

Институт проблем управления
им. В.А. Трапезникова РАН

УПРАВЛЕНИЕ БОЛЬШИМИ СИСТЕМАМИ

Выпуск 108
Март 2024

**СБОРНИК
ТРУДОВ**

ISSN 1819-2467

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

Москва – 2024

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

**УПРАВЛЕНИЕ
БОЛЬШИМИ
СИСТЕМАМИ**

СБОРНИК ТРУДОВ

Выпуск 108

Москва – 2024

КООРДИНАЦИОННЫЙ СОВЕТ

Академики: Васильев С.Н., Желтов С.Ю., Каляев И.А., Куржанский А.Б., Новиков Д.А., Попков Ю.С., Черноусько Ф.Л.; члены-корреспонденты РАН: Галяев А.А., Савватеев А.В.; д-ра техн. наук: Кузнецов О.П. (ИПУ РАН), Кульба В.В. (ИПУ РАН), Павлов Б.В. (ИПУ РАН), Рутковский В.Ю. (ИПУ РАН).

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

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

Д-ра техн. наук: проф. Алескеров Ф.Т. (ГУ ВШЭ), проф. Алчинов А.И. (ИПУ РАН), проф. Андриевский Б.Р. (ИТМО), проф. Афанасьев В.Н. (МИЭМ, ГУ ВШЭ), проф. Бахтадзе Н.Н. (ИПУ РАН), проф. Бурков В.Н. (ИПУ РАН), проф. Вишневский В.М. (ИПУ РАН), д-р физ.-мат. наук проф. Ерешко Ф.И. (ВЦ РАН), д-ра техн. наук: проф. Калянов Г.Н. (ИПУ РАН), проф. Каравай М.Ф. (ИПУ РАН); д-р экон. наук, проф. Клочков В.В. (ИПУ РАН); д-р техн. наук доцент Коргин Н.А. (ИПУ РАН); д-ра физ.-мат. наук: проф. Кушнер А.Г., проф. Лазарев А.А. (ИПУ РАН); д-ра техн. наук: проф. Лебедев В.Г. (ИПУ РАН), проф. Мандель А.С. (ИПУ РАН); д-р биол. наук проф. Михальский А.И. (ИПУ РАН); д-р экон. наук, проф. Нижегородцев Р.М. (ИПУ РАН); д-ра физ.-мат. наук: проф. Райгородский А.М. (МФТИ), проф. Рапопорт Л.Б. (ИПУ РАН); д-ра техн. наук: проф. Самуйлов К.Е. (РУДН), проф. Сидельников Ю.В. (МАИ), проф. Совлуков А.С. (ИПУ РАН); д-ра физ.-мат. наук: проф. Соловьев С.Ю. (МГУ), проф. Угольницкий Г.А. (ЮФУ); д-ра техн. наук: проф. Уткин В.А. (ИПУ РАН), проф. Хоботов Е.Н. (ИПУ РАН); д-ра физ.-мат. наук: 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. (АПИ НГТУ); **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. (ВГУ), д-р техн. наук, проф. Подвальный С.Л. (ВГТУ); **Иркутск** – д-р техн. наук проф. Зоркальцев В.И. (ЛИН СО РАН), 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., проф. Сараяв П.В. (ЛГТУ); **Самара** – д-ра экон. наук: проф. Богатырев В.Д., проф. Гераськин М.И.; **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. Фургат И.Б. (ИПМ РАН).

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

Адрес в интернете: ubs.mtas.ru.

СОДЕРЖАНИЕ

Системный анализ

Горбунова А.В., Лебедев А.В. <i>О новом подходе к оценке квантилей времени отклика системы с разделением и параллельным обслуживанием заявок.....</i>	6
Назаров А.А., Рожкова С.В., Титаренко Е.Ю. <i>Асимптотический анализ системы $M^{[n]}/GI/1$ с учетом остаточного времени обслуживания.....</i>	22
Полин Е.П., Моисеева С.П., Моисеев А.Н. <i>Применение отрицательного биномиального распределения для аппроксимации стационарного распределения числа заявок в СМО с входящим МАР-поток, интенсивность которого зависит от состояния системы</i>	40
Шумов В.В. <i>Модели и методики обоснования плотностей пограничных сил, обеспечивающих надежную охрану государственной границы</i>	57

Анализ и синтез систем управления

Парсегов С.Э., Пугач М.А., Ерофеева В.А. <i>Методы вычисления состояния заряда ванадиевых проточных аккумуляторов: анализ взаимосвязей</i>	78
--	----

Информационные технологии в управлении

Диане С.А.К., Выговтов К.А., Барабанова Е.А. <i>Алгоритм анализа мультиспектральных аэрофотоснимков для идентификации загрязнений водоемов с использованием аналитических методов и нейросетевых подходов.....</i>	98
--	----

Управление в социально-экономических системах

Козицин И.В.

Построение прогноза динамики общественного мнения при помощи SCARDO-модели..... 124

Нижегородцев Р.М., Рослякова Н.А., Горидько Н.П.

Логистические модели жизненного цикла технологий как инструмент оценки эффективности затрат на НИОКР..... 137

Шульц В.Л., Чернов И.В.

Использование виртуальных структур при формировании сценарно-когнитивных моделей на основе использования экспертных знаний..... 156

Управление техническими системами и технологическими процессами

Брокарев И.А., Васьковский С.В., Фархадов М.П.

Автоматизированная информационная система анализа качества природного газа..... 174

Управление подвижными объектами и навигация

Романова В.Р., Зувев С.В.

Адаптивная система траекторного управления АНПА на базе нейронной сети прямого распространения..... 192

О НОВОМ ПОДХОДЕ К ОЦЕНКЕ КВАНТИЛЕЙ ВРЕМЕНИ ОТКЛИКА СИСТЕМЫ С РАЗДЕЛЕНИЕМ И ПАРАЛЛЕЛЬНЫМ ОБСЛУЖИВАНИЕМ ЗАЯВОК

Горбунова А. В.¹

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

Лебедев А. В.²

(Московский государственный университет
имени М.В. Ломоносова, Москва)

Предлагается новый подход к оценке квантилей распределения времени отклика системы массового обслуживания с разделением и параллельным обслуживанием заявок, для обозначения которой в англоязычной литературе используется термин fork-join. Рассматривается классический вариант данной системы с пуассоновским входным потоком и экспоненциальными временами обслуживания на однородных приборах. Заявки при поступлении в систему мгновенно разделяются на фиксированное число подзаявок и отправляются на обслуживание в соответствующие подсистемы с накопителем неограниченной емкости и одним прибором. Заявка считается обслуженной после обслуживания всех ее компонентов. Данная система позволяет смоделировать множество реальных процессов, для которых с целью увеличения эффективности характерно разделение крупных задач на более мелкие составляющие, например, системы параллельных или распределенных вычислений. Сложность анализа систем заключается в наличии зависимости между временами пребывания подзаявок, что значительно затрудняет анализ всех показателей производительности таких систем. Основным вкладом статьи является подход к определению квантилей распределения времени отклика, оценка которых является не менее ценной по сравнению с оценкой среднего значения времени отклика. При этом вычислению математического ожидания посвящено гораздо большее количество работ в данной области, что объясняется в том числе сложностью проведения подобного анализа даже для данной характеристики, а оценка квантилей представляется еще более трудоемкой задачей.

