объемов информации по времени. Конечно, абсолютные значения скоростей зависят от единиц измерения, и это вносит определённую путаницу в сравнение оценок скоростей, полученных в различных исследованиях (привести все оценки к одним единицам измерения в общем случае представляется невозможным, так как зачастую отсутствуют точные данные об экспериментальных условиях). Относительно вывода информации считается, что максимальная скорость работы голосовых связок человека на продолжительном отрезке времени соответствует $R_{out} = 5$ слогов/с, что соответствует $R_{out} = 12,5$ звуков/с для английского языка (для других языков это значение может отличаться) [4]. Для случая, когда человек использует какой-либо инструмент, значения скорости вывода обычно представляются в битах или символах в секунду. Для рукописного текста (авторучка) верхняя граница равна примерно 3,5 символа/с, средняя скорость печати составляет около 6,7 символов/с, а её рекордные значения достигают 10 символов/с. В целом оценка максимальной скорости человека по выдаче информации составляет 40 бит/с. Значения для скорости приёма информации хоть и выше, но не принципиально. R_{in} для восприятия речи на слух оценивается в 3–4 слова/с, для внимательного чтения с пониманием – 2,5 слова/с, или 18 бит/с, для чтения вслух – 30 бит/с, для скорочтения (на русском языке) – до 11,7 слов/с или 80 бит/с. Большинство исследователей соглашаются, что хотя R_{in} для чтения может достигать 45 бит/с, возможности визуального

восприятия человека в длительном периоде скорее соответствуют лишь 8 бит/с [4].

Скорость может быть увеличена до определённой степени посредством использования многомерных сигналов вместо простых одномерных, – например, тон и громкость вместо просто громкости, или же размер, яркость и цвет вместо лишь одного изменяемого визуального стимула, но эффект не является аддитивным. Известно, что ещё одним фактором, повышающим скорость, является тренировка, которая позволяет опытным операторам воспринимать целевую информацию на более высоком уровне, имея возможность обрабатывать меньшее количество единиц информации. Например, в то время как пилотам-новичкам приходится проверять каждое из показаний приборов в отдельности, опытным требуется лишь общий взгляд на приборную панель, чтобы понять, всё ли в порядке с объектом управления (летательным аппаратом).

Это отчасти связано с тем, что оператор в системах управления может обрабатывать информацию без выраженного задействования памяти или выступать как «канал с памятью», что неизбежно для многих типов задач, – например, отмеченного выше внимательного чтения с пониманием, в ходе которого в оперативной (рабочей) памяти может удерживаться около 160 бит информации [2, т. 7]. Соответственно, общая пропускная способность в основном зависит от особенностей извлечения информации из памяти и во втором случае она обычно составляет в районе 2–4 бит/с, т.е. на порядок ниже, чем в первом [2, с. 34], [3]. Приведённый в [4, с. 201–206] сравнительный анализ значительного количества работ, исследовавших работу кратковременной памяти, свидетельствует, что в большинстве исследований скорость обработки оказывалась равной 3–5 ед./с (не бит, а именно разнородных единиц, в том числе фигур).

В хорошо известной с 1980-х гг. модели человека как обработчика информации [11] (см. рис. 1) работа когнитивной подсистемы, включающей оперативную память, долговременную память и когнитивный процессор, является наиболее сложной, и по сей день привлекает внимание исследователей. Для органов чувств давно известна экспериментальная статистика, для мо-

торной системы широко признанной моделью является закон Фиттса, но для когнитивного процессора существуют лишь примерные оценки, что цикл его работы занимает 25–170 мс [11]. При этом остаётся неясным, как определять количество таких циклов, в каких случаях и как идёт обмен информацией с памятью, как он различается для различных групп пользователей и контекстов использования, и т.п.

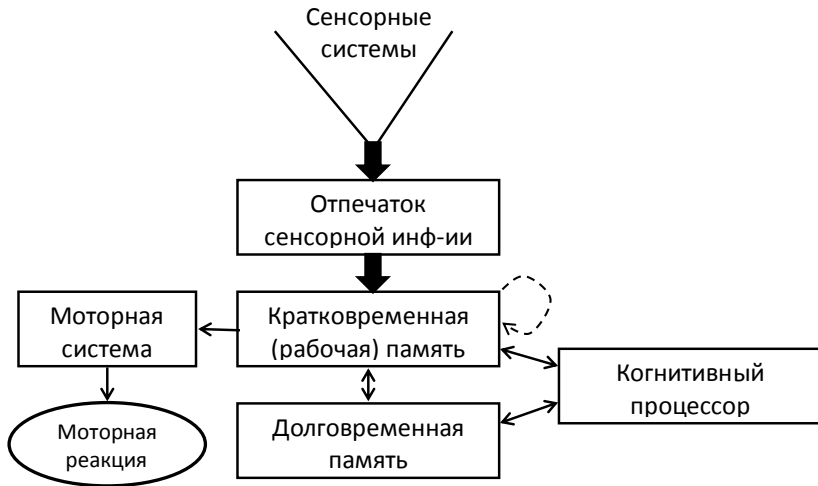


Рис. 1. Основные компоненты модели человека как обработчика информации

Выше было отмечено, что в современных графических интерфейсах основным каналом получения информации является визуальный, а основной тип взаимодействия – выбор графических элементов на мониторе мышью или пальцами на сенсорном экране. При этом основным инструментом пользователя или оператора является рабочая память, в частности её зрительно-пространственный компонент. Известно, что объём рабочей памяти и скорость восприятия объектов, отражая способности временного хранения и манипуляций с информацией для организации целенаправленного действия, представляют важные компоненты интеллекта, который в свою очередь рассматрива-

ется как психометрический показатель успешности адаптации человека [13]. Оценка потенциального диапазона ресурсов рабочей памяти, одним из важных компонентов которой является контур обработки зрительно-пространственной информации [6], важна для эффективной операторской деятельности в условиях нарастающей информационной емкости в задачах управления.

Понятия «рабочая память» и «кратковременная память» до сих пор не получили ясных критериев разграничения, так как и функциональное значение, и временной диапазон сохранения информации существенно зависят от ее характера и способа обработки. В связи с этим предлагается рассматривать емкость рабочей памяти как динамический процесс с учетом типа предъявленного при тестировании материала, способа его организации, контекста и сложности когнитивной нагрузки [21, 30]. Кратковременная память обычно противопоставляется долговременной, объем и время сохранения информации в которой не имеет пределов. Рабочая память представляет структуры и процессы для манипуляций с информацией и ее трансформации в долговременную память [14]. Однако, на наш взгляд, такое определение в большей степени отражает функции кратковременной памяти, тогда как рабочая или оперативная память выполняет функции организации и исполнительного контроля поведения, в том числе формирования ответа при решении поставленной проблемы.

Хорошо известно, что объем кратковременной памяти при последовательном предъявлении информации (букв, цифр, слов, образов) составляет 7 ± 2 усл. ед. [1]. Имеется, однако, другая точка зрения с ограничением емкости памяти до 3–4 ед. [15, 16]. Это заключение сделано на основе результатов серий экспериментов, требующих при тестировании опознать целевые зрительно предъявленные объекты, имеющие различия в цвете, форме и пространственном расположении; причем стимулы предъявлялись достаточно короткое время, исключающее возможность повторения или какого либо структурирования информации. В таком случае оказывается, что в рабочей памяти может сохраняться только до четырех элементарных признаков информации.

Сопоставление приведенных данных указывает на перспективность изучения параметров предъявления аллоцентрической информации для оптимизации зрительно-пространственных процессов ее обработки с учетом индивидуально комфортных или предельно возможных временных параметров безошибочных действий в системе человек-компьютер.

Таким образом, наша статья посвящена исследованию факторов, влияющих на эффективность выполнения человеком задач, задействующих зрительно-пространственную память, которые характерны при практическом взаимодействии с графическими интерфейсами. В разделе 2 мы проводим обзор существующих подходов и уточняем проблему, рассматриваемую в нашей статье, а также предлагаем методы для определения пропускной способности человека-оператора и описываем экспериментальное исследование, проведенное с 88 участниками. Раздел 3 посвящён анализу полученных экспериментальных данных, в нём приводится итоговая формулировка индекса сложности для использованной задачи зрительно-пространственной памяти и вычисляется пропускная способность. В Заключение мы перечисляем основные результаты, полученные в ходе исследования, а также вкратце обсуждаем его ограничения и дальнейшие перспективы.

2. Проблема определения количества воспринимаемой информации

2.1. ЗАКОН ХИКА-ХАЙМАНА

Вероятно, наиболее очевидным подходом к количественному измерению воспринимаемой человеком-оператором информации явилось бы вычисление информационной энтропии. У. Хик, будучи одним из первых, кто применил положения теории информации К. Шеннона в когнитивной психологии, предложил выражение для времени реакции человека (RT) в зависимости от количества рассматриваемых альтернатив (N_H):

$$(1) \quad RT \sim b \cdot \log_2(N_H + 1),$$

где коэффициент b – время для обработки одного бита информации (около 0,15 с). Таким образом, $1/b$ – это скорость перера-

ботки информации, которая считается равной около 6,7 бит/с, но может значительно варьировать в зависимости от характеристик конкретного человека, типа и интенсивности стимулов, обстановки и т.п. [24]. Конечно же, (1) является частным случаем (для равновероятных альтернатив) более фундаментального линейного соотношения между RT и энтропией набора стимулов (H_T), которое было несколько позднее предложено Р. Хайманом:

$$(2) \quad RT = a_H + b_H \cdot H_T,$$

где a_H и b_H – константы, определяемые эмпирическим путём. Принципиальная применимость подхода, основанного на расчёте и минимизации количества обрабатываемой информации, впрочем, была продемонстрирована для некоторых элементов графических интерфейсов – например, выбора из особым образом организованных меню [22]. Были также получены расширенные модели для некоторых задач, возникающих в ходе взаимодействия в системе человек-компьютер, например, для визуального поиска [7, 12], причём в обоих исследованиях в модель вводился как собственно показатель сложности поиска, так и фактор, отражающий опыт пользователя. Некоторые из популярных рекомендаций в сфере проектирования человеко-машинного взаимодействия и удобства использования (юзабилити) также, фактически, вытекают из принципа минимизации энтропии: группировка элементов интерфейса и расположение их «по сетке», важность стандартизации и согласованности проектных решений, эстетичность минимализма и т.п.

Тем не менее, остаётся в общем случае неясным, как именно вычислять H_T на практике, и это признаётся самой существенной проблемой, препятствующей широкому применению закона Хика–Хаймана (2) в сфере человеко-машинного взаимодействия [29]. Нам представляется, что чисто аналитический метод, соответствующий классической шенноновской теории информации, не является адекватным для моделирования восприятия и запоминания человеком сложных графических объектов. Соответственно, для описания работы зрительно-пространственной памяти мы предлагаем рассмотреть объеди-

няющий подход с применением принципов гештальт-психологии.

2.2. ГЕШТАЛЬТ-ПРИНЦИПЫ ВОСПРИЯТИЯ

Гештальт-психология – школа, зародившаяся в начале прошлого века в Германии из исследований процессов восприятия (прежде всего визуального) и способности человека организовывать отдельные элементы в целое, отличающееся от суммы частей. Подробный обзор истории развития этой школы, её основных положений и «заката» в 1960-х годах дан в [32], причём отмечается, что в последние 20 лет интерес к гештальт-подходу заметно вырос – на наш взгляд, это связано с усложнением человеко-машинного взаимодействия в ходе масштабной информатизации, развитием технологий компьютерного зрения и пр. Перечень гештальт-принципов восприятия является не совсем организованным (критики даже обвиняли представителей этой школы, что они на каждый обнаруженный новый фактор, влияющий на восприятие, придумывают новый принцип), но кратко его можно изложить следующим образом [32]:

- «близость»: расположенные рядом или в едином регионе (например, внутри рамки) объекты воспринимаются как единая группа;
- «схожесть»: объекты, совпадающие по цвету, размеру, поведению (например, направлению или скорости движения) и т.п. воспринимаются как единая группа;
- «хорошая фигура» и «дополнение»: объекты подразделяются на группы так, чтобы образовывать более простые и привычные фигуры (например, крест воспринимается как два пересекающихся прямых отрезка, а не как два соприкасающихся прямых угла; отдельные точки и штрихи воспринимаются как прерывистое начертание буквы);
- гештальт-принципы, относящиеся к фону (переднему и заднему плану).

Последователи гештальт-школы утверждали, что эти принципы объясняются свойством человеческого сознания: поиском закономерностей, шаблонов, правил, – однако единой формаль-

ной теории, объединяющей принципы, разработано не было. Это давало основание критикам говорить о чисто описательном характере гештальт-подхода, слабой его применимости для понимания механизмов когнитивных процессов.

Центральное положение в обосновании гештальт-принципов занимало общее правило «Prägnanz» («простота»), которое гласит, что человек организует воспринимаемое так, чтобы оно было регулярным, упорядоченным, симметричным и простым. Уже позднее был предпринят ряд попыток подведения теоретического базиса под понятие «простоты» и его квантификации (например, в [33] описаны четыре подхода, а в [32] – подход на основе теории информации Шеннона). Мы полагаем, что концепции простоты/сложности или «экономии нейронных ресурсов» человека (см. в [33]) действительно являются адекватным базисом, поскольку группировка воспринимаемых объектов уменьшает энтропию (хаос, неопределенность) и мозг имеет возможность воспринимать и обрабатывать меньшее количество информации. Таким образом, общая сущность гештальт-принципов – это уменьшение количества информационных единиц (способность, приобретенная человеком в ходе эволюции и личного опыта) для более эффективной обработки поступающих сигналов.

Как мы уже отмечали, шенноновское (вероятностное) определение количества информации, по сути, основанное на «суммировании» общего из неорганизованных элементов, не совсем пригодно для применения в сфере человеко-машинного взаимодействия. Альтернативами выступают теория алгоритмической (информационной) сложности Колмогорова и схожая с ней «теория структурной информации» в психологии [33], которая подходит к проблеме квантификации сложности в восприятии через описательную сложность конкретных объектов. При определении такой информационной сложности для графического интерфейса необходимо, безусловно, учитывать и принципы обработки информации человеком, на чём мы остановимся чуть далее.

Среди основных гештальт-принципов, оказывающих влияние на рабочую память при визуальном восприятии, называют

«связность», «схожесть» и «близость» [26]. Если привести данные понятия к визуальной сложности графического интерфейса, то можно использовать следующие аналогии:

- «связность» соответствует количественной характеристике воспринимаемых объектов (N), насколько они могут обрабатываться индивидуально или группой;
- «схожесть» соответствует характеристике гетерогенности объектов, размеру «словаря» элементов интерфейса (K);
- «близость» соответствует характеристике масштаба интерфейса (S_0), возможной распределенности объектов по области визуального восприятия.

При этом известно, что гештальт-принципы не равны по силе влияния на восприятие, и пока не разработано убедительной структуры, описывающей значение отдельных принципов и их сочетаний [17, 20, 26] (например, в каких случаях объекты разного цвета на двумерном поле будут группироваться по признаку близости, а в каких – схожести). Соответственно, конкретная форма показателя сложности (3), сформированного на основе трёх перечисленных выше факторов, по-видимому, должна определяться в зависимости от особенностей конкретной задачи по восприятию и обработке информации:

$$(3) \quad IDM = f(N, K, S_0)$$

Существует значительное количество свидетельств, что процесс обработки информации у человека реализован на базе двоичного поиска (дихотомий), что подтверждается, в частности, двоичным логарифмом как функцией зависимости времени реакции от количества альтернатив в законе Хика–Хаймана (1). Действительно, предлагая итоговую форму своей модели, Хик заметил, что она соответствует количеству шагов в иерархической последовательности решений о бинарной классификации, ведущей к верному выбору [27].

Предположим, что для задач по зрительно-пространственному восприятию (запоминанию и визуальному поиску объектов) можно выделить два компонента: информационный и пространственный. Первый из них соответствует объему информации в «сообщении», удерживаемом в памяти, а вто-

рой – пространственному расположению элементов этого сообщения. В одном из наших прошлых исследований [8] результативность выполнения задач по поиску и выбору объектов на двумерном поле (выраженная затратами времени) лучше всего объяснялась индексом сложности выбора, включающим в себя количество объектов и двоичный логарифм размера словаря объектов. При этом влияние размера поля восприятия в совокупности с двумя другими факторами также было значимым, однако остался открытым вопрос о порядке взаимодействия этих двух компонентов, которое, по-видимому, зависит от особенностей задачи – т.е. где именно ищутся альтернативы: а) в запомненной информации или б) среди пространственных позиций. Случаю а) соответствует, например, эксперимент [34], где эффективность выполнения задания зависела от логарифма количества хранимых в памяти фотографий друзей. Для случая б) можно выдвинуть гипотезу, что результативность будет зависеть от двоичного логарифма размера области восприятия, но линейно – от количества соотносимой с ней информации, хранимой в памяти, т.е. $N \cdot \log(K)$.

Необходимо заметить, что проверить обоснованность предложенной формулировки показателя сложности напрямую затруднительно (неясно, с чем сравнивать полученное значение сложности задания, равное, например, 50 битам). Однако мы могли бы опираться на хорошо известные показатели скорости обработки информации человеком, рассмотренные в разделе 1. Для связи сложности и эффективности (скорости) выполнения задач обратимся к такой хорошо зарекомендовавшей себя фундаментальной модели в человеко-машинном взаимодействии как закон Фиттса.

2.3. ЗАКОН ФИТТСА И ПРОПУСКНАЯ СПОСОБНОСТЬ МОТОРНОЙ СИСТЕМЫ ЧЕЛОВЕКА

Закон Фиттса фактически проводит аналогию между выполнением человеком быстрых прицельных движения и передачей информации в электронной системе и базируется на 17-й теореме Шеннона (также называемой теоремой Шеннона–

Хартли), постулирующей, что ёмкость канала C равна разнице между энтропиями сигнала и шума:

$$(4) \quad C = H_S - H_N = B \log_2((P_S + P_N)/P_N) = B \log_2(P_S/P_N + 1),$$

где C – ёмкость канала, означающая теоретическую верхнюю границу скорости передачи данных (бит в секунду); H_S – энтропия сигнала; H_N – энтропия шума; B – полоса пропускания канала (в герцах); P_S – мощность сигнала (в ваттах); P_N – мощность белого гауссовского шума (в ваттах).

Фиттс предложил измерять сложность совершаемого человеком движения в битах, а сам человек (точнее, его нервная система), таким образом, выступает аналогом канала, выполняющего передачу данных. Далее, он предположил, что мощность сигнала аналогична расстоянию, на которое производится движение (A), а мощность шума – допустимому отклонению или размеру региона (W), в пределах которого движение должно быть завершено. Исходя из этого он задал индекс сложности (ID) движения, для которого позднее была обоснована [31] предпочтительная формулировка в следующем виде:

$$(5) \quad ID = \log_2((A + W)/W) = \log_2(A/W + 1).$$

По аналогии с ёмкостью канала, передающего электронный сигнал, Фиттс стремился получить возможность вычисления «ёмкости» двигательной системы человека и в качестве её показателя ввёл индекс производительности (IP), равный отношению индекса сложности движения ко времени, затраченному на его выполнение (MT):

$$(6) \quad IP = ID / MT.$$

Нетрудно заметить, что выражение (6) аналогично (4), причем IP соответствует C , а MT – величина, обратная B . В экспериментальных исследованиях A и W выступают в качестве независимых переменных, а MT является зависимой. Тогда возможно построение зависимости MT от ID при помощи регрессии, где a и b – регрессионные коэффициенты:

$$(7) \quad MT = a + b \cdot ID = a + b \log_2(A/W + 1).$$

На практике коэффициенты детерминации (R^2), получаемые для регрессий (7), как правило, превышают 0,9, причем применимость закона Фиттса была продемонстрирована в широком

диапазоне условий, и он является общепризнанной и широко используемой моделью в сфере человеко-машинного взаимодействия [29]. Индекс производительности (6), введенный Фиттсом, в настоящее время чаще называют пропускной способностью (англ. throughput) и обозначают как TP .

Важным добавлением в модель стал учёт фактора «эффективного» размера цели (W_e), так как очевидно, что время, затрачиваемое на движение, не должно рассматриваться в отрыве от точности его выполнения. «Эффективный» размер определяется из фактического распределения конечных точек движения так, чтобы в его область попадало 96% из них: теорема Шеннона–Хартли имеет предпосылкой то, что шум является белым гауссовским, следовательно отклонения от цели движения должны быть распределены нормально (подробнее вывод W_e представлен, например, в [31]). В отличие от «номинального» ID , который показывает сложность «запланированного» движения, «эффективный» индекс сложности (ID_e), полученный на основе подстановки W_e в (5), соответствует реальному выполнению заданий:

$$(8) \quad MT = a' + b' \cdot ID_e = a' + b' \log_2(A_e / W_e + 1),$$

где A_e – фактическая («эффективная») дистанция между точкой начала и окончания движения. Соответственно, именно ID_e рекомендуется использовать при регрессионном анализе данных экспериментов, изучающих закон Фиттса, и при вычислении пропускной способности моторной системы для участников. Как правило, TP рассчитывается как усредненная пропускная способность каждого из S участников, которая, в свою очередь, вычисляется как среднее значение отношения ID_e к MT , для каждого из M возможных сочетаний значений независимых переменных (A и W) в эксперименте:

$$(9) \quad TP = \frac{1}{S} \sum_{i=1}^S \left(\frac{1}{M} \sum_{j=1}^M \frac{ID_{eij}}{MT_{ij}} \right).$$

Рассчитанная с использованием эффективного индекса сложности пропускная способность является объективной мерой для заданных условий эксперимента, учитывающей как

скорость, так и точность выполнения движений участниками. Следует отметить, что при изучении механизмов памяти показатели точности воспроизведения и времени реакции рассматриваются как наиболее информативные, причём рекомендуется их использование именно в комплексе, во избежание искаженной интерпретации исследуемых явлений [19]. Было даже продемонстрировано, что отсутствие учёта времени выполнения приводит к неверному пониманию изучаемых когнитивных процессов [28]. С другой стороны, характеристики нагрузки памяти и итоговая результативность выполнения задания зависимы от его сложности, включающей как информационные, так и временные переменные [23].

Соответственно, для применения аналогичного закону Фиттса аппарата с целью определения сложности задач для зрительно-пространственной памяти и пропускной способности оператора как «канала с памятью», нам необходимо: 1) сформулировать индекс сложности для таких задач и проверить обоснованность формулировки и 2) выбрать способ определения результативности выполнения задачи участником и также оценить его обоснованность. Для достижения поставленных целей мы провели экспериментальное исследование и проанализировали полученные в ходе него данные.

2.4. ОПИСАНИЕ ЭКСПЕРИМЕНТАЛЬНОГО ИССЛЕДОВАНИЯ

Экспериментальное задание заключалось в запоминании пространственного распределения одинаковых по содержанию стимулов (окрашенных клеток, расположенных на двумерном поле в случайном режиме) и воспроизведения этого набора. При случайном пространственном распределении объектов они могут формировать фигуры, которые воспринимаются в соответствии с гештальт-принципами как интегральные гностические единицы: буквы, геометрические фигуры и т.д. Известно, что группировка элементов предъявленной информации повышает результативность рабочей памяти [17, 20, 26].

Из теории клеточных автоматов, в анализе сложности которых используются как шенноновская, так и колмогоровская ме-

ры информации [35], известно два типа окрестностей («близости») – фон Неймана и Мура. В нашем экспериментальном исследовании мы будем определять количество информационных элементов тремя разными способами:

- 1) как простое количество «элементарных» объектов;
- 2) как количество фигур, определяемое по правилу окрестности фон Неймана 1-го порядка (элементы, соприкасающиеся диагоналями, не являются одной фигурой);
- 3) как количество фигур, определяемое по правилу окрестности Мура 1-го порядка (элементы, соприкасающиеся диагоналями, являются одной фигурой).

Таким образом, основные задачи эксперимента включают в себя:

1. Проверку концептуальной валидности подхода и собранных данных.

2. Выбор предпочтительного способа определения информационной единицы при запоминании: «элементарные» объекты, фигуры по правилу фон Неймана или фигуры по правилу Мура. Для этого мы проанализируем связь их количества (N^*) с результативностью запоминания и соответствие числу Миллера (7 ± 2 информационных единицы, эффективно удерживаемые в рабочей памяти). Выбор такого критерия обусловлен использованием в нашем эксперименте однотипных стимулов, эффективность запоминания которых определялась только точностью их пространственного расположения (окрашенных квадратов на белом поле), отмеченного при тестировании.

3. Формулировка индекса сложности задач для зрительно-пространственной памяти (IDM^*) и проверка её обоснованности. Исходя из изложенных ранее соображений, следующих из гештальт-принципов и обработки информации на базе двоичного поиска, можно предположить:

$$(10) \quad IDM^* = N^* \cdot \log_2 S_0 \cdot \log_2 K,$$

где S_0 – характеристика размера двумерного поля (соответствует «длине сообщения» в теории информации), K – характеристика гетерогенности (размер словаря) запоминаемых объектов. Помимо теоретического обоснования, критерием удачности фор-

мулировки индекса сложности будет являться величина значения R^2 в уравнении регрессии, соответствующем (7).

4. Предложение способа определения результативности выполнения задания и проверка его обоснованности. Как мы отмечали выше, в рамках определения пропускной способности человека-оператора необходим учёт как точности воспроизведения предъявленного для запоминания материала, так и фактического времени выполнения задания. Критерием обоснованности будет являться сила связи предложенного показателя с индексом сложности задания.

5. Вычисление пропускной способности человека-оператора как «канала с памятью» и сравнение полученных значений с аналогичными, известными из литературы.

В экспериментальном исследовании, проведённом в течение одной недели, приняли участие 88 студентов (из них 26 мужчин) бакалавриата и магистратуры Новосибирского государственного технического университета в возрасте от 16 до 22 лет (средний возраст 19,4 лет; среднеквадратическое отклонение – СКО – 1,1). Все участники имели нормальное или скорректированное до нормального зрение и опыт в использовании компьютерных технологий и компьютерной мыши. Все они принимали участие добровольно и использовали оборудование и системное ПО компьютерных классов университета, аналогичное по своему составу и параметрам (в частности, размер монитора и разрешение экрана).

Для проведения эксперимента использовалось специально разработанное программное обеспечение, реализованное на веб-платформе (размещено для авторизованного доступа по адресу <http://psyttest.nstu.ru/tests/23/>). Участнику демонстрировалось двумерное поле, разделённое на клетки, некоторое количество которых, расположенных случайным образом, было окрашено синим цветом (рис. 2). Известно, что для сохранения информации в кратковременной памяти имеет значение не только объём и структура предъявляемой информации, но и временные параметры ее переработки: например, возможности трансформации с учетом имеющихся ресурсов и/или повторения [10]. Чтобы не вносить в эксперимент дополнительный фактор, стимулы

предъявлялись в течение фиксированного времени (2 с) вне зависимости от сложности задания, после чего участник должен был указателем мыши кликнуть на те места, где находились закрашенные клетки. Правильно отмеченные клетки окрашивались зеленым цветом, неправильные – красным, и в этом случае выбор считался ошибочным (рис. 3).

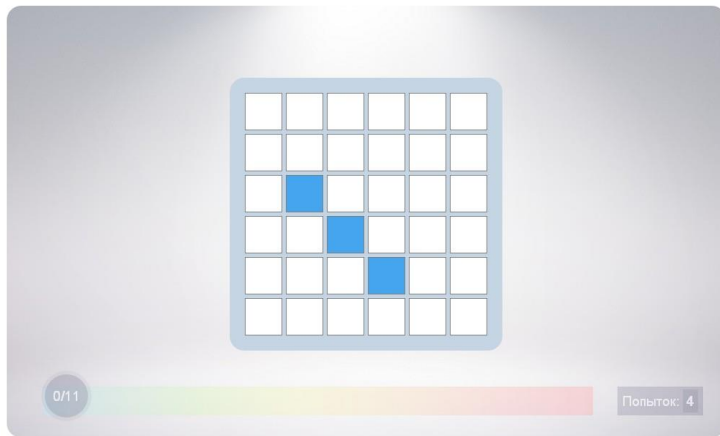


Рис. 2. Экран с предъявляемыми стимулами (3 фигуры по правилу фон Неймана или 1 фигура по правилу Мура)

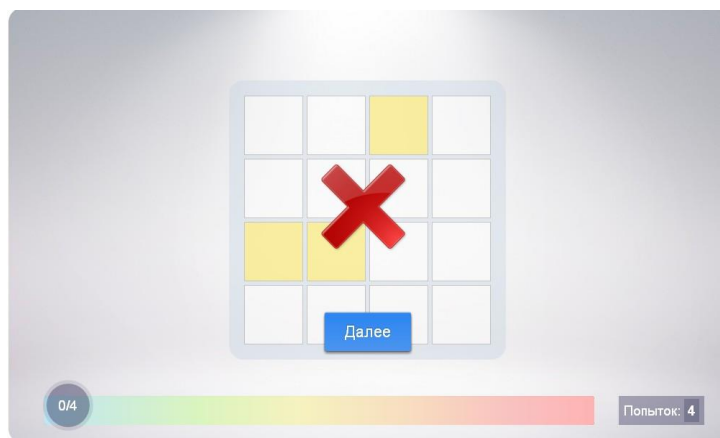


Рис. 3. Экран с завершением исхода (тренировочное задание на поле 4×4 было выполнено с ошибкой, показано верное расположение фигур)

В случае ошибочного выбора, участнику предлагалось ещё раз выполнить задание с тем же количеством закрашенных клеток (снова расположенных случайным образом), однако общее количество таких дополнительных попыток на сессию было ограничено. В случае верного выбора всех клеток участник переходил к следующему набору стимулов, в котором количество закрашенных клеток увеличивалось на 1. Сессия завершалась, когда количество закрашенных клеток достигало установленного максимума или количество попыток обнулялось из-за частых ошибочных выборов. В ходе нашего эксперимента каждый участник должен был выполнить 5 сессий с различными размерами двумерного поля.

Программное обеспечение, разработанное для проведения эксперимента, фиксировало для каждого исхода:

1) независимые переменные:

- размер квадратного поля (S_0) – в эксперименте использовались размеры полей, равные 25 (5×5) и 36 (6×6) клеток;
- количество закрашенных клеток (S_p) варьировалось от 3 до 13 (соответствует количеству информационных единиц на базе элементарных объектов);
- количество получившихся из клеток «фигур» согласно правилу окрестности фон Неймана 1-го порядка (F_N);
- количество фигур согласно правилу окрестности Мура 1-го порядка (F_M);

2) зависимые переменные:

- количество верно выбранных клеток (S_C);
- время, затраченное на выполнение задачи (T_M), т.е. не включающее в себя 2 с, в течение которых производилась демонстрация закрашенных клеток в начале исхода.

Для каждого исхода нами вычислялся уровень ошибки (E_M), поскольку для характеристики работы зрительно-пространственной памяти точность запоминания традиционно считается является наиболее важным показателем:

$$(11) E_M = 1 - \frac{S_C}{S_P}.$$

Отметим также, что в данном эксперименте размер словаря для запоминаемых элементов не варьировался и равнялся 2 (закрашена клетка или нет), т.е. $\log_2 K = 1$.

3. Анализ данных и результаты

3.1. ВАЛИДНОСТЬ ДАННЫХ И ОПИСАТЕЛЬНАЯ СТАТИСТИКА

В общей сложности нами было получено 5243 набора данных, из которых количество валидных исходов составило 4654 (88,77 %). Были исключены ошибочные данные и выбросы, когда итоговое время превышало стандартный срок нахождения информации в рабочей памяти ($T_M > 14$ с), когда выполнение задания фактически не производилось ($T_M \leq 0$ с), или же когда слишком большая ошибка при выборе свидетельствовала о недобросовестном подходе участника ($E_M > 0,75$).

В таблице 1 приведены диапазоны изменений, средние значения и среднеквадратичные отклонения для основных зависимых и независимых переменных в эксперименте.

Таблица 1. Описательная статистика для основных переменных в эксперименте

Переменная	Диапазон	Среднее (СКО)
S_P	3–13	7,36 (2,03)
F_N	1–10	4,55 (1,42)
F_M	1–8	3,2 (1,19)
E_M	0–0,75	0,079 (0,14)
T_M	1–14 с	6,01 (2,22)

Корреляция между зависимыми переменными (E_M и T_M) оказалась значимой и положительной ($r = 0,143$; $p < 0,001$), что объяснимо, поскольку более трудные задания, в которых возникали ошибки, требовали и больше времени на выполнение. Кроме того, это может свидетельствовать о том, что дополнительные затраты времени не увеличивали шансы успешного извлечения информации из рабочей памяти (скорее наоборот, учитывая ограниченный срок хранения информации вследствие освобождения памяти для новых операций с информацией).

Корреляция T_M с S_P составила $r = 0,403$ ($p < 0,001$), T_M с F_N — $r = 0,233$ ($p < 0,001$), T_M с F_M — $r = 0,078$ ($p < 0,001$). Наиболее тесная взаимосвязь времени с количеством закрасенных клеток была ожидаема, поскольку в ходе выполнения задания участник должен был выполнить клик мышью на каждой из запомненных клеток, т.е. чем больше стимулов предъявлялось, тем больше времени требовалось для воспроизведения информации.