Ключевые слова: система с параллельным обслуживанием заявок, fork-join система массового обслуживания, время отклика, квантили распределения, имитационное моделирование.

¹ Анастасия Владимировна Горбунова, к.ф.-м.н. (avgorbunova@list.ru).

² Алексей Викторович Лебедев, д.ф.-м.н. (avlebed@yandex.ru).

1. Введение

В работе исследуется классическая fork-join система массового обслуживания (СМО) с пуассоновским входным потоком и экспоненциальными временами обслуживания. Fork-join СМО является математической моделью для множества реально существующих систем, в которых происходит распараллеливание задач. При поступлении в систему заявка разделяется (fork point) на число подзаявок, равное числу подсистем $K \geq 2$. Все подсистемы фактически представляют собой самостоятельные системы массового обслуживания с бесконечной очередью и единственным прибором. Каждая из подзаявок поступает в одну из подсистем, обслуживается там, после чего попадает в условный буфер синхронизации (join point), где ожидает обслуживания оставшихся частей заявки. После окончания обслуживания всех подзаявок происходит мгновенная сборка целой заявки, и она может покинуть систему.

Ранее такие системы изучались авторами в работах [10, 11]. Так, в статье [11] с помощью искусственных нейронных сетей аппроксимируются математическое ожидание и среднее квадратическое отклонение времени отклика fork-join СМО для различного числа подсистем K . При этом был смоделирован большой объем данных (который используется далее в [10] и в настоящей работе) в форме наборов из 20 времен пребывания подзаявок от каждой заявки при значениях загрузки от 0,1 до 0,9 с шагом 0,05, по 5 миллионов заявок при загрузке менее 0,5 и по 10 миллионов при загрузке от 0,5 и выше. Наличие таких наборов позволило оценить характеристики систем для всех $3 \leq K \leq 20$. Приближение нейронной сетью показало среднюю абсолютную относительную ошибку для среднего времени отклика 0,74%, а для среднее квадратического отклонения – 0,36%, для рассмотренных в работе выборов.

В [10] для оценки аналогичных характеристик использовалось расширение предложенного подхода на основе комбинации различных методов машинного обучения, кроме того обсужда-

лись вопросы имитационного моделирования системы, а также построения доверительных интервалов. С помощью выведенных формул удалось повысить точность оценки среднего времени отклика более чем в 11 раз по сравнению с классической формулой Нельсона – Тантави [16], а точность оценки среднеквадратического отклонения – более чем в 25 раз по сравнению с приближением на основе независимых случайных величин. Была изучена также корреляционная функция последовательности времен отклика заявок, зависимость которых приводит к увеличению дисперсии оценки среднего, и выведена приближенная формула этой дисперсии.

Текущая работа продолжает исследование зависимости времен пребывания подзаявок (частей одной заявки). С анализом fork-join системы, но с более сложной архитектурой, можно ознакомиться, например, в одной из последних работ [9]. Заметим, что погрешность приближений различных характеристик оценивается с помощью результатов имитационного моделирования системы, точность которых обеспечивается большим числом прогонов имитационной модели.

Интерес к исследованию fork-join систем объясняется наличием множества реальных физических систем, функционирующие которых может быть ими смоделировано, особенно в области моделирования информационных систем и протекающих в них процессов. Разумеется, область применения fork-join моделей не ограничивается только информационно-вычислительными системами. Вопросы оптимизации процессов в производственных системах (например, сборка заказов в складских системах, изготовление многокомпонентных изделий) или повышение эффективности организации процесса пребывания пациентов в медицинских учреждениях и т.д. продолжают оставаться востребованными до настоящего момента [1, 2, 7, 12, 20].

Другой причиной актуализации исследований fork-join систем является появление новых методов и подходов к анализу сложных систем массового обслуживания, в частности, подхода, основанного на применении методов машинного обучения и его

различные модификации [3, 4, 5, 19]. В данном случае речь идет о дальнейшем развитии названного подхода и включении в его состав графического анализа и методов оптимизации.

Несмотря на кажущуюся простоту функционирования, исследование fork-join СМО относится к труднорешаемым задачам. Основная причина сложности заключается в общности моментов появления подзаявок в подсистемах, что делает времена их пребывания зависимыми случайными величинами. В этом и есть основное отличие fork-join системы от просто параллельно функционирующих СМО того же типа, что и подсистемы fork-join СМО. Поэтому точные результаты были получены только для среднего времени отклика в случае двух подсистем с пуассоновским входным потоком и экспоненциальными временами обслуживания [16]. Для других вариантов архитектур fork-join СМО, под которыми подразумевается, например, увеличение числа подсистем или приборов в них, ограниченная емкость накопителей или другие типы распределений для входных и обслуживающих потоков, получены только аппроксимации математического ожидания времени отклика различными способами.

Что касается оценки других характеристик производительности fork-join системы, то исследований в этом направлении гораздо меньше [6, 11]. Однако помимо первых или вторых моментов случайной величины времени отклика интерес представляют квантили ее распределения. Установление квантиля заданного уровня вероятности означает, что система гарантирует обслуживание заявки не более чем за установленное время с данной вероятностью. Такой подход к оценке работы системы уместен, если важна срочность обслуживания (например, в медицине) или если долгое обслуживание вызывает недовольство клиентов и это не компенсируется быстрым обслуживанием других. Таким образом, оценка квантилями является более тонкой, чем оценка средним, в смысле качества обслуживания в современном мире.

В [18] с помощью векторно-матричной техники и фазовых распределений получены теоретические оценки хвоста и квантилей высоких уровней. В [17] были получены оценки для кванти-

лей высоких уровней в условиях высокой загрузки системы для нескольких типов распределений и архитектур fork-join СМО.

В данной работе предлагается подход для нахождения квантилей времени отклика различного уровня для более широкого диапазона загрузок, что позволяет составить полноценное представление о поведении исследуемой случайной величины. Несмотря на то, что объектом исследования статьи является классическая fork-join СМО, предложенный в статье подход может быть распространен и на другие архитектуры fork-join систем и не только.

2. Математическая модель fork-join СМО

Опишем более подробно процесс функционирования fork-join системы. Будем рассматривать частный случай двух подсистем ($K = 2$), однако заметим, что количество подсистем никак не влияет на зависимость в любой паре времен пребывания подзаявок одной заявки (рис. 1). В систему поступает пуассоновский поток заявок с интенсивностью $\lambda > 0$. В момент поступления в систему заявка мгновенно разделяется на 2 подзаявки, каждая из которых попадает в соответствующую подсистему, имеющую накопитель неограниченной емкости и один прибор. Все приборы являются однородными, время обслуживания имеет экспоненциальное распределение с параметром $\mu > 0$.

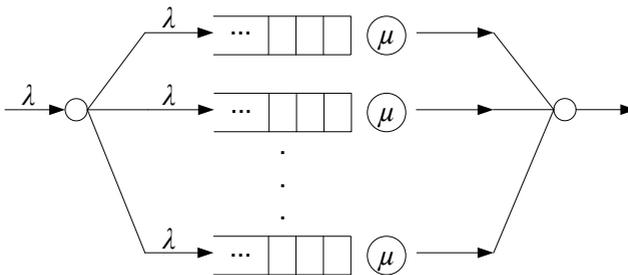


Рис. 1. Модель fork-join системы массового обслуживания с подсистемами типа $M_\lambda | M_\mu | 1$

Таким образом, подсистемы представляют собой две идентичных СМО типа $M|M|1$. Поскольку заявка считается обслуженной только после окончания обслуживания обеих составляющих ее подзаявок, то случайное время пребывания заявки в СМО (время отклика) R является максимумом из двух случайных времен пребывания подзаявок ξ_i , $i = 1, 2$, в каждой из двух подсистем:

$$(1) \quad R = \max\{\xi_1, \xi_2\}.$$

Случайные величины ξ_1 и ξ_2 являются коррелированными в силу того, что все подзаявки (части одной заявки) поступают в подсистемы в одно и то же время.

Обозначим через $\rho = \lambda/\mu$ коэффициент загрузки системы и далее для простоты положим $\lambda = 1$, $\mu = 1/\rho$. Такие параметры использовались при имитационном моделировании в [10, 11].

Отметим, что различные коэффициенты корреляции в общем случае хотя и отражают зависимость, но лишь частично. В полной мере отражают зависимость только копулы.

3. Подход к оценке диагонального сечения и квантилей распределения времени отклика

Для приближения квантилей распределения случайной величины времени отклика $R = \max\{\xi_1, \xi_2\}$ воспользуемся элементами теории копул.

Копулой C называется функция многомерного распределения на $[0, 1]^d$, $d \geq 2$, если все частные распределения являются равномерными на $[0, 1]$. Согласно знаменитой теореме Склера, любая функция многомерного распределения в \mathbb{R}^d представима в виде

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)),$$

где F_i , $1 \leq i \leq d$, — функции частных распределений. Таким образом, всякому многомерному распределению можно поставить в соответствие его копулу. Если частные распределения непрерывны, то такое представление единственно. В качестве классического учебника по копулам укажем [15], а в качестве работ авторов,

связанных с копулами, можно указать [8, 13, 14]. Далее ограничимся случаем двумерных копул ($d = 2$).

Диагональным сечением (двумерной) копулы называется функция $\delta(u) = C(u, u)$, $u \in [0, 1]$. Она обладает следующими (необходимыми и достаточными) свойствами:

$$(2) \quad \begin{aligned} \max\{2u - 1, 0\} &\leq \delta(u) \leq u; \\ 0 &\leq \delta(u_2) - \delta(u_1) \leq 2(u_2 - u_1), \quad 0 \leq u_1 \leq u_2 \leq 1. \end{aligned}$$

Смысл изучения диагональных сечений, например, в следующем. Если заданы случайные величины X_1 и X_2 с одинаковыми частными распределениями $F_1 = F_2 = F$ и копулой совместного распределения C , то их максимум $X_{max} = \max\{X_1, X_2\}$ имеет функцию распределения

$$(3) \quad F_{max}(x) = P(X_1 < x, X_2 < x) = C(F(x), F(x)) = \delta(F(x)),$$

так что для ее вычисления достаточно знать только диагональное сечение, а не всю копулу.

Легко заметить, что условиям (2) удовлетворяет степенная функция

$$\delta(u) = u^\alpha, \quad 1 \leq \alpha \leq 2,$$

тогда случаю $\alpha = 1$ соответствует совершенная положительная зависимость (комонотонность), а случаю $\alpha = 2$ – независимость случайных величин.

Будем рассматривать двумерную копулу $C(u_1, u_2)$ случайных векторов времен пребывания в подсистемах (ξ_1, ξ_2) . Каждая компонента случайного вектора имеет экспоненциальное распределение с функцией распределения $F(x) = 1 - e^{-(\mu-\lambda)x}$, $x \geq 0$. Тогда в соответствии с теоремой Склера представление с помощью копулы совместного распределения (ξ_1, ξ_2) существует и единственно:

$$F_{\xi_1, \xi_2}(x_1, x_2) = P(\xi_1 < x_1, \xi_2 < x_2) = C(F(x_1), F(x_2)).$$

В силу (3) получаем

$$(4) \quad F_R(x) = C(F(x), F(x)) = \delta(F(x)),$$

где $\delta(u) = C(u, u)$ — диагональное сечение копулы, что дает нам уравнение для квантили уровня p распределения времени отклика

$$F_R(x_p) = \delta(F(x_p)) = p,$$

поэтому

$$(5) \quad x_p = F_R^{-1}(p) = F^{-1}(\delta^{-1}(p)).$$

Учитывая метод обратного преобразования для генерации случайных величин с заданной функцией распределения, рассмотрим

$$U_i = 1 - e^{-(\mu-\lambda)\xi_i}, \quad i = 1, 2.$$

Эти случайные величины будут иметь равномерное распределение на отрезке $[0, 1]$, т.е. $U_i \sim R[0, 1]$. Тогда

$$(6) \quad V = \max\{U_1, U_2\} = 1 - e^{-(\mu-\lambda) \cdot \max\{\xi_1, \xi_2\}} = 1 - e^{-(\mu-\lambda)R}.$$

Диагональное сечение копулы можно оценить следующим образом. Имеем

$$\begin{aligned} \delta(u) &= C(u, u) = P(U_1 < u, U_2 < u) = \\ &= P(\max(U_1, U_2) < u) = P(V < u) = p, \end{aligned}$$

т.е.

$$\delta(u_p) = P(V < u_p) = p,$$

где u_p — это квантиль распределения с. в. V . С помощью реализаций V_i случайной величины V , полученных посредством имитационного моделирования значений случайных времен пребывания в fork-join СМО R_i и дальнейшей подстановкой их в формулу (6), строим оценку диагонального сечения $\delta(u)$, а фактически вероятностей p . Иными словами, строим эмпирическую оценку диагонального сечения с помощью квантилей распределения V .