Корреляция E_M с S_P составила $r = 0,238$ ($p < 0,001$), E_M с F_N — $r = 0,336$ ($p < 0,001$), E_M с F_M — $r = 0,240$ ($p < 0,001$). Таким образом, самая тесная взаимосвязь с уровнем ошибок была обнаружена не для количества закрасенных клеток, а для количества фигур, определяемых по правилу окрестности фон Неймана. Корреляция E_M с общим размером поля (S_0) также была значимой, хотя и более слабой: $r = 0,179$ ($p < 0,001$).

Далее мы приводим диаграммы, отображающие средние значения E_M в зависимости от значений независимых переменных, соответствующих количеству информационных единиц N^* : S_P (рис. 4), F_N (рис. 5) и F_M (рис. 6). Средние ошибки для крайних правых значений не приводятся, так как исходов с максимальным количеством закрасенных клеток и фигур было слишком мало для статистических выводов. Из диаграмм можно заметить, что числу Миллера 7 ± 2 соответствует F_N в большей степени, чем S_P (про F_M выводы сделать затруднительно вследствие диапазона принимаемых значений).

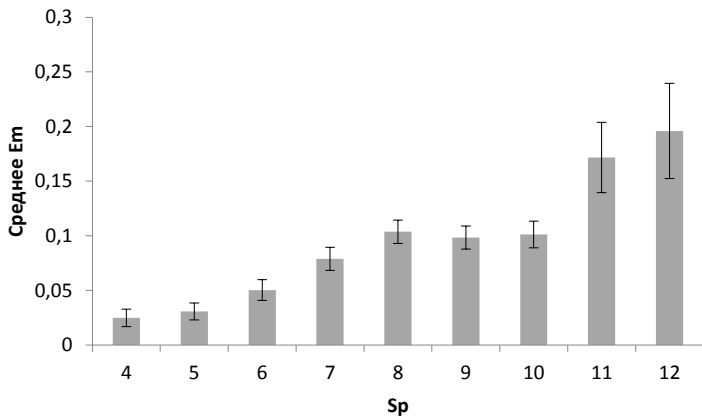


Рис. 4. Средний уровень ошибок E_M по значениям S_P

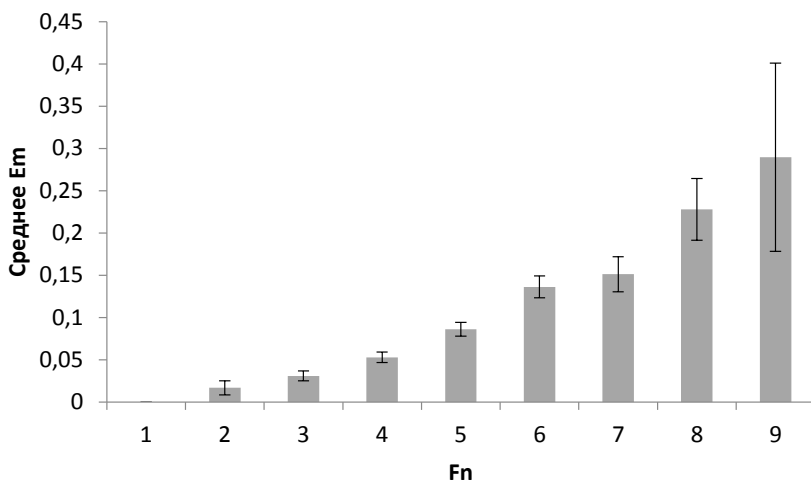


Рис. 5. Средний уровень ошибок E_M по значениям F_N

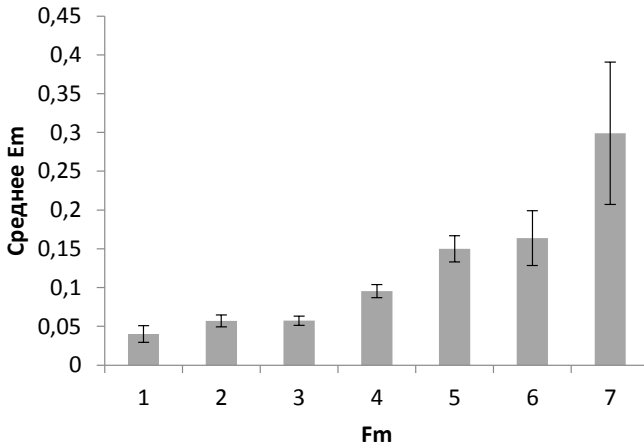


Рис. 6. Средний уровень ошибок E_M по значениям F_M

3.2. ИНДЕКС СЛОЖНОСТИ ДЛЯ ЗРИТЕЛЬНО-ПРОСТРАНСТВЕННОЙ ПАМЯТИ

Исходя из (10) и в соответствии с целями эксперимента, зададим три возможных формулировки индекса сложности (как мы отмечали ранее, в данном исследовании мы не рассматриваем влияние K):

$$(12) \quad IDM_{SP} = S_P \cdot \log_2 S_0,$$

$$(13) \quad IDM_{FN} = F_N \cdot \log_2 S_0,$$

$$(14) \quad IDM_{FM} = F_M \cdot \log_2 S_0.$$

В таблице 2 для каждого из индексов приведен диапазон значений, среднее, среднеквадратичное отклонение и корреляция с E_M и T_M . Можно заметить, что наибольшее значение корреляции с уровнем ошибок при выборе запомненных клеток отмечается для IDM_{FN} .

Таблица 2. Характеристики для индексов сложности

Индекс	Диапазон	Среднее (СКО)	Корр. с E_M	Корр. с T_M
IDM_{SP}	13,9–67,2	36,2 (10,44)	$r = 0,265$ $p < 0,001$	$r = 0,386$ $p < 0,001$
IDM_{FN}	4,0–51,7	22,48 (7,57)	$r = 0,345$ $p < 0,001$	$r = 0,216$ $p < 0,001$
IDM_{FM}	4,0–41,4	15,86 (6,29)	$r = 0,252$ $p < 0,001$	$r = 0,074$ $p < 0,001$

Для проведения регрессионного анализа мы предварительно вычислили средние значения E_M для каждого из значений, индексов, определённых согласно (12), (13) и (14). В итоге регрессионные уравнения приняли следующий вид:

$$(15) E_M = -0,038 + 0,003 \cdot IDM_{SP}, R^2 = 0,694; F_{1,20} = 45,3; p < 0,001,$$

$$(16) E_M = -0,065 + 0,007 \cdot IDM_{FN}, R^2 = 0,895; F_{1,20} = 170; p < 0,001,$$

$$(17) E_M = -0,04 + 0,007 \cdot IDM_{FM}, R^2 = 0,815; F_{1,17} = 74,7; p < 0,001.$$

Наилучшее значение R^2 было получено в регрессионной модели (16), однако оно содержит наибольший по модулю отрицательный свободный член, содержательный смысл которого труднообъясним.

У регрессионной модели, построенной для средних значений времени выполнения задачи (без учёта 47 исходов, где участники делали паузу неопределённой продолжительности между сессиями) и IDM_{FN} , коэффициент R^2 оказался существенно ниже, чем в (16) для E_M :

$$(18) T_M = 6,5 + 0,058 \cdot IDM_{FN}, R^2 = 0,488; F_{1,20} = 19,1; p < 0,001.$$

То, что T_M в меньшей степени, чем E_M объясняется фактором сложности задания по запоминанию, ожидаемо, так как этот показатель отражает не только объем кратковременной памяти, но и навыки владения мышью, которая использовалась для выбора нужных пространственных меток.

3.3. «ЭФФЕКТИВНОЕ» ВРЕМЯ И ПРОПУСКНАЯ СПОСОБНОСТЬ

Как мы отмечали ранее при описании модели Фиттса, для определения пропускной способности (учитывающей как скорость, так и точность выполнения задания), использовался «эффективный» индекс сложности [31], соответствующий реальному заданию, выполненному участником. Для нашей формулировки IDM^* (10) «эффективные» компоненты будут следующими:

- N_e^* – количество информационных единиц, правильно запомненных и воспроизведённых участником (может быть получено как произведение номинального, т.е. предъявленного количества информационных единиц на точность выполнения задания);
- K_e – гетерогенность объектов, предъявленных в конкретном исходе (в нашем эксперименте оно не отличается от номинального);
- S_{0e} – размер участка поля, который реально рассматривался при воспроизведении положения запомненных элементов (распределённости элементов по полю и, предположительно, хаотичности их расположения, что мы не будем учитывать в данном исследовании, используя номинальное значение S_0).

Тогда для нашего эксперимента эффективный индекс сложности, определённый на основе (13), примет вид:

$$(19) \quad IDM_{FN}^e = F_N \cdot (1 - E_M) \cdot \log_2 S_0.$$

Рассчитанный согласно (19) по данным нашего эксперимента IDM_{FN}^e оказался в диапазоне от 3,45 до 46,53 бит (среднее 20,35; СКО = 6,74).

Используя формулировку (19) и учитывая соотношение номинального индекса сложности с IDM_{FN}^e , мы можем переписать формулу для пропускной способности (9) следующим образом:

$$(20) \quad \overline{TP}_M = \frac{\overline{IDM}_{FN}^e}{T_M} = \frac{\overline{IDM}_{FN}}{T_M(1 - E_M)} = \frac{\overline{IDM}_{FN}}{T_M^e},$$

где T_M^e – «эффективное» время, т.е. затраченное только на воспроизведение правильно запомненной информации. Среднее

значение рассчитанного согласно (20) «эффективного» времени в эксперименте (без учёта 47 исходов, где участники делали паузу неопределённой продолжительности между сессиями) составило 6,86 с (СКО = 3,34 с). Корреляция между T_M^E и IDM_{FN} составила $r = 0,302$ ($p < 0,001$), т.е. выше, чем корреляция между T_M и IDM_{FN} ($r = 0,216$, см. в таблице 2). Регрессионная модель для T_M^E и IDM_{FN} (т.е. по смыслу аналогичная закону Фиттса) показала высокую значимость и коэффициент R^2 значительно выше, чем для T_M в (18), хотя и несколько ниже, чем для E_M в (16): (21) $T_M^E = 3,8 + 0,132 \cdot IDM_{FN}$, $R^2 = 0,884$; $F_{1,20} = 152$; $p < 0,001$.

Значение TP_M в эксперименте, вычисленное по всем валидным исходам и усреднённое, оказалось равным 3,75 бит/с (СКО = 1,69 бит/с). Метод статистического дисперсионного анализа (ANOVA) показал, что TP_M , усреднённое по активным участникам, существенно различалось в зависимости от пола ($F_{1,86} = 3,87$; $p = 0,05$; $\eta^2 = 0,0435$), при этом среднее значение для участников-мужчин составило 3,9 бит/с, а для участниц-женщин – 3,61 бит/с. Можно отметить, что все полученные значения для пропускной способности довольно хорошо согласуются с известной скоростью работы человека-оператора как «канала с памятью» (2–4 бит/с [4]) и пропускной способностью моторной системы человека при использовании мыши (3,7–4,9 бит/с [31]). Преимущество мужчин как в пространственной ориентации, так и в скорости моторной реакции также часто отмечается в различных исследованиях [9, 18]. При этом различие в точности между полами может отсутствовать [25] (в нашем эксперименте эффект фактора пола на уровень ошибки также не был значим: $F_{1,86} = 2,91$; $p = 0,09$).

4. Заключение

Результаты анализа данных проведенного нами экспериментального исследования свидетельствуют, что объединение подходов на основе гештальт-принципов восприятия и информационной энтропии быть полезно для вычисления пропускной способности человека-оператора. В частности, регрессионные модели, построенные с использованием предложенной нами формулы индекса сложности задач для зрительно-пространственной памяти (13), показали высокое значение R^2 . Результаты корреляционного анализа и сравнения с известным объемом рабочей памяти (7 ± 2) также свидетельствуют о корректности полученных в ходе эксперимента данных.

Основной вклад данного исследования мы видим в следующем:

1. Показано, что информационными единицами при запоминании объектов, расположенных на двумерном поле (соответствующем графическому интерфейсу пользователя), могут быть не отдельные элементы, а составленные из них фигуры, определяемые по правилу окрестности фон Неймана 1-го порядка.

2. В рамках предложенного подхода на основе гештальт-принципов и закона Фиттса разработана формулировка индекса сложности рассмотренных задач для зрительно-пространственной памяти:

$$(22) \text{IDM}_{FN} = F_N \cdot \log_2 K \cdot \log_2 S_0 .$$

3. Предложен способ определения пропускной способности оператора (20), который может применяться для различных типов задач, возникающих при человеко-машинном взаимодействии в системах управления.

Следует отметить некоторые ограничения нашего исследования. Прежде всего, задание включало в себя стадию не только запоминания, но и воспроизведения (указания объектов мышью в ходе моторной деятельности), что фактически означает смешение этих двух задач. Молодые участники, имеющие большой опыт взаимодействия с компьютерной техникой, выполняли эту

стадию относительно быстро и без затруднений, однако для других категорий участников ситуация может быть иной – время выполнения заданий может превысить типичный срок хранения информации в рабочей памяти. Впрочем, в одном из наших предыдущих исследований мы установили, что пожилые люди воспроизводят след зрительной памяти даже быстрее, чем молодые, так как объем кратковременной памяти у них меньше из-за более выраженных процессов торможения [5]. Следует также заметить, что в ходе эксперимента от участников не требовалось «выполнять задание настолько быстро», поэтому найденная в нашем эксперименте пропускная способность может не являться максимальной. С другой стороны, молодые студенты-участники в действительности стремились закончить эксперимент поскорее.

Примечательно также то, что более резкий рост ошибок начался со сложности менее 160 бит (объема, который удерживается в рабочей памяти при чтении, согласно [2, т. 7]), но следует иметь в виду, что текстовую, как и вообще семантическую информацию большинству участников проще удерживать в памяти, чем абстрактные фигуры и клетки. Кроме того, единицы, получаемые в результате взаимодействия информационной и пространственной компонент сложности (22) – «пространственные биты» – могут иметь иную природу, нежели биты теории информации. Это предположение представляется обоснованным на нейрофизиологическом уровне, поскольку обработка информации в зрительно-пространственной памяти осуществляется за счет двух разных систем мозга (вентральной и дорсальной), но требует дальнейшего исследования.

Наконец, в данном исследовании мы не варьировали словарь объектов, поэтому изучение его эффекта и оценку обоснованности формулировки индекса сложности согласно (22) мы планируем в дальнейшей работе. Кроме того, среди перспектив исследования мы видим изучение связи предложенного индекса сложности с колмогоровской (описательной) сложностью, а также спецификация его пространственного компонента на основе энтропии и информационной сложности клеточных автоматов.

Литература

1. АТКИНСОН Д. *Когнитивная психология*. – СПб.: Питер, 2002.
2. ГАСОВ В.М., СОЛОМОНОВ Л.А. *Организация взаимодействия человека с техническими средствами АСУ* (в 7 томах). – М.: Высшая школа, 1990.
3. ЛОМОВ Б.Ф. *Справочник по инженерной психологии* / под ред. Б.Ф. Ломова. – Москва, Машиностроение, 1982.
4. ПРИСНЯКОВ В.Ф., ПРИСНЯКОВА Л.М. *Математическое моделирование переработки информации оператором человеко-машинных систем*. – Москва, Машиностроение, 1990. – 248 с.
5. РАЗУМНИКОВА О.М., ВОЛЬФ Н.В. *Реорганизация связи интеллекта с характеристиками внимания и памяти при старении* // Журн. высш. нервн. деят. – 2017. – Т. 67, №1. – С. 55–67.
6. BADDELEY A.D., LOGIE R.H. *Working memory: The multiple-component model* // In: A. Miyake & P. Shah (Eds.), *Models of working memory*. – Cambridge, UK: Cambridge University Press, 1990. – P. 28–61.
7. БАКАЕВ М., АВДЕЕНКО Т., ЧЕНГ Н.И. *Modelling selection tasks and assessing performance in web interaction* // IADIS Int. J. Computer Science and Inf. Systems. – 2012. – Vol. 7(1). – P. 94–105.
8. БАКАЕВ М., АВДЕЕНКО Т. *A quantitative measure for information transfer in human-machine control systems* // Proc. of the International Siberian Conference Control and Communications (SIBCON-2015), May 21–25, Omsk, Russia. – 2015 – P. 1–4.
9. BLOUGH P.M., SLAVIN L.K. *Reaction time assessments of gender differences in visual-spatial performance* // Percept Psychophys. – 1987. – Vol. 41. – P. 276–281.
10. CAMOS V., PORTRAT S. *The impact of cognitive load on delayed recall* // Psychonomic bulletin & review. – 2015. – Vol. 22(4). – P. 1029–1034.

11. CARD S.K., MORAN T.P., NEWELL A. *The psychology of human-computer interaction*. – Hillsdale, NJ: Erlbaum, 1983.
12. COCKBURN A., GUTWIN C., GREENBERG S. A predictive model of menu performance // Proc. of the SIGCHI Conference on Human Factors in Computing Systems, Apr 28–May 3, San Jose, California, USA. – 2007. – P. 627–636.
13. CONWAY A.R.A., COWAN N., BUNTING M.F., THERIAULT D.J., MINKOFF S.R.B. *A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence* // Intelligence. – 2002. – Vol. 30(2). – P. 163–183.
14. COWAN N. *What are the differences between long-term, short-term, and working memory?* // Progress in Brain Research, 2008. – Vol. 169. – P. 323–338.
15. COWAN N. *The Magical Mystery Four: How is Working Memory Capacity Limited, and Why?* // Current Directions in Psychological Science. – 2010. – Vol. 19(1). – P. 51–57.
16. COWAN N. *Working Memory Maturation: Can We Get at the Essence of Cognitive Growth* // Perspectives on Psychological Science. – 2016. – Vol. 11(2). – P. 239–264.
17. GAO Z., GAO Q., TANG N., SHUI R., SHEN M. *Organization principles in visual working memory: Evidence from sequential stimulus display* // Cognition. – 2016. – Vol. 146. – P. 277–288.
18. HALPERN D.F. *Sex differences in cognitive abilities* (3rd. ed.). – Mahwah, N-Y, London: Lawrence Erlbaum Ass. Publ, 2000.
19. KAHANA M., LOFTUS G. *Response Time versus Accuracy in Human Memory* // In: The nature of cognition (ed. by R.J. Sternberg), 1999 – P. 323–384.
20. KAŁAMAŁA P., SADOWSKA A., ORDZINIĄK W., CHUDERSKI A. *Gestalt effects in visual working memory* // Experimental Psychology. – 2017. – Vol. 64(1). – P. 5–13.
21. MACKEN B., TAYLOR J., JONES D. *Limitless capacity: a dynamic object-oriented approach to short-term memory* // Frontiers in Psychology. – 2015. – Vol. 6, Iss. 293. – P. 1–15.

22. LANDAUER T., NACHBAR D. *Selection from alphabetic and numeric trees using a touch screen: Breadth, depth, and width* // ACM SIGCHI Bulletin. – 1985. – Vol. 16(3). – P. 73–78.
23. LOGAN G.D. *Short-Term Memory Demands of Reaction-Time Tasks That Differ in Complexity* // J. of Experimental Psychology: Human Perception and Performance. – 1980. – Vol. 6(2). – P. 375–389.
24. LONGSTRETH L.E., EL-ZAHHAR N., ALCORN M. B. *Exceptions to Hick's law: Explorations with a response duration measure* // J. of Experimental Psychology: General. – 1985. – Vol. 114(4). – P. 417–434.
25. LORING-MEIER S., HALPERN D.F. *Sex differences in visuospatial working memory: components of cognitive processing* // Psychonomic Bulletin & Review, 1999. – Vol. 6(3). – P. 464–471.
26. PETERSON D.J., BERRYHILL M.E. *The Gestalt principle of similarity benefits visual working memory* // Psychonomic bulletin & review. – 2013. – Vol. 20.61. – P. 1282–1289.
27. PROCTOR R.W., SCHNEIDER D.W. *Hick's Law for Choice Reaction Time: A Review* // The Quarterly J. of Experimental Psychology. – 2017. – Vol. 24(1). – P. 1–56.
28. RATCLIFF R., STARNS J.J. *Modeling confidence and response time in recognition memory* // Psychological Review. – 2009. – Vol. 116(1). – P. 59–83.
29. SEOW S. *Information theoretic models of HCI: A comparison of the Hick-Hyman law and Fitts' law* // J. of Human-Computer Interaction. – 2005. – Vol. 20(3). – P. 315–352.
30. SIMMERING V.R., PERONE S. *Working Memory Capacity as a Dynamic Process* // Frontiers in Psychology. – 2013. – Vol. 3, Iss. 567. – P. 1–26.
31. SOUKOREFF R.W., MACKENZIE I.S. *Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research* // Int. J. of Human-Computer Studies. – 2004. – Vol. 61(6). – P. 751–789.

32. WAGEMANS J., ELDER J.H., KUBOVY M., PALMER S.E., PETERSON M.A., SINGH M., VON DER HEYDT R. *A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization* // Psychological Bulletin. – 2012. – Vol. 138(6). – P. 1172–1217.
33. WAGEMANS J., FELDMAN J., GEPSHTEIN S., KIMCHI R., POMERANTZ J.R., VAN DER HELM P.A., VAN LEEUWEN C. *A century of Gestalt psychology in visual perception: II. Conceptual and theoretical foundations* // Psychological Bulletin. – 2012. – Vol. 138(6). – P. 1218–1252.
34. WOLFE J.M. *Saved by a log: How do human perform hybrid visual and memory search* // Psychological Science. – 2012 – Vol. 23(7). – P. 698–703.
35. ZENIL H., VILLARREAL-ZAPATA E. *Asymptotic behavior and ratios of complexity in cellular automata* // Int. J. of Bifurcation and Chaos. – 2013. – Vol. 23.09. – P. 1350159.

DEFINING COMPLEXITY FOR VISUAL-SPATIAL MEMORY TASKS AND HUMAN OPERATOR'S THROUGHPUT

Maxim Bakaev, Cand.Sc., Novosibirsk State Technical University, Novosibirsk, Senior Researcher (maxis81@gmail.com).

Olga Razumnikova, Doctor of Science, Head Researcher, Scientific Research Institute of Physiology and Basic Medicine, Novosibirsk, Professor, Novosibirsk State Technical University, Novosibirsk, (razum@physiol.ru).

Abstract: In this paper we study effectiveness and efficiency of HCI tasks that involve visual-spatial memory. Based on the Fitts' law model that is widely used to model human motor reactions and inspired by Gestalt principles of human perception, we proposed formulation for index of difficulty for these tasks. The index includes the number of memorized information chunks, the size of vocabulary for the displayed objects (representing the diversity of elements), and the size of the working area, where the visual search and memorization are performed. We performed experiment with 88 subjects and the analysis of the collected data suggested that the information chunks were not elemental objects (squares), but figures composed based on range 1 von Neumann neighborhood known from cell automata. Based on the proposed index of visual-spatial memorization difficulty, we calculated throughput, which in our experiment was 3.75 bit/s (4.07 bit/s for male subjects and 3.61 bit/s for female ones).

Keywords: human-computer interaction, information complexity, cognitive psychology, gestalt principles, graphical user interface.

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

*Поступила в редакцию 31.07.2017.
Опубликована 30.11.2017.*

УДК 021.8 + 025.1
ББК 78.34

О ТОПОЛОГИЧЕСКОЙ ОТКАЗОУСТОЙЧИВОСТИ МАСШТАБИРУЕМЫХ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ¹

Мелентьев В. А.²

(ФГБУН Институт физики полупроводников
им. А.В. Ржанова СО РАН, Новосибирск)

Рассматриваются проблемы анализа топологической отказоустойчивости масштабируемой вычислительной системы (ВС) и обеспечения ее устойчивости к отказам заданной кратности. Предложен критерий топологической отказоустойчивости, напрямую связывающий топологию с потенциальным параллелизмом системы при заданной кратности допускаемых отказов. Определена взаимосвязь функций топологической масштабируемости и топологической отказоустойчивости систем. Показана обусловленность минимума топологической отказоустойчивости обхватом графа вычислительной системы. Модель параллельных вычислений, а также функции топологической отказоустойчивости и масштабируемости адаптированы к наличию уникальных узлов в информационной топологии решаемой задачи. Предложен способ конфигурирования отказоустойчивых подсистем при дефицитной топологической отказоустойчивости ВС, при этом обеспечение заданной для приложения кратности отказов достигается дублированием подсистем, сконфигурированных для меньшей, чем заданная, кратности отказов.

Ключевые слова: масштабируемые вычислительные системы, их топологическая отказоустойчивость.

¹ Работа выполнена при поддержке Российского фонда фундаментальных исследований, проект №14-07-00169а.

² Виктор Александрович Мелентьев, кандидат технических наук (melva@isp.nsc.ru).

1. Введение

Масштабирование системы, как правило, имеет целью увеличение производительности системы на задачах, критичных к времени их решения. Однако сопутствующий горизонтальному масштабированию экстенсивный рост количества используемых компонентов ведет к уменьшению времени наработки на отказ: порядка 20% вычислительной мощности высокопроизводительных систем теряется по причине возникших неисправностей и восстанавливается после них [2]. Поэтому в [15] особо выделена необходимость оперативного реконfigurирования мультипроцессорных сред и изоляции отказавших в процессе счета элементов, сохраняющего при этом эффект масштабирования; там же указывается на определяющее при этом значение системной топологии. Взаимосвязь и топологическая обусловленность масштабирования и отказоустойчивого функционирования ВС определяют потребность в совместном исследовании этих проблем именно в топологическом аспекте.

Несмотря на то, что известные проекты суперкомпьютеров класса «экзаскейл» (например, проект КНР [2]) предполагают использование межблочных шин с производительностью более 200 Гбит/с и латентностью менее 1,5 мкс, ключевой и наиболее насущной проблемой эксплуатации таких систем считается масштабирование и портирование ранее созданных приложений [3] с тем, чтобы они могли демонстрировать параллельную эффективность на вычислительном поле, не меньшем 20% используемых мощностей. В 2011 году это соответствовало 20 тысячам ядер, а к 2013 году это число увеличилось уже на порядок – до 200 тысяч [14].

Понятно, что в связи с нарастающей глобализацией информационных и вычислительных ресурсов интенсивность и кратность возникающих в таких системах отказов существенно возрастает. Эффективное их парирование требует адекватной и своевременной реакции на возникающие в связи с этим ситуации и заключается в поддержании функциональной и топологической целостности системы. Для исследования и реализации топологической целостности параллельных систем используют, как правило, две основные модели: модель с полной

k -отказоустойчивостью подсистем [20] и модель k -отказоустойчивых подсистем с амортизацией отказов.

Для первой модели характерна неизменность ранга p решаемой в подсистеме задачи даже при наличии k отказов, и k -отказоустойчивость достигается выявлением в графе ВС подграфа с минимизированной для заданных p и k вершинной избыточностью, достаточной для изоморфного (даже при удалении любых k вершин) вложения в него информационного графа задачи ранга p . Использование основанных на этой модели подходов в усовершенствовании и сопоставлении топологий вычислительных систем с позиции отказоустойчивости достаточно широко представлено как зарубежными [21, 22, 26], так и отечественными исследованиями [1, 4, 5]. В качестве основного критерия при этом используют, как правило, максимальный при заданной кратности отказов размер подсистемы, образуемой для решения задач с некоторой вполне определенной информационной топологией, или максимальную при заданном размере подсистемы кратность допускаемых отказов.

В данной работе проблема топологической отказоустойчивости вычислительных систем рассматривается в контексте модели с амортизацией отказов, согласно которой при возникновении в процессе решения задачи отказов ее ранг (размер подсистемы) может изменяться до приемлемого значения. При этом, в отличие от обговоренной выше модели, анализ обусловленной топологией отказоустойчивости ВС обычно осуществляют для системной сети связи в целом, не ограничивая ее рамками подсистем, образуемых для решения определенных, обладающих характерными информационными топологиями, задач и используя, как правило, сетевые показатели, статистически усредненные на множестве возможных при заданной кратности отказов конфигураций. Из критериев, имеющих отношение к проблеме отказоустойчивости, используют также критерий синхронизируемости [19], связывающий топологию сети с возможностью бесконфликтного (в пределах заданных временных ограничений, ограничений на приращение трафика и т.п.) общения произвольных вершин. В [18] предложен основанный на спектральных методиках подход к определению устойчивости этого критерия.

Отметим, что в силу экспоненциальной сложности точного вычисления большинство таких показателей являются стохастическими. Использование многокритериальной оптимизации в практике построения ВС, как правило, еще более усложняет процессы анализа и синтеза обладающих совокупностью заданных свойств топологий при том, что условия совместности/несовместности критериев также не безусловны и обосновываются стохастически. К тому же использование таких показателей в анализе топологической отказоустойчивости крупномасштабных систем представляется нам не вполне правомерным не только из-за их «комбинаторной взрывоопасности», связанной с непропорциональным (масштабированию системы и увеличению кратности отказов) ростом числа возможных конфигураций.

Во-первых, это связано с тем, что потенциалы эффективного масштабирования решаемых в системе задач совпадают с потенциалом системы лишь до некоторых пределов ее масштабирования. Поэтому, если для задачи существует предел эффективного распараллеливания, масштабирование системы сверх этого предела не приводит к аналогичному масштабированию ранга этой задачи, и число задач из решаемого набора, полностью использующих вычислительные (соответственно, и сетевые) ресурсы, с масштабированием системы будет сокращаться. Понятно при этом, что чем меньше в сравнении с системой размер подсистемы, тем в меньшей степени будет справедливым отнесение к ней сетевых показателей системы в целом. О недостаточной достоверности характеристики отказоустойчивости сетей статистически усредненными сетевыми критериями (в частности, средним расстоянием между узлами) говорится, например, в [17]. То же самое можно сказать и об использовании в сопоставлении топологий ВС таких критериев устойчивости к отказам как стабильность нагрузки узлов [25], стабильность диаметра, стабильность средних межузловых расстояний [17], стабильность связности [16] и о других традиционно используемых в анализе сетевой отказоустойчивости [13, 23] показателях.

Во-вторых, – и это, пожалуй, наиболее существенно – оценки влияния отказов на изменения потенциала параллелизма

вычислительной системы являются опосредованными через изменения ее сетевых характеристик. Отсутствие при этом формального соответствия значений сетевых показателей ВС ее потенциалу допускает, таким образом, не более чем качественную оценку топологической отказоустойчивости – на уровне «лучше/хуже».

Основная цель настоящей работы состоит в разработке критериев, не опосредованно, а напрямую связывающих топологию системы с потенциальным ее параллелизмом при заданной кратности допускаемых отказов. Приведено описание используемой при этом модели параллельных вычислений, обуславливающей получение требуемых при реализации приложений значений ускорения и эффективности в решении задачи ее рангом и предельным расстоянием между информационно смежными ветвями. Представлены способы описания лимитированных достижимостью и заданными значениями кратности отказов топологий ВС. Дано определение свойства топологической адекватности системы решаемым на ней задачам, предложен критерий оценки топологической отказоустойчивости, определена обусловленность топологической отказоустойчивости ВС обхватом ее графа. Установлена формальная взаимосвязь функций топологической отказоустойчивости и масштабируемости. Модель параллельных вычислений и эти функции адаптированы к наличию в приложениях уникальных ветвей. Предложен способ отказоустойчивой реализации приложений при дефицитной топологической отказоустойчивости системы.

2. Моделирование топологически отказоустойчивой масштабируемой вычислительной системы

2.1. ОПИСАНИЕ МОДЕЛИ

Зависимость масштабируемости параллельных систем и решаемых на них задач от топологии оценивается в [9] на модели параллельных вычислений, разделенной на две составляющие: первая (пп. 1–4) отнесена к параллельным приложениям и приписывает им свойства неограниченной распараллеливаемости, вторая (пп. 5–10) характеризует систему, ограничения параллелизма в которой обусловлены дефицитным быстродей-

ствием интерконнекта. Суть применения такой модели состоит не в том чтобы получить основанные на игнорировании архитектуры реальные оценки максимально возможного ускорения практически больших задач, а в формальном обосновании вносимых архитектурными компонентами системы (в данном случае – топологией интерконнекта) ограничений основного ее качества – параллелизма и в получении инструмента сопоставления топологий, исходя именно из этого качества. К сожалению, традиционно используемые в исследовании топологий интерконнекта сетевые критерии с указанным качеством систем непосредственного формального соответствия не имеют.

Понятно, что необходимость единообразного для всех топологий сопоставления по их влиянию на ограничения параллелизма потребовало выбора единой «точки отсчета» – эталонного приложения, обладающего идеальным в отношении параллелизма потенциалом. Только при этом условии допустимо считать, что все ограничения параллелизма в системе связаны исключительно с ее топологией, и соответственно, характеризовать топологию достигаемым с ее использованием потенциалом. Идеализация только приложений (пп. 1–4) обеспечивает возможность **единообразного** сопоставления топологий в отношении максимально достигаемого при их использовании потенциала параллелизма.

Заметим также, что принцип **единообразного** сопоставления систем по оценкам их пиковой производительности реализован выбором соответствующих тестов (например, LINPACK, HPL). Неполные соответствия тестов практическим задачам в части используемых параллельных алгоритмов, в части форматов представления данных и/или их объемов и т.п. не позволяют перенести полученные с их помощью абсолютные значения производительности систем на решаемые задачи. Однако эти данные повсеместно используют для **единообразной** рейтинговой оценки производительности систем в целом (в действующих их конфигурациях, включающих типы используемых процессоров, их число, интерконнектные топологии и технологии, интерфейсы программирования и т.п.).