Для этого упорядочиваем полученные посредством симуляции величины V : $V_{(1)}, V_{(2)}, \dots, V_{(N)}$, где $V_{(k)}$ — это k -я порядковая статистика, $k = 1, \dots, N$, и по точкам $(V_{(k)}, k/(N + 1))$ определяем оценки (u_p, p) для значений вероятности из интересующего нас интервала $p \in \{0,2; 0,25; 0,30; \dots; 0,90\}$, при конкретном фиксированном значении коэффициента загрузки

$\rho \in \{0,10; 0,15; 0,20; \dots; 0,90\}$. Выбор значений ρ обусловлен тем, что, как правило, больший интерес представляют квантили более высокого уровня, поэтому значения ρ начинаем рассматривать с 0,2. Далее на основе имеющихся данных будем строить прогноз вероятностей p в зависимости от квантилей u_p и коэффициента загрузки ρ :

$$p \approx \hat{p} = \hat{\delta}(u_p, \rho).$$

Теперь для определения вида функциональной зависимости проведем графический анализ полученных данных. Прежде всего было замечено, что зависимость p от u_p хорошо описывается степенной функцией, что соответствует линейной зависимости для логарифмов (см. рис. 2).

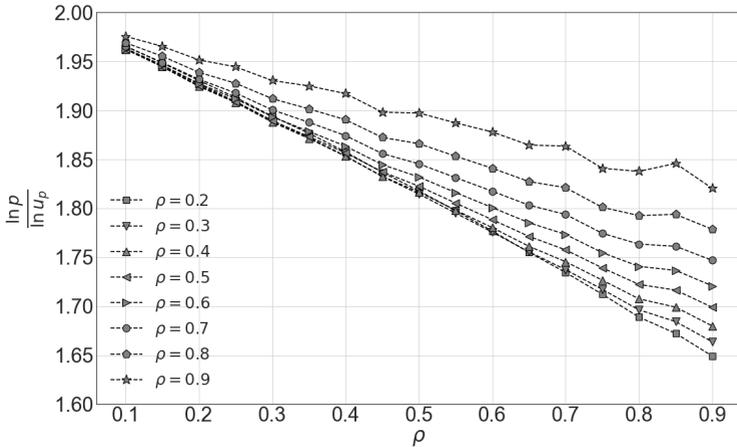


Рис. 2. Зависимость $(\ln p / \ln u_p)$ от ρ .

Зависимость показателя степени α от ρ также оказалась близка к линейной (см. рис. 3). Отметим, что при $\rho \rightarrow 0$ времена пребывания подзаявок асимптотически независимы, откуда $\alpha \rightarrow 2$.

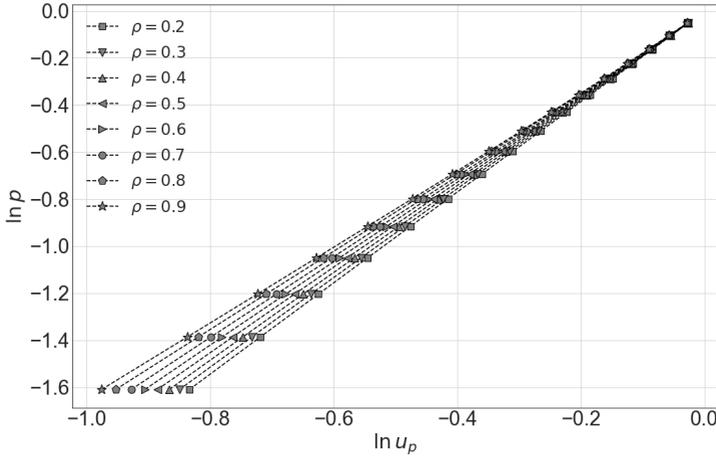


Рис. 3. Зависимость $\ln p$ от $\ln u_p$

Как видно из рис. 3, график зависимости напоминает пучок близких прямых, проходящих через точку $(0, 2)$, поэтому естественно предположить (в качестве первого приближения), что

$$\frac{\ln p}{\ln u_p} \approx 2 - C \cdot \rho,$$

а следовательно

$$(7) \quad p = \delta(u_p, \rho) \approx u_p^{2-C \cdot \rho}.$$

Остается только вычислить значение коэффициента C . Для этого будем минимизировать методом Нелдера – Мида модуль относительной погрешности аппроксимации относительно данных имитационного моделирования, в результате чего получим значение

$$(8) \quad C \approx 0,370608.$$

Таким образом, имеем

$$(9) \quad p = \delta(u_p, \rho) \approx u_p^{2-0,370608 \cdot \rho}.$$

Сравнение результатов имитационного моделирования вероятностей или уровней p квантилей u_p случайной величины

$V = F(R)$ с результатами вычислений по аналитической формуле (9) в диапазоне $[0,20; 0,95]$ с шагом 0,05, т.е. фактически для 272 рассчитанных значений p , позволяет говорить о хорошем уровне приближения. Так, средняя погрешность приближения составляет около 0,44%, в то время как максимальное и минимальное значение погрешности аппроксимации составляет 1,68% и 0,003% соответственно.

Теперь, учитывая (5), можем записать
 (10)
$$\delta^{-1}(p) = F(x_p),$$
 при этом из (7) следует, что $\delta^{-1}(p) \approx p^{\frac{1}{2-C \cdot \rho}}$. Подставляем оценку $\delta^{-1}(p)$ в (10) и получаем соотношение

$$p^{\frac{1}{2-C \cdot \rho}} = 1 - e^{-(\mu-\lambda)x_p},$$

откуда следует, что квантиль уровня p распределения случайной величины времени отклика fork-join СМО R определяется выражением

$$(11) \quad x_p \approx -\frac{\ln(1 - p^{\frac{1}{2-C \cdot \rho}})}{\mu - \lambda}.$$

Далее аналогично оценим качество аппроксимации полученного выражения (11) для 272 рассчитанных значений квантилей при $\rho \in \{0,10; 0,15; \dots; 0,90\}$ и $p \in \{0,20; 0,25; \dots; 0,90\}$. Получаем, что максимум модуля относительной ошибки составляет около 3,12%, а среднее значение этого модуля равно примерно 0,73%.

Ради уточнения аппроксимации квантилей, возвращаясь к рис. 2, можно отметить зависимость наклона прямых от p . Это наводит на мысль вместо константы C в (11) использовать выражения вида $(C_1 - C_2 p)$ или $(C_1 - C_2 p^2)$. Подбор констант методом Нелдера – Мида и сравнительный анализ точности показывают, что лучше второй вариант, а именно, приближение

$$(12) \quad x_p \approx -\frac{\ln(1 - p^{\frac{1}{2-(C_1-C_2 p^2)\rho}})}{\mu - \lambda},$$

где $C_1 \approx 0,390327$, $C_2 \approx 0,237842$. При этом погрешность (максимум модуля относительной ошибки) составляет всего 0,62%, что меньше прежнего в 4,6 раза.

4. Заключение

Данная работа продолжает авторский цикл, посвященный исследованию характеристик fork-join-систем с пуассоновским входным потоком и экспоненциальными временами обслуживания. Несмотря на простоту систем и давность их исследования (с 1980-х годов) в этой области все еще много неясного. Точных результатов здесь мало, многие оценки нуждаются в улучшении. Есть вопросы, которыми мало кто или вообще никто не занимался. Исследования в основном сосредоточены на среднем времени отклика, в то время как интерес представляют также дисперсия, квантили и др.

Ключевой проблемой является наличие зависимости между временами пребывания подзаявок (частей одной заявки), обусловленной общностью входного потока в подсистемы. Эта зависимость хотя и не очень сильная, но оказывает существенное влияние на характеристики, при этом она далека от описания классическими моделями (например, многомерным нормальным распределением, линейной регрессией и т.п.). Поэтому авторы в последних работах сосредоточились на изучении данной зависимости. Случай двух подсистем рассматривался исходя из того, что при любом числе подсистем для любой их пары времена пребывания подзаявок будут иметь такое же совместное распределение. В настоящей работе изучались приближения совместного распределения времен пребывания подзаявок с помощью теории копул. Получено хорошее соответствие с данными для степенных диагональных сечений. На основе оценок диагональных сечений выведены оценки квантилей времени отклика в широком диапазоне уровней и загрузок.

Развитый при этом подход, основанный на теории копул, можно попытаться обобщить на системы с большим числом подсистем или случай более сложных подсистем (например, с тяжелыми хвостами распределений времен обслуживания).

Литература

1. ARMONY M., ISRAELIT S., MANDELBAUM A., MARMOR Y.N. et al. *Patient flow in hospitals: a data-based queueing-science perspective* // Stochastic Systems. – 2015. – Vol. 5, No. 1. – P. 146–194.
2. BACCELLI F., MAKOWSKI A.M. *Queueing models for systems with synchronization constraints* // Proc. of the IEEE. – 1989. – Vol. 77, No. 1. – P. 138–161.
3. BARON O., KRASS D., SHERZER E., SENDEROVICH A. *Can machines solve general queueing problems?* // Winter Simulation Conference (WSC) – 2022. – P. 2830–2841.
4. CHOCRON E., COHEN I., FEIGIN P. *Delay prediction for managing multiclass service systems: An investigation of queueing theory and machine learning approaches* // IEEE Trans. on Engineering Management. – 2022. – P. 1–11.
5. DIELEMAN A., BERKHOUT A., HEIDERGOTT B. *A neural network approach to performance analysis of tandem lines: The value of analytical knowledge* // Computers & Operations Research. – 2023. – Vol. 152. – P. 106–124.
6. ENGANTI P., ROSENKRANTZ T., SUN L., WANG Z. et al. *ForkMV: Mean-and-Variance Estimation of Fork-Join Queueing Networks for Datacenter Applications* // IEEE Int. Conf. on Networking, Architecture and Storage (NAS). – 2022. – P. 1–8.
7. GALLIEN J., WEIN L.M. *A simple and effective component procurement policy for stochastic assembly Systems* // Queueing Systems. – 2001. – Vol. 38. – P. 221–248.
8. GORBUNOVA A.V., LEBEDEV A.V. *Bivariate distributions of maximum remaining service times in fork-join infinite-server queues* // Probl. Inf. Transm. – 2020. – Vol. 56, No. 1. – P. 73–90. – DOI: <https://doi.org/10.1134/S003294602001007X>.

9. GORBUNOVA A.V., LEBEDEV A.V. *Nonlinear approximation of characteristics of a fork-join queueing system with Pareto service as a model of parallel structure of data processing* // Mathematics and Computers in Simulation. – 2023. – Vol. 214. – P. 409–428. – DOI: <https://doi.org/10.1016/j.matcom.2023.07.029>.
10. GORBUNOVA A.V., LEBEDEV A.V. *On estimating the characteristics of a fork-join queueing system with Poisson input and exponential service times* // Advances in Systems Science and Applications. – 2023. – Vol. 23, No. 2. – P. 99–114. – DOI: <https://doi.org/10.25728/assa.2023.23.2.1351>.
11. GORBUNOVA A.V., VISHNEVSKY V.M. *Estimating the response time of a cloud computing system with the help of neural networks* // Advances in Systems Science and Applications. – 2020. – Vol. 20, No. 3. – P. 105–112.
12. JIANG L., GIACHETTI R.E. *A queueing network model to analyze the impact of parallelization of care on patient cycle time* // Health Care Management Science. – 2008. – Vol. 11. – P. 248–261.
13. LEBEDEV A.V. *Upper bound for the expected minimum of dependent random variables with known Kendall's tau* // Theory of Probability and Its Applications. – 2019. – Vol. 64, No. 3. – P. 465–473.
14. LEBEDEV A.V. *On the Interrelation between Dependence Coefficients of Bivariate Extreme Value Copulas* // Markov Proc. Relat. Fields. – 2019. – Vol. 25, No. 4. – P. 639–648.
15. NELSEN R.B. *An introduction to copulas*. – Springer Science & Business Media, 2007. – 272 p.
16. NELSON R., TANTAWI A.N. *Approximate analysis of fork/join synchronization in parallel queues* // IEEE Trans. Comput. – 1988. – Vol. 37, No. 6. – P. 739–743.
17. NGUYEN M., ALESAWI S., LI S., CHE S. et al. *A black-box fork-join latency prediction model for data-intensive applications* // IEEE Trans. on Parallel and Distributed Systems. – 2020. – Vol. 31, No. 9. – P. 1983–2000.

18. QIU ZH., PEREZ J.F., HARRISON P.G. *Beyond the mean in fork-join queues: Efficient approximation for response-time tails* // Performance Evaluation. – 2015. – Vol. 91. – P. 99–116.
19. VISHNEVSKY V.M., GORBUNOVA A.V. *Application of machine learning methods to solving problems of queuing theory* // Communications in Computer and Information Science. – 2022. – Vol. 1605. – P. 304–316. – DOI: https://doi.org/10.1007/978-3-031-09331-9_24.
20. SCHOL D., VLASIOU M., ZWART B. *Large fork-join queues with nearly deterministic arrival and service times* // Mathematics of Operations Research. – 2022. – Vol. 47, No. 2. – P. 1335–1364. – DOI: <https://doi.org/10.1287/moor.2021.1171>.

ON A NEW APPROACH TO ESTIMATING RESPONSE TIME QUANTILES OF A FORK-JOIN QUEUEING SYSTEM

Anastasia Gorbunova, V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, Cand.Sc., senior researcher (avgorbunova@list.ru).

Alexey Lebedev, Lomonosov Moscow State University, Moscow, Doctor of Science, professor (avlebed@yandex.ru).