Итак, исходные пункты описания модели содержат следующие обозначения: W и w – измеряемые временем объемы

вычислений при решении произвольной задачи на одном и на p процессорах вычислительной системы; Q и q – измеряемые информационными единицами (байтами) объемы подлежащих обмену данных, соответствующие одному и p задействованным в системе процессорам. Итак,

1. Задача допускает разбиение на произвольное число p информационно связанных параллельных ветвей: $1 \leq p \leq \infty$. Информационный граф распараллеленной на p ветвей задачи может быть нерегулярным, но обязательно связан.

2. Масштабирование данных в задаче с коэффициентом m увеличивает объем вычислений W и объем Q подлежащих обмену данных в m раз.

3. Общий объем вычислений W и объем Q подлежащих обмену данных при разбиении задачи на p параллельных ветвей не зависят от числа процессоров p и распределяются по ним равномерно: $w = W/p$ и $q = Q/p$.

4. Параллельный алгоритм не содержит скалярных фрагментов – это свойство вытекает из сформулированных выше пп. 1 и 3.

5. Все процессоры системы идентичны, их общее число n достаточно для реализации на них p параллельных ветвей, и первоначальное предварительное распределение входных данных по задействованным в параллельном приложении процессорам не требуется.

6. Топология ВС является регулярной¹ и неполносвязной.

7. Общие объемы W и Q не зависят от топологии сети связи и от используемой NT , и ограничения на минимальные объемы w и q отсутствуют.

8. Вычислительные и коммуникационные элементы ВС допускают совмещенную во времени работу.

9. Временные затраты на обмены пропорциональны расстояниям между информационно-связанными задачей вершинами графа ВС.

¹ Степени всех вершин графа ВС одинаковы.

10. Совокупность используемых в вычислительной системе топологии и NT гарантирует отсутствие сетевых коллизий и связанных с ними задержек.

Использование предложенной в [9] модели позволило установить формальную связь объемов W вычислений и Q обмениваемых данных распараллеливаемой на p ветвей задачи с предельно допускаемыми расстояниями δ между информационно смежными ее ветвями при заданных значениях ускорения – $S = T_1/T_p$ (здесь T_1 и T_p – времена решения задачи на одном и на p процессорах соответственно) или эффективности использования задействованных в решении процессоров – $E = S/p$:

$$(1a) \quad \delta_S(p) = \left\lfloor \frac{W}{S \cdot t_{NT}(Q/p)} \right\rfloor,$$

$$(1b) \quad \delta_E(p) = \left\lfloor \frac{W}{p \cdot E \cdot t_{NT}(Q/p)} \right\rfloor.$$

Естественно, что предельные (для рассматриваемых в конкретных ВС приложений) значения достижимостей δ различаются между собой тем значительнее, чем больше отличия в быстродействии используемых в этих ВС сетевых технологий (далее NT – Network Technology). Формулы (1a) и (1b), в которых t_{NT} – известная для NT зависимость времени задержки от объема передаваемой информации, достаточно наглядно демонстрируют это. Они же устанавливают зависимость предельного расстояния между информационно смежными ветвями распараллеливаемой задачи от заданного ускорения и присущих ей объемов вычислительных и обменных операций. Таким образом, реализуемые в исследуемой ВС¹ параллельные приложения могут быть классифицированы по значениям предельных для информационно смежных ветвей расстояний, соответствующих требуемым значениям ускорения S и/или эффективности E при условии топологической адекватности системы этим приложениям.

¹ Здесь и далее понятие *ВС* предполагает использование в ней вполне определенной *NT* с характерной для нее функцией $t_{NT}(Q/p)$.

Определение. Вычислительная система топологически адекватна задаче с присущими ей объемами W вычислительных операций и Q обрабатываемых данных, если топология системы позволяет сконфигурировать подсистему, в которой число процессоров p и предельное расстояние δ между информационно смежными процессорами соответствуют заданным значениям ускорения S и/или эффективности E ее решения.

Понятие топологической адекватности ВС и реализуемых на ней параллельных приложений неоднократно используется в последующем тексте и базируется на описанной выше модели и формальном ее выражении зависимостями (1а), (1б). Напомним, что если решение (W, Q) -задачи обусловлено получаемым для нее ускорением S , то число параллельных ветвей p в подсистеме определяет нижнюю границу числа используемых при этом процессоров, если же критерием является эффективность E использования процессоров, то число параллельных ветвей p в подсистеме определяет верхнюю границу их числа [9].

При необходимости одновременного удовлетворения критериям ускорения и эффективности, требования к топологии и к используемой системой сетевой технологии существенно ужесточаются: топология системы с действующей в ней NT должна быть привязана к информационной топологии задачи так, чтобы обеспечить превышение верхней (определяемой требуемой эффективностью E) границы числа p над его нижней (обусловленной заданным ускорением S) границей. Такое превышение может быть обеспечено, например, использованием более быстрой действующей NT , что в соответствии с (1а) позволяет использовать меньшее число процессоров. При конструктивной невозможности такого рода модификаций технологии сети связи и ее топологии, или при недостаточности этого для совместности E и S следует прибегнуть к поиску компромисса между ними, возможно, с масштабированием обрабатываемых задач и передаваемых между ее ветвями данных [9].

Как видно из (1а), (1б), связанное с отказами уменьшение числа исправных процессоров в реализующих параллельные приложения подсистемах отрицательно сказывается на ускорении, но не на эффективности решения параллельных задач. Поэтому в нашей работе по умолчанию предполагается, что

предельно допустимое для информационно смежных процессоров подсистемы расстояние δ определено из (1а), т.е. в соответствии с требуемым при отказоустойчивом решении задачи ускорением S .

Понятие изоморфизма в теории графов отражает взаимную однозначность (биективность) отображения вершин сопоставляемых графов одного порядка – при этом каждой вершине одного из них соответствует ровно одна вершина другого, и отношения смежности вершин и их образов совпадают. В теории вычислительных систем проблема изоморфизма состоит в выявлении в графе G системы подграфа, изоморфного информационному графу параллельной задачи $W(p)$ с заданным числом p параллельных ветвей. Имеющая при этом место дискретность в отношении изоморфизма вложения графа $W(p)$ в G (да/нет) не позволяет численно оценивать степень топологической адекватности вычислительной системы и реализуемых ею параллельных приложений. Такую возможность предоставляет введенный в [11] показатель масштабируемости $\mu(W, G_\delta)$ системы с G -топологией¹ в реализации на ней задачи с W -топологией при определяемом из (1а), (1б) значении достижимости δ :

$$\mu_\delta(W, G) = n_\delta(W, G)/n(G);$$

здесь $n(W, G_\delta) \equiv n_\delta(W, G)$ – порядок максимального, изоморфно вкладываемого в граф δ -достижимости системы G_δ , подграфа задачи с W -топологией.

Равенство единице показателя $\mu_\delta(W, G)$, говорит о том, что такой максимально вкладываемый подграф или является суграфом² графа G_δ , или изоморфен ему. Понятно, что если G_δ -граф полон, то масштабируемость в нем δ -задач с любой топологией максимальна и равна единице, т.е. он адекватен любым δ -задачам независимо от их информационной топологии. Меньшие значения этого показателя для рассматриваемых в системе задач соответствуют меньшей топологической ее адекватности этим задачам.

¹ Здесь топология системы задана графом G .

² Суграф, или остовный (вершинно-порожденный) подграф – часть графа, имеющая то же множество вершин, что и сам граф.

Ниже даны определение, способ получения и примеры использования графа δ -достижимости в исследовании топологической адекватности систем и решаемых на них задач.

2.2. ГРАФ δ -ДОСТИЖИМОСТИ

Напомним определение введенного в [6] понятия δ -достижимости как отношения смежности вершин в графе G_δ , расстояние (длина простого пути) между которыми в графе G вычислительной системы не превышает значения δ . Проиллюстрируем это простым примером сопоставления кольцевого графа G (рис. 1) и соответствующих ему графов 2- и 3-достижимости (рисунки 2 а и 2 б соответственно).

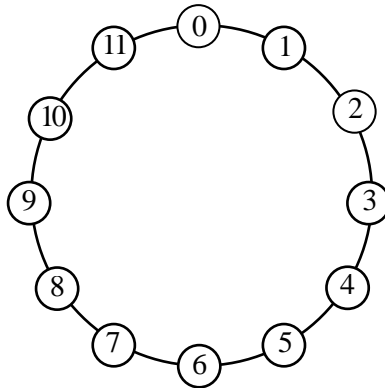


Рис. 1. Кольцевой граф G порядка $n(G) = 12$

Отметим, что все вершины рассматриваемого кольцевого графа дистанционно эквивалентны: i -окрестности любой пары вершин $u, v \in V$, вплоть до равного диаметру $d(G)$ индекса i , — попарно равномогны. Поэтому при построении графа δ -достижимости $G_\delta(G)$ можно ограничиться рассмотрением лишь одной вершины и ее расстояний до остальных вершин графа G . Напомним, что i -окрестность вершины $u \in V$ есть подграф $[u]_i \equiv G_i(u)$, индуцированный множеством вершин,

находящихся от вершины $u \in G$ на расстоянии¹ i . Множество вершин, составляющих i -окрестность вершины u называют ее i -окружением [8].

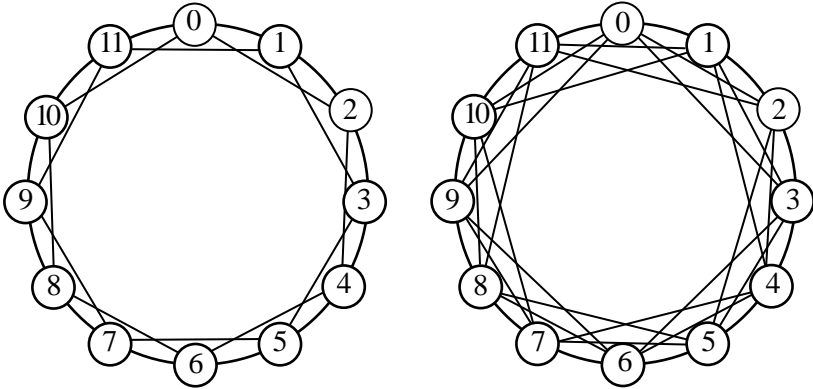


Рис. 2. Графы δ -достижимости кольцевого графа G :
 а – граф 2-достижимости G_2 ; б – граф 3-достижимости G_3

Построенная из любой вершины x кольца (рис. 1) минимально полная² его проекция $P_6(x)$ является 6-уровневой, и диаметр такого графа $d(G) = 6$. Для определенности в проекции (2) в качестве ракурсной выбрана вершина 0, она же составляет нулевой уровень этой проекции. Плотность рассматриваемого графа $\varphi(G) = \varphi(G_1) = 2$, это означает, что число p процессоров, которые могут быть задействованы в решении информационно полносвязных задач, лимитированных достижимостью $\delta = 1$, не превышает двух: $p \leq 2$. Для информационно неполносвязных задач это число может быть большим двух, вплоть до $p = 12$ – для задач с кольцевой топологией.

¹ Расстояние $d(u, v)$ между вершинами u и v равно длине кратчайшего соединяющего их пути.

² Проекция графа является полной, если она содержит в себе описания всех его вершин и инцидентных им ребер. Проекция с минимальным для ее полноты числом уровней является минимально полной.

$$(2) \quad P_6(0) = 0^{(1^{(2^{(3^{(4^{(5^{(6^{\dots}))})})})})})}, 11^{(10^{(9^{(8^{(7^{(6^{\dots}))})})})})}.$$

Построение графа ∂ -достижимости $G_\partial(G)$ можно осуществлять с помощью матрицы смежности исходного графа G , однако, в сравнении с использованием описаний графа в виде проекций [10], такие построения, наряду с большей сложностью, еще и менее иллюстративны. Поэтому в изложении последующего материала использованы проективные описания топологии ВС.

Полные, построенные из ракурсной вершины x , проекции $P_{(x)}(G_\partial)$ графа ∂ -достижимости G_∂ при $1 < \partial \leq d(G)$ получим «сжатием» в ∂ раз полной проекции $P_{(x)}(G)$ исходного графа G . Первый уровень выстраиваемой проекции $P_{(x)}(G_\partial)$ при этом составят вершины, входящие в составы i -окружений ($1 \leq i \leq \partial$) ракурсной вершины. На каждый последующий j -й уровень проекции поместим вершины, составляющие i -окружения вершин $(j - 1)$ -го уровня и связанные таким образом с вершинами j -го уровня графа G_∂ отношениями смежности. Напомним, что вершины, составляющие в получаемой проекции пути из ракурсной вершины x в вершину y , из i -окружения вершины y необходимо исключать [10].

Используя для взятого нами в качестве примера кольцевого графа G модульную арифметику с модулем m , равным порядку «кольца» $n(G) = 12$, получим одноуровневые проекции i -окрестностей $[x]_i \equiv G_i(x)$ любой из его вершин. Для 1- и 2-окрестностей это будут

$$(3) \quad P_1([x]_1) = x^{(x+1, x-1)},$$

$$(4) \quad P_1([x]_2) = x^{(x+2, x-2)}.$$

Сжав вдвое проекцию (2) объединением 1- и 2-окружений из окрестностей (3) и (4), получим полную проекцию графа 2-достижимости для нулевой ракурсной вершины:

$$P_{(0)}(G_2) = 0^{(1^{(2^{(3,4)}, 3^{(2,4,5)}, 11^{(9,10)}, 2^{(1^{(3,11)}, 3^{(1,4,5)}, 4^{(3,5,6)}, 10^{(8^{(6,7,9)}, 9^{(7,8,11)}, 11^{(1,9,10)}, 11^{(1^{(2,3)}, 9^{(7,8,10)}, 10^{(8,9)}, \dots)})})})})}.$$

Как видим, полная проекция $P_{(0)}(G_2)$ графа 2-достижимости кольца в отличие от (2) содержит всего 3 уровня, соответственно уменьшается диаметр — $d(G_2) = 3$, и увеличивается плотность — $\varphi(G_2) = 3$ (максимальная, она же наибольшая, клика

графа G_2 порождается любыми тремя подряд расположенными вершинами $\{x, x + 1, x + 2\}$). Такое увеличение плотности означает, что на задачах с лимитируемой достижимостью $\delta = 2$ число параллельно используемых процессоров p может быть увеличено, как минимум, от трех (для полносвязных задач) до $n(G) = 12$ (для кольцевых задач).

Построенная из произвольной ракурсной вершины $x \in V$ проекция 3-окрестности «кольца» G имеет вид:

$$(5) P_1[x]_3 = x^{(x+3, x-3)}$$

Сжав (2) с учетом (3), (4) и (5), получим полную проекцию $P_{(0)}(G_3)$ графа 3-достижимости:

$$P_{(0)}(G_3) = 0^{(1(2^{(3,4,5,11)}, 3^{(2,4,5,6)}, 4^{(2,3,5,6,7)}, 10^{(7,8,9,11)}, 11^{(2,8,9,10)}), 2^{(1(3,4,10,11)}, 3^{(1,4,5,6)}, 4^{(1,3,5,6,7)}, 5^{(3,4,6,7,8)}, 11^{(2,8,9,10)}), 0^{(3^{(2,4,10,11)}, 2^{(1,4,5,11)}, 4^{(1,2,5,6,7)}, 5^{(2,4,6,7,8)}, 6^{(4,5,7,8,9)}, 9^{(6^{(3,4,5,7,8)}, 7^{(4,5,6,8,10)}, 8^{(5,6,7,10,11)}, 10^{(1,7,8,11)}, 11^{(1,2,8,10)}), 0^{(10^{(1(2,3,4,11)}, 7^{(4,5,6,8,9)}, 8^{(5,6,7,9,11)}, 9^{(6,7,8,11)}, 11^{(1,2,8,9)}), 11^{(1(2,3,4,10)}, 2^{(1,3,4,5)}, 8^{(5,6,7,9,10)}, 9^{(6,7,8,10)}, 10^{(1,7,8,9)})}$$

В этой проекции все вершины графа 3-достижимости определены двумя уровнями, и, хотя диаметр $d(G_3) = 2$, полнота описания ребер в G_3 и, соответственно, полнота его проекции, достигается не на двух, а, как и в графе 2-достижимости, на трех уровнях. Степень этого графа в сравнении с G_2 возрастает с четырех до шести, δ -плотность при этом повышается с $\varphi(G_2) = 3$ до $\varphi(G_3) = 4$.

Общеизвестно, что кольцевая топология обладает крайне ограниченными возможностями в реализации параллельных задач с отличной от кольцевой информационной топологией. Это подтверждается и рассмотренными примерами, из которых видно, что смещение вверх предельного расстояния δ между информационно смежными в задаче процессорами (от $\delta = 2$ до $\delta = 3$), которое при тех же (W, Q) -параметрах задачи может быть достигнуто только повышением быстродействия используемой в системе сетевой технологии, но это даст лишь незначительное увеличение параллелизма приложений: для информационно полносвязных задач число параллельных ветвей может быть увеличено с трех (при $\delta = 2$) всего лишь до четырех ($\delta = 3$).

3. Топологическая отказоустойчивость масштабируемых вычислительных систем

Как мы уже указывали в описании используемой здесь модели (раздел 2.1), реализуемые исследуемой ВС параллельные приложения могут быть классифицированы по значениям предельных для информационно смежных ветвей расстояний δ . Подобные ограничения длины пути, связанные с недостаточным для актуальности информационных обменов быстродействием сетевых технологий, имеют место и в других, например, в распределенных информационных системах. Поэтому ценность исследований, отнесенных к сети в целом и игнорирующих при этом потребности актуализации в ней информационных обменов, с увеличением масштаба сетей связи существенно снижается. Традиционное представление сетей связи графами, в которых отношения смежности вершин соответствуют физической смежности сетевых узлов, также не способствует исследованию сетей в реализации лимитированных транзитных взаимодействий, доля которых в общем объеме информационных обменов растет непропорционально сетевым масштабам. В описанных в разделе 2.2 графах δ -достижимости смежность вершин обусловлена заданным предельно допускаемым расстоянием $\delta \geq 1$ между ними, и это позволяет коррелировать характеристики сети допускаемыми в ней задержками. Кроме того, применение таких графов в конфигурировании в ВС подсистем позволяет исключить возможность превышения допустимых задержек между информационно смежными ветвями задач.

Для исследования топологических аспектов устойчивости ВС к отказам кратности l в данной работе мы используем графы $\delta(k)$ -достижимости, в которых смежность вершин, в отличие от графов δ -достижимости, лимитирована не только допускаемым расстоянием δ между ними, но и числом $k = l + 1$ не превышающих этого расстояния независимых путей. Отметим при этом, что понятие отказоустойчивости имеет различную окраску при отнесении его к сети связи ВС или к реализующей параллельное приложение подсистеме. В первом случае вполне допустимо использование традиционных сетевых показателей, отнесенных,

однако, не только к исходному графу ВС (с $\partial = 1$), а к производным от него графам $\partial(k)$ -достижимости (с $\partial > 1, k \geq 1$). Если при этом в качестве критерия $\partial(l)$ -отказоустойчивости сети связи ВС выбрать, например, связность, то отношение порядка $\partial(k)$ -компоненты связности (компоненты связности графа $\partial(k)$ -достижимости) к порядку исходного графа ВС даст достаточно наглядное представление о присущем рассматриваемой топологии ВС уровне ее отказоустойчивости. Однако в данной работе нас интересуют не топологические возможности системы в целом, а возможности отказоустойчивой реализации в ней параллельных приложений, и именно в этом аспекте приведено изложение последующего материала.

3.1. ГРАФ $\partial(k)$ -ДОСТИЖИМОСТИ

Определим граф $\partial(k)$ -достижимости $G_{\partial(k)}(G)$, он же – граф $k(\partial)$ -соединимости $G_{k(\partial)}(G)$ [6], как надграф графа $G(V, E)$, дополненный ребрами между вершинами $u, v, \in V$ при наличии между ними в графе G не менее k независимых путей с длиной, не превышающей значение заданной достижимости ∂ . Очевидно, что число k таких путей (соединимость информационно смежных в подсистеме вершин подграфа системы) должно по меньшей мере превышать заданную кратность l отказов: $k \geq l + 1$. В этом случае при удалении l каких-либо вершин из соответствующего выполняемой задаче подграфа ВС любые две из оставшихся в нем информационно смежных вершин по-прежнему останутся соединенными путем с длиной, не превышающей предельной для этой задачи достижимости ∂ .

Построить полную проекцию графа $\partial(k)$ -достижимости можно сжатием всего одной полной проекции исходного графа. При этом критерием помещения вершины на j -й уровень выстраиваемой проекции будет не просто присутствие ее в составе i -окружений (при $1 \leq i \leq \partial$) вершин $(j - 1)$ -го уровня, а ее повторяемость с кратностью $m \geq k$. Вместе с этим должен быть проведен анализ на независимость всех m ∂ -путей между вершинами $(j - 1)$ -го уровня и порожденными ими вершинами j -го уровня сопоставлением промежуточных на каждом пути вершин, при этом пути с непустым множеством совпадений считаются зави-

симыми. Большее или равное k число выявленных таким образом независимых путей между вершинами u и v в графе G соответствует наличию ребра между ними в графе $G_{\hat{\alpha}(k)}(G)$.

Как мы уже упоминали выше, все вершины рассматриваемого нами кольцевого графа G – дистанционно эквивалентны, поэтому для построения графа $G_{\hat{\alpha}(k)}(G)$ достаточно единственной проекции графа G . Однако более наглядным будет построение графа $\hat{\alpha}(k)$ -достижимости $G_{\hat{\alpha}(k)}(G)$ с использованием $n(G)$ $\hat{\alpha}$ -уровневых проекций исходного графа $G(V, E)$. В этом случае единственный, стоящий над нулевым, уровень проекции $P_1[x]_i \mid i \in \infty^{\hat{\alpha}}$, определяет перечень смежных по отношению к x вершин в графе $G_{\hat{\alpha}(k)}(G)$. Здесь $\infty^{\hat{\alpha}}$ – последовательность всех натуральных чисел от единицы до $\hat{\alpha}$. Далее условие перебора i -окрестностей от $i = 1$ до $i = \hat{\alpha}$ в проекции $P_1[x]_i \mid i \in \infty^{\hat{\alpha}}$ для компактности опускаем – $P_1[x]_i$.

Из приведенной в (2) проекции $P_6(0)$ понятно, что граф $G_{\hat{\alpha}(2)}(G)$ $\hat{\alpha}(2)$ -достижимости исходного (рис. 1) кольцевого графа G при $\hat{\alpha} < 6$ не содержит ребер и, таким образом, является пустым – $\hat{\alpha} < 6 \Rightarrow \forall x \in V P_1[x]_i = x$. Единичное значение плотности при этом указывает на отсутствие в системе возможностей конфигурирования отказоустойчивых подсистем для задач с $\hat{\alpha} < 6$. Если же допускаемое задачей расстояние между информационно смежными процессорами $\hat{\alpha} = 6$, то $P_1[x]_i = x^{(x+6)}$, граф $G_{6(2)}(G)$ образован шестью 2-вершинными (со смежными в этом графе вершинами x и $x + 6$) компонентами связности – плотность графа и максимальная степень параллелизма полносвязных задач при этом равны двум. Для $\hat{\alpha} = 8$ получим $P_1[x]_i = x^{(x+6, x+7, x+8, x-7, x-8)}$. Полная проекция графа $8(2)$ -достижимости, построенная из нулевой вершины, при этом имеет вид:

$$P_{(0)}(G_{8(2)}) = 0^{(4^{(8(1,2,3)}_9(1,2,3,5)_{10}(2,3,5,6)_{11}(3,5,6,7)}_5(6,7,8,9)_{9(1,2,3,4)_{10}(2,3,4,6)_{11}(3,4,6,7)}),}$$

$$0^{(6^{(1(5,7,8,9)}_2(7,8,9,10)_{10}(2,3,4,5)_{11}(3,4,5,7)}_7(5,6,8,9)_{2(6,8,9,10)}_3(8,9,10,11)_{11}(3,4,5,6)}),}$$

$$0^{(8^{(1(5,6,7,9)}_2(6,7,9,10)_{3(7,9,10,11)}_4(9,10,11)}),}$$

Соответствующие полученным при $\hat{\alpha} = 6$ и $\hat{\alpha} = 8$ проекциям графы $G_{6(2)}(G)$ и $G_{8(2)}(G)$ приведены на рис. 3а и 3б. Как нетруд-

но заметить, диаметры их – $d(G_{6(2)}) = \infty$ и $d(G_{8(2)}) = 2$, а плотности – $\varphi(G_{6(2)}) = 2$ и $\varphi(G_{8(2)}) = 3$.

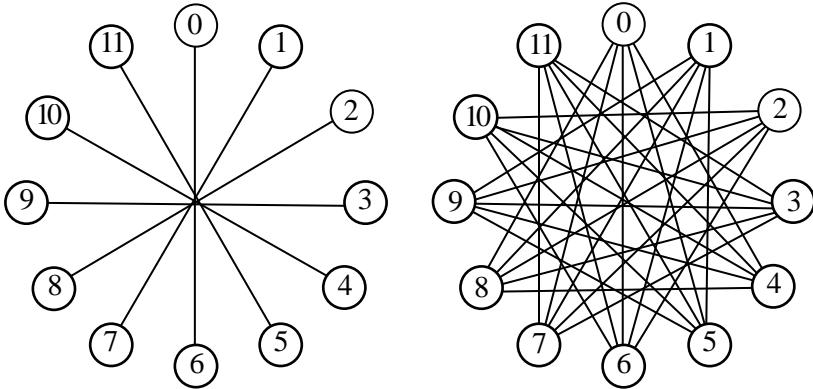


Рис. 3. Графы $\partial(2)$ -достижимости кольцевого графа G :
 а – граф $6(2)$ -достижимости $G_{6(2)}(G)$;
 б – граф $8(2)$ -достижимости $G_{8(2)}(G)$.

3.2. ОБХВАТ ГРАФА ВС И ЕЕ ТОПОЛОГИЧЕСКАЯ ОТКАЗОУСТОЙЧИВОСТЬ

Рассмотренные в предыдущем разделе примеры подтверждают общеизвестный факт крайне низкого потенциала кольцевых топологий в реализации на ВС отказоустойчивых приложений: даже для топологически адекватных системе задач возможность отказоустойчивой их реализации может появиться только при сравнимой с порядком «кольца» достижимости: $\partial \geq \lfloor n/2 \rfloor$. Это существенно ограничивает возможности масштабирования требующих отказоустойчивости приложений даже при использовании в ВС сверхбыстродействующих сетевых технологий: каким бы ни было быстродействие NT в кольцевой ВС, число независимых цепей между любыми двумя вершинами не превысит двух, таким образом, кратность допускаемых в ней отказов не может превышать единицы: $l \leq 1$.

Однако из этих же примеров видно, что ограничения топологической отказоустойчивости ВС связаны с обхватом¹ ее графа, который для кольцевых систем имеет предельное, равное порядку «кольца» значение. Но очевидно, что подобная обусловленность имеет место и для систем с любой другой топологией.

Утверждение. Обхват g графа G вычислительной системы не должен превышать удвоенного для отказоустойчивого решения задачи W значения достижимости $\delta(W)$:

$$(3) \quad g(G) \leq 2\delta.$$

Доказательство. Если обхват $g(G)$ графа ВС превышает удвоенное значение регламентируемой задачей достижимости δ , то в этом графе по определению обхвата не найдется и пары вершин $u, v \in V$, связанных хотя бы двумя независимыми цепями с длинами, не превышающими этой достижимости: $\delta < \lceil g(G)/2 \rceil \Rightarrow k(u, v) \leq 1$. Следовательно, граф $\delta(k)$ -достижимости $G_{\delta(k)}$ будет пустым при отличной от нуля кратности l отказов, порядок компоненты связности в нем равен единице, что отрицает возможность конфигурирования отказоустойчивых подсистем с $\delta < \lceil g(G)/2 \rceil$ в принципе.

Понятно, что для набора реализуемых в ВС отказоустойчивых приложений обхват графа ВС обусловлен минимальным в наборе таких приложений значением достижимости δ_{\min} :

$$g(G) \leq 2\delta_{\min}.$$

Из (3) и отсутствия кратных ребер между вершинами графа ВС ясно, что для задач, в решении которых из-за дефицитного быстрогодействия NT не может быть обеспечено большее единицы значение достижимости – $\delta > 1$, свойство топологической отказоустойчивости не реализуемо в принципе: даже при использовании полностью связной топологии: $\delta = 1 \Rightarrow l = 0$. Поэтому для того чтобы, например, гиперкубическая ВС была способна парировать однократные отказы, быстродействие ее NT должно быть достаточным для обеспечения достижимости $\delta \geq 2$.

¹ Обхватом графа называется длина его кратчайшего цикла.

Очевидно, что при кратности $l \geq 1$ отказов ($k \geq 2$) обусловленное задачами ограничение (3) обхвата графа ВС дополняется ограничением снизу его степени, которая, как и обхват, также влияет на диаметр графа и на его порядок. Это объясняется тем, что в s -регулярном графе с заданными значениями порядка n и степени s наибольшей компактностью¹ обладают графы с наибольшими обхватами [12], и тем, что число k независимых цепей между двумя вершинами регулярного графа не может превышать его степени: $k \leq s$. Поэтому утверждение о том, что «снижение диаметра графа межмашинных связей ведет к увеличению эффективности и повышению надежности системы...» [12], может быть справедливым лишь в отношении эффективности, но не в отношении отказоустойчивости, так как при заданных значениях порядка и степени графа стремление к минимизации его диаметра $d_{\min}(G)$ требует увеличения обхвата до максимального значения $g_{\max}(G)$, которое, как известно [7], равно удвоенному диаметру $g_{\max}(G) = 2d_{\min}(G)$. Это нарушает условие (3) для всех задач с меньшей этого диаметра достижимостью $\partial \leq d_{\min}(G)$ и, таким образом, приводит к полной утрате свойства топологической отказоустойчивости системы при решении таких задач.

3.3. ФУНКЦИЯ ТОПОЛОГИЧЕСКОЙ ОТКАЗОУСТОЙЧИВОСТИ ВС

В принципе, введенная в [9] функция топологической масштабируемости $\mu_k(G)$ способна дать представление и о топологической отказоустойчивости вычислительной системы. Для этого необходимо функцию плотности $\varphi(G_\partial)$ графа G системы заменить определенной в разделе 2.3 плотностью $\varphi(G_{\alpha(k)})$ и аргументировать $\varphi(G_{\alpha(k)})$ не числом $n \equiv n(G)$ узлов в ней (при нулевой кратности отказов $l = 0$), как это делается в $\mu_k(G)$, а кратностью $l = k - 1$ отказов. Изменив при этом обозначение

¹ Компактные структуры ВС определены как s -регулярные заданного порядка графы с минимально возможным диаметром.

функции топологической масштабируемости с $\mu_c(G)$ на $\mu_{c,l}(n)$, получим:

$$(6) \quad \mu_{c,l}(n) = \varphi(G_{\partial(k)})/n(G).$$

Давая, однако, количественную характеристику потенциала параллелизма ВС (какая часть общего числа $n(G)$ процессоров может быть использована в l -отказоустойчивом решении ∂ -задач) функция топологической масштабируемости отказоустойчивой ВС (6) дает лишь качественное представление об изменении этого потенциала в сравнении с изначальным (при $l=0$) потенциалом – $\varphi_c(G)$, который у топологически различающихся ВС одного порядка, как правило, тоже различен.

Поэтому топологическую l -отказоустойчивость $\theta_{n,\partial}(l)$ системы с G -топологией ($n \equiv n(G)$) в решении ∂ -задач определим отношением плотности $\varphi_{\partial(k)}(G) \equiv \varphi(G_{\partial(k)})$ графа $\partial(k)$ -достижимости $G_{\partial(k)}$, соответствующей определенной для задач из (1а), (1б) достижимости ∂ , и заданной кратности $l \geq 1$ отказов ($k = l + 1$), к исключаяющей наличие отказов ($l = 0$) плотности $\varphi_c(G) \equiv \varphi(G_\partial)$ графа ∂ -достижимости G_∂ :

$$(7) \quad \theta_{n,\partial}(l) = \varphi_{\partial(k)}(G)/\varphi_c(G).$$

Сравнивая (6) и (7), нетрудно заметить, что для заданных значений n , ∂ и l функция топологической l -отказоустойчивости $\theta_{n,\partial}(l)$ равна отношению функции топологической масштабируемости l -отказоустойчивой ВС (6) к функции топологической масштабируемости системы $\mu_c(G) = \varphi_c(G)/n(G)$, определяемой в предположении отсутствия ($l = 0$) отказов:

$$\theta_{n,\partial}(l) = \mu_{c,l}(n)/\mu_c(G).$$

В отличие от традиционно используемых в сопоставлении ВС с различными топологиями сетевых показателей (диаметра, ширины бисекции), позволяющих судить лишь о качестве (лучше/хуже, больше/меньше) влияния топологии на потенциал распараллеливаемости задач в системе, показатели топологической отказоустойчивости ВС – (6) и топологической ее масштабируемости – (7) позволяют численно оценивать нижнюю (при распараллеливании в системе информационно ∂ -полносвязных задач) границу такого потенциала при возникновении в процессе их решения отказов кратности l . Следует помнить, однако,

что такая оценка справедлива лишь при паритетности (идентичности) всех параллельных ветвей задачи.

3.4. НЕПАРИТЕТНОСТЬ ВЕТВЕЙ ЗАДАЧИ И ТОПОЛОГИЧЕСКАЯ ОТКАЗОУСТОЙЧИВОСТЬ ВС

В практике параллельных вычислений далеко не редкими являются задачи, в которых отдельные параллельные ветви функционально или топологически отличны от остальных и в этом смысле такие ветви уникальны. К примеру, в задачах со звездной топологией уникальной является центральная вершина «звезды». То же можно сказать и в отношении используемых в решении задач схем информационного взаимодействия: наличие центрального узла свойственно трансляционной и коллекторной схемам обменов. Понятно, что для наделения свойством l -отказоустойчивости сконфигурированных под такие задачи подсистем процессоры, соответствующие x уникальным ветвям, должны быть l -кратно продублированы. Поэтому в сравнении с плотностью $\varphi(G_{\alpha(k)})$ число процессоров, задействованных в решении таких задач при их информационно полной связности, уменьшится на произведение $x \cdot l$ – соответственно уменьшаются значения топологической l -отказоустойчивости и топологической масштабируемости l -отказоустойчивой системы:

$$\theta_{\alpha(l)}(G) = (\varphi_{\alpha(k)}(G) - l \cdot x) / \varphi_{\alpha}(G),$$

$$\mu_{\alpha(l)}(G) = (\varphi_{\alpha(k)}(G) - l \cdot x) / n(G).$$

Для информационно неполносвязных задач, параллельные алгоритмы решения которых предполагают наличие уникальных ветвей ($x > 0$), информационный граф W исследуемой задачи следует дополнить такими $l \cdot x$ вершинами и инцидентными им ребрами, что окружение дублирующей вершины совпадает с окружением дублируемой уникальной вершины. Модифицированный таким образом граф задачи W обозначим W_l . Для определения топологической отказоустойчивости и масштабируемости $\mu_{\alpha(k)}(W, G) \equiv \mu(W, G_{\alpha(k)})$ системы при l -отказоустойчивой реализации решения такой W -задачи следует определить порядок $n(W_l, G_{\alpha(k)}) \equiv n_{\alpha(k)}(W_l, G)$ максимального изоморфного графу W_l подграфа в графе $\partial(k)$ -достижимости $G_{\alpha(k)}$, здесь $-k = l + 1$. Способ выявления изоморфного вложения информационного

графа задачи в такие графы предложен и подробно рассмотрен в [6] и [8]. Тогда топологическая отказоустойчивость и масштабируемость системы в решении таких задач определяются выражениями:

$$(8) \quad \theta_{\alpha(l)}(W, G) = (n(W_l, G_{\alpha(k)}) - l \cdot x) / \varphi_{\partial}(G),$$

$$(9) \quad \mu_{\alpha(l)}(W, G) = (n(W_l, G_{\alpha(k)}) - l \cdot x) / n(G).$$

Из (8), (9) ясно, что использование алгоритмов с непаритетными ветвями, существенно ухудшает потенциал отказоустойчивого и параллельного решения задач даже при их полной топологической адекватности системе.

Резюмируя вышесказанное, отметим, что требование топологической отказоустойчивости в решении задач ограничивает потенциал параллелизма в их решении тем в большей степени, чем выше поднята планка кратности допускаемых при решении задач отказов, так как $n(W_{l+1}, G_{\alpha(l+2)}) < n(W_l, G_{\alpha(l+1)})$ (для полностью связанных задач – $\varphi(G_{\alpha(k+1)}) < \varphi(G_{\alpha(k)})$). При этом может случиться так, что необходимое для обеспечения ускорения S задачи число процессоров p , соответствующее достижимости ∂ и заданной кратности l отказов, топологически не может быть обеспечено: $p > n(W_l, G_{\alpha(l+1)})$. Один из возможных способов решения этой проблемы приведем ниже.

Пусть топология ВС задана графом G , информационная топология решаемой задачи – графом W . Кратность допускаемых при решении этой задачи отказов отлична от нуля. Необходимое для обеспечения топологической l -отказоустойчивости число k независимых путей между процессорами, реализующими информационно смежные в задаче ветви, $k = l + 1$. Пусть полученное из (1а) минимальное число p процессоров, необходимое для получения заданного в решении данной задачи ускорения превышает порядок максимального изоморфного графу W_l подграфа в графе $\partial(k)$ -достижимости $G_{\alpha(k)} - p > n(W_l, G_{\alpha(k)})$. Пусть при этом существует некоторая отличная от нуля величина $y \leq l$, дополняющая до l значение кратности $l - y$ отказов, при которой $p \leq n(W_{l-y}, G_{\alpha(k-y)})$. Тогда если в графе $G_{\alpha(k-y)}$ может быть организовано $y + 1$ подграфов W_{l-y} с $(l - y)$ -отказоустойчивой для рассматриваемой задачи топологией, то устойчивость к кратности l отказов в решении задачи может быть достигнута u -кратным

резервированием топологически $(l - y)$ -отказоустойчивых подсистем.

4. Заключение

Наращивание ресурсов в системе и стремление к увеличению их загрузки неизбежно приводят к увеличению вероятности отказов и их кратности [24]. В этой связи основное требование к масштабируемым системам, состоящее в том, что они должны допускать наращивание ресурсов, обеспечивающее пропорциональный прирост производительности без глобальной перестройки архитектуры, следует дополнить требованием повышения устойчивости к отказам или сохранения ее как минимум на уровне, предшествующем масштабированию.

Потенциал параллелизма системы во многом ограничен топологическим потенциалом обеспечения информационно-логической целостности решаемых задач. Традиционно используемые при сопоставлении вычислительных систем с различными топологиями сетевые характеристики квалифицируют лишь, собственно, топологии без учета потенциалов параллелизма, достигаемых системой при их использовании, и могут быть, таким образом, не более чем качественными (лучше/хуже). То же можно сказать и об использовании этих характеристик в сопоставлении отказоустойчивости топологически отличающихся систем.

В работе дано краткое описание модели, формально связывающей объемные параметры параллельной задачи (объемы вычислений и обрабатываемых данных) и требуемые значения критериев актуальности (ускорения) или эффективности ее решения, с необходимыми при этом и обусловленными быстрым действием используемой в системе сетевой технологии числом процессоров и предельными расстояниями (достижимостью) между ними. Основанное на такой модели свойство топологической отказоустойчивости ВС предложено оценивать относительно, при увеличении кратности l отказов, изменением плотности графа $\partial(k)$ -достижимости, где $k = l + 1$ – число независимых с длиной, не превышающей ∂ , цепей между процессо-

рами, соответствующими информационно смежным ветвям задач.

Показана формальная обусловленность топологической отказоустойчивости системы обхватом ее графа. Установлена формальная взаимосвязь топологической отказоустойчивости и топологической масштабируемости систем. Модель параллельных вычислений, функции топологической отказоустойчивости и топологической масштабируемости адаптированы к задачам, в которых отдельные параллельные ветви функционально или топологически отличны от остальных и в этом смысле являются уникальными. Рассмотрен случай с дефицитной в сравнении с заданной топологической отказоустойчивостью ВС, и предложен способ решения этой проблемы, когда обеспечение заданной для приложения кратности отказов может быть достигнуто дублированием подсистем, сконфигурированных для меньшей кратности.

Результаты работы могут быть полезны при исследовании деградации параллелизма действующих в условиях отказов систем, при выборе или генерации топологий вычислительных систем, ориентированных на отказоустойчивую реализацию набора решаемых задач, а также при распараллеливании таких задач и при конфигурировании соответствующих им подсистем. Использование полученных в работе результатов открывает также возможности детерминированного исследования и проектирования крупномасштабных информационных сетей с заданной устойчивостью к отказам в отношении сохранности и актуальности предоставления тех или иных информационных ресурсов, а также оптимизированного в соответствии с этим размещения их в сети.

Литература

1. АБРОСИМОВ М.Б. *Графовые модели отказоустойчивости*: дисс. докт. физ.-мат. наук. – URL: https://cs.msu.ru/sites/smcc/files/theses/doktorskaya_abrosimov.pdf (дата обращения: 18.10.2016).
2. ГЕРГЕЛЬ В.П., ЛИНЁВ А.В. *Проблемы и перспективы достижения экзафлопного уровня производительности суперкомпьютерных систем* // Информационные технологии. Вестник Нижегородского университета им. Н.И. Лобачевского. – 2012. – №3(1). – С. 189–198.
3. *Китай планирует построить суперкомпьютер уровня «экзаскейла» в 2020 году*. – URL: http://www.thg.ru/technews/20160503_110013.html/esk_tex.pdf (дата обращения: 03.06.2016).
4. КАРАВАЙ М.Ф. Применение теории симметрии к анализу и синтезу отказоустойчивых систем // Автоматика и телемеханика. – 1996. – №6. – С. 159–173.
5. КАРАВАЙ М.Ф., ПОДЛАЗОВ В.С. *Расширенный обобщенный гиперкуб как отказоустойчивая системная сеть для многопроцессорных систем* // Управление большими системами. – 2013. – Вып. 45. – С. 344–371.
6. МЕЛЕНТЬЕВ В.А. *Вложение подсистем, лимитирующих длину и число путей между вершинами графа вычислительной системы* // Управление большими системами. – 2015. – Вып. 47. – С. 212–246.
7. МЕЛЕНТЬЕВ В.А. *Компактные структуры вычислительных систем и их синтез* // Управление большими системами. – 2011. – Вып. 32. – С. 241–261.
8. МЕЛЕНТЬЕВ В.А. *Предельное конфигурирование подсистем в гиперкубических вычислительных системах* // Информационные технологии и вычислительные системы. – М.: Федеральный исследовательский центр «Информатика и управление» РАН, 2015. – №2. – С. 20–30.
9. МЕЛЕНТЬЕВ В.А. *О топологической масштабируемости вычислительных систем* // Управление большими системами. – 2015. – Вып. 58. – С. 115–143.

10. МЕЛЕНТЬЕВ В.А. *Формальные основы скобочных образов в теории графов* // Труды Второй Междунар. конф. «Параллельные вычисления и задачи управления» (РАСО), 4-6 октября 2004 г., Москва, Россия. – 2004. – С. 694–706.
11. МЕЛЕНТЬЕВ В.А., ШУБИН В.И., ЗАДОРОЖНЫЙ А.Ф. *Топологическая масштабируемость гиперкубических параллельных систем и задач* // ISJ Theoretical & Applied Science. – 2015. – №11. – С. 122–129.
12. ПИНЧУК В.П. *Базовые графы для построения топологии управляющих сетей многопроцессорных систем* // Искусственный интеллект. – 2004. – №4. – С. 46–58.
13. СНАРСКИЙ А.А., ЛАНДЭ Д.В. *Моделирование сложных сетей. Учебное пособие.* – Киев: НТУУ «КПИ», 2015. – 212 с. – URL: <http://freescb.info/sites/freescb.info/files/mss-new.pdf> (дата обращения: 20.10.2016)
14. СТЕГАЙЛОВ В.В., НОРМАН Г.Э. *Проблемы развития суперкомпьютерной отрасли в России: взгляд пользователя высокопроизводительных систем* // Программные системы: теория и приложения. – 2014. – №1(19). – С. 111–152.
15. *Эксафлопные технологии. Концепция по развитию технологии высокопроизводительных вычислений на базе суперэвм эксафлопного класса (2012–2020 гг.).* – URL: http://filearchive.cnews.ru/doc/2012/03/esk_tex.pdf (дата обращения: 01.06.2016).
16. ADHIKARI N., TRIPATHY C.R. *The Folded Crossed Cube: A New Interconnection Network for Parallel Systems* // Int. Journal of Computer Applications. – July, 2010. – Vol. 4(3). – P. 43–50. – URL: <http://www.ijcaonline.org/volume4/number3/rxc3871147.pdf> (дата обращения: 20.10.2016)
17. ALBERT R., JEONG H., BARABASI A. *Error and attack tolerance of complex networks* // Nature. – 2000. – Vol. 406. – P. 378–382. – URL: <http://barabasi.com/f/77.pdf> (дата обращения: 20.10.2016).
18. BARAHONA M., PECORA L.M. *Synchronization in Small-world Systems* // Phys. Rev. Lett. – 2002. – Vol. 89, Iss. 5. – URL: <https://doi.org/10.1103/PhysRevLett.89.054101> (дата обращения: 20.10.2016).

19. DONETTI L., HURTADO P.I., MUFIOZ M.A. *Entangled networks, synchronization and optimal network topology* // Phys. Rev. Lett. – 2005. – Vol. 95, Iss. 18. – URL: <https://doi.org/10.1103/PhysRevLett.95.188701> (дата обращения: 20.10.2016).
20. HAYES J.P. *A graph model for fault-tolerant computing system* // IEEE Trans. Comput. – 1976. – Vol. C-25, No. 9. – P. 875–884.
21. HARARY F., HAYES J.P. *Edge fault tolerance in graphs* // Networks. – 1993. – No. 23. – P. 135–142.
22. HO T.-Y., SUNG T.-Y., HSU L.-H. *A note on edge fault tolerance with respect to hypercubes* // Applied Mathematics Letters. – 2005. – No. 18. – P. 1125–1128.
23. KLAU G.W., WEISKIRCHER R. *Robustness and Resilience* // Network analysis. – 2005. – P. 417–437. – URL: <http://homepages.cwi.nl/~klaus/pubs/kw-rr:2005.pdf> (дата обращения: 20.10.2016).
24. MELENT'EV V.A. *Reliability of elements of the computing system and its fault tolerance* // ISJ Theoretical & Applied Science. – 2014. – №9(17). – P. 34–45.
25. NEWMAN M.E.J. *A measure of betweenness centrality based on random walks* // Social Networks. – 2005. – Vol. 27. – P. 39–54. – URL: <https://pdfs.semanticscholar.org/0a57/5498f9e6bc0cc43b977c6e952101f89be90c.pdf> (дата обращения: 20.10.2016).
26. ZIMMERMAN G.W., ESFAHANIAN A.-H. *Chordal rings as fault-tolerant loops* // Discrete Applied Mathematics. – 1992. – No. 37/38. – P. 563–573.

ON TOPOLOGICAL FAULT-TOLERANCE OF SCALABLE COMPUTING SYSTEMS

Victor Melentiev, Rzhanov Institute of Semiconductor Physics Siberian Branch of RAS, Novosibirsk, Cand. Sc., senior research associate (melva@isp.nsc.ru).

Abstract: Problems of the analysis of topological fault tolerance of the scalable computing system and ensuring its sustainability to fault of the given multiplicity are considered. The measure of topological fault tolerance is offered, which connects the computing system topology with its potential parallelism for the given fault multiplicity. The relationship between the functions of topological scalability and topological fault tolerance is defined. The dependence of the minimum of a topological fault tolerance by the girth of the system graph is shown. Model of parallel computings, and functions of the topological fault tolerance and scalability are adapted to the existence of unique nodes in information topology of the solved task. A method for configuring fault-tolerant subsystems for a deficient topological fault tolerance of a computing system is proposed, while providing the preassigned fault multiplicity for the solved task is achieved by duplicating subsystems which are configured for less, than the preassigned, fault multiplicity.

Keywords: scalable computing systems, their topological fault-tolerance.

*Статья представлена к публикации
членом редакционной коллегии Н.Н. Непейвода.*

*Поступила в редакцию 20.09.2016.
Опубликована 30.11.2017.*

УДК 33
ББК 65

ОПТИМАЛЬНАЯ ФУНКЦИЯ ИЗДЕРЖЕК ПРЕДОТВРАЩЕНИЯ ПРОМЫШЛЕННЫХ РИСКОВ ФИРМЫ

Ростова Е. П.¹, Гераськин М. И.²

(Самарский национальный исследовательский
университет им. академика С.П. Королёва, Самара)

Рассмотрена задача определения функции издержек фирмы на предотвращение промышленных рисков (добровольных рисков издержек), оптимальной по критерию общих издержек. Для степенной функции производственных издержек и экспоненциально убывающей функции промышленного ущерба определен вид функции добровольных рисков издержек. Проанализировано влияние параметров функции промышленного ущерба на область существования решения задачи максимизации прибыли фирмы при оптимальной функции добровольных рисков издержек.

Ключевые слова: промышленный риск, промышленный ущерб, издержки, оптимизация, риск-менеджмент.

1. Введение

Промышленные риски, трактуемые как возникающие в коммерческой деятельности промышленных фирм, всесторонне рассматривались в экономической теории. Иерархия задач управления промышленными рисками охватывает различные уровни рисков: от рисков, связанных с выходом фирм на мировой рынок (risk of globalization) [42] и рисков воздействия ком-

¹ Елена Павловна Ростова, кандидат экономических наук, доцент (el_rostova@mail.ru).

² Михаил Иванович Гераськин, доктор экономических наук, профессор, заведующий кафедрой математических методов в экономике (innovation@ssau.ru).

мерческой деятельности фирм на региональные экономические системы [45, 47, 48] до рисков, обусловленных невыполнением отдельных технологических операций. В контексте дальнейшего анализа будут рассматриваться риски на уровне фирмы. В этом аспекте сопоставлялись традиционный реактивный экологический менеджмент и активный риск-менеджмент предприятия (ERM)¹ [51] как варианты поведения руководства фирмы в условиях неопределенности. Анализировалось влияние периодичности ремонта оборудования и организационной среды (organizational environment) фирмы на промышленный ущерб [39]: анализ риска проводился в рамках моделей UML² [38] и MADS³, затем моделировался риск-менеджмент в целях обеспечения безопасности персонала от вероятных промышленных угроз. Эффективность применения системы ERM исследована на материалах девяти фирм различных отраслей промышленности [36]. Исследовались риски возникновения разрывов в цепях поставок товаров и ресурсов как в рамках технологического процесса внутри фирмы [50], так и в цепях внешних поставок [45]. Исследования моделей принятия решений в условиях неопределенности [46] обосновали применимость системного подхода к моделированию сложных задач анализа рисков несовершенства технического обслуживания промышленного оборудования на основе модели UML и метода MOSAR⁴ [41]. Отметим, что если в конце XX века необходимость управления промышленными рисками фирм требовала доказательства [36, 37], то сейчас актуальность этой проблемы не вызывает сомнений.

Среди отечественных исследователей вопросов риска в экономике спектр интересов довольно широк: от международных аспектов до отдельно взятых фирм. Рассматриваются меж-

¹ *Enterprise Risk Management (англ.) – риск-менеджмент предприятия.*

² *Unified Modeling Language (англ.) – унифицированный язык моделирования.*

³ *Analysis Method of Dysfunctional Systems (англ.) – метод анализа дисфункциональных систем.*

⁴ *Method Organized for a Systemic Analysis of Risk (англ.) – метод системного анализа риска.*

дународные финансовые риски [8], риски операций с ценными бумагами на международном рынке [2], логистические риски [18, 26]. Ряд авторов исследуют риски регионального уровня [35]. Например, Шелков А.Б. значительное внимание в своих работах уделяет проблемам региональной безопасности и ее анализу с помощью сценарного подхода [23, 34, 43]. Риски на промышленных предприятиях анализируются с учетом человеческого фактора [1], с помощью деловых игр [6] и механизмов штрафов [15], с точки зрения внедрения инноваций [3, 16]. В целом показано, что производственная деятельность в любой сфере экономики в той или иной степени связана с неопределенностью и риском [37, 51].

Следует отдельно отметить ряд работ отечественных ученых, посвященных вопросам промышленного или производственного риска. Р.М. Качалов в своих работах [19-24] детально рассматривает вопросы управления рисками на различных уровнях, в том числе на региональном и отдельного предприятия. В соавторстве с Ю.А. Слепцовой [21-23] анализ риска социально-экономических систем рассматривается через призму декомпозиции и структурирования экономического пространства предприятия. Рассмотрено выделение «подцелей деятельности» подсистем и выявление присущих им факторов риска [18]. О.Н. Гримашевич уделяет в своих работах [10-15] значительное внимание промышленным рискам и рассматривает все этапы управления данным риском от идентификации до разработки программы управления рисками. О.П. Черникова предлагает использовать операционный рычаг, учитывающий изменение факторов (объема продаж, цены, удельных переменных затрат, суммы постоянных затрат), влияющих на прибыль предприятия, а также на производственный риск предприятия [25]. Затраты в данной статье рассматриваются в целом без выделения затрат на управление риском.

В дальнейшем промышленный риск трактуется как некоторый техногенный фактор вероятностного характера, приводящий к потере активов фирмы¹ [27, 33, 49]. Следовательно, про-

¹ *Коммерческие риски, приводящие к потере доходов, в данной статье не рассматриваются.*

мышленный риск, в отличие от коммерческого, приводит к дополнительным издержкам фирмы (далее «*рисковым издержкам*»), но не приводит к снижению доходов. С учетом вероятностной природы фактора риска будем считать мерой риска математическое ожидание ущерба от фактора риска как некоторого случайного события (аварии, инцидента, остановки оборудования, сбоя производственного процесса и т.д.), измеряемое в денежных единицах издержек фирмы; степень риска представляет собой вероятность этого события.

Рисковые издержки фирмы (ущерб) декомпозируются на два компонента [40]: 1) непредвиденные расходы, обусловленные указанными событиями, в том числе штрафы за превышение размера ущерба окружающей среде над существующими нормами; 2) планируемые расходы, обусловленные необходимостью предотвращения рискованной ситуации или снижения риска ее возникновения.

Кроме того, классифицируем ущерб от риска на внутренний и внешний по отношению к фирме. *Внутренний ущерб* фирмы от фактора риска включает в себя утраченные активы, а также расходы на ликвидацию последствий и восстановительные работы. Размер возможного внутреннего ущерба, очевидно, пропорционален объему производства фирмы, с ростом которого увеличиваются производственные мощности фирмы (производственные площади, оборудование, персонал), т.е. растет объем активов, которые могут потенциально пострадать. Поэтому снижение возможного внутреннего ущерба может быть достигнуто за счет дополнительных расходов на мероприятия по снижению риска, которые в дальнейшем будем называть *добровольными рисковыми издержками* (ДРИ). Структура ДРИ включает в себя расходы на мероприятия по обучению персонала, установке более нового и совершенного оборудования, установке очистных сооружений, систем контроля и сигнализации и пр. *Внешний ущерб* фирмы, называемый далее *обязательными рисковыми издержками*, примем равным сумме штрафов, налагаемых на фирму за превышение уровня ущерба окружающей среде над установленными государством нормами. Таким образом, общие издержки промышленной фирмы, наряду с затратами на производство продукции, также включают в себя добро-

вольные и обязательные рисковые издержки и внутренний ущерб.

Выбор величины ДРИ предопределен следующими противоречивыми факторами. С одной стороны, снижение этих издержек приводит к сокращению общих издержек фирмы, что способствует росту прибыли и дальнейшему расширению производства; однако при этом могут возрасти внутренний ущерб и обязательные рисковые издержки, что, наоборот, повышает общие издержки. С другой стороны, увеличение ДРИ приводит к снижению внутреннего ущерба и обязательных рисковых издержек, однако общие издержки фирмы могут вследствие этого возрасти. В дальнейшем рассматривается модель фирмы без учета обязательных рисковых издержек.

Таким образом, актуальной задачей управления промышленными рисками является выбор функции ДРИ фирмы, оптимальной по критерию общих издержек фирмы. При оптимальной функции ДРИ решение основной задачи фирмы, заключающейся в оптимизации объема производства по критерию прибыли, представляет собой модель управления фирмой с учетом промышленных рисков. Эта проблема является предметом данной статьи.

2. Постановка задачи

Рассмотрим задачу определения неотрицательной вещественной ограниченной сверху функции ДРИ фирмы, минимизирующей функцию общих издержек фирмы:

$$(1) \quad f^*(Q) = \arg \min_{f(\bullet) \in A_f} C_\Sigma(Q, f(\bullet)),$$

$$A_f = \{f(\bullet) \in R^+ : f(\bullet) \leq f^{\max}, f^{\max} \in (0, C_\Sigma)\},$$

$$(2) \quad C_\Sigma(Q, f) = C(Q) + f(Q) + X(Q, f), C_\Sigma(\bullet) \leq^{\max},$$

где Q – объём производства; $C_\Sigma(Q, f)$ – функция общих издержек фирмы; $C(Q)$ – функция издержек на производство продукции; $f(Q)$ – функция ДРИ; $X(Q, f)$ – функция внутреннего ущерба; f^{\max} – максимально возможная сумма ДРИ; C^{\max} – максимально возможная сумма общих издержек исходя из производственной

мощности фирмы; символом « $*$ » обозначены оптимальные значения.

Основную задачу фирмы сформулируем как задачу выбора ограниченного сверх объема производства, максимизирующего прибыль с учетом оптимальной по (1) функции ДРИ:

$$(3) \quad Q^* = \arg \max_{Q \in A_Q} \Pi(Q, f^*(Q)),$$
$$A_Q = \{Q \in R^+ : Q \leq Q^{\max}, Q^{\max} > 0\},$$

$$(4) \quad \Pi(Q, f) = R(Q) - C_{\Sigma}(Q, f^*(Q)),$$

$$(5) \quad R(Q) = p Q,$$

где Π , R – прибыль и доход фирмы; p – цена продукции; Q^{\max} – максимально возможный объем производства с учетом производственной мощности фирмы.

Следует отметить математические модели оптимального управления промышленной безопасностью [4], разработанные для регионального уровня и для отдельного предприятия. Решение задачи происходит в два этапа: на первом этапе определяется стратегия повышения уровня промышленной безопасности на региональном уровне, на втором этапе – определяется уровень промышленной безопасности каждого предприятия региона, сумма которых равна полученному на предыдущем шаге региональному уровню. Для решения задачи применяются аддитивно-мультипликативные модели, использующие показатели затрат на достижение и поддержание определенного уровня промышленной безопасности и показатели уровня промышленной безопасности.

Также в [7] рассмотрены математические модели механизмов снижения рисков возникновения чрезвычайных ситуаций, применяемые для решения задач управления экологической безопасностью. Проанализированы как уровень безопасности, так и уровень риска, представлены различные механизмы управления риском, произведена оценка их экономической эффективности.

Поскольку вопросы управления рисками тесно связаны со сферой страхования, математические модели, используемые для анализа снижения ущерба от непредвиденных ситуаций, встречаются при моделировании механизмов страхования. Например, с точки зрения теоретико-игрового моделирования при решении

задач определения страховых тарифов, взаимного и смешанного страхования [5].

Введем следующие предположения, определяющие границы применимости моделей (1), (3).

1. Гипотеза совершенной конкуренции на рынке товаров фирмы: цена товара для фирмы является экзогенной переменной, т.е. фирма не влияет на рыночную цену:

$$p'_Q = 0.$$

2. Гипотеза убывающей отдачи от расширения масштаба производства соответствует относительно крупной фирме:

$$C''_Q(Q) > 0.$$

3. Гипотеза влияния параметров управления на внутренний ущерб: с увеличением объема производства растут производственные фонды, что приводит к росту возможного ущерба от фактора риска; с увеличением ДРИ снижается внутренний ущерб; функция внутреннего ущерба ограничена сверху вследствие особенностей технологии и ограниченности объема производства:

$$X'_Q(Q, f) > 0, X'_f(Q, f) < 0, X(Q, f) \in [0, X^{max}], X^{max} > 0,$$

где X^{max} – максимально возможный внутренний ущерб.

Функция производственных издержек и функция ущерба, удовлетворяющие гипотезам 2,3, имеют вид

$$(6) C(Q) = BQ^\beta, \beta \in (1, \beta^{max}], \beta^{max} \in (1, 2], B > 0, [28, 29],$$

$$(7) X(Q, f) = \chi(Q)e^{-\xi f}, \xi \in (0, \xi^{max}], \xi^{max} \in (0, 1], \chi(Q)'_Q \geq 0.$$

Будем называть функцию $X(Q, f)$ функцией ожидаемого внутреннего ущерба, в которой функция $\chi(Q)$ есть детерминированный компонент («функция ущерба»), а множитель $e^{-\xi f}$ – плотность экспоненциального распределения вероятности значений f как случайной величины затрат на возмещение потерь от рискованных ситуаций в том случае, если бы систематические затраты на предупреждение риска не осуществлялись. Показательное распределение в приложении к моделированию социально-экономических показателей чаще всего используется в теории надежности. Величина ξ описывает время ожидания при техническом обслуживании, срок службы оборудования. Требование гипотезы 3 ($X'_Q(Q, f) > 0$) обусловлено увеличением ожидаемого ущерба при наступлении неблагоприятного события

в связи с увеличением стоимости имущества, подверженного данному риску [17, 30, 32]. Другими словами, это функция плотности вероятности наступления рискованных ситуаций, измеренных в стоимостном выражении.

Рассмотрим задачу оптимального управления: найти пару $\langle f^*(\cdot), Q^* \rangle$, оптимальную по критериям (1), (3) на соответствующих допустимых множествах для функций издержек и ущерба вида (6), (7).

3. Результаты

На первом этапе определим функцию ДРИ $f^*(\cdot)$ по условию (1) в виде следующего утверждения, доказательство которого приведено в приложении.

Утверждение 1. Для непрерывно дифференцируемых функций $C(\cdot)$, $f(\cdot)$ и функции $X(\cdot)$ вида (7) функция

$$(8) \quad f^*(Q) = \frac{1}{\xi} \ln[\xi \chi(Q)]$$

является решением задачи (1) $\forall Q \in A_Q$ при $\xi \chi(Q) \geq 1$ и $\chi'' \chi - \chi'^2 > 0$.

Логарифмическая функция от объема выпуска фирмы (8) минимизирует общие издержки фирмы при экспоненциально убывающей функции промышленного ущерба.

На втором этапе определим объем производства, при котором прибыль фирмы будет максимальной с учетом функции (8).

Утверждение 2. Для непрерывно дифференцируемых функций $C(\cdot)$, $f(\cdot)$, $\chi(\cdot)$ уравнение

$$(9) \quad p = B\beta Q^{\beta-1} + \frac{\chi'_Q}{\xi \chi(Q)} + \chi'_Q e^{-\xi f}$$

является решением задачи (3), если выполняется

$$(10) \quad B\beta(\beta-1)Q^{\beta-2} - \frac{\chi''_Q}{\xi \chi(Q)^2} < 0 \text{ для } Q \in A_Q.$$

Таким образом, при Q^* , удовлетворяющем (9) и (10), функция прибыли (1) будет иметь максимальное значение.

Исследуем влияние вида функции $\chi(\cdot)$ на границы области допустимых значений параметров функций издержек B , β , ξ ,

в пределах которых уравнение (9) имеет неотрицательное решение. Рассмотрим степенную и экспоненциальную функции $\chi(\cdot)$:

$$\chi(Q) = Q^n, \quad \chi(Q) = e^Q.$$

Утверждение 3. Для функции $\chi(Q)$ вида $\chi(Q) = Q^n$ утверждение 2 имеет силу, если выполняются условия:

$$(11) \quad Q < \beta \sqrt{\frac{2n}{B\beta(\beta-1)\xi}} \quad \text{при } n \geq 2 \text{ для } Q \in A_Q,$$

$$(12) \quad Q < \beta \sqrt{\frac{n^2}{B\beta(\beta-1)\xi}} \quad \text{при } n < 2 \text{ для } Q \in A_Q.$$

Нетрудно заметить, что при $n < 2$, наиболее строгим из двух неравенств является второе, а при $n \geq 2$ нижнюю границу задает первое неравенство.

Утверждение 4. Для функции $\chi(Q)$ вида $\chi(Q) = e^Q$ утверждение 2 имеет силу, если выполняется условие:

$$(13) \quad Q < \beta^{-2} \sqrt{\frac{P}{2B\beta(\beta-1)}} \quad \text{для } Q \in A_Q.$$

4. Анализ результатов моделирования

Рассмотрим некоторые возможные виды нелинейной функции ущерба, удовлетворяющие гипотезе роста ущерба с увеличением масштаба производства, такой, чтобы в эти виды, как частный случай, входила линейная функция. Рассмотрим два варианта функции $\chi(Q)$ – степенную и экспоненциальную. Если $\chi(Q) = Q^n$, тогда в силу (8)

$$(8') \quad f^*(Q) = \frac{1}{\xi} \ln(\xi Q^n) = \frac{\ln \xi}{\xi} + \frac{n}{\xi} \ln Q.$$

Получили функцию вида $f^*(Q) = K_1 + K_2 \ln(Q)$, где K_1 и K_2 – константы.

Рассмотрим влияние параметра ξ на вид функции (8'). В силу (7), ξ может принимать значения из диапазона (0, 1].

$$f^*(Q) = \lim_{\xi \rightarrow 0} \left(\frac{\ln \xi}{\xi} + \frac{n}{\xi} \ln Q \right) = \infty,$$

$$f^*(Q) \Big|_{\xi=1} = n \ln Q.$$

То есть с уменьшением параметра ζ функция (8') неограниченно возрастает, а при увеличении параметра ζ до значений, близких к 1, функция (8') принимает вид логарифмической функции. В данном случае значение объема производства $Q \geq 1$.

Рассмотрим $\chi(Q) = e^Q$, тогда в силу (8)

$$(8'') \quad f^*(Q) = \frac{1}{\xi} \ln(\xi e^Q) = \frac{\ln \xi}{\xi} + \frac{Q}{\xi}.$$

Получили линейную функцию вида $f^*(Q) = K_1 + K_2 Q$, где K_1 и K_2 – константы.