Abstract: The article proposes a new approach to estimating the quantiles of the response time distribution of the fork-join queueing system. We consider a classic version of this system with a Poisson input flow and exponential service times on homogeneous servers. Upon receipt of tasks into the system, they are instantly divided into a fixed number of subtasks and sent for service to the appropriate subsystem with an unlimited capacity storage device and one server. The task is considered served after all its components have been serviced. This system allows you to simulate many real processes, which, in order to increase efficiency, are characterized by dividing large tasks into smaller components, for example, parallel or distributed computing systems. The difficulty of analyzing systems lies in the presence of a dependence between the sojourn times of subtasks, which significantly complicates the analysis of all performance characteristics of such systems. The main contribution of the article is the approach to determining the quantiles of the response time distribution, the assessment of which is no less valuable than the assessment of the mean response time. At the same time, a much larger number of works in this area are devoted to calculating the mean, which is explained, among other things, by the complexity of carrying out such an analysis even for a given characteristic, and estimating quantiles seems to be an even more laborious task.

Keywords: system with parallel service of tasks, fork-join queueing system, response time, distribution quantiles, simulation modeling.

УДК 519.2

ББК 22.17

DOI: 10.25728/ubs.2024.108.1

*Статья представлена к публикации
членом редакционной коллегии Я.И. Квинто.*

Поступила в редакцию 27.02.2024.

Дата опубликования 31.03.2024.

АСИМПТОТИЧЕСКИЙ АНАЛИЗ СИСТЕМЫ $M^{[N]}/GI/1$ С УЧЕТОМ ОСТАТОЧНОГО ВРЕМЕНИ ОБСЛУЖИВАНИЯ

Назаров А. А.¹

(Томский государственный университет, Томск)

Рожкова С. В.², Титаренко Е. Ю.³

(Томский государственный университет, Томск)

Томский политехнический университет, Томск)

Рассматривается задача исследования одноканальной системы массового обслуживания с повторными вызовами, мгновенными и отложенными обратными связями. Такие системы моделируют ситуации повторной передачи данных в компьютерных сетях в случае занятости сервера или повреждения данных. Входящий поток является неординарным пуассоновским. Время обслуживания заявок – неотрицательная случайная величина с произвольной функцией распределения вероятностей и конечными моментами первого и второго порядка. Когда сервер занят, поступающие заявки отправляются на орбиту, где осуществляют случайную задержку и повторно принимают попытку обслужиться. Исследуется число заявок на орбите. При составлении уравнений Колмогорова для системы используется дополнительная переменная – остаточное время обслуживания, – которая позволяет получить многомерный марковский случайный процесс. Полученная система уравнений решается методом асимптотического анализа в условиях большой задержки заявок на орбите. В работе найдено стационарное распределение вероятностей числа заявок на орбите. Проведено сравнение полученного асимптотического распределения с распределением, найденным для случая экспоненциально распределенного времени обслуживания. Рассмотрен численный пример для системы, в которой длительность обслуживания имеет гамма-распределение с различными параметрами.

Ключевые слова: RQ-система, обратные связи, рекуррентное обслуживание, остаточное время.

1. Введение

Системы с повторными вызовами, или RQ-системы, широко используются для моделирования многих задач, возникающих в компьютерных сетях и в повседневной жизни. Такие системы

¹ Анатолий Андреевич Назаров, д.т.н., профессор (nazarov.tsu@gmail.com).

² Светлана Владимировна Рожкова, д.ф.-м.н., профессор (rozhkova@tpu.ru).

³ Екатерина Юрьевна Титаренко, старший преподаватель (teu@tpu.ru).

предполагают повторные попытки получить обслуживание для заявок, поступивших в систему, когда обслуживающий прибор занят [3, 9], при этом заявки в течение некоторого случайного времени ожидают повторного обслуживания на орбите. Если в случае потери или повреждении блока данных при передаче в компьютерных сетях происходит повторная передача, то могут применяться модели с обратными связями [12, 19, 20], в которых заявки, уже получившие обслуживание, повторно поступают на обслуживание. В таких моделях рассматриваются мгновенные [13, 14, 17] и отсроченные обратные связи [1, 11]. В моделях, учитывающих и повторные вызовы, и обратную связь, орбита формируется не только вновь поступившими, но и обслуженными заявками.

В системах сотовой связи иногда возникают ситуации, когда многие пользователи практически одновременно совершают звонки. Для трафика в интернете также характерно групповое поступление сообщений, когда пользователь открывает несколько интернет-сессий. В этих случаях образуется неординарный поток заявок, при котором заявки поступают в систему пачками случайного объема [8, 10]. Групповое поступление заявок существенно усложняет исследование системы [2, 5], при этом могут рассматриваться как фиксированные, так и произвольные размеры пакетов, а моменты поступления пакетов определяются различными законами распределения. Что касается длительности обслуживания заявок, то во многих исследованиях рассматриваются модели с экспоненциальным распределением, которые являются очень грубым приближением. Более точными являются модели с произвольным обслуживанием $M/G/1$. Такие модели исследуются либо с помощью встроенной цепи Маркова, либо с помощью метода дополнительных переменных [16].

В [15] рассмотрена RQ-система $M^{[n]}/M/1$ с обратными связями, исследован случайный процесс числа заявок на орбите, найдено распределение вероятностей. Задача решена методом асимптотического анализа при условии растущего среднего времени ожидания на орбите. Данная работа является обобщением

на случай произвольного закона распределения длительности обслуживания. В таком случае при исследовании систем вводятся дополнительные переменные, истекшее или остаточное время обслуживания [6, 18]. В исследовании [7] истекшее время позволило получить аналитическое решение системы уравнений Колмогорова.

В данной работе исследуется RQ-система $M^{[n]}/GI/1$ с обратной связью и вводится остаточное время обслуживания в качестве дополнительной переменной. Поскольку полученная в этом случае система уравнений Колмогорова не может быть решена аналитически, то для ее решения применен метод асимптотического анализа.

2. Математическая модель

Рассмотрим RQ-систему $M^{[n]}/GI/1$ с обратной связью и рекуррентным обслуживанием (см. рис. 1).

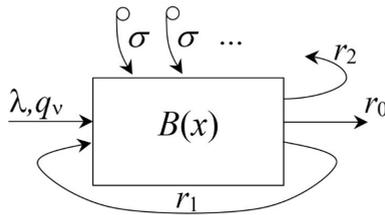


Рис. 1. RQ-система $M^{[n]}/GI/1$ с обратной связью

Входящий поток является неординарным пуассоновским. Полагаем, что моменты прихода заявок образуют пуассоновский поток с параметром λ . Заявки приходят группами. Размер группы является дискретной случайной величиной с заданным рядом распределения вероятностей q_ν появления ν заявок в группе, $\nu \geq 1$, $q_0 = 0$, $\sum_{\nu=1}^{\infty} q_\nu = 1$. Если прибор свободен, одна заявка из группы поступает на обслуживание, а остальные – на орбиту, иначе все заявки поступают на орбиту.