При бесконечно малом значении параметра ζ получаем:

$$f^*(Q) = \lim_{\xi \rightarrow 0} \left(\frac{\ln \xi}{\xi} + \frac{Q}{\xi} \right) = \infty,$$

$$f^*(Q)|_{\xi=1} = Q.$$

Как и в случае (8'), при уменьшении ζ , функция (8'') бесконечно возрастает, а при ζ , близком к 1, принимает вид линейной функции, а именно, биссектрисы угла первой четверти. Выбор в качестве функции $\chi(Q)$ степенной функции накладывает некоторые ограничения на возможные значения объема производства, в то время как экспоненциальная функция $\chi(Q)$ позволяет использовать в качестве значения Q любое действительное число.

Рассмотрим моделирование оптимальных механизмов (8), (9) на примере реальной промышленной фирмы (Самарского подшипникового завода) при различных значениях параметров объекта управления p, B, β, ζ .

В результате моделирования были получены значения, представленные в таблицах 1 и 2. Для функции $\chi(Q) = Q^3$ при всех значениях параметров β и ζ , удовлетворяющих (6) и (7) соответственно, задача (1) не имеет решения при достаточно малых значениях цены p и параметра B (модель 1). То есть при низкой рыночной цене и малом коэффициенте функции производственных издержек фирма выпускает настолько низкий объем продукции, что не имеет дохода, достаточного для совершения ДРИ. С увеличением цены доход фирмы возрастает, и в результате решения (9) получаем два возможных значения Q^* , из которых только $Q^* < 1$ удовлетворяет (10) и (11) (модель 2). Об этом говорилось выше при обсуждении утверждения 3. При

более высоких значениях параметра B производственные издержки фирмы становятся больше, поэтому задача (1) не имеет решения (модели 3-5) при более высоких уровнях внутреннего ущерба и ДРИ. То есть производственные издержки настолько высоки, что доход, получаемый фирмой, не покрывает расходов на управление рисками.

Таблица 1. Результаты моделирования для $\chi(Q) = Q^3$

№ модели	p	B	ξ	β	Q^*
1	1	1	(0; 1]	(1; 2]	$Q^* \notin R$
2	100	1	0,7	1,2	$Q^*_1 = 0,08$ $Q^*_2 = 4018775720$
3	100	[80; $+\infty$)	0,7	1,2	$Q^* \notin R$
4	1000	1000	0,7	1,2	$Q^*_1 = 0,018$ $Q^*_2 = 0,3557$
5	1000	(1257; $+\infty$)	0,7	1,2	$Q^* \notin R$

Таблица 2. Результаты моделирования для $\chi(Q) = e^Q$

№ модели	p	B	ξ	β	Q^*
6	(0; 2)	1	(0; 1]	2	$Q^* < 0$
7	10	1	1	1,3	8
8	10	1	1	(1,8; 2]	
9	10	[6; $+\infty$)	0,7	1,2	
10	100	3	0,7	1,2	$14,306 \cdot 10^6$

Для функции $\chi(Q) = e^Q$ при малых значениях цены p (модель 6) задача (1) не имеет решения по причине, описанной выше для модели 1. С увеличением цены задача (1) имеет решение, что говорит о возросшем доходе, который позволяет осуществлять ДРИ (модель 7). Увеличение параметров β и B ведет к возрастанию производственных издержек, что также не оставляет фирме дохода для совершения ДРИ, поэтому задача (1) не имеет решения (модели 8, 9). Модель 10 иллюстрирует соотношение параметров состояния, при которых цена достаточно

высока, чтобы принести фирме доход, способный покрыть расходы на управление рисками.

Графическая иллюстрация результатов моделирования позволяет наглядно показать исследуемые функции.

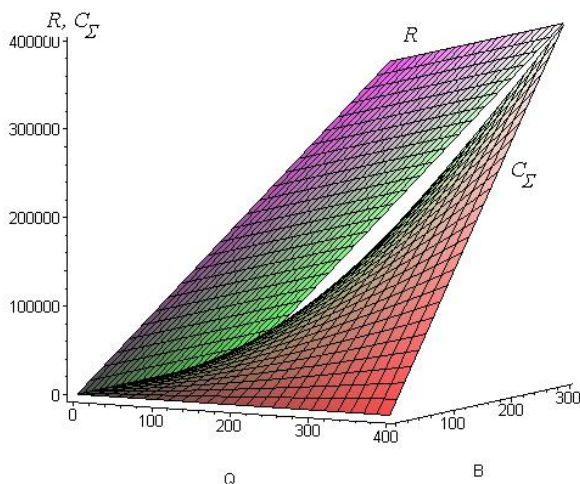


Рис. 1. Зависимость дохода и общих затрат от объема производства и параметра B

На графике (рис. 1) видно, что с увеличением параметра B при возрастающем объеме производства увеличиваются общие затраты. Причем дальнейшее их увеличение ведет к превышению дохода. Таким образом, увеличение общих издержек при постоянной цене не является целесообразным.

Изменение параметра ξ незначительно влияет на значения дохода и общих издержек (рис. 2). Его изменение практически не отражается на графиках функций.

Увеличение параметра β совместно с возрастающим объемом производства приводит к стремительно возрастающим общим издержкам. На графике (рис. 3) видна линия, в каждой точке которой выполняется равенство дохода и общих издержек. Таким образом, фирма может определить границы для

объема производства и параметра β , при которых фирма работает с прибылью.

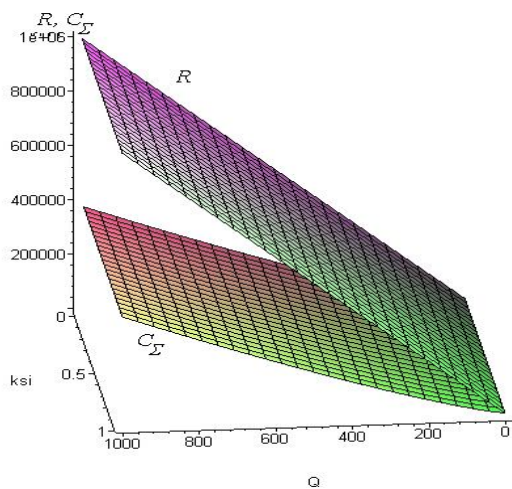


Рис. 2. Зависимость дохода и общих издержек от объема производства и параметра ζ

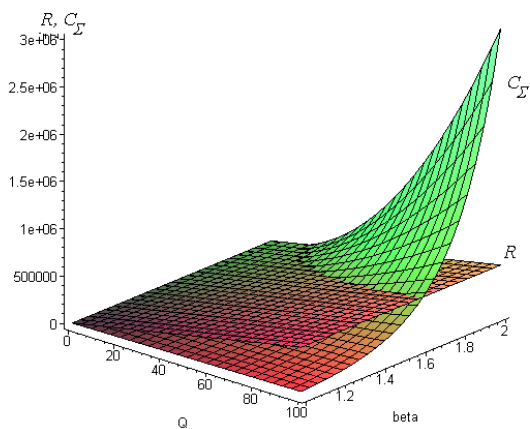


Рис. 3. Зависимость дохода и общих издержек от объема производства и параметра β

5. Выводы

Рассмотрена проблема управления издержками фирмы с учетом промышленного ущерба от непредвиденных производственных ситуаций. Классификация вероятных в этих ситуациях издержек фирмы (рисковых издержек) на добровольные и обязательные позволила выделить задачу оптимизации ДРИ как самостоятельный компонент общей задачи управления издержками фирмы, которая, в свою очередь, является компонентом задачи максимизации прибыли фирмы.

Оптимальная по критерию общих издержек функция ДРИ определена для степенной функции производственных издержек и экспоненциально убывающей функции промышленного ущерба. Оптимальная функция ДРИ представляет собой логарифмическую зависимость от некоторой монотонной функции объема выпуска фирмы. Следовательно, с ростом объема производства рациональная стратегия управления промышленными рисками фирмы заключается в плавном увеличении издержек на предотвращение рискованных ситуаций. Также следует отметить, что данное увеличение целесообразно производить до определенных границ, как показано на рис. 1-3.

Анализ влияния параметров функции издержек и функции промышленного ущерба на область существования решения задачи максимизации прибыли фирмы при оптимальной функции ДРИ позволил сделать следующие выводы. Низкая рыночная цена на товар фирмы приводит к низкому доходу, что в совокупности с высокими значениями параметров функции издержек, во-первых, не позволяет фирме осуществлять ДРИ; во-вторых, вследствие высоких уровней оптимального выпуска при этом значительны размеры внутреннего ущерба. При достаточно высокой рыночной цене, наоборот, показано существование максимизирующей прибыль стратегии (выпуска) фирмы, имеющей возможность осуществлять ДРИ, т.е. наличие дохода фирмы, достаточного для проведения мероприятий по снижению риска. Таким образом, полученная модель позволяет для каждого конкретного случая, описывающего производственный процесс предприятия, установить границы рыночной цены и параметров функции издержек, при которых управление про-

мышленным риском является критическим для существования фирмы.

6. Приложение

Доказательство утверждения 1. Решим задачу (1) методом подстановки в целевую функцию $C_{\Sigma}(Q, f)$ функции $X(Q, f)$. Целевая функция непрерывно дифференцируема и функция $X(Q, f) = \chi(Q) e^{-\xi f}$ непрерывно дифференцируема с частными производными, не равными нулю одновременно (уравнение $X(Q, f) - \chi(Q) e^{-\xi f} = 0$ задает гладкую кривую).

Запишем полученную в результате подстановки функцию

$$C_{\Sigma}(Q, f) = C(Q) + f + \chi(Q) e^{-\xi f}$$

и продифференцируем её

$$(14) C_{\Sigma}(Q, f)'_f = 1 - \xi \chi(Q) e^{-\xi f} = 0.$$

Решив (14), получим функцию ДРИ при условии минимизации общих издержек:

$$f^*(Q) = \frac{1}{\xi} \ln(\xi \chi(Q)).$$

Проверим выполнение достаточного условия минимума функции $C_{\Sigma}(Q, f)$ при $f = f^*$. Для этого определим знак

$$\Delta = \begin{vmatrix} C''_{\Sigma QQ} & C''_{\Sigma Qf} \\ C''_{\Sigma Qf} & C''_{\Sigma ff} \end{vmatrix} \text{ при } f = f^*.$$

$$\begin{aligned} \Delta|_{f=f^*} &= \begin{vmatrix} C''_{\Sigma QQ} & C''_{\Sigma Qf} \\ C''_{\Sigma Qf} & C''_{\Sigma ff} \end{vmatrix} \Big|_{f=f^*} = \begin{vmatrix} V\beta(\beta-1)Q^{\beta-2} + \frac{\chi''}{\xi\chi} & -\frac{\chi'}{\chi} \\ -\frac{\chi'}{\chi} & \xi \end{vmatrix} = \\ &= V\beta(\beta-1)\xi Q^{\beta-2} + \frac{\chi''}{\chi} - \left(\frac{\chi'}{\chi}\right)^2. \end{aligned}$$

Так как $V\beta(\beta-1)\xi Q^{\beta-2} > 0$ в силу (6) и $\frac{\chi''}{\chi} - \left(\frac{\chi'}{\chi}\right)^2 > 0$ по

условию утверждения 1, то $\Delta|_{f=f^*} > 0$.

Данное условие действительно $\forall Q \in A_Q$. Тогда функция $f^*(Q)$ в случае $\xi\chi(Q) \geq 1$ является решением задачи $\min_{f \in A_f} C_\Sigma(X, f)$. ■

Доказательство утверждения 2. Определим значение объема производства Q^* , при котором целевая функция достигает максимального значения. Для этого найдем частную производную (4) по Q :

$$P'_Q(Q, f) = R'_Q(Q) - C'_\Sigma(Q, f) = 0,$$

$$R'_Q(Q) = C'_\Sigma(Q, f).$$

То есть прибыль будет максимальной, если темп роста издержек будет равен темпу роста дохода, т.е. рыночной цене.

$$p = C(Q)'_Q + f^*(Q)'_Q + X(Q, f^*)'_Q,$$

$$p = B\beta Q^{\beta-1} + \frac{\chi'_Q}{\xi\chi(Q)} + \chi'_Q e^{-\xi f}.$$

С учетом $f = f^*$ по (8), получим:

$$p = B\beta Q^{\beta-1} + \frac{2\chi'_Q}{\xi\chi(Q)}.$$

Проверим выполнение достаточного условия максимума функции $\Pi(Q, f)$:

$$\Pi''_{QQ}(Q, f) = R''_{QQ}(Q) - C''_{\Sigma QQ}(Q, f) < 0.$$

$$\Pi''_{QQ}(Q, f) = B\beta(\beta-1)Q^{\beta-2} + \frac{\chi''_{QQ}\xi\chi(Q) - \xi\chi'^2_Q}{(\xi\chi(Q))^2} - \chi''_{QQ}e^{-\xi f}.$$

$$\Pi''_{QQ}(Q, f) = B\beta(\beta-1)Q^{\beta-2} + \frac{\chi''_{QQ}}{\xi\chi(Q)} - \frac{\chi'^2_Q}{\xi\chi(Q)^2} - \chi''_{QQ}e^{-\xi f}.$$

С учетом (8) получим:

$$\Pi''_{QQ}(Q, f) = B\beta(\beta-1)Q^{\beta-2} - \frac{\chi'^2_Q}{\xi\chi(Q)^2}.$$

Достаточное условие максимума функции действует при Q , удовлетворяющих (10):

$$B\beta(\beta-1)Q^{\beta-2} - \frac{\chi'^2_Q}{\xi\chi(Q)^2} < 0. \blacksquare$$

Доказательство утверждения 3. Если $\chi(Q) = Q^n$, тогда (9) примет вид

$$p = B\beta Q^{\beta-1} + 2 \frac{nQ^{n-1}}{\xi Q^n}.$$

Преобразуем полученное уравнение

$$p = \frac{B\beta Q^\beta}{Q} + 2 \frac{n}{\xi Q},$$

$$p = \frac{\xi B\beta Q^\beta + 2n}{\xi Q}.$$

Рассмотрим решение полученного уравнения графическим способом как отыскание точки пересечения горизонтальной прямой $y_1 = p$ с графиком функции $y_2 = \frac{\xi B\beta Q^\beta + 2n}{\xi Q}$. Исследуем функцию y_2 :

$$\lim_{Q \rightarrow 0} \frac{\xi B\beta Q^\beta + 2n}{\xi Q} = \infty.$$

Графики y_1 и y_2 будут иметь точку пересечения, если y_2 будет убывать, т.е. должно выполняться условие $y_2' < 0$:

$$y_2' = B\beta(\beta-1)Q^{\beta-2} - 2 \frac{n}{\xi Q^2} < 0.$$

После преобразования, получим:

$$\frac{\xi B\beta(\beta-1)Q^\beta - 2n}{\xi Q^2} < 0$$

или в силу (6), (7)

$$\xi B\beta(\beta-1)Q^\beta < 2n,$$

$$Q^\beta < \frac{2n}{\xi B\beta(\beta-1)}.$$

Поскольку с учетом (6), (7) $\frac{2n}{\xi B\beta(\beta-1)} > 0$, то можно записать:

$$0 < Q < \sqrt[\beta]{\frac{2n}{\xi B \beta (\beta - 1)}}.$$

Получили условие существования решения уравнения (9).
Рассмотрим далее условие (10) для $\chi(Q) = Q^n$:

$$V\beta(\beta - 1)Q^{\beta-2} - \frac{n^2 Q^{2n-2}}{\xi Q^{2n}} < 0.$$

Преобразуем:

$$\frac{\xi B \beta (\beta - 1) Q^\beta - n^2}{\xi Q^2} < 0.$$

или

$$\xi B \beta (\beta - 1) Q^\beta - n^2 < 0,$$

$$Q^\beta < \frac{n^2}{\xi B \beta (\beta - 1)},$$

$$0 < Q < \sqrt[\beta]{\frac{n^2}{\xi B \beta (\beta - 1)}}.$$

Таким образом, для функции $\chi(Q) = Q^n$ утверждение 2 имеет силу, если выполняются два условия

$$\begin{cases} Q < \sqrt[\beta]{\frac{2n}{\xi B \beta (\beta - 1)}} & \text{при } n \geq 2, \\ Q < \sqrt[\beta]{\frac{n^2}{\xi B \beta (\beta - 1)}} & \text{при } n < 2. \end{cases}$$

Доказательство утверждения 4. Для $\chi(Q) = e^Q$ уравнение (9) принимает вид

$$p = V\beta Q^{\beta-1} + \frac{2}{\xi}.$$

Рассмотрим опять решение данного уравнения с помощью графического метода. Пусть $y_1 = p$ задает горизонтальную прямую и $y_2 = V\beta Q^{\beta-1} + \frac{2}{\xi}$. Точка пересечения графиков функций y_1 и y_2 соответствует решению (9). Проанализируем функцию y_2 :

$$y_2|_{Q=0} = \frac{2}{\xi}.$$

Определим характер монотонности функции y_2 .

$$y_2' = V\beta(\beta-1)Q^{\beta-2} \geq 0 \quad \forall Q \in A_Q.$$

То есть функция y_2 возрастает при $\forall Q \in A_Q$. Значит, если $y_2(0) > p$, то графики функций y_1 и y_2 не пересекаются и (9) не имеет решения. Если $y_2(0) \leq p$, т.е. выполняется условие $\frac{2}{\xi} \leq p$,

тогда графики функций y_1 и y_2 пересекаются в единственной точке, которая является решением (9).

Рассмотрим далее условие (10) применительно к функции $\chi(Q) = e^Q$:

$$V\beta(\beta-1)Q^{\beta-2} - \frac{1}{\xi} < 0$$

или

$$V\beta(\beta-1)Q^{\beta-2} < \frac{1}{\xi}.$$

Объединим условие существования решения уравнения (9) и (10):

$$2V\beta(\beta-1)Q^{\beta-2} < \frac{2}{\xi} \leq p. \blacksquare$$

Литература

1. АБРАМОВА Н.А., КОВРИГА С.В., МАКАРЕНКО Д.И. *Принципы управления качеством проектирования сложных программно-технических комплексов с учетом оценки рисков ошибок человека* // Надежность. – 2014. – №2(49). – С. 73–80.
2. АНОХИН А.А., КАКАЛОВА Е.С. *Управление рисками в операциях на международном рынке ценных бумаг* // Математика. Информатика. Естествознание в экономике и обществе (МИЕСЭКО–2015) труды Всероссийской научной конференции: в 2-х томах. – 2015. – С. 48–52.
3. БАТЬКОВСКИЙ А.М., БАТЬКОВСКИЙ М.А., БОЖКО В.П., КАЛАЧИХИН П.А. и др. *Управление рисками инновационного развития базовых высокотехнологичных отраслей.* – М.: Тезарус, 2015. – 332 с.
4. БУРКОВ В.Н., ГРИЩЕНКО А.Ф., КУЛИК О.С. *Задачи оптимального управления промышленной безопасностью.* – М.: ИПУ РАН. – 2000. – 70 с.
5. БУРКОВ В.Н., ЗАЛОЖНЕВ А.Ю., КУЛИК О.С., НОВИКОВ Д.А. *Механизмы страхования в социально-экономических системах.* – М.: ИПУ РАН, 2001. – 109 с.
6. БУРКОВА И.В., ПОЛОВИНКИНА А.И., СИДОРОВ Е.А., АКАМСИНА Н.В. *Деловые игры как метод исследования систем управления производственными рисками* // Экономика и менеджмент систем управления. – 2016. – Том 19, №1. – С. 87–93.
7. БУРКОВ В.Н., ЩЕПКИН А.В. *Экологическая безопасность.* – М.: ИПУ РАН, 2003. – 92 с.
8. ВЕРЕМЕЕВА О.В. *Институциональные предпосылки управления международными финансовыми рисками* // Федерализм. – 2015. – №3. – С. 141–150.
9. ГРИМАШЕВИЧ О.Н. *Идентификация рисков промышленных предприятий* // Наука и общество. – 2015. – №2(21). – С. 4–9.

10. ГРИМАШЕВИЧ О.Н. *Методологические основы исследования рискообразующих факторов промышленных предприятий* // Вестник Саратовского государственного социально-экономического университета. – 2012. – №1(40). – С. 68–72.
11. ГРИМАШЕВИЧ О.Н. *Оценка качества системы управления рисками промышленных предприятий* // В сб.: Современный взгляд на проблемы качества и управления конкурентоспособностью в условиях внешних вызовов. Материалы международной (очно-заочной) научно-практической конференции / Под ред. А.В. Гугелева, Н.С. Яшина, И.П. Степановой, П.В. Старцева. – 2015. – С. 89–92.
12. ГРИМАШЕВИЧ О.Н. *Понятие риска применительно к качеству продукции промышленного предприятия* // Вестник Тверского государственного университета. Серия: Экономика и управление. – 2013. – №17. – С. 164–172.
13. ГРИМАШЕВИЧ О.Н. *Разработка программы управления рисками промышленных предприятий* // Наука и общество. – 2013. – №3(12). – С. 45–49.
14. ГРИМАШЕВИЧ О.Н. *Система управления рисками промышленного предприятия: методология и практика*. – Саратов: КУБиК, 2012.
15. ДИНОВА Н.И., ЩЕПКИН А.В. *Управление деятельностью предприятий механизмом штрафов* // XII Всероссийское совещание по проблемам управления ВСПУ-2014, 16–19 июля 2014 г., Москва. – С. 5643–5647.
16. ДРОНОВА Ю.В. *Вероятностная оценка рисков инноваций в условиях неопределенности* // Экономика и управление: проблемы, решения. – 2016. – №1. – С. 69–72.
17. ЕРМАСОВ С.В., ЕРМАСОВА Н.Б. *Страхование*: учебник. 2-е изд., перераб. и доп. – М.: Высшее образование. 2008. – 613 с.

18. ЗАВЬЯЛОВА О.В. *Международные перспективы управления таможенными рисками цепей поставки на территории Евразийского экономического союза* // Сборник материалов Всероссийской научно-практической конференции «Таможенные чтения – 2015. Евразийский экономический союз в условиях глобализации: вызовы, риски, тенденции», 23–27 ноября 2015 г., Санкт-Петербург / Под общ. ред. С.Н. Гамидуллаева. – 2015. – С. 51–60.
19. КАЧАЛОВ Р.М. *Моделирование феномена экономического риска как объекта управления* // Сборник научных трудов XVIII Международной научно-практической конференции «Системный анализ в проектировании и управлении», 1–3 июля 2014 г., Санкт-Петербург. – 2014. – С. 151–154.
20. КАЧАЛОВ Р.М. *Операционная концепция управления экономическим риском в системном пространстве* // В кн.: Управление рисками в экономике: проблемы и решения / Под ред. С.Г. Опарина. – СПб: ФГАОУВО «Санкт-Петербургский политехнический университет Петра Великого», 2015. – С. 9–31.
21. КАЧАЛОВ Р.М., СЛЕПЦОВА Ю.А. *Идентификация факторов риска на основе декомпозиции экономического пространства предприятия* // Вестник Челябинского государственного университета. – 2016. – №14(396). – С. 86–94.
22. КАЧАЛОВ Р.М., СЛЕПЦОВА Ю.А. *Качество управления предприятием и феномен экономического риска* // Управленческие науки в современном мире. – 2016. – Т. 2, №2. – С. 15–19.
23. КАЧАЛОВ Р.М., СЛЕПЦОВА Ю.А. *Цели и факторы риска в задаче управления социально-экономической системой* // Сборник трудов IV Международной научно-практической конференции-биеннале «Системный анализ в экономике – 2016», 9–11 ноября 2016 г., Москва / Под ред. Г.Б. Клейнера, С.Е. Щепетовой. – 2016. – С. 118–122.
24. КАЧАЛОВ Р.М., СТАВЧИКОВ А.И. *Управление экономическим риском инновационного развития предприятия* // Вестник РАЕН. – 2015. – №2. – С. 44–48.

25. КУЛЬБА В.В., ЧЕРНОВ И.В., ШЕЛКОВ А.Б. *Управление региональной безопасностью* // Материалы Восьмой международной конференции «Управление развитием крупномасштабных систем» (MLSD'2015), 29 сентября – 01 октября 2015 г., Москва / Под общ. ред. С.Н. Васильева, А.Д. Цвиркуна. – 2015. – С. 150–152.
26. ЛЕВКИН Г.Г., ТЫРНОВА Е.А. *Управление логистическими рисками при организации доставки грузов в международном сообщении* // Инновационная экономика и общество. – 2015. – №4(10). – С. 75–80.
27. СМИТ А. *Исследование о природе и причинах богатства народов*. – М.: Эксмо, 2007. – 960 с.
28. УОЛТЕРС А.А. *Производственные функции и функции затрат* // Вехи экономической мысли. Т.2 Теория фирмы / Под ред. В.М. Гальперина. – СПб: Экономическая школа, 1999. – 534 с.
29. ХЕЙ Д., МОРРИС Д. *Теория организации промышленности*. В 2 т. / Пер. с англ. под ред. А.Г. Слуцкого. – Санкт-Петербург: Экономическая школа, 1999. – Т. 1. – 384с., Т. 2. – 592 с.
30. ХОХЛОВ Н.В. *Управление риском: Учеб. пособие для вузов*. – М.: Изд-во «Юнити-Дана», 2001. – 239 с.
31. ЧЕРНИКОВА О.П. *Методика расчета показателей операционного рычага для управления производственным риском промышленного предприятия* // Science Time. – 2015. – №8(20). – С. 239–245.
32. ШАХОВ В.В. *Страхование: Учебник для вузов*. – М.: ЮНИТИ, 2003. – 311 с.
33. ШУМПЕТЕР Й. *История экономического анализа*. Т. 2. –, СПб: Экономическая школа, 2004.
34. ШУЛЬЦ В.Л., КУЛЬБА В.В., ШЕЛКОВ А.Б., ЧЕРНОВ И.В. *Сценарий анализа в управлении региональной безопасностью* // Вопросы безопасности. – 2016. – №3. – С. 41–79.
35. ЩЕПКИН А.В., ГОЛЕВ С.А. *Управление уровнем риска в регионе механизмом штрафа* // XII Всероссийское совещание по проблемам управления ВСПУ-2014, 16–19 июля 2014 г., Москва. – 2014. – С. 5682–5689.

36. ARENA M., ARNABOLDI M., AZZONE J. *Is enterprise risk management real?* // J. of Risk Research. – 2011. – Vol. 14. – P. 779–797.
37. BERKELEY D., HUMPHREYS P.C., THOMAS R.D. *Project risk action management* // Construction Management and Economics. – 1991. – Vol. 9. – P. 3–17.
38. BOULOIZ H., TKIOUAT M., GARBOLINO E., BENDAHA T. *Contribution to risk management in industrial maintenance* // Proc. of IEEE International Conference on Industrial Engineering and Systems Management – 2013,(IEEE IESM–2013), October 28–30, 2013, Rabat, Morocco. – Article number 6761470.
39. BOULOIZ H., GARBOLINO E., TKIOUAT M. *Contribution of UML model to risk analysis of an industrial system* // Reliability, Risk and Safety: Back to the Future. – 2010. – P. 127–134.
40. CROPLEY C.H. *The case for truly integrated cost and schedule risk analysis* // Handbook of Research on Leveraging Risk and Uncertainties for Effective Project Management, 29 November 2016. – 2016. – P. 76–108.
41. GALLAB M., TKIOUAT M., BOULOIZ H., GARBOLINO E. *Model for developing a database for risk analysis* // Proc. of IEEE International Conference on Industrial Engineering and Systems Management – 2015, (IEEE IESM–2015), October 21–23, 2015, Seville, Spain. – P. 833–841.
42. Hanson D., White R. *Regimes of risk management in corporate annual reports: a case study of one globalizing Australian company* // J. of Risk Research. – 2004. – Vol. 7. – P. 445–460.
43. KULBA V., SHELKOV A., CHERNOV I., ZAIKIN O. *Scenario analysis in the management of regional security and social stability* // Intelligent systems reference library. – 2016. – Vol. 98. – P. 249–268.
44. Oehmen J., ZIEGENBEIN A., ALARD R. SCHÖNSLEBEN P. *System-oriented supply chain risk management* // Production Planning & Control The Management of Operations. – 2009. – Vol. 20. – P. 343–361.
45. PAZDNIKOVA N.P., SHIPITSYNA S.Y. *Stress analysis in managing the region's budget risks, Risk factors for the regional economic growth* // Economy of Region. – 2014 – Vol. 3(39). – P. 208–217.

46. De ROCQUIGNY E. *Modelling Under Risk and Uncertainty: An Introduction to Statistical // Phenomenological and Computational Methods.* – 2012. – No 1. – 434 p.
47. SAPIRO E.S., MIROLJUBOVA T.V. *Risk factors for the regional economic growth // Economy of Region.* – 2008. – Vol. 1. – P. 39–49.
48. SHORIKOV A.F. *Dynamic model of minimax control over economic security state of the region in the presence of risks // Economy of Region.* – 2012. – Vol. 2(30). – P. 258–266.
49. STUART MILL J. *Essays on economics and society. Collected works of John Stuart Mill.* – Toronto: University of Toronto Press, 1967. – 404 p.
50. THUN J.-H., DRÜKE M., HOENIG D. *Managing uncertainty – an empirical analysis of supply chain risk management in small and medium-sized enterprises // Int. J. of Production Research.* – 2011. – Vol. 49. – P. 5511–5525.
51. VARCHOLOVÁ T. *Active enterprise risk management // Ekonomicky casopis.* – 2003. – Vol. 51. – P. 997–1010.

THE OPTIMAL COST FUNCTION OF THE FIRM TO PREVENT INDUSTRIAL RISKS

Elena Rostova, Samara National Research University, Samara, Cand.Sc., assistant professor (el_rostova@mail.ru).

Mihail Geraskin, Samara National Research University, Samara, Doctor of Science, professor (innovation@ssau.ru).

Abstract: The problem of determining the cost function of the firm to prevent industrial risks (voluntary risk costs), the optimal by the criterion of total costs. For the exponential function of production costs and exponentially decreasing functions of the industrial damage is determined as a function of voluntary risk costs. The analysis of influence of parameters of the function of industrial damage to the area of existence of the solution of the profit maximization problem of the firm under optimal functions of voluntary risk costs. Produced by testing the theoretical results on numerical examples, provides a graphic illustration of simulation results. Analyzed the impact of the market price of commodities, production costs of firms in the voluntary risk costs. The resulting model allows for each case, describing the production process of the enterprise, to establish the boundaries of market prices and parameters of the cost function, in which the management of industrial risk is critical for the existence of the company.

Keywords: industrial risk, industrial damages, costs, optimization, risk management.

*Статья представлена к публикации
членом редакционной коллегии А.Г. Чхартишвили.*

*Поступила в редакцию 08.06.2017.
Опубликована 30.11.2017.*

УДК 519.23
ББК 2.2.22.172

ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ТЕРМОМЕТРИЧЕСКИХ ДАННЫХ В ДИАГНОСТИКЕ МОЛОЧНЫХ ЖЕЛЕЗ

Лосев А. Г.¹, Левшинский В. В.²

(Волгоградский государственный университет, Волгоград)

Работа посвящена разработке нового метода интеллектуального анализа медицинских термометрических данных, предназначенного для создания на его основе консультативных интеллектуальных диагностических систем. Рассмотрены подходы к анализу данных микроволновой радиотермометрии, полученных для диагностики заболеваний молочной железы. Представлен метод формирования информационных признаков на базе количественного описания медицинских знаний о поведении температурных полей молочных желез.

Ключевые слова: интеллектуальный анализ данных, микроволновая радиотермометрия, консультативные интеллектуальные системы.

1. Введение

Разработка диагностических технологий, основанных на интеграции современных достижений в инженерии, информационных технологиях, медицине и математических методах искусственного интеллекта, является крайне актуальной задачей. В первую очередь это связано с необходимостью создания новых методик функциональной диагностики, основанных на динамическом картировании физических полей и излучений

¹ Александр Георгиевич Лосев, доктор физико-математических наук, профессор (alexander.losev@volsu.ru).

² Владислав Викторович Левшинский, студент (vladi.lev.email@gmail.com).

организма человека. Однако применение современного медицинского оборудования в процессе решения одних проблем зачастую порождает другие. В настоящее время трудности диагностики возникают не из-за дефицита информации, а из-за недостаточной эффективности методов ее обработки. В какой-то мере решение указанных проблем обеспечивается созданием систем интерпретации и анализа медицинских данных. При этом наибольший интерес вызывает разработка консультативных интеллектуальных систем, т.е. экспертных систем, содержащих механизм объяснения и обоснования предлагаемых решений на языке, понятном пользователю [5].

Целью данного исследования является разработка некоторых методов анализа и интерпретации медицинских данных, получаемых с помощью микроволновой радиотермометрии.

2. Обзор литературы

Первые попытки обоснования возможности применения радиотермометрии для диагностики рака молочной железы были предприняты в середине 70-х годов 20 века [11, 12, 14]. В указанных работах были предложены первые математические модели, описывающие температурные поля молочных желез в терминах решений уравнений с частными производными. Дальнейшие исследования показали, что в подобных моделях следует учитывать физико-химические свойства биотканей [21], выбор реологической модели крови [15], сложное внутреннее строение молочных желез и ряд других параметров [9]. Однако, учитывая достаточно широкий разброс даже основных параметров указанных моделей, их применение в диагностических системах пока не представляется возможным.

Одним из наиболее популярных средств повышения эффективности диагностики является автоматизация обработки данных с использованием методов искусственного интеллекта. Наиболее популярным здесь является применение искусственных нейронных сетей [10, 13], байесовского классификатора [19], генетических алгоритмов [16], алгоритмов символьного обучения [18], нечеткой логики [20].

В настоящее время большинство экспертных систем предлагает врачу свои решения либо в детерминированной форме однозначного заключения, либо в виде вероятностных оценок каждого из возможных диагнозов. Одновременно идет процесс формирования новых подходов к созданию диагностических систем, «объясняющих» предлагаемые ими решения [4, 5].

3. Описание задачи

Микроволновая радиотермометрия – биофизический метод неинвазивного обследования, заключающийся в измерении внутренних и поверхностных температур тканей по интенсивности их теплового излучения в микроволновом (РТМ) и инфракрасном (ИК) диапазонах соответственно. В течение последнего десятилетия данный метод получил распространение в различных областях медицины. Одновременно возник и ряд проблем. Существующий на данный момент диагностический комплекс РТМ-01-РЭС является системой поддержки принятия решений специалиста высокой квалификации. Сложность восприятия информации, возникающая у медицинского персонала без специальной длительной подготовки, значительно снижает потенциальную возможность использования термометрической аппаратуры в скрининге. Таким образом, актуальной задачей является создание экспертной системы, обладающей возможностями обоснования предполагаемого диагностического решения.

Особую сложность в данной проблеме вызывает нахождение высокоинформативных признаков заболеваний. Как отмечается большинством специалистов [6], на будущее качество алгоритмов классификации влияют качественный и количественный составы пространства информационных признаков.

Основной задачей данного исследования является разработка метода формирования пространства информационных признаков.

4. Качественные составляющие информационных признаков

Вначале опишем подробнее существующую методiku диагностики рака молочной железы по данным микроволновой радиотермометрии. Комплекс РТМ-01-РЭС позволяет оценивать функциональное состояние тканей путем измерения внутренней температуры (РТМ) на глубине до 5 см и температуры кожи (ИК). Обследование пациентки начинается с измерения температур в опорных точках $T1$ и $T2$: первая расположена в центре грудной клетки сразу под и между молочными железами, вторая – непосредственно под мечевидным отростком. Далее измерения проводятся в 10 точках на каждой железе и в аксиллярной области (схема представлена на рис. 1).

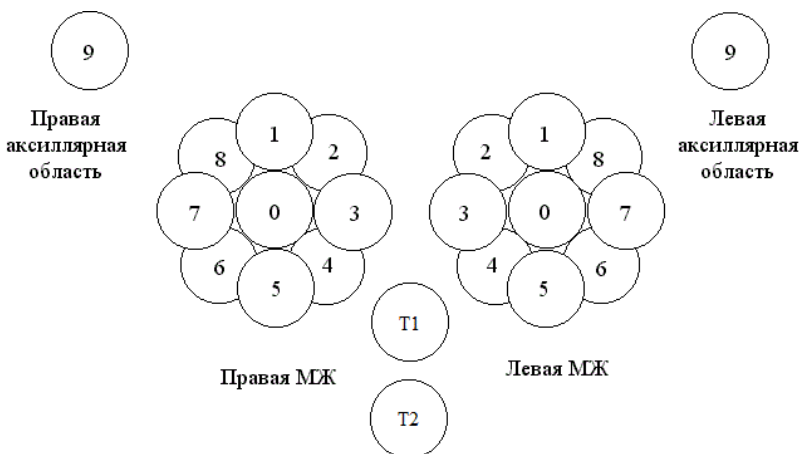


Рис. 1. Схема обследования молочной железы

На основе данных, предоставленных онкологическими центрами России, была сформирована экспертная база термометрических данных. В настоящее время она включает в себя информацию о 734 молочных железах пациенток, которые делятся на два контрольных класса: «Здоровые» – 148 молочных желез и «Больные» – 586 молочных желез. Статистический анализ используемых термометрических данных был проведен в [8].

В том числе доказано, что существующий объем выборки является достаточным для проведения исследований и анализа данных.

В ходе исследований и анализа данных специалистами были выявлены следующие признаки рака молочной железы [1, 2, 7] (далее будем называть их качественными):

- повышенная величина термоасимметрии между одноименными точками молочных желез;
- повышенный разброс температур между отдельными точками в пораженной молочной железе;
- разница температур сосков;
- повышенная температура соска в пораженной молочной железе по сравнению со средней температурой молочной железы с учетом возрастных изменений температуры;
- соотношение кожной и глубинной температур и некоторые другие.

Важным этапом создания эффективной консультационной интеллектуальной системы является математическое описание данных признаков, а также выявление их количественных характеристик.

На первом этапе, опираясь на существующие медицинские знания, а также модели поведения температурных полей, попытаемся классифицировать существующие термометрические диагностические признаки, а также дополнить их.

1. Группа признаков, характеризующая асимметрию температурных полей молочных желез.

Заметим, что данный класс признаков исходит из гипотезы о «зеркальной» симметрии температурных полей правой и левой молочных желез здоровых пациенток. Данная гипотеза используется при анализе термометрических данных не только молочных желез, но практически всех парных органов человека [9]. Опишем вначале известные признаки заболеваний, относящиеся к данному классу.

1.1. Повышенное значение разности температур между одноименными точками правой и левой молочных желез. В качестве характеристик, описывающих этот эффект, могут быть использованы функции вида $|t_{i,пр} - t_{i,л}|$ или $(t_{i,пр} - t_{i,л})$, где

$t_{i,пр}$ и $t_{i,л}$ – температуры в i -х точках правой и левой молочных желез соответственно.

Отметим, что в данном признаке медики традиционно используют $i = 1, \dots, 8$, выделяя разность температур сосков в отдельный признак.

1.2. Повышенная разница температур сосков правой и левой молочных желез: $|t_{0,пр} - t_{0,л}|$, или $(t_{0,пр} - t_{0,л})$.

1.3. Повышенное среднеквадратичное значение разностей температур между одноименными точками правой и левой молочных желез:

$$(1) \sqrt{\sum_{i=0}^8 \frac{(t_{i,пр} - t_{i,л})^2}{9}}.$$

На самом деле асимметрия полей температур молочных желез может быть описана различными функциями вида

$$(2) g(f(t_{0,пр}, \dots, t_{n,пр}) - f(t_{0,л}, \dots, t_{n,л})),$$

где $f(t_{0,пр}, \dots, t_{n,пр})$ – значение функции температур для точек t_0, \dots, t_n правой молочной железы, $f(t_{0,л}, \dots, t_{n,л})$ – значение аналогичной функции температур для точек t_0, \dots, t_n левой молочной железы, а $g(x)$ – некоторая функция одного переменного. Таким образом, значительно расширяется множество исследуемых параметров. В частности, таким способом можно описать следующие характеристики.

1.4. Разница средних значений температур «зеркально-симметрично» расположенных подобластей молочных желез, например:

$$(3) \left| \frac{t_{0,пр} + t_{i,пр} + t_{i(\text{mod}8)+1,пр}}{3} - \frac{t_{0,л} + t_{i,л} + t_{i(\text{mod}8)+1,л}}{3} \right|,$$

где $i = 1, \dots, 8$.

1.5. Разница среднеквадратичных отклонений температур молочных желез:

$$(4) \left| \left(\sqrt{\sum_{i=1}^8 \frac{(t_i - t_{cp})^2}{8}} \right)_{пр} - \left(\sqrt{\sum_{i=1}^8 \frac{(t_i - t_{cp})^2}{8}} \right)_{л} \right|,$$

где

$$(5) \quad t_{\text{cp}} = \sum_{i=1}^8 \frac{t_i}{8}.$$

2. Группа признаков, характеризующих повышенный разброс температур в пораженной молочной железе.

2.1. Повышенное среднеквадратичное отклонение температур в одной из молочных желез:

$$(6) \quad \sqrt{\sum_{i=0}^8 \frac{(t_i - \overline{t_{\text{cp}}})^2}{9}},$$

где

$$(7) \quad \overline{t_{\text{cp}}} = \sum_{i=0}^8 \frac{t_i}{9}.$$

2.2. Повышенный разброс температур между отдельными точками в пораженной молочной железе: $t_{\text{cp}} - t_i$, где t_i – температура в i -й точке молочной железы.

На самом деле разброс температур молочных желез может быть описан различными функциями вида

$$(8) \quad g(f_1(t_0, \dots, t_n) - f_2(t_0, \dots, t_n)),$$

где $f_1(t_0, \dots, t_n)$ и $f_2(t_0, \dots, t_n)$ – функции температур точек t_0, \dots, t_n соответствующей молочной железы, а $g(x)$ – некоторая функция одного переменного. В частности, таким способом можно описать следующие характеристики.

2.3. Повышенные средние значения температур соседних точек вплоть до средней температуры молочной железы, т.е. функции вида:

$$(9) \quad t_{\text{cp}} - \frac{t_i + t_{i(\text{mod}8)+1} + t_{i(\text{mod}8)+2}}{3}.$$

3. Группа признаков, характеризующих повышенное значение температуры соска в пораженной молочной железе.

3.1. Аномальная разность температуры соска и средней температуры молочной железы: $t_0 - t_{\text{cp}}$.

3.2. Аномальная разность температуры соска и температур отдельных точек молочной железы: $t_0 - t_i$, $i = 1, \dots, 8$.

На самом деле аномальные значения температуры соска по отношению к другим параметрам можно описать функциями вида

$$(10) g(t_0 - f_1(t_1, \dots, t_n)),$$

где $f_1(t_1, \dots, t_n)$ – функция температур точек t_1, \dots, t_n молочной железы, а $g(x)$ – некоторая функция одного переменного. В частности, таким способом можно описать следующие характеристики.

3.3. Аномальная разность температуры соска и средней температуры различных подобластей молочной железы, например:

$$(11) t_0 - \frac{t_i + t_{i(\text{mod}8)+1}}{2}.$$

4. Группа признаков, характеризующих соотношение кожной и глубинной температур:

4.1. Аномальное значение разности между кожной и глубинной температурами точки пораженной молочной железы (внутренний градиент): $(t_{i,\text{РТМ}} - t_{i,\text{ИК}})$, где $t_{i,\text{РТМ}}$ – глубинные и $t_{i,\text{ИК}}$ – кожные температуры в i -й точке молочной железы.

Разность температур молочных желез, измеренных в РТМ и ИК диапазонах (так называемый внутренний градиент) может быть описана функциями вида

$$(12) g(f(t_{0,\text{РТМ}}, \dots, t_{n,\text{РТМ}}) - f(t_{0,\text{ИК}}, \dots, t_{n,\text{ИК}})),$$

где $f(t_0, \dots, t_n)$ – функция температур точек t_0, \dots, t_n молочной железы, а $g(x)$ – некоторая функция одного переменного.

Отметим, что сформированный выше набор функций представляет собой расширенное описание известных качественных признаков и получен на базе уже известных медицинских фактов. Но особенно важной, хотя и достаточно сложной задачей, является выявление новых знаний.

5. Признаки, базирующиеся на параметрах физико-математических моделей поведения температурных полей.

Как было отмечено выше, за последние годы было построено несколько математических моделей, описывающих поведение температурных полей молочных желез с помощью уравнений в частных производных второго порядка [9]. Таким образом, для обнаружения аномалий температурных полей пациенток может оказаться интересным изучение поведения разностных аналогов вторых производных функции температур.

Заметим, что в предъявленном выше наборе функций присутствуют как функции температур, так и разностные аналоги их производных по различным направлениям.

Например, величина $(t_0 - t_i)$ является разностным аналогом производной в радиальном направлении (радиальный градиент). Аналогично величина $(t_{i,р\text{тм}} - t_{i,и\text{к}})$ является разностным аналогом производной во внутреннем направлении (внутренний градиент). Учитывая вышесказанное, в работе было решено рассматривать и разностные аналоги вторых производных функций температур, т.е. функции вида:

$$(13) \quad \begin{aligned} & \nabla(t_0 - f(t_1, \dots, t_n)) = \\ & = t_{0,р\text{тм}} - f(t_{1,р\text{тм}}, \dots, t_{n,р\text{тм}}) - (t_{0,и\text{к}} - f(t_{1,и\text{к}}, \dots, t_{n,и\text{к}})) \end{aligned}$$

Были предложены и иные математические описания возможных аномалий поведения температурных полей. Всего при исследовании рассмотрено порядка 900 числовых функций.

Перейдем к формальному описанию задачи поиска информативных признаков. Введем следующие обозначения.

Пусть t_i^j , $i = 0, \dots, 9$ – температуры i -й точки правой молочной железы j -й пациентки в диапазоне РТМ;

t_i^j , $i = 10, \dots, 19$ – температура $[i - 10]$ точки правой молочной железы j -й пациентки в диапазоне ИК;

t_i^j , $i = 20, \dots, 29$ – температура $[i - 20]$ точки левой молочной железы j -й пациентки в диапазоне РТМ;

t_i^j , $i = 30, \dots, 39$ – температура $[i - 30]$ точки левой молочной железы j -й пациентки в диапазоне ИК;

$t_{40}^j = T1$, $t_{41}^j = T2$ – глубинные температуры опорных точек j -й пациентки;

$t_{42}^j = T1$, $t_{43}^j = T2$ – кожные температуры опорных точек j -й пациентки.

Тогда обучающую выборку можно представить в виде матрицы

$$(14) S = \begin{pmatrix} t_0^1 & \cdots & t_{43}^1 \\ \vdots & \ddots & \vdots \\ t_0^k & \cdots & t_{43}^k \\ t_0^{k+1} & \cdots & t_{43}^{k+1} \\ \vdots & \ddots & \vdots \\ t_0^n & \cdots & t_{43}^n \end{pmatrix},$$

где t_i^j при $j = 1, \dots, k$ – температуры молочных желез пациенток класса «Здоровые», t_i^j при $j = k + 1, \dots, n$ – температуры молочных желез пациенток класса «Больные».

Пусть f^q – q -я исследуемая функция, рассматриваемая на множестве векторов $\Omega_q = \{(t_0^j, \dots, t_{43}^j)\}$, где $j = 1, \dots, n$. Обозначим $f_j^q = f^q(t_0^j, \dots, t_{43}^j)$.

Пусть X_α – некоторое множество числовой прямой R , такое что $f_j^q \in X_\alpha$ для некоторых $j = k + 1, \dots, n$, и $f_j^q \in R \setminus X_\alpha$ для всех $j = 1, \dots, k$. Тогда «характерной» областью множества значений f^q будем называть такое подмножество числовой прямой X , что $X_\alpha \subset X$ для всех возможных значений индекса α , но при этом $f_j^q \in R \setminus X$ для всех $j = 1, \dots, k$.

Характеристическим признаком будем называть пару (f^q, X) , где X – «характерная» область множества значений f^q .

Алгоритм поиска характеристических признаков заключается в нахождении «максимальных» подобластей, которым принадлежат значения функции, соответствующие молочным железам одного класса, но не принадлежат значения функции, соответствующие молочным железам другого. Отметим также, что зачастую удобнее вместо всей «характерной» области множества значений f^q рассматривать ее связные подмножества.

Для оценки эффективности алгоритмов классификации, в частности и диагностических методов в медицине, традиционно используются следующие показатели.

Чувствительность (Sensitivity), также известная как полнота (Recall), вычисляется по формуле

$$(15) Sens = \frac{TP}{TP + FN},$$

где TP – количество молочных желез класса «Больные», которые были классифицированы как «Больные», FN – количество молочных желез класса «Больные», которые не были классифицированы как «Больные».

Специфичность (Specificity), вычисляется по формуле

$$(16) \text{Spec} = \frac{TN}{TN + FP},$$

где TN – количество молочных желез класса «Здоровые», которые не были классифицированы как «Больные», FP – количество молочных желез класса «Здоровые», которые были классифицированы как «Больные».

Точность (Precision), вычисляется по формуле

$$(17) \text{Prec} = \frac{TP}{TP + FP}.$$

Достоверность (Accuracy), вычисляется по формуле

$$(18) \text{Acc} = \frac{TP + TN}{TP + FP + FN + TN}.$$

Также отметим, что в медицине важнейшими характеристиками диагностического метода считают чувствительность и специфичность. Для характеристических признаков, очевидно, выполнено $\text{Spec} = 1$, поэтому их эффективность оценивается чувствительностью и достоверностью.

Несколько характеристических признаков с лучшими показателями истинно-положительных результатов диагностики приведены в таблице 1.

Таблица 1. Характеристические признаки

Признак	Sens	Acc
$\nabla \left(t_0 - \frac{t_1 + t_2 + t_3}{3} \right), (-\infty; -1, 0)$	0,08	0,27
$t_{1, \text{np}} - t_{1, \text{л}}, (-\infty; -1, 0)$	0,07	0,26
$\sqrt{\sum_{i=0}^8 \frac{(t_{i, \text{np}} - t_{i, \text{л}})^2}{9}}, (1, 08; \infty)$	0,07	0,26

Во втором и третьем признаках используются значения температур в РТМ-диапазоне.

Очевидно, что характеристические признаки в состоянии обнаружить лишь достаточно серьезные аномалии поведения температурных полей. При использовании набора наиболее значимых характеристических признаков было обнаружено 43% молочных желез класса «Больные».

Далее, основываясь на методике поиска характеристических признаков, попробуем построить набор высокоинформативных признаков.

Высокоинформативным признаком будем называть тройку (f^q, V, X) , где f^q – функция, описывающая поведение температурных полей, $V = I(f^q, X)$ – информативность признака, X – «информативная» область множества значений функции f^q .

Определим указанные выше характеристики. Под информативностью понимается количественный параметр, определяющий, насколько хорошо закономерность описывает различия между искомой и отделяемой группами. В качестве $I(f^q, X)$ в данной работе использовались следующие характеристики.

Статистическая информативность, вычисляемая по формуле

$$(19) \quad ST(f^q, X) = -\ln \left(\frac{C_k^h C_{(n-k)}^s}{C_{k+(n-k)}^{h+s}} \right),$$

где n – количество векторов с температурными данными пациенток в обучающей выборке; k – количество векторов с температурными данными пациенток класса «Здоровые» в обучающей выборке; h – количество молочных желез класса «Здоровые», для которых $f^q \in X$, а s – количество молочных желез класса «Больные», для которых $f^q \in X$.

Эвристическая информативность, вычисляемая по формуле

$$(20) \quad HR(f^q, X) = \frac{\max \left\{ \frac{k}{h}, \frac{n-k}{s} \right\}}{\min \left\{ \frac{k}{h}, \frac{n-k}{s} \right\}}.$$

Энтропийная информативность, вычисляемая по формуле

$$(21) \quad EN(f^q, X) = H\left(\frac{k}{k+(n-k)}, \frac{(n-k)}{k+(n-k)}\right) - \frac{k+(n-k)-h-s}{k+(n-k)} H\left(\frac{k-h}{k+(n-k)-h-s}, \frac{(n-k)-s}{k+(n-k)-h-s}\right),$$

где $H(q_0, q_1) = -q_0 \log_2 q_0 - q_1 \log_2 q_1$ – математическое ожидание количества информации.

Комбинированная информативность, которая вычисляется по формуле

$$(22) \quad CI(f^q, X) = \sqrt[n]{I_1(f^q, X) I_2(f^q, X) \dots I_n(f^q, X)},$$

где I_1, I_2, \dots, I_n – другие информативности;

Среднее гармоническое информативностей, которое вычисляется по формуле

$$(23) \quad HM(f^q, X) = \frac{n}{\frac{1}{I_1(f^q, X)} + \dots + \frac{1}{I_n(f^q, X)}}.$$

Далее, пусть

$$(24) \quad I_s = \sup_{X_\alpha} I(f^q, X_\alpha),$$

где X_α – все возможные подобласти числовой прямой, на которых указанная информативность имеет смысл. «Информативной» областью множества значений функции f^q будем называть подмножество числовой прямой, обеспечивающее достижение (или приближение с заданной точностью) величины I_s .

Для j -х молекулярных признаков (f_j^q, V, X) считается выполненным, если $f_j^q \in X$.

Некоторые примеры высокоинформативных признаков приведены в таблице 2. В данном случае при поиске использовался показатель комбинированной информативности статистической и эвристической информативностей.

Во втором признаке используются значения температур в ИК-диапазоне, в четвертом и пятом – в РТМ-диапазоне.

Таблица 2. Высокоинформативные признаки

Признак	Sens	Spec	Acc
$t_{1,np} - t_{1,l}, (-\infty, -0,9)$	0,14	0,99	0,31
$\nabla\left(t_0 - \frac{t_1 + t_2 + t_3}{3}\right), (-\infty, -0,89)$	0,13	0,99	0,30
$\nabla(t_0 - t_9), (-\infty, -1,49)$	0,11	0,99	0,29
$t_0 - t_6, (1,4, \infty)$	0,1	0,99	0,28
$t_{cp} - \frac{t_4 + t_5 + t_6 + t_7}{4}, (0,3, \infty)$	0,1	0,99	0,28

Отметим, что применяя более сложные конструкции из данных признаков, например, на основе генетических алгоритмов [3], можно получить признаки, обладающие большей информативностью, специфичностью и чувствительностью. Однако резко осложняется возможность их применения в блоке обоснования предлагаемого диагностического решения.

5. Алгоритм классификации

Полученные наборы характеристических и высокоинформативных признаков в первую очередь предназначены для обозначения найденных аномалий поведения температурных полей и обоснования предполагаемого диагноза в консультативной интеллектуальной системе. Предлагаемая их структура вполне позволяет обнаруживать и описывать особенности поведения температурных полей молочных желез на языке, понятном пользователю системы, т.е. врачу-диагносту. Однако вполне обоснованным выглядит их дальнейшее использование в различных алгоритмах классификации. Для иллюстрации проверим данную возможность на простейших алгоритмах, в которых найденные признаки используются непосредственно (в отличие от более мощных и эффективных алгоритмов типа нейронных сетей или генетических алгоритмов, использующих достаточно сложные конструкции из базовых элементов [3, 9]).

Опишем кратко используемый алгоритм классификации.

На первом этапе проводится проверка выполнения характеристических признаков. При этом если выполнен хотя бы один из них, то молочная железа относится к классу «Больные».

На втором этапе проверяется выполнение высокоинформативных признаков для неклассифицированных на первом этапе молочных желез. При этом для каждой молочной железы считается сумма информативностей выполненных признаков. Если найденная сумма выше некоторого критического значения (будем называть его критерием классификации), то молочная железа относится к классу «Больные», в противном случае – к классу «Здоровые».

Критерий классификации можно определять различными способами. Опишем один из возможных вариантов. Введем следующие обозначения: H_m – среднее арифметическое множества сумм информативностей выполненных признаков для каждой молочной железы класса «Здоровые»; S_m – среднее арифметическое множества сумм информативностей выполненных признаков для каждой молочной железы класса «Больные»; H_d – стандартное отклонение множества сумм информативностей выполненных признаков для каждой молочной железы класса «Здоровые»; S_d – стандартное отклонение множества сумм информативностей выполненных признаков для каждой молочной железы класса «Больные».

Набору высокоинформативных признаков будем ставить в соответствие критерий классификации:

$$(25) S_0 = \frac{x_1 + x_2}{2},$$

где

$$(26) x_1 = [\min\{H_m, H_d\}, \max\{H_m, H_d\}],$$

$$(27) x_2 = [\min\{S_m, S_d\}, \max\{S_m, S_d\}].$$

Другими словами, S_0 – точка пересечения интервалов, концами которых являются среднее арифметическое и стандартное отклонение сумм информативностей выполненных признаков для каждой молочной железы из определенного класса.

Сформированные наборы высокоинформативных признаков не являются оптимальными для данного алгоритма классификации, поскольку в них содержится большое количество

«родственных» признаков, обозначающих одну и ту же аномалию температурного поля. Поиск оптимального набора высокоинформативных признаков является актуальной проблемой. В данном случае использовалось достаточно простое эвристическое решение, состоящее из двух этапов. На этапе фильтрации из набора исключаются малоинформативные признаки, информативность которых меньше 3.

Для дальнейшей работы понадобилось использование метрики качества классификации. Чаще всего используются различные варианты F -мер, в частности – среднее гармоническое между точностью ($Prec$) и полнотой, иначе называемой чувствительностью ($Sens$ в (15)):

$$(28) F_1 = 2 \frac{Prec \cdot Sens}{Prec + Sens}.$$

Учитывая, что в медицинской диагностике наиболее значимыми являются параметры $Sens$ и $Spec$, было решено в качестве метрики качества классификации использовать среднее гармоническое между $Sens$ и $Spec$:

$$(29) D = 2 \frac{Sens \cdot Spec}{Sens + Spec}.$$

Отметим, что аналогичный подход применяется не впервые [17]. Ниже будет приведена сравнительная таблица эффективности диагностики с наборами признаков, сформированными при использовании разных метрик.

Итак, на втором этапе применялась «жадная» схема формирования набора признаков. А именно, в результирующий набор, который изначально был пуст, поочередно помещалось по одному признаку из исходного набора. С каждым новым признаком выполнялась диагностика и вычислялась метрика эффективности классификации D . Если полученное значение D было выше предыдущего, то признак запоминался и перемещался в результирующий набор в конце итерации. Результирующий набор сформирован, если в конце итерации не было добавлено нового признака.

В процессе исследования были проведены вычислительные эксперименты с различным выбором информативности, различными наборами высокоинформативных признаков, различными

методами формирования обучающей и тестовой выборок. Затем полученный алгоритм классификации применялся к тестовой выборке. В таблице 3 приведены результаты расчетов для случая, когда обучающая выборка составляла 62%, а тестовая – 38% от всей имеющейся базы данных.

Таблица 3. Результаты работы алгоритма классификации

Информативность	Выборка	<i>Sens</i>	<i>Spec</i>	<i>Acc</i>	<i>D</i>	F_1
Среднее гармоническое статистической, энтропийной, эвристической	Обучающая	0,87	0,86	0,87	0,86	0,92
	Тестовая	0,84	0,59	0,79	0,69	0,86
Среднее гармоническое статистической, эвристической	Обучающая	0,84	0,78	0,83	0,81	0,89
	Тестовая	0,79	0,52	0,73	0,62	0,82
Эвристическая	Обучающая	0,75	0,79	0,76	0,77	0,83
	Тестовая	0,73	0,53	0,69	0,62	0,79
Среднее гармоническое энтропийной, эвристической	Обучающая	0,87	0,77	0,85	0,82	0,9
	Тестовая	0,82	0,43	0,74	0,57	0,84
Комбинированная статистической, энтропийной	Обучающая	0,77	0,67	0,75	0,72	0,83
	Тестовая	0,79	0,45	0,72	0,57	0,82

Также на базе найденных высокоинформативных признаков была проведена классификация с помощью логистической регрессии. Результаты приведены в таблице 4.

Таким образом, в обоих случаях на тестовой выборке получалась чувствительность в районе 80%, а специфичность, соответственно, в районе 60%. Лучшие критерии эффективности алгоритма получались в районе 65% (при проверке по всей экспертной базе, естественно, результаты существенно улучшаются и становятся выше 90%). Полученные результаты оказались достаточно стабильными, т.е. изменение состава обучающей и тестовой выборок практически не меняло получаемый критерий эффективности алгоритма.

Таблица 4. Результаты работы алгоритма логистической регрессии

Информативность	Выборка	<i>Sens</i>	<i>Spec</i>	<i>Acc</i>	<i>D</i>	F_1
Эвристическая	Обучающая	0,82	0,81	0,82	0,81	0,88
	Тестовая	0,73	0,6	0,71	0,66	0,8
Среднее гармоническое статистической, эвристической	Обучающая	0,66	0,83	0,7	0,74	0,78
	Тестовая	0,65	0,59	0,64	0,62	0,74
Комбинированная статистической, эвристической	Обучающая	0,66	0,81	0,69	0,73	0,77
	Тестовая	0,64	0,59	0,63	0,61	0,74
Комбинированная энтропийной, эвристической	Обучающая	0,69	0,89	0,73	0,77	0,8
	Тестовая	0,68	0,55	0,65	0,61	0,76
Среднее гармоническое энтропийной, эвристической	Обучающая	0,83	0,82	0,83	0,83	0,89
	Тестовая	0,73	0,52	0,68	0,6	0,78

Как было отмечено выше, в качестве метрики качества классификации используется целый ряд параметров: различные варианты F -мер, коэффициент корреляции Мэтьюса, а также самые различные усреднения множественных метрик. При этом наиболее популярным является использование F_1 , т.е. среднего гармонического между точностью и чувствительностью. Обоснованность использования в нашем случае среднего гармонического между чувствительностью и специфичностью показывают приведенные в таблице 5 расчеты. Из них видно, что использование F_1 в большинстве случаев слишком сильно минимизирует специфичность.

Таблица 5. Эффективность диагностики на тестовой выборке при использовании наборов признаков, полученных с разными метриками качества классификации

Информативность	Метрика	<i>Sens</i>	<i>Spec</i>	<i>Acc</i>	<i>D</i>	F_1
Среднее гармоническое статистической, энтропийной, эвристической	F_1	0,88	0,24	0,75	0,38	0,85
	<i>D</i>	0,78	0,59	0,74	0,67	0,83
Среднее гармоническое статистической, эвристической	F_1	0,91	0,29	0,78	0,44	0,87
	<i>D</i>	0,73	0,52	0,69	0,61	0,79
Эвристическая	F_1	0,64	0,5	0,61	0,56	0,73
	<i>D</i>	0,62	0,53	0,6	0,57	0,71
Среднее гармоническое энтропийной, эвристической	F_1	0,9	0,24	0,76	0,38	0,86
	<i>D</i>	0,8	0,43	0,72	0,56	0,82

6. Заключение

В работе предложен новый метод формирования информационных признаков на базе количественного описания качественных признаков, получаемых с помощью анализа медицинских знаний и физико-математических моделей температурных полей молочных желез. Предлагаемый подход позволил получить новую медицинскую информацию об особенностях поведения температурных полей пациенток. А именно, с помощью исследования разностных аналогов вторых производных функции температур по различным направлениям была обнаружена целая группа качественно новых диагностических признаков.

Применение на базе полученных признаков более мощных алгоритмов классификации улучшает результаты диагностики. Например, применение нейронных сетей каскадной корреляции с методом имитации отжига в качестве обучения [3] позволяет получить чувствительность и специфичность в районе 80%. Более того, в разрабатываемой консультационной системе предполагается применять алгоритмы классификации на базе нейронных сетей, генетических алгоритмов, нечеткой логики и т.д. Однако обоснование получаемых результатов будет про-

изводиться на основе полученных характеристических и высокоинформативных признаков.

Отдельно заметим, что реальная чувствительность и специфичность консультационной системы должна выявляться путем опытной эксплуатации, т.е. более интересным является вопрос, насколько полученные признаки улучшат диагностику, осуществляемую специалистами.

Отметим, что предложенный метод нахождения диагностических признаков может быть применен в других областях медицины, использующих данные микроволновой радиотермометрии.

Работа выполнена при финансовой поддержке РФФИ (проект № 15-47-02475-р_поволжье_a).

Литература

1. ВАЙСБЛАТ А.В., ВЕСНИН С.Г., КОНКИН М.А. и др. *Использование микроволновой радиотермометрии в диагностике рака молочной железы* [Электронный ресурс]. – URL: <http://www.resltd.ru/rus/literature/cancer.htm>.
2. ВЕСНИН С.Г., КАПЛАН М.А., АВАКЯН Р.С. *Современная микроволновая радиотермометрия молочных желез // Опухоли женской репродуктивной системы*. – 2008. – №3. – С. 28–33.
3. ЗЕНОВИЧ А.В., ГЛАЗУНОВ В.А., ОПАРИН А.С., ПРИМАЧЕНКО Ф.Г. *Алгоритмы принятия решений в консультативной интеллектуальной системе диагностики молочных желез // Вестник Волгоградского государственного университета*. – Серия 1: Математика. Физика. – 2016. – №6. – С. 129–142.
4. КОБРИНСКИЙ Б.А. *Системы искусственного интеллекта в медицине: состояние, проблемы и перспективы // Новости искусственного интеллекта*. – 1995. – №2. – С. 65–79.
5. КОБРИНСКИЙ Б.А. *Консультативные интеллектуальные медицинские системы: классификация, принципы построения, эффективность // Врач и информационные технологии*. – 2008. – №2. – С. 38–47.

6. КОРЕНЕВСКИЙ Н.А., ЛУКАШОВ М.И., АРТЕМЕНКО М.В., АГАРКОВ Н.М. *Синтез гибридных нечетких решающих правил для классификации клинических вариантов течения генитального герпеса на основе моделей системных взаимосвязей // Фундаментальные исследования.* – 2014. – №10. – С. 901–907.
7. ЛОСЕВ А.Г., МАЗЕПА Е.А., ЗАМЕЧНИК Т.В. *О некоторых характерных признаках в диагностике патологии молочных желез по данным микроволновой радиотермометрии // Современные проблемы науки и образования.* – 2014. – №6. – С. 254.
8. ЛОСЕВ А.Г., МАЗЕПА Е.А., СУЛЕЙМАНОВА Х.М. *О взаимосвязи некоторых признаков РТМ-диагностики заболеваний молочных желез // Вестник Волгоградского государственного университета.* – Серия 1: Математика. Физика. – 2015. – №4(29). – С. 35–44.
9. ЛОСЕВ А.Г., ХОПЕРСКОВ А.В., АСТАХОВ А.С., СУЛЕЙМАНОВА Х.М. *Проблемы измерения и моделирования тепловых и радиационных полей в биотканях: анализ данных микроволновой радиотермометрии // Вестник Волгоградского государственного университета.* – Серия 1: Математика. Физика. – 2015. – №6. – С. 31–71.
10. ЯСНИЦКИЙ Л.Н. *Введение в искусственный интеллект: учебное пособие для студентов вузов.* – М.: Academia. – 2005. – 176 с.
11. BARRETT A.H., MYERS P.C. *Subcutaneous Temperature: A method of Noninvasive Sensing // Science.* – 1975. – Vol. 190. – P. 669–671.
12. BARRETT A.H., MYERS P.C., SADOWSKY N.L. *Microwave Thermography in the Detection of Breast Cancer // Am. J. Roengenol.* – 1980. – №34. – P. 365–368.
13. BURKE H.B. *Artificial neural networks for cancer research: outcome prediction // Seminars in Surgical Oncology.* – 1994. – Vol. 10. – №1. – P. 73–79.
14. GAUTHERIE M. *Temperature and Blood Flow Patterns in Breast Cancer During Natural Evolution and Following Radiotherapy // Biomedical Thermology.* – 1982. – Vol. 107. – P. 21–64.