Через $B(x)$ обозначим функцию распределения времени обслуживания. По окончании обслуживания с вероятностью r_0 заявка покидает систему, с вероятностью r_1 мгновенно поступает на повторное обслуживание, с вероятностью r_2 переходит на орбиту, $r_0 + r_1 + r_2 = 1$.

Длительность пребывания на орбите распределена по экспоненциальному закону с параметром σ . По истечении времени ожидания заявки повторяют попытку занять прибор. Если прибор занят, то заявки остаются на орбите.

Обозначим $i(t)$ – число заявок на орбите в момент времени t ; $n(t)$ – состояние прибора: $n(t) = 0$, если прибор свободен, $n(t) = 1$, если занят. Задача: найти стационарное распределение вероятностей числа заявок на орбите.

Добавим переменную $z(t)$ – остаточное время обслуживания [4], т.е. длину интервала времени от момента t до момента окончания обслуживания заявки. Если прибор свободен, то процесс $z(t)$ не определяется. Тогда случайный процесс с переменным числом компонент $\{i(t), n(t), -/z(t)\}$ является марковским. Он состоит из двух компонент, если прибор свободен, и из трех компонент, если прибор занят.

Рассмотрим вероятности

$$P_0(i, t) = P\{i(t) = i, n(t) = 0\},$$

$$P_1(i, z, t) = P\{i(t) = i, n(t) = 1, z(t) < z\}, i = 0, 1, 2 \dots$$

С использованием формулы полной вероятности запишем допредельные равенства для распределения вероятностей. Получим следующую систему:

$$\begin{aligned} P_0(i, t + \Delta t) &= (1 - \lambda \Delta t) (1 - i \sigma \Delta t) P_0(i, t) + r_0 P_1(i, \Delta t, t) + \\ &\quad + r_2 P_1(i - 1, \Delta t, t) + o(\Delta t); \\ P_1(i, z - \Delta t, t + \Delta t) &= (1 - \lambda \Delta t) [P_1(i, z, t) - P_1(i, \Delta t, t)] + \\ &\quad + (i + 1) \sigma \Delta t B(z) P_0(i + 1, t) + B(z) r_1 P_1(i, \Delta t, t) + \\ &\quad + \sum_{\nu=1}^{i+1} \lambda q_{\nu} \Delta t B(z) P_0(i - \nu + 1, t) + \sum_{\nu=1}^i \lambda q_{\nu} \Delta t P_1(i - \nu, z, t) + o(\Delta t). \end{aligned}$$

Переходим к пределу при $\Delta t \rightarrow 0$, получим систему дифференциальных уравнений Колмогорова:

$$\begin{aligned} \frac{\partial P_0(i, t)}{\partial t} &= -(\lambda + i\sigma) P_0(i, t) + r_0 \frac{\partial P_1(i, 0, t)}{\partial z} + r_2 \frac{\partial P_1(i-1, 0, t)}{\partial z}; \\ \frac{\partial P_1(i, z, t)}{\partial t} - \frac{\partial P_1(i, z, t)}{\partial z} &= -\lambda P_1(i, z, t) - \frac{\partial P_1(i, 0, t)}{\partial z} + \\ &+ r_1 \frac{\partial P_1(i, 0, t)}{\partial z} B(z) + (i+1)\sigma B(z) P_0(i+1, t) + \\ &+ \sum_{\nu=1}^{i+1} \lambda q_\nu B(z) P_0(i-\nu+1, t) + \sum_{\nu=1}^i \lambda q_\nu P_1(i-\nu, z, t). \end{aligned}$$

Здесь обозначено $\left. \frac{\partial P_1(i, 0, t)}{\partial z} = \frac{\partial P_1(i, z, t)}{\partial z} \right|_{z=0}$.

Для стационарного распределения вероятностей $P_0(i) \equiv P_0(i, t)$, $P_1(i, z) \equiv P_1(i, z, t)$ перепишем систему в виде

$$\begin{aligned} (1) \quad & -(\lambda + i\sigma) P_0(i) + r_0 \frac{\partial P_1(i, 0)}{\partial z} + r_2 \frac{\partial P_1(i-1, 0)}{\partial z} = 0; \\ & \frac{\partial P_1(i, z)}{\partial z} - \lambda P_1(i, z) - \frac{\partial P_1(i, 0)}{\partial z} + r_1 \frac{\partial P_1(i, 0)}{\partial z} B(z) + \\ & + (i+1)\sigma B(z) P_0(i+1) + \\ & + \sum_{\nu=1}^{i+1} \lambda q_\nu B(z) P_0(i-\nu+1) + \sum_{\nu=1}^i \lambda q_\nu P_1(i-\nu, z) = 0. \end{aligned}$$

Сложим уравнения системы (1). Обозначим вероятности $P_1(i) = \lim_{z \rightarrow \infty} P_1(i, z)$ и при $z \rightarrow \infty$ получим

$$\begin{aligned} (2) \quad & -(\lambda + i\sigma) P_0(i) + (i+1)\sigma P_0(i+1) - \lambda P_1(i) - r_2 \frac{\partial P_1(i, 0)}{\partial z} + \\ & + r_2 \frac{\partial P_1(i-1, 0)}{\partial z} + \sum_{\nu=1}^{i+1} \lambda q_\nu P_0(i-\nu+1) + \sum_{\nu=1}^i \lambda q_\nu P_1(i-\nu) = 0. \end{aligned}$$

Введем частичные характеристические функции $H_n(u) = \sum_{i=0}^{\infty} e^{ju_i} P_n(i)$, $n = 0, 1$, $\tilde{H}_1(u, z) = \sum_{i=0}^{\infty} e^{ju_i} P_1(i, z)$ и характеристическую функцию объема группы $h(u) = \sum_{\nu=1}^{\infty} e^{j\nu u} q_\nu$, где

$j = \sqrt{-1}$. Преобразуем систему (1) к виду

$$(3) \quad -\lambda H_0(u) + (r_0 + r_2 e^{ju}) \frac{\partial \tilde{H}_1(u, 0)}{\partial z} + j\sigma \frac{\partial H_0(u)}{\partial u} = 0;$$

$$\lambda B(z) e^{-ju} h(u) H_0(u) + \lambda (h(u) - 1) \tilde{H}_1(u, z) +$$

$$+ \frac{\partial \tilde{H}_1(u, z)}{\partial z} + (r_1 B(z) - 1) \frac{\partial \tilde{H}_1(u, 0)}{\partial z} - j\sigma B(z) \frac{\partial H_0(u)}{\partial u} e^{-ju} = 0.$$

Тогда характеристическая функция числа заявок на орбите определяется как $H(u) = H_0(u) + H_1(u)$, где $H_1(u) = \lim_{z \rightarrow \infty} \tilde{H}_1(u, z)$.