15. HAMLIN S.K., STRAUSS P.Z. *Basic Concepts of Hemorheology in Microvascular Hemodynamics* // Critical care nursing clinics of North America. – 2014. – Vol. 26. – P. 337–344.
16. JAIN L.C., MARTIN N.M. *Fusion of Neural Networks, Fuzzy Systems and Genetic Algorithms: Industrial Applications* // CRC Press. – 1998. – 368 p.
17. KENNEDY K., MAC NAMEE B., DELANY S. *Learning without default: a study of one-class classification and low-default portfolio problem* // Proc. of the 20th Irish conference on Artificial intelligence and cognitive science (AICS'09), August 19-21, 2009, Dublin, Ireland. – 2009. – P. 174–187.
18. KONONENKO I., BRATKO I., KUKAR M. *Application of machine learning to medical diagnosis* // Machine Learning, Data Mining and Knowledge Discovery: Methods and Applications. – 1998. – P. 389–408.
19. KONONENKO I. *Machine Learning for Medical Diagnosis: History, State of the Art and Perspective* // Artificial Intelligence in Medicine. – 2001. – Vol. 23(1). – P. 89–109.
20. LESMO L., SAITTA L., TORASSO P. *Learning of Fuzzy Production Rules for Medical Diagnoses* // Approximate Reasoning in Decision Analysis. – 1982. – P. 249–260.
21. RODRIGUES D.B., MACCARINI P.F., SALAH S., COLEBECK E., TOPSAKAL E., PEREIRA P.J., LIMA-OVIEIRA P., STAUFFER P.R. *Numerical 3D modeling of heat transfer in human tissues for microwave radiometry monitoring of brown fat metabolism* // Proc. of SPIE 8584 (SPIE BIOS), February 2-7, 2013, San Francisco, California, United States. – 2013. – DOI: 10.1117/12.2004931.

THE THERMOMETRY DATA MINING IN THE DIAGNOSTICS OF MAMMARY GLANDS

Alexander Losev, Volgograd State University, Volgograd, Doctor of Science, professor (allosev59@gmail.com).

Vladislav Levshinsky, Volgograd State University, Volgograd, student (vladi.lev.email@gmail.com).

Abstract: The paper deals with the microwave thermometry data and proposes a method for forming informative features, based on qualitative descriptions of medical knowledge about the behavior of the temperature fields of mammary glands. There are some already known qualitative characteristics of breast cancer, which were form a basis for quantitative features, e.g. feature 'a large thermal asymmetry between mammary glands' can be described by temperature differences between corresponding points of right and left mammary glands. If the value of such difference is sufficiently large, then it may be an indication of pathology. After preprocessing, which involves weighting and cleaning, informative features may be applied in various classification algorithms, such as logistic regression, which yields about seventy percents of accuracy on a test sample, or the more complex ones that yield better accuracy: neural networks, genetic algorithm, and fuzzy classification. The significance of proposed features consists of the fact that they were formed from qualitative characteristics and each of them has a qualitative description, therefore they are of interest for further study and can be applied in diagnosis-advisory systems.

Keywords: data mining, microwave radiothermometry, intelligent advisory systems.

*Статья представлена к публикации
членом редакционной коллегии А.И. Михальским.*

Поступила в редакцию 01.11.2016.

Опубликована 30.11.2017.

УДК 519.711

ББК 22.18

ОРГАНИЗАЦИЯ СТРОЯ АГЕНТОВ С ПОМОЩЬЮ КЛЕТОЧНОГО АВТОМАТА

Кузнецов А. В.¹,

(Воронежский государственный университет, Воронеж)

Рассматривается алгоритм распределенной организации заданного графом строя агентов и численная симуляция такого алгоритма. Описывается клеточный автомат, моделирующий движение агентов, и исследуются его особенности в связи с типом ландшафта, по которому перемещаются агенты. Указанный клеточный автомат имеет два представления – одномерное и двумерное.

Ключевые слова: клеточный автомат, автономные агенты, рефлексивные агенты, управление строем агентов.

1. Введение

Статья является итогом работ автора по исследованию поиска оптимального пути группами автономных агентов по ландшафтам разной степени сложности в рамках клеточно-автоматного подхода. В работах [1, 4, 14] был сконструирован клеточный автомат, позволяющий моделировать разные виды взаимодействия между иерархически организованными агентами, движущимися по ландшафту переменной сложности, и исследованы некоторые характеристики этих взаимодействий. Эти работы возникли в результате переосмысления автором моделей, предложенных в [6] и [12]. Однако, в отличие от [6], агенты двигались иерархически вложенными друг в друга роями по местности, состоящей из препятствий разной проходимости, причем

¹ Александр Владимирович Кузнецов, кандидат физико-математических наук, доцент (avkuz@bk.ru).

способ их передвижения отличался гораздо большей самостоятельностью, чем в [12], где агенты двигались, в сущности, по набору опорных точек.

В данной работе анализируется эффективность предложенного клеточного автомата для поиска оптимального маршрута в ландшафтах разного вида. Также предлагается алгоритм организации строя агентов с учетом необходимости обхода препятствий. Отметим, что автомат, предложенный автором в цитированных выше работах, уже предполагал возможность самоорганизации агентов в иерархические рои с помощью аналога социального потенциала, предложенного в [13]. В отличие от работ [7, 16, 18, 19], посвященных построению строя роботов с использованием одномерного клеточного автомата, используется автомат, имеющий одновременно и одномерное, и двумерное представление в зависимости от типа решаемой задачи. Также, в отличие от упомянутых работ, рассматривается строй, порожденный графом, а не набором математических функций, а в качестве окрестностей в одномерном представлении клеточного автомата рассматриваются окрестности, порожденные геодезическим расстоянием на графе строя. Наконец, агенты в настоящей статье не обмениваются координатами для поддержания строя и не безусловно следуют за лидером (что может быть крайне невыгодным на пересеченной местности), а стараются поддерживать строй, выбирая при этом направление движения самостоятельно, исходя из своих предположений о передвижении других агентов в наблюдаемой ими окружающей среде.

2. Непрерывная постановка задачи

В работе далее будут использоваться следующие стандартные обозначения: \mathbb{R} , \mathbb{Z} и \mathbb{N} – множества вещественных, целых и натуральных чисел соответственно, $\mathbb{R}_{\geq 0}$, $\mathbb{Z}_{\geq 0}$ – множества неотрицательных вещественных и целых чисел, \mathbb{R}^n – пространство n -мерных векторов с вещественными координатами, $\|\cdot\|$ – произвольная (например, евклидова) норма в \mathbb{R}^n .

Пусть заданы

1. Область $\Omega \subset \mathbb{R}^2$.

2. Множество агентов $Ag = \{ag_k | k = \overline{1, n}\}$.

3. Точки начала и конца движения агента $ag_k \in Ag$, $A_k, B_k \in \Omega$, $k = \overline{1, n}$

4. Функция проходимости $u^c : [0, T] \times \Omega \rightarrow \mathbb{R}_{\geq 0}$, $u^c(t, r)$ – это максимально возможная норма скорости в момент времени $t \in [0, T]$ в точке $r \in \Omega$.

5. Желаемый граф строя $\Phi = (Ag, E, \text{соо}_\Phi^c)$, $E \subseteq Ag^2$ – множество ребер, $\text{соо}_\Phi^c : [0, T] \times Ag \rightarrow \Omega$ – функция, дающая желаемые для сохранения строя координаты агента в Ω .

6. Фактический граф строя $\Gamma = (Ag, E, \text{соо}_\Gamma^c)$, $\text{соо}_\Gamma^c : [0, T] \times Ag \rightarrow \Omega$ – функция, дающая реальные координаты агента в Ω .

Траектория k -го агента $r_k : [0, T] \rightarrow \Omega$ отвечает условиям

$$(1) \quad \|\dot{r}_k(t)\| = u^c(t, r_k(t)),$$

$$(2) \quad r_k(0) = A_k, \quad r_k(t) = B_k, \quad t \geq T_k, \quad T_k \leq T,$$

$$(3) \quad r_k(t) \neq r_l(t), \quad k \neq l, t \in [0, T],$$

где

$$T_k(r_k) = \min\{t \in \mathbb{R}_{\geq 0} | r_k(t) = B_k\},$$

$\|\dot{r}_k(t)\|$ – норма скорости k -го агента в момент времени t ,
 $\mathcal{Y} \subset C([0, T]; \Omega)$ – множество допустимых траекторий. В явном виде выписать функционалы T_k в общем случае невозможно. Отметим, что уравнение (1) означает, что агент движется в каждой точке области с максимальной (по норме) возможной в этот момент времени скоростью.

Необходимо найти такие траектории агентов, чтоб время $T = \max_{k=\overline{1, n}} T_k$ было минимальным, т.е. надо минимизировать все функционалы T_k , сопоставляющие траектории r_k агента ag_k продолжительность прохождения по этой траектории. Более формально, нужно решить задачу минимизации

$$(4) \quad T(r_1, \dots, r_n) = \max_{k=\overline{1, n}} T_k(r_k) \rightarrow \min, \quad r_k \in \mathcal{Y}, k = \overline{1, n}.$$

Также необходимо, чтоб в любой момент времени $t \in [0, T]$ отличие графов Φ и Γ было бы минимально:

$$(5) \quad \text{dist}(\Gamma(t), \Phi(t)) \rightarrow \min.$$

Формальное описание функции несходства графов dist будет дано ниже в разделе 6. Таким образом, необходимо найти траектории $r_k \in \mathcal{Y}$, $k = \overline{1, n}$, являющиеся решениями задачи оптимизации с двумя целевыми функциями (4)–(5).

3. Дискретная модель

Для поиска приближенных решений задачи оптимизации, приведенной в предыдущем параграфе, будет использоваться клеточный автомат, или, в отечественной терминологии, однородная структура со входами и выходами [2, с. 159].

Определение 1. Однородной структурой со входами и выходами называется набор $(\mathbb{Z}^k, \mathcal{W}, V)$, где \mathcal{W} – конечный автомат с n основными и p боковыми входными каналами, m основными и r боковыми выходными каналами, l состояниями, помещенный в каждом узле целочисленной решетки \mathbb{Z}^k и задаваемый системой уравнений:

$$\begin{cases} y(t) = \varphi_1(x(t), b(t), \mathbf{s}(t)), \\ a(t) = \varphi_2(\mathbf{s}(t)), \\ \mathbf{s}(t+1) = \varphi(x(t), b(t), \mathbf{s}(t)), \end{cases}$$

$$\begin{aligned} \varphi_1 : \mathcal{E}^n \times \mathcal{E}^p \times \mathcal{S} &\rightarrow \mathcal{E}^m, \\ \varphi_2 : \mathcal{S} &\rightarrow \mathcal{E}^p, \\ \varphi : \mathcal{E}^n \times \mathcal{E}^p \times \mathcal{S} &\rightarrow \mathcal{E}^m, \end{aligned}$$

где $x = (x_1, \dots, x_n)$, $b = (b_1, \dots, b_p)$, $a = (a_1, \dots, a_p)$, $y = (y_1, \dots, y_m)$ обозначают соответственно основной вход, боковой вход, боковой выход и основной выход, \mathbf{s} – состояние автомата, $t = \overline{0, N}$ – такт дискретного времени, \mathcal{E} – входной и выходной алфавит, \mathcal{S} – множество состояний автомата \mathcal{W} , $V = (\alpha_1, \dots, \alpha_p)$, $\alpha_i \in \mathbb{Z}^k$, $i = \overline{1, p}$, – шаблон соседства: для автомата в узле α каждый $\alpha_i \in V$ определяет конечный автомат с координатами $\alpha + \alpha_i$, с i -м боковым выходным каналом которого соединен i -й боковой входной канал автомата в узле α .

В клеточных автоматах, моделирующих движение агентов, конечному автомату \mathcal{W} будет соответствовать ячейка «местности», содержащая агента, или пустая ячейка, боковым каналам будут соответствовать связи между ячейками, задающие закономерности перемещения агентов между ячейками, а основным каналом – например, модели каналов связи между агентами.

Клеточные автоматы, описывающие движение автономных агентов, можно условно поделить на два типа. Первый из них – двумерный клеточный автомат (так называемый *world-space cellular automaton*, рис. 1a [17]), например, с правилами, близкими к *Game of Life* Конвея, приведенный в работе [13] для моделирования образования роя агентов.

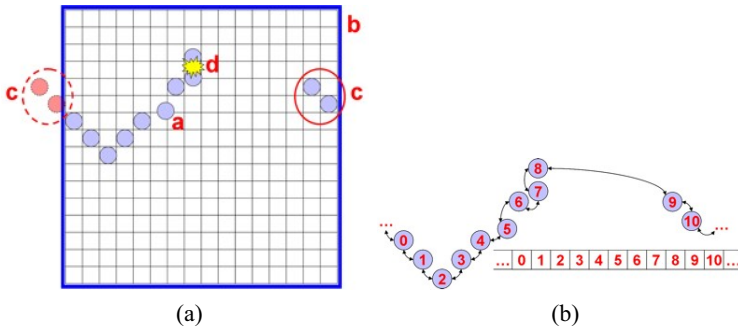


Рис. 1: *World-space cellular automaton* и *robot-space cellular automaton*

Второй тип – это одномерный клеточный автомат (так называемый *robot-space cellular automaton*), каждая клетка которого содержит координаты агента на плоскости, его реальное и желаемое взаимное расположение с соседними агентами (см. рис. 1b [19]).

Такие автоматы применяются в работах для построения строя реальных роботов и для компьютерного моделирования данного процесса, например в [19].

Автор в своих предыдущих работах подробно описывал конструкцию клеточного автомата (далее – КА), используемого им

для моделирования движения агентов. КА, предлагаемый в настоящей работе, незначительно отличается от ранее описанных, поэтому приведем его краткую характеристику. Пусть заданы

1. Замоещение Ω_h области $\Omega \subset \mathbb{R}^2$ (для простоты можно считать, что это правильный квадратный паркет $\Omega_h = \{\omega_{ij} | i, j \in Q_h \subset \mathbb{Z}\}$ и не различать клетку ω_{ij} и ее координаты (i, j)).

2. Взаимно-однозначная функция получения уникального идентификатора (УИД) агента $\text{uid} : Ag \rightarrow \mathbb{N}$.

3. Множество возможных направлений движения агентов $\mathcal{D} = \{(i, j) | i, j = \overline{-1, 1}\}$.

4. Дискретная функция непроходимости $u : \mathbb{Z}_{\geq 0} \times \Omega_h \rightarrow \mathbb{Z}$, устанавливающая, сколько тактов функционирования КА необходимо для преодоления клетки (i, j) . При этом предполагается, что для совершенно непроходимой в момент времени t клетки (i, j) $u(t, (i, j)) = +\infty$. Эта функция получается, по своей сути, дискретизацией функции проходимости, например:

$$u_{ij}(t) = u(t, (i, j)) = \left\lfloor \frac{\max_{r \in \Omega} u^c(\tau t, r)}{\min_{r \in \omega_{ij}} u^c(\tau t, r)} \right\rfloor,$$

где $\lfloor x \rfloor$ означает целую часть x , а τ – продолжительность одного такта.

5. Дискретные функции, дающие желаемые для сохранения строя, предусмотренного графом Φ , координаты агента в текущий момент дискретного времени $\text{соо}_\Phi : \mathbb{Z}_{\geq 0} \times Ag \rightarrow \Omega_h$ и реальные координаты агента $\text{соо}_\Gamma : \mathbb{Z}_{\geq 0} \times Ag \rightarrow \Omega_h$.

6. Шаблон формации $\Phi_h(ag) = \{(i, j, agId) | (i, j) \in \mathbb{Z}^2, agId \in \text{uid}(Ag) \cup \{0\}\}$, содержащий относительные координаты соседей агента по строю (строгое определение соседства дано в разделе 6), причем $agId = 0$, если нет требований к тому, какой именно сосед должен находиться, и $agId > 0$, если в клетке с координатами $\text{соо}(t, ag) + (i, j)$ в момент времени t должен находиться именно сосед с УИД $agId$. Таким образом, для элемента $(i, j, agId) \in \Phi_h(ag)$, $agId \neq 0$, справедливо соотношение

$$(i, j) + \text{соо}_\Phi(ag) = \text{соо}_\Phi(ag'), \quad \text{uid}(ag') = agId.$$

Пусть $r_h \subseteq \Omega_h$ – маршрут (на языке перечислительной комбинаторики – решеточный путь), начальная клетка маршрута $r_h[1] = (i, j) + d_1$, $d_1 \in \mathcal{D}$, k -я клетка маршрута $r_h[k] = r_h[k-1] + d_k$, $d_k \in \mathcal{D}$, $k > 1$. Обозначим $(k-1)$ -ю клетку маршрута r_h с координатами (i', j') , в которой агент находится в момент времени t , как $r_h(t)$. Тогда дискретным аналогом уравнения (1), определяющего движение агента, будет уравнение

$$(6) \quad \|r_h(t+k) - r_h(t)\| = \|d_k\| \left\lfloor \frac{k}{\|d_k\|u_{i'j'}(t)} \right\rfloor, \quad k \leq \|d_k\|u_{i'j'}(t).$$

Детальное формальное определение функций локального такового функционирования описываемого КА (без задания весов, необходимых для поддержания строя) приведено в работе [4]. Вкратце, агент на каждом такте выбирает случайный маршрут движения, удовлетворяющий уравнению (6) и отвечающий ряду других ограничений (например, два агента не могут одновременно находиться в одной клетке). Распределение случайного выбора маршрута таково, что вероятность выбрать минимизирующий нижеприведенные функционалы (7), (9) или (10) маршрут максимальна.

В связи с особенностями реализации, КА должен быть представлен как в виде *robot-space cellular automaton*, так и в виде *world-space cellular automaton*, так как первый тип автомата удобен для построения строя, а второй – для работы алгоритма поиска оптимального маршрута и для построения тестовых ландшафтов. Поэтому агент (как ячейка *robot-space cellular automaton*) $ag \in Ag$ в момент времени $t \in \mathbb{Z}_{\geq 0}$ представляет собой объект вида

$$ag = \{selfId, leadId, lead, w, \mathcal{D}', trgtAct, trgtTmp, formTmp, @cell\},$$

где $selfId = uid(ag)$ – УИд агента; $leadId \in \mathbb{N}$ – УИд лидера данного агента; $lead \in \{true, false\}$ указывает на то, что агент является ведущим; w – количество тактов, которое агент уже простоял в текущей клетке; \mathcal{D}' – упорядоченный по убыванию желательности список возможных направлений из \mathcal{D} для

агента на следующий такт; $trgtAct \in \Omega_h$ – текущая целевая клетка агента; $trgtTmp \in \Omega_h$ – временная целевая клетка агента; $formTmpl = \Phi_h(ag)$ – шаблон строя; $@cell$ – указатель на ячейку, в которой находится агент ag в момент времени t . Ячейка $cell$, соответствующая клетке $(i, j) \in \Omega_h$ для *world-space cellular automaton*, представляет собой

$$cell = \{u, соо, @ag\},$$

где $u = u(t, соо)$ – значение функции непроходимости клетки (т.е. минимальное количество тактов, нужное для преодоления клетки); $соо = (i, j)$; $@ag$ – указатель на агента, находящегося в клетке (i, j) в момент времени t или нуль, если в (i, j) в момент времени t нет агента. Будем обозначать агента, на которого указывает $@ag$ как ag_{ij} и писать

$$ag_{ij} = \{selfId_{ij}, leadId_{ij}, w_{ij}, \mathcal{D}_{ij}, trgtAct_{ij}, trgtTmp_{ij}, formTmpl_{ij}, @cell_{ij}\}.$$

Если $@ag_{ij} = 0$, то для единообразия будем считать, что

$$ag_{ij} = \{0, 0, 0, 0, 0, 0, 0, 0\}.$$

Также будем обозначать ячейку, на которую ссылается $@cell_k$, как $cell_k$ и писать $cell_k = (u_k, соо_k, @ag_k)$. Помимо этого будем обозначать, что $cell_{ij} = \{u_{ij}, (i, j), @ag_{ij}\}$.

В результате использования ячеек, содержащих указатели, возможно при реализации КА сэкономить объем памяти, нужной для хранения конфигурации КА, так как количество агентов обычно существенно меньше количества клеток и при хранении состояния агента в каждой клетке *world-space cellular automaton* потребовалось бы хранение большого количества нулевых значений. Также возможно увеличить быстродействие КА, так как при перемещении агента в новую клетку не нужно копировать его состояние целиком, а нужно лишь изменить значение пары указателей, и при доступе к ячейке $cell$ уже сразу известен адрес

@*ag* и наоборот. В псевдокоде алгоритмов мы также будем обозначать поля объектов *cell* и *ag* принятым в языках программирования способом, т.е. в формате «имя_объекта.имя_поля», опуская числовые индексы.

Будем обозначать как $V_r(i, j)$ окрестность клетки (i, j) радиуса r

$$V_r(i, j) = \{c \in \mathbb{Z}^2 \mid \|c - (i, j)\| \leq r\}.$$

Определим функцию

$$\theta(x) = \begin{cases} 0, & x \leq 0, \\ 1, & x > 0. \end{cases}$$

Введем дискретный аналог функционалов T_k :

$$(7) \quad T_h(r_h) = \alpha_0 \sum_{(i,j) \in r_h} \|d_{ij}\| u_{ij} + \alpha_1 \sum_{(i,j) \in r_h} \theta(agId_{ij}),$$

где $\alpha_0, \alpha_1 \geq 0$ – параметры модели, отвечающие за то, предпочитает ли агент менее непроходимые или же менее населенные другими агентами маршруты, $d_{ij} = d_k$, если $r_h[k] = (i, j)$.

Обозначим множество всех маршрутов $\mathcal{M}_d(i, j; i_r, j_r)$ агента ag_{ij} , соединяющих (i, j) и (i_r, j_r) , причем (i, j) не входит в маршрут и если $(i_r, j_r) = \text{trgtAct}_{ij}$, то (i_r, j_r) не входит в маршрут, $d = r_h[1] - (i, j)$.

Общая идея состоит в том, чтоб каждый агент в клетке (i, j) искал оптимальный по времени маршрут в некоторой своей окрестности $V_r(i, j)$, а потом собирал из таких локально оптимальных маршрутов квазиоптимальный маршрут. Поскольку непроходимость области Ω_h может меняться со временем, а также не быть полностью разведанной и проходы могут блокироваться другими агентами, то поиск глобального оптимального маршрута может быть невозможен или лишен смысла. Такое функционирование *world-space cellular automaton*, в котором агенты ищут кратчайший маршрут до точки назначения, описывается с помощью алгоритма 1, в который входят алгоритмы 3 и 4, приведенные в Приложении.

Следует подчеркнуть, что для агента движение в сторону скорейшего маршрута наиболее вероятно, но не обязательно. Точнее, агент в клетке (i, j) назначает каждому направлению $d \in \mathcal{D}$ вес $weights[d]$:

$$(8) \quad weights[d] = \min_{r_h \in \mathcal{M}_d(i, j; i_r, j_r)} T_h(r_h).$$

Далее \mathcal{D} перемешивается в соответствии с весами из $weights$. Процедура взвешенного случайного перемешивания нетривиальна: используется предложенный в [9] алгоритм. Таким образом, получается упорядоченный список направлений. Если движение в первом по данному списку направлении маршрута невозможно (например, путь заблокирован), то агент выбирает следующий по времени прохождения маршрут и т.д. Это требуется для разрешения коллизий между агентами и для симуляции элемента случайности в передвижении агентов.

Алгоритм 1 (Поиска квазиоптимального по времени пути).

- 1: **for all** $ag \in Ag$ **do**
- 2: $weights = \text{sortDirections}(ag)$ ▷ Сортировка направлений, алгоритм 3
- 3: Случайно перемешать $ag.D'$ в соответствии с весами из $weights$
- 4: **end for**
- 5: **for all** $ag' \in Ag$ **do**
- 6: $\text{Actualize}(ag')$ ▷ Актуализация, алгоритм 4
- 7: **end for**
- 8: $t = t+1$

Можно обратить внимание, что в алгоритме разделен поиск наилучшего (в смысле кратчайшего по времени маршрута) направления движения агентов (алгоритм 3) и действительное обновление конфигурации КА. Это сделано для двух целей: для возможности вставки между этими блоками алгоритма, отвечающего за построение строя, и для параллельного вычисления наилучших направлений перемещения. При этом поиск маршрута является самым трудным вычислительно моментом работы КА и занимает, при вычислении алгоритмом Дейкстры, минимум

$O((2r + 1)^4)$ времени каждый ход для каждого агента, так как по сути кратчайший путь в окрестности $V_r(i, j)$ – это кратчайший путь на графе с $(2r + 1)^2$ вершинами для всех агентов.

Поиск локально оптимального маршрута может повторяться каждый такт времени, как в предыдущих работах автора, альтернативно агент ag_{ij} может рассчитать, например, алгоритмом Дейкстры или A^* локально оптимальный в $V_r(i, j)$ маршрут, следовать по нему и производить повторный пересчет лишь при достижении (i_r, j_r) или при обнаружении по ходу маршрута непроходимых препятствий или скоплений других агентов. Очевидно, в этом случае список предпочитаемых направлений движения $ag_{ij}.D'$ заменяется на цепной путь d_1, d_2, \dots, d_p , соединяющий (i, j) и (i_r, j_r) . Выбор между этими методами зависит от скорости изменения ландшафта. Также возможно производить поиск пути с учетом знаний агента о местности и предыдущего опыта агента, как это показано в [14]. В этом случае, помимо слоя разделяемой всеми агентами реальности $Cell = \{cell\}$, для каждого агента задается свой слой «субъективной реальности». В данный слой, который также описывается как множество ячеек, по мере поступления агент сохраняет информацию об уже посещенных клетках, количестве и времени посещения клеток. При поиске локально оптимального пути необходимо, например, увеличивать «веса» переходов в уже посещенные клетки, для чего рассматривать вместо функционала (7) функционал

$$(9) \quad \tilde{T}_h(r_h) = T_h(r_h) + \alpha_2 \sum_{(i,j) \in r_h} visit_{ij}, \quad \alpha_2 \geq 0,$$

где $visit_{ij}$ – количество предыдущих посещений клетки (i, j) , аналогично [14], моделировать конфликт агентов, как в [4] и т.д. Если агенту известно расстояние от coo_n – ближайшей к клетке назначения клетке из множества $r_h \cap V_r(i, j)$ до точки назначения coo_{target} , то для повышения качества алгоритма поиска можно рассматривать вместо (9) функционал

$$(10) \quad \hat{T}_h(r_h) = \tilde{T}_h(r_h) + \alpha_3 \|coo_n - coo_{target}\|, \quad \alpha_3 \geq 0.$$

3.1. Автоматы и теория категорий

Отметим, что агенты, движущиеся по решетке, могут соответствовать не только реальным роботам, перемещающимся по пересеченной местности, но и моделировать некоторый процесс неклассических вычислений. Можно рассматривать конфигурации *world-space cellular automaton* $Cell = \{cell\}$ как категорию, в которой объектами являются ячейки, а морфизмами – решеточные пути между ячейками. Аналогично можно рассмотреть конфигурации *robot-space cellular automaton* Ag как категорию, в которой объектами являются состояния агентов ag_i в момент времени t , а морфизмами – отображения $\gamma_{t_0} : ag_i(t) \mapsto ag_i(t + t_0)$.

Операция @ порождает пару функторов между этими категориями:

$$\begin{aligned} \mathcal{F}_1 : ag &\mapsto \{u, \text{coo}, @ag\}, \\ \mathcal{F}_2 : cell &\mapsto \{selfId, leadId, lead, w, \mathcal{D}', \text{trgtAct}, \text{trgtTmpl}, \\ &\quad \text{formTmpl}, @cell\}. \end{aligned}$$

Далее, в рамках теории категорий всякую клетку $(i, j) \in \Omega_h$ можно интерпретировать как *итератор*. В этом случае $ag \in Ag$ является *коитератором* [11], который порождает поток клеток, через которые он проходит.

4. Случайные ландшафты

Назовем ландшафтом в момент времени t множество непроеходимостей клеток определенного подмножества замощения $\mathcal{L}(\Omega_h, l) = \{u_{ij} | u_{ij} = u(t, \omega_{ij}), \omega_{ij} \in \Omega_h\}$, такого что u принимает на $\{\omega_{ij} | i = 1, n, j = 1, m\}$ не более l значений, причем к классу i принадлежит N_i клеток, т.е. $\sum_{i=1}^l N_i = M$, $M = |\mathcal{L}(\Omega_h, l)|$. Введем следующие характеристики ландшафта, известные из ландшафтной экологии (см., например, [8, 10, 20]).

Определение 2. Конфигурационная энтропия ландшафта $\mathcal{L}(\Omega_h, l)$ определяется как

$$S(\mathcal{L}(\Omega_h, l)) = - \sum_{i=1}^l \frac{N_i}{M} \ln \frac{N_i}{M}.$$

Определение 3. Total Edge (TE) определяется как общее количество соприкосновений сторон клеток в $\mathcal{L}(\Omega_h, l)$, принадлежащих к разным классам. Будем далее обозначать Total Edge ландшафта $\mathcal{L}(\Omega_h, l)$ как $TE(\mathcal{L}(\Omega_h, l))$.

Определение 4. Total Edge Density (TED) для ландшафта $\mathcal{L}_{n \times m}(l)$ определяется как отношение $TE(\mathcal{L}(\Omega_h, l))$ к общему количеству клеток $\mathcal{L}(\Omega_h, l)$:

$$TED(\mathcal{L}(\Omega_h, l)) = TE(\mathcal{L}(\Omega_h, l))/M.$$

Специально для исследования алгоритмов нахождения кратчайшего пути и построения строя в [15] были разработаны методы построения случайных ландшафтов с заданной конфигурационной энтропией, TE и TED и исследованы зависимости между упомянутыми характеристиками.

5. Поиск квазиоптимальной траектории

Получившийся в результате работы алгоритма 1 маршрут сравнивался с глобально оптимальным, найденным с помощью алгоритма Дейкстры. Для разных типов ландшафтов результаты сравнения получились довольно разными. Определим отклонение длины Λ маршрута $r_h(t) = (i(t), j(t))$ от оптимального по времени маршрута $r_{opt}(t) = (i_{opt}(t), j_{opt}(t))$, $t \in \mathbb{Z}_{\geq 0}$:

$$\delta\Lambda = \frac{|\Lambda(r) - \Lambda(r_{opt})|}{\Lambda(r_{opt})} \cdot 100\%,$$

и отклонение времени T прохождения маршрута r_h от времени T_{opt} прохождения оптимального маршрута r_{opt} :

$$\delta T = \frac{|T - T_{opt}|}{T_{opt}} \cdot 100\%.$$

Зададим функцию $N(S) = 0,00160518e^{4,22769S}$ зависимости количества препятствий от конфигурационной энтропии ландшафта. Для ландшафтов одного типа это будет в действительности количество препятствий, для ландшафтов другого типа эта величина применяется для того, чтобы их можно было легче сравнивать с другими.

Исследовались квадратные случайные ландшафты $\mathcal{L}(n \times m, l)$ со сторонами в $n = m = 100$ клеток, с $l = 9$ различными классами клеток. Агент имел радиус обзора $r = 5$ и функционал поиска оптимального маршрута вида (10).

Было произведено по 1000 экспериментов для каждого значения $N(S)$ в случайном ландшафте $\mathcal{L}(100 \times 100, 9)$. В каждом эксперименте агент перемещался из клетки $(1, 1)$ в клетку $(98, 98)$ ². Оказалось, что при возрастании конфигурационной энтропии S в ландшафте, состоящем из равномерно распределенных клеток разной непроходимости (см. рис. 7с в Приложении), среднее отклонение длины и времени от оптимальных (рис. 2а) медленно убывает, причем в районе $N(S) = 70$ наблюдается небольшая ступенька (более выраженная у ландшафтов малой площади), очевидно связанная со ступенькой в зависимости среднего значения TED от $N(S)$ (рис. 2b).

Однако при возрастании количества препятствий в ландшафте, состоящем из нескольких максимально труднопроходимых клеток, которые окружены менее непроходимыми, которые, в свою очередь, окружены еще менее непроходимыми и т.д. (рис. 7а, 7б в Приложении), характеристики алгоритма поиска несколько иные. При росте количества максимально непроходимых клеток N_{obst} и соответствующем росте конфигурационной энтропии (для такого ландшафта справедливо соотношение $N_{obst} = N(S)$) среднее отклонение длины и времени от оптимальных медленно убывает, причем в районе $N_{obst} = 70$ наблюдается резкий скачок (рис. 3а и 3б), после чего опять наблюдается

² Данные доступны по адресу https://www.researchgate.net/publication/319998893_Eksperimentalnye_dannye_po_sravneniu_kvazioptimalnoj_i_optimalnoj_traektorij_agenta.

медленное убывание среднего отклонения. Автор полагает, что именно такие ландшафты, а не полностью равномерно случайные, более соответствуют встречающимся в приложениях и что отклонение от оптимального маршрута не более 10% по длине и порядка 30% по времени вполне достаточно для многих практических нужд. Следует учесть, что при поиске локально оптимального маршрута алгоритмом Дейкстры требуется всего примерно $O((2r + 1)^4)$ времени. При «склежке» квазиоптимального маршрута длиной L клеток из локально оптимальных временная сложность будет, таким образом, примерно $O((2r + 1)^4 L)$ при пересчете направления маршрута каждый ход, при котором изменяется положение агента, или $O((2r + 1)^3 L)$ при пересчете маршрута лишь по достижении границы окрестности вместо $O((nm)^2)$ при поиске глобально оптимального маршрута, что дает определенный выигрыш при $r^2 \ll nm$.

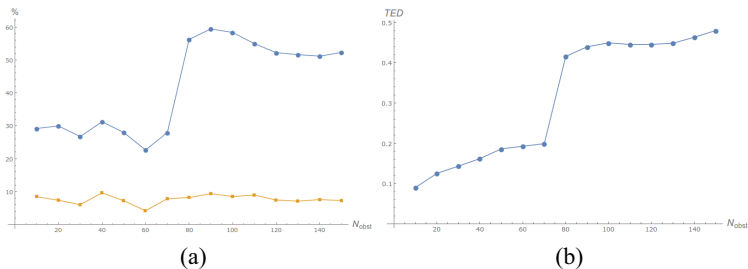


Рис. 2: Проигрыш времени (круглые маркеры) и отклонение от траектории (квадратные) в процентах по сравнению с оптимальным маршрутом для равномерного случайного ландшафта

На рис. 4а приводятся гистограммы распределения времени прибытия в конечную точку маршрута для равномерного ландшафта ($N_{obs} = 130$). Слева на рисунке изображена гистограмма распределения времени прибытия в конечную точку для оптимального маршрута, найденного алгоритмом Дейкстры, справа – для квазиоптимального маршрута. Данные гистограммы при
150

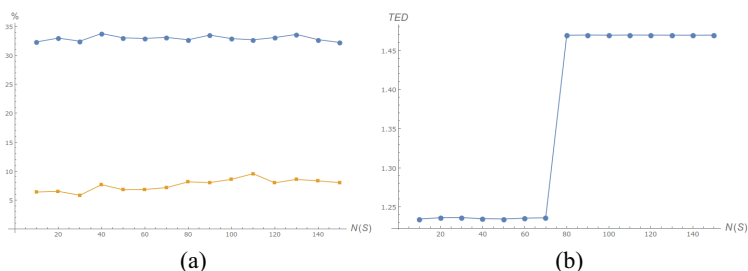


Рис. 3: Проигрыши времени (круглые маркеры) и отклонение от траектории (квадратные) в процентах по сравнению с оптимальным маршрутом для «естественного» случайного ландшафта

мерно соответствуют (с p -значениями около 0,7) распределению Нортона–Райса с функцией плотности

$$p(m, \alpha, \beta; x) = \begin{cases} \frac{\alpha m \left(\frac{x}{\alpha}\right)^m e^{-\frac{m(\alpha^2+x^2)}{2\beta^2}} I_{m-1}\left(\frac{mx\alpha}{\beta^2}\right)}{\beta^2}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

где I_k – модифицированная функция Бесселя первого рода.

Аналогичные гистограммы приведены на рис. 4б для распределения времени прибытия в конечную точку маршрута для «естественного» ландшафта ($N_{obst} = 130$).

6. Метрика сходства графов и метрика, порожденная графом

Определение 5. Определим расстояние $gd : Ag^2 \rightarrow \mathbb{R}_{\geq 0}$, следующим образом. Если между агентами $ag_1, ag_2 \in Ag$ существует в Φ кратчайший путь длиной l , то $gd(ag_1, ag_2) = l$, иначе $gd(ag_1, ag_2) = \infty$, $gd(ag_1, ag_1) = 0$.

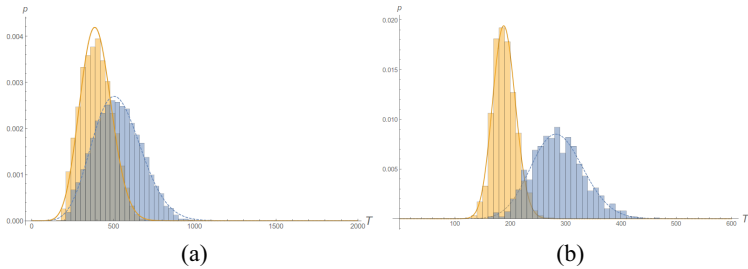


Рис. 4: Гистограммы распределения времени прибытия в конечную точку маршрута для равномерного (а) и «естественного» (б) ландшафтов

Это расстояние принято называть геодезическим расстоянием на графе. Будем считать окрестностью агента $ag \in Ag$ радиуса ρ множество $W_\rho(ag) \subseteq Ag$, такое что $\forall (ag' \in W_\rho(ag)) \text{ gd}(ag, ag') \leq \rho$. Соответственно, будем считать соседями агента ag всех агентов $ag' \in W_\rho(ag)$.

Пусть $\Gamma_i = (Ag_i, E_i, \text{coo}_i)$, $Ag_i \subseteq Ag$, $E_i \subseteq Ag^2$, $\text{coo}_{\Gamma_i} : \mathbb{Z}_{\geq 0} \times Ag_i \rightarrow \Omega_h$ – функция соответствия вершины графа Γ_i и клетки Ω_h , $i = 1, 2$, – два графа, описывающие строи агентов. Определим расстояние несходства между ними в момент времени t как

$$\text{dist}(\Gamma_1(t), \Gamma_2(t)) = m_0 + m_1 + \kappa \sum_{ag \in Ag_c} \|\text{coo}_{\Gamma_1}(t, ag) - \text{coo}_{\Gamma_2}(t, ag)\|,$$

где $Ag_c \subseteq Ag$ – множество вершин наибольшего общего подграфа Γ_1 и Γ_2 , $m_0 = |Ag_0 \setminus Ag_c|$, $m_1 = |Ag_1 \setminus Ag_c|$, $\kappa = (\text{diam}(\Omega_h))^{-1}$ – диаметр множества Ω_h в клетках. Это расстояние в своей сущности является частным случаем ранее введенного в [3] расстояния графов.

7. Синтез локальной функции перехода

Обратим внимание на то, что у двумерного *robot-space cellular automaton* окрестность $V_r(i, j)$ каждой ячейки в клетке (i, j) порождена окрестностью (i, j) в обычной евклидовой метрике \mathbb{R}^2 , и алгоритм нахождения скорейшего маршрута учитывает такую окрестность. Однако алгоритм построения строя основан именно на *robot-space cellular automaton* и в качестве множества ячеек рассматривает Ag , а в качестве окрестности агента $ag \in Ag$ – окрестности $W_\rho(ag)$, порожденные геодезическим расстоянием на графе строя Φ .

Доработаем алгоритм 1 так, чтоб агенты стремились в каждый момент времени поддерживать формацию, заданную графом Φ . Отметим, что без существенных потерь во времени прохождения поддерживать в точности заданную формацию можно лишь на ландшафтах с относительно малым количеством препятствий (иначе говоря, с малой конфигурационной энтропией).

Если предположить, что строй агентов в момент времени t описывается графом $\Gamma(t)$, то агенты стараются выбрать такое направление движения, которое было бы компромиссным между направлением, минимизирующим $\text{dist}(\Phi(t), \Gamma(t))$ (точнее, минимизирующим расстояние между графом строя и той части реального графа строя, которая известна агенту) и направлением в сторону кратчайшего по времени пути.

В рамках такой логики агентов, существуют следующие виды взаимодействия:

1. $\text{predict} : Ag \rightarrow \Omega_h$ – функция, с помощью которой агент предсказывает новое положение другого агента на следующем такте функционирования КА. Поскольку каждый агент каждый такт вычисляет и сортирует возможные направления движения по желательности (алгоритмы 3 и 7), то $\text{predict}(ag)$ может возвращать, например, $d_0 + (i, j)$, где d_0 – самое желательное для агента в клетке (i, j) направление. Проще говоря, все агенты содержат одинаковый алгоритм поведения. Поэтому каждый агент может «мысленно» подставить себя на место другого и просчи-

тать, куда бы пошел он сам в этой ситуации. Результат такого подсчета и возвращает функция `predict`. Более формально, агент вызывает функцию вычисления весов не со своими координатами на плоскости и не со своей позицией в графе строя, а с координатами и позицией в строю того агента, для которого он хочет предсказать положение.

2. Рекурсивный запрос остановки. Агент просит остановиться своего лидера, который просит остановиться своего лидера и т.п., пока запрос не дойдет до агента, не имеющего лидера.

3. Рекурсивный запрос продолжения движения. Агент просит продолжить движение своего лидера, который просит продолжить движение у своего лидера и т.п., пока запрос не дойдет до агента, не имеющего лидера.

Отметим, что поскольку $\text{dist}(\Phi, \Gamma) \geq 0$ и $T_h(r_h) > 0$ для любых графов Φ, Γ и маршрута r_h , то для скаляризации вышеупомянутых целевых функций их проще всего сложить. Идея алгоритма состоит в том, что

1. Находящиеся в одной окрестности $V_r(i, j)$ и в одной окрестности $W_\rho(ag)$ агенты выбирают лидера (алгоритм б), за которым следуют, причем так, чтоб лидер агентов одной окрестности обязательно бы следовал за лидером агентов в другой, кроме лидера самого высокого уровня. Множество агентов из $W_\rho(ag)$, найденных агентом ag в окрестности $V_r(i, j)$, обозначим как $\text{found}(ag)$.

2. Каждому возможному направлению $d \in \mathcal{D}$ движения агента ag в клетке (i, j) присваиваются веса, учитывающие как направление в сторону кратчайшего пути, аналогично (8), так и в сторону сохранения строя, например

$$\begin{aligned} \text{weights}[d] = & \min_{r_h \in \mathcal{M}_d(i, j; i_r, j_r)} T_h(r_h) + \\ & + \sum_{ag' \in \text{found}(ag)} \|\text{coo}_\Phi(t, ag') + d - \text{predict}(ag')\|. \end{aligned}$$

3. Множество \mathcal{D} случайно перемешивается в соответствии с весами *weights*, получая множество \mathcal{D}' . Используется предложенный в [9] алгоритм случайного взвешенного перемешивания. Без перемешивания возможна ситуация, когда веса направлений движения в нескольких направлениях будут равны и неясно, какой вес выбрать. Также во многих случаях агенту будет тогда неясно, что предпочесть – удерживать строй и проиграть в скорости или же отклониться от строя и выиграть в скорости. Более того, перемешиванием моделируется ситуация, когда люди, например, неверно угадывают направление движения соседа и сталкиваются с ним.

4. Агент выбирает лучшее направление из перемешанного \mathcal{D}' , если не слишком отстал от лидера. Поскольку перемешивание не равномерно, а учитывает веса (т.е. самые оптимальные маршруты имеют самый большой шанс оказаться первыми) с помощью особого алгоритма, то поддержание строя от такого перемешивания практически не страдает. У экспоненциального распределения, которое используется для перемешивания, очень легкий хвост и вероятность выбора направления, сильно нарушающего строй, близка к нулю.

5. Если агент слишком отстал, то для движения за лидером агент в клетке (i, j) вычисляет положение лидера (i_f, j_f) согласно своему шаблону формации, и ищет в окрестности $V_\varepsilon(i_f, j_f)$ своего лидера. При обнаружении лидера ag_l в $(i_l, j_l) \in V_\varepsilon(i_f, j_f)$ агент выбирает такое направление движения d , чтоб минимизировать $\|(i, j) + d - \text{predict}(ag_l)\|$.

6. Если в окрестности $V_r(i, j)$ ведущего агента ag слишком много агентов, не принадлежащих $W_\rho(ag)$ или, наоборот, не хватает агентов, принадлежащих $W_\rho(ag)$, то ag пытается выбрать другое направление (строка 24 алгоритма 2) или запросить «лишних» или «недостающих» агентов выбрать другое направление. Подробный разбор механизма обмена сообщениями в среде «Психодод» [5] выходит за рамки настоящей работы, но, в целом, когда агент обнаруживает, что ему пришло сообщение типа «запрос другого направления», то он генерирует веса направлений

заново, причем устанавливает вес ранее выбранного направления в 0, делая невозможным выбор этого направления.

Рекурсивные запросы от ведомых агентов к лидерам нужны для восстановления сильных нарушений строя или на очень неоднородных ландшафтах, когда самоорганизация для необобщающихся агентов не работает. При этом, если лидер ag заметил отклонение от строя ведомых агентов из множества $found$, то он вместо остановки выбирает направление движения

$$\text{def}(ag) = \frac{\sum_{ag' \in found} (\text{соог}(ag') - \text{сооф}(ag'))}{\|\sum_{ag' \in found} (\text{соог}(ag') - \text{сооф}(ag'))\|}.$$

Подробно вышеуказанное поведение агентов может быть описано как алгоритм 2, в который входят алгоритмы 5-7, приведенные в Приложении.

Алгоритм 2 (Поддержания строя).

- 1: **for all** $ag \in Ag$ **do**
- 2: $weights = \text{SORTDIRECTIONS}(ag)$ ▷ Алгоритм 3
- 3: $predicted, less, more, found = \text{FINDNEIGHBOURS}(ag, \varepsilon)$ ▷
 Алгоритм 5
- 4: **if** $t == 0$ **then**
- 5: $\text{FINDLEADER}(ag, found)$ ▷ Алгоритм 6
- 6: **end if**
- 7: $distances = \text{FINDDIRECTION}(ag, found, predicted)$ ▷
 Алгоритм 7
- 8: Случайно перемешать $ag.D'$ в соответствии с весами из $weights$ и $distances$.
- 9: $d_0 = ag.D'[1]$ ▷ Лучшее направление
- 10: **if** $ag.lead \wedge distances[d_0] \gg r_0$ **then** ▷ r_0 – допустимое
 расхождение со строем
- 11: $trgtTmp = trgtAct$ ▷ Назначить временную цель
- 12: $trgtAct = (ag.@cell \rightarrow coo) + \text{def}(ag)$
- 13: Рекурсивный запрос остановки агента с УИД $ag.leadId$
- 14: **else if** $ag.lead \wedge distances[d_0] \approx r_0 \wedge$ был рекурсивный
 запрос остановки **then**

```

15:      Продолжить движение
16:      Рекурсивный запрос продолжить движение агента с
      УИД  $ag.leadId$ 
17:      else if  $\neg ag.lead \wedge distances[d_0] \gg r_0$  then
18:      Рекурсивный запрос остановки агента с УИД
       $ag.leadId$ 
19:       $trgtTmp = trgtAct$        $\triangleright$  Назначить временную цель
20:       $trgtAct = found[leadId] - (i_s, j_s)$        $\triangleright$ 
       $(i_s, j_s, ag.selfId) \in ag.formTmpl$  – собственное положение
      агента в строю
21:      else if  $distances[d_0] \approx r_0 \wedge$  была назначена временная
      цель then
22:       $trgtAct = trgtTmp$ 
23:      Рекурсивный запрос продолжить движение агента с
      УИД  $ag.leadId$ 
24:      else if длина more или less слишком велика then
25:      Случайно перемешать  $ag.D'$  в соответствии с весами
      из weights и distances.
26:      end if
27: end for
28: for all  $ag' \in Ag$  do       $\triangleright$  Актуальное перемещение агентов
29:   ACTUALIZE( $ag'$ )       $\triangleright$  Алгоритм 4
30: end for
31:  $t = t+1$ 

```

8. Результаты симуляции

Предложенный алгоритм организации строя был смоделирован в программе «Психолод»³, разработанной в [5], и при наличии небольших одиночных препятствий агенты ведут себя так, как показано на рис. 5а. На указанном рисунке изображено шесть агентов ag_i , $i = \overline{1, 6}$, имеющих граф строя, показанный на рис. 5б. Для ag_2 , ag_4 лидером является агент ag_1 , для ag_3 , ag_5 – агент ag_2 , для ag_6 – агент ag_3 . Агенты пытаются наименее времязатратным

³ <https://bitbucket.org/bokohodteam/bokohod>.

образом обойти препятствие (показанное в виде прямоугольника), после чего восстанавливают строй.

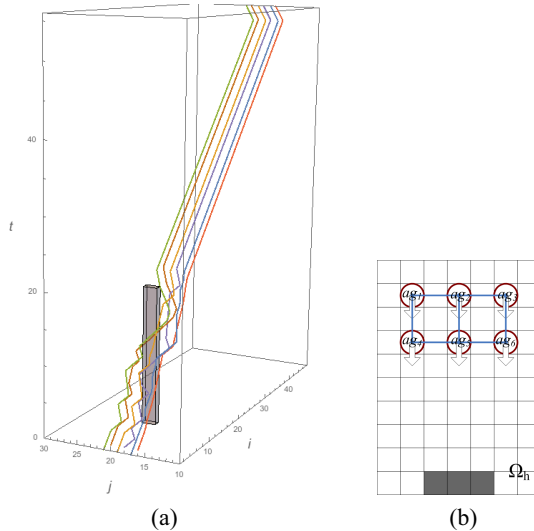


Рис. 5: Обход небольшого одиночного препятствия строем из шести агентов

Однако когда ландшафт становится более неоднородным, эффективность построения строя снижается. Причины у этого следующие: агенты теряют способность адекватно предсказывать направления движения друг у друга и агентам негде развернуть правильный строй после его нарушения из-за препятствия, так как препятствия встречаются слишком часто.

Обозначим

$$\Delta\Phi = \frac{1}{T} \sum_{ag \in Ag} \sum_{t=\overline{1, T}} \|\text{соо}_\Phi(t, ag) - \text{соо}_\Gamma(t, ag)\|,$$

где $\text{соо}_\Phi(t, ag)$ – желаемое положение агента ag в такт $t = \overline{1, T}$ согласно графу строя Φ , $\text{соо}_\Gamma(t, ag)$ – реальное положение агента

ag в такт $t = \overline{1, T}$. Моделирование⁴ вышеупомянутого строя из шести агентов на «естественном ландшафте», типа указанного в разделе 5, дает результат, приведенный на рис. 6. На этом рисунке показана зависимость $\Delta\Phi/6$ (множитель $1/6$ взят, потому как агентов шесть, и смысл величины $\Delta\Phi/6$ – это среднее отклонение агента от своей позиции в строю) от количества препятствий N_{obst} , которая аппроксимируется с коэффициентом детерминации $R^2 = 0,992729$ как

$$\Delta\Phi/6 \approx 0,901073 \ln(N_{obst} + 1) - 0,673763.$$

Выяснилось, что уменьшить отклонение агентов от строя

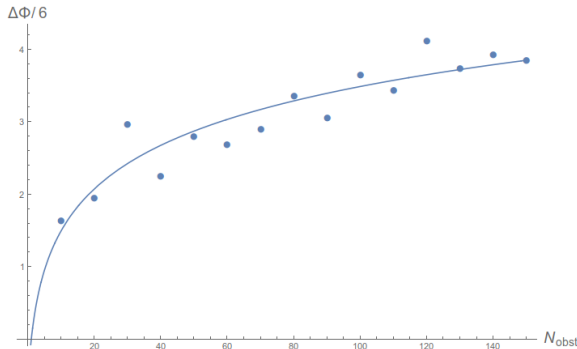


Рис. 6: Зависимость среднего отклонения позиции агентов от предусмотренной строем и количества препятствий

возможно, увеличивая параметр ε , указывающий, в насколько широкой окрестности $V_\varepsilon(i_f, j_f)$ заданной шаблоном строя, $(i_f, j_f, agId)$ следует искать агента с УИД $agId$.

⁴ Данные доступны по адресу https://www.researchgate.net/publication/318663064_Eksperimentalnye_dannye_otnositelno_obhoda_prepastsvij_stroem_iz_sesti_agentov.

9. Заключение

Получены зависимости эффективности построения квазиоптимального маршрута от типа ландшафта и временная вычислительная сложность этого построения. Разработан алгоритм построения и поддержания строя агентов,двигающихся по ландшафту с препятствиями, в котором в качестве критерия близости получившегося строя к заданному используется определенная метрика несходства графов. Также описана новая методика тестирования алгоритмов нахождения оптимального маршрута с помощью случайных ландшафтов с заданными характеристиками типа конфигурационной энтропии.

Клеточный автомат, сконструированный в статье, может быть использован как часть искусственного интеллекта наземного робота, передвигающегося в составе строя других подобных ему роботов. Хочется подчеркнуть, что клеточный автомат был положен в основу предлагаемой модели в силу нелинейности уравнений (1)–(3), определяющих движение агента, и невозможности получить в явном виде функционал времени прохождения к цели T для меняющегося со временем ландшафта. Перечисленные обстоятельства, а также тот факт, что функция проходимости u^c обычно задана в табличном виде (т.е. изначально разрывна) сильно препятствуют применению классических методов.

Далее планируется получить результат, аналогичный приведенному в статье, для организации телекоммуникационной *ad hoc* сети агентов, движущихся по пересеченной местности.

Приложение

В данном разделе приведены примеры ландшафтов различного вида. Подробное описание способов генерации ландшафтов приведено в работах [14, 15]. Также приведен псевдокод алгоритмов, необходимых для построения строя и поиска оптимального по времени маршрута.

Алгоритм 3 (Сортировки направлений движения).

```

1: function SORTDIRECTIONS(ag)
2:   (i, j) = ag.@cell → coo
3:   weights = {}
4:   if ag.w > 0 then ag.w = ag.w - 1
5:   end if
6:   if ag.w == 0 ∧ (i, j) ≠ ag.trgtAct then
7:     
$$direction = \frac{(i, j) - ag.trgtAct}{\|(i, j) - ag.trgtAct\|}$$

8:     (ir, jr) = клетка, в которой с границей  $V_r(i, j)$  пересекается луч с началом в (i, j) и направленный в направлении direction
9:     ag.D' =  $\mathcal{D}$ 
10:    for all d ∈  $\mathcal{D}$  do
11:      weights[d] =  $\min_{r_h \in \mathcal{M}_d(i, j; i_r, j_r)} J(r_h)$ 
12:    end for
13:  end if
14:  return weights
15: end function

```

Алгоритм 4 (Актуализации новых положений).

```

1: function ACTUALIZE(ag')
2:   k = 1
3:   (i, j) = ag'.@cell → coo
4:   repeat
5:     (inew, jnew) := (i, j) + ag'.D'[k]
6:     k = k + 1
7:   until ag'_{inew, jnew}} ≠ 0 ∨ (inew, jnew) ≠ (i, j)
8:   cell_{inew, jnew}}.@ag = @ag'
9:   ag'.@cell = @cell_{inew, jnew}}.coo
10: end function

```

Алгоритм 5 (Обнаружения соседей).

```

1: function FINDNEIGHBOURS( $ag, \varepsilon$ )
2:    $predicted = \{\}, less = \{\}, more = \{\}, found = \{\}$ 
3:    $(i, j) = ag.@cell \rightarrow coo$ 
4:   for all  $(s, p, agId) \in ag.formTmpl$  do ▷ Поиск в
      окрестности  $W_\rho(ag)$ 
5:     Искать в окрестности  $V_\varepsilon(i + s, j + p)$  агента
6:     if агент  $ag'$  найден и  $agId == 0$  then
7:        $agId = ag'.selfId$ 
8:        $predicted[agId] = predict(ag')$ 
9:       Добавить в  $found$   $agId$ 
10:    else if агент  $ag'$  найден и  $agId == ag'.selfId$  then
11:       $predicted[agId] = predict(ag')$ 
12:      Добавить в  $found$   $agId$ 
13:    else if агент  $ag'$  найден и  $agId \neq ag'.selfId \wedge agId \neq 0$ 
      then
14:      Добавить в  $more$   $ag'.selfId$  ▷ длина  $more$  – это
       $m_1$  из определения метрики
15:    end if
16:    if Длина( $ag.formTmpl$ ) > Длина( $predicted$ ) then
17:      Добавить в  $less$  УИД не найденных агентов ▷
      длина  $less$  – это  $m_0$  из определения метрики
18:    end if
19:  end for
20:  return  $predicted, less, more, found$ 
21: end function

```

Алгоритм 6 (Назначения и обнаружения лидера).

```

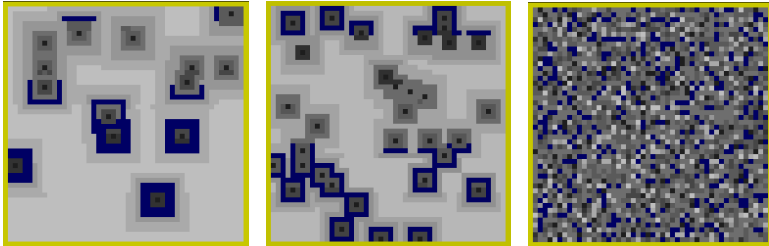
1: function FINDLEADER( $ag, found$ )
2:    $ag.leadId = 0$ 
3:   for all  $id \in found$  do ▷ агент выбирает лидером агента с
      самым маленьким УИД из своих соседей
4:     if  $id < ag.selfId$  then
5:        $ag.leadId = id$ 
6:     end if
7:   end for

```

8: **end function**

Алгоритм 7 (Поиска направления для поддержания формации).

```
1: function FINDDIRECTION(ag, found, predicted)
2:   dist = 0, distances = {}
3:   (i, j) = ag.@cell → coo
4:   for all d ∈  $\mathcal{D}$  do
5:     for all id ∈ found do
6:       dist = dist +  $\|d + (i, j) - \text{predicted}[id]\|$ 
7:     end for
8:     distances[d] = dist
9:   end for
10:  return distances
11: end function
```



(a) $S = 1,6$, $N_{obst} = 14$ (b) $S = 1,8$, $N_{obst} = 33$ (c) $S = 1,84877$

Рис. 7: Примеры ландшафтов. Чем клетка темнее, тем больше ее непроходимость

Литература

1. КУЗНЕЦОВ А.В. *Модель совместного движения агентов с трехуровневой иерархией на основе клеточного автомата* // Журнал вычислительной математики и математической физики. – 2017. – Т. 57, № 2. – С. 339–349. – URL: <https://elibrary.ru/item.asp?id=28918677>.
2. КУДРЯВЦЕВ В.Б., ПОДКОЛЗИН А.С., БОЛОТОВ А.А. *Основы теории однородных структур*. – М.: Наука, 1990. – 296 с.
3. КУЗНЕЦОВ А.В. *Мера несходства на множестве графов и ее приложения* // Вестник ВГУ. Серия: Системный анализ и информационные технологии. – 2017. – Т. 1. – С. 125–131. – URL: <https://elibrary.ru/item.asp?id=29185115>.
4. КУЗНЕЦОВ А.В. *Упрощенная модель боевых действий на основе клеточного автомата* // Известия РАН. Теория и системы управления. – 2017. – Т. 56, № 3. – С. 59–71. – URL: <https://elibrary.ru/item.asp?id=29369822>.
5. КУЗНЕЦОВ А.В., ЛЕЩЕВ А.С. *Программная среда многоагентного моделирования «Психодод»*. Пр. для ЭВМ № 2017619605. Дата регистрации: 28.08.2017, номер и дата поступления заявки: 2017616880 11.07.2017. Правообл. А.В. Кузнецов, А.С. Лещев // Программы для ЭВМ. Базы данных. Топологии интегральных микросхем. – 2017. – № 9.
6. МАЛИНЕЦКИЙ Г. Г., СТЕПАНЦОВ М. Е. *Применение клеточных автоматов для моделирования движения группы людей* // Журнал вычислительной математики и математической физики. – 2004. – Т. 44, № 11. – С. 2094–2098.
7. BEER B., MEAD R., WEINBERG J.B. *A Distributed Method for Evaluating Properties of a Robot Formation* // Proc. of the 24th Conference on Artificial Intelligence (AAAI-2010), July 11–15, 2010, Atlanta, Georgia, USA. – URL: <http://www.aaai.org/ocs/index.php/AAAI/AAAI10/paper/view/1677>.

8. CUSHMAN S.A. *Calculating the configurational entropy of a landscape mosaic* // Landscape Ecology. – 2016. – Vol. 31, No. 3. – P. 481–489. – URL: <http://dx.doi.org/10.1007/s10980-015-0305-2>.
9. EFRAIMIDIS P., SPIRAKIS P. *Weighted Random Sampling* // Encyclopedia of Algorithms / Ed. by Ming-Yang Kao. – Boston, MA: Springer US, 2008. – P. 1–99. – ISBN: <http://isbndb.com/search-all.html?kw=978-0-387-30162-4978-0-387-30162-4>. – URL: https://doi.org/10.1007/978-0-387-30162-4_478.
10. FRAGSTATS: *Spatial Pattern Analysis Program for Categorical Maps. Documentation.* – URL: http://www.umass.edu/landeco/research/fragstats/documents/fragstats_documents.html.
11. GEUVERS H. *Inductive and coinductive types with iteration and recursion* // Proc. of the 1992 workshop on Types for Proofs and Programs. – Bastad: Chalmers University of Technology, 1992. – P. 183–207. – URL: http://www.cs.ru.nl/herman/PUBS/BRABasInf_RecTyp.pdf.
12. ILACHINSKI A. *Artificial War: Multiagent-Based Simulation of Combat.* – Singapore: World Scientific Publishing Company, 2004. – P. 747.
13. JONES M.P., DUDENHOEFFER D.D. *A formation behavior for large-scale micro-robot force deployment* // Winter Simulation Conference. – Vol. 01. – Los Alamitos, CA, USA: IEEE Computer Society, 2000. – P. 972–982. – URL: doi.ieeecomputersociety.org/10.1109/WSC.2000.899900.
14. KUZNETSOV A.V. *Cellular automata-based model of group motion of agents with memory and related continuous model* // Mathematical Modeling. Information Technology and Nanotechnology – 2017 / Ed. by S. Sazhin, E. Shchepakina, V. Sobolev, D. Kudryashov. – CEUR Workshop Proc. No. 1904. – Aachen, 2017. – P. 223–231. – URL: <http://ceur-ws.org/Vol-1904/paper38.pdf>.

15. KUZNETSOV A. *Generation of a Random Landscape by given Configuration Entropy and Total Edge* // Computational Technologies. – 2017. – Vol. 22, No. 4. – P. 4–10. – URL: <https://elibrary.ru/item.asp?id=30053459>.
16. LONG R.L., MEAD R., WEINBERG J.B. *Distributed Auction-Based Initialization of Mobile Robot Formations* // Proc. of the 24th Conference on Artificial Intelligence (AAAI-2010), July 11-15, 2010, Atlanta, Georgia, USA. – 2010. – URL: <http://www.aaai.org/ocs/index.php/AAAI/AAAI10/paper/view/1673>.
17. MEAD R. *Cellular Automata for Control and Interactions of Large Formations of Robots*. – 2008. – URL: http://robotics.usc.edu/rossmead/docs/2008/2008Mead_Thesis.pdf.
18. MEAD R., WEINBERG J.B. *2-Dimensional Cellular Automata Approach for Robot Grid Formations* // Proc. of the 23rd Conference on Artificial Intelligence (AAAI-2008), July 13-17, 2008, Chicago, Illinois, USA. – 2008. – P. 1818–1819. – URL: <http://www.aaai.org/Library/AAAI/2008/aaai08-299.php>.
19. MEAD R., WEINBERG J.B. *A Single- and Multi-Dimensional Cellular Automata Approach to Robot Formation Control* // Proc. of IEEE Int. Conference on Robotics and Automation (ICRA-08). – 2008. – URL: http://robotics.usc.edu/rossmead/docs/2008/2008WeinbergMead_ICRA08.pdf.
20. RODRÁGUEZ-ITURBE I., D'ODORICO P., RINALDO A. *Configuration entropy of fractal landscapes* // Geophysical Research Letters. – 1998. – Vol. 25, No. 7. – P. 1015–1018. – URL: <http://dx.doi.org/10.1029/98GL00654>.

ORGANIZATION OF AN AGENTS' FORMATION THROUGH A CELLULAR AUTOMATON

Alexander Kuznetsov, Voronezh State University, Voronezh, Cand.Sc., associate professor (avkuz@bk.ru).

Abstract: The article deals with the algorithm for a distributed organization of the agents' formation defined by a graph and the numerical simulation of such algorithm. Agents move through terrain with many random obstacles ("random landscape"). At first, we describe the continuous statement of the two-criteria minimization problem. The first criterion is the agent's route time. The second criterion is the closeness of agents' formation to the desired one. Next, we introduce a cellular automaton simulating the movement of agents for obtaining quasi-optimal solutions of the problem. The cellular automaton has one-dimensional and two-dimensional representations. Agents use reflexion to predict the motion of other agents. Then we compare obtained solutions with optimal ones for different types of random landscapes via numerical experiment. At this point, we obtain the empirical distribution for the time of an agent's exit to finish point. Finally, we find the relation between a type of random landscape through which agents move and the quality of agents' formation.

Keywords: cellular automaton, autonomous agents, reflexive agents, agents' formation control.

Статья представлена к публикации членом редакционной коллегии А.А. Шевляковым.

*Поступила в редакцию 08.08.2017.
Дата опубликования 30.11.2017.*