Аналогично уравнение (2) преобразуем к виду

$$(4) \quad \lambda (e^{-ju} h(u) - 1) H_0(u) + \lambda (h(u) - 1) H_1(u) +$$

$$+ j\sigma (1 - e^{-ju}) H'_0(u) + r_2 (e^{ju} - 1) \frac{\partial \tilde{H}_1(u, 0)}{\partial z} = 0.$$

3. Асимптотический анализ системы при условии растущего среднего времени ожидания на орбите

Для решения уравнений (3), (4) применим метод асимптотического анализа в предельном условии большой задержки заявок на орбите, т.е. при $\sigma \rightarrow 0$.

Теорема 1 (Асимптотика первого порядка). Пусть $i(t)$ – число заявок на орбите в системе $M^{[n]}/GI/1$ с обратной связью, рекуррентным обслуживанием, тогда выполняется равенство $\lim_{\sigma \rightarrow 0} M \{e^{jwi(t)\sigma}\} = e^{jw\kappa_1}$, где

$$(5) \quad \kappa_1 = \lambda \bar{v} \frac{r_0 + r_2}{r_0 - \lambda \bar{v} b} - \lambda,$$

$\bar{v} = \sum_{\nu=1}^{\infty} \nu q_{\nu}$ – средний размер группы, $b = \int_0^{\infty} (1 - B(x)) dx$ – среднее время обслуживания.

Доказательство. Обозначим $\sigma = \varepsilon$. В системе (3) и уравнении (4) сделаем следующие замены $u = \varepsilon w$, $H_n(u) = F_n(w, \varepsilon)$,

$n = 0, 1$, $\tilde{H}_1(u, z) = \tilde{F}_1(w, \varepsilon, z)$, тогда система (3) имеет вид

$$(6) \quad -\lambda F_0(w, \varepsilon) + (r_0 + r_2 e^{j\varepsilon w}) \frac{\partial \tilde{F}_1(w, \varepsilon, 0)}{\partial z} + j \frac{\partial F_0(w, \varepsilon)}{\partial w} = 0;$$

$$\lambda B(z) e^{-j\varepsilon w} h(w, \varepsilon) F_0(w, \varepsilon) + \lambda (h(w, \varepsilon) - 1) \tilde{F}_1(w, \varepsilon, z) +$$

$$+ \frac{\partial \tilde{F}_1(w, \varepsilon, z)}{\partial z} + (r_1 B(z) - 1) \frac{\partial \tilde{F}_1(w, \varepsilon, 0)}{\partial z} -$$

$$- j B(z) e^{-j\varepsilon w} \frac{\partial F_0(w, \varepsilon)}{\partial w} = 0;$$

а уравнение (4) имеет вид

$$(7) \quad \lambda (e^{-j\varepsilon w} h(w, \varepsilon) - 1) F_0(w, \varepsilon) + \lambda (h(w, \varepsilon) - 1) F_1(w, \varepsilon) +$$

$$+ r_2 (e^{j\varepsilon w} - 1) \frac{\partial \tilde{F}_1(w, \varepsilon, 0)}{\partial z} + j (1 - e^{-j\varepsilon w}) \frac{\partial F_0(w, \varepsilon)}{\partial w} = 0.$$

Пусть $\varepsilon \rightarrow 0$. Обозначим $F_n(w) = \lim_{\varepsilon \rightarrow 0} F_n(w, \varepsilon)$, $n = 0, 1$, $\tilde{F}_1(w, z) = \lim_{\varepsilon \rightarrow 0} \tilde{F}_1(w, \varepsilon, z)$ и учитывая, что $\lim_{\varepsilon \rightarrow 0} h(w, \varepsilon) = 1$,

$h'(0) = j \sum_{\nu=1}^{\infty} \nu q_\nu = j\bar{\nu}$, преобразуем систему (6) в систему

$$(8) \quad -\lambda F_0(w) + (r_0 + r_2) \frac{\partial \tilde{F}_1(w, 0)}{\partial z} + j \frac{\partial F_0(w)}{\partial w} = 0;$$

$$\lambda B(z) F_0(w) + \frac{\partial \tilde{F}_1(w, z)}{\partial z} + (r_1 B(z) - 1) \frac{\partial \tilde{F}_1(w, 0)}{\partial z} -$$

$$- j B(z) \frac{\partial F_0(w)}{\partial w} = 0.$$

Аналогично преобразуем уравнение (7), затем разделим на ε и при $\varepsilon \rightarrow 0$ получим

$$(9) \quad \lambda (\bar{\nu} - 1) F_0(w) + \lambda \bar{\nu} F_1(w) + r_2 \frac{\partial \tilde{F}_1(w, 0)}{\partial z} + j \frac{\partial F_0(w)}{\partial w} = 0.$$

Решение системы (8), (9) найдем в виде

$$(10) \quad F_n(w) = R_n \Phi(w) + O(\varepsilon), n = 0, 1,$$

$$\tilde{F}_1(w, z) = R_1(z) \Phi(w) + O(\varepsilon),$$

тогда

$$(11) \quad -\lambda R_0 + (r_0 + r_2) R'_1(0) + jR_0 \frac{\Phi'(w)}{\Phi(w)} = 0;$$

$$\lambda B(z)R_0 + R'_1(z) + (r_1 B(z) - 1) R'_1(0) - jB(z)R_0 \frac{\Phi'(w)}{\Phi(w)} = 0;$$

$$\lambda(\bar{\nu} - 1) R_0 + \lambda\bar{\nu}R_1 + r_2 R'_1(0) + jR_0 \frac{\Phi'(w)}{\Phi(w)} = 0.$$

Из системы (11) видно, что $\Phi'(w)/\Phi(w)$ не зависит от w , поэтому запишем $\Phi'(w)/\Phi(w) = j\kappa_1$, т.е. $\Phi(w) = \exp\{jw\kappa_1\}$. Далее найдем κ_1 . Для этого подставим выражение для $\Phi(w)$ в систему уравнений (11), получим

$$(12) \quad -(\lambda + \kappa_1)R_0 + (r_0 + r_2)R'_1(0) = 0;$$

$$(13) \quad (\lambda + \kappa_1)B(z)R_0 + R'_1(z) + (r_1 B(z) - 1)R'_1(0) = 0;$$

$$(14) \quad (\lambda\bar{\nu} - \lambda - \kappa_1)R_0 + \lambda\bar{\nu}R_1 + r_2 R'_1(0) = 0.$$

Заметим, что уравнения (12) и (13) совпадают при $z \rightarrow \infty$.

Из уравнения (12) выразим

$$(15) \quad R'_1(0) = R_0(\lambda + \kappa_1)/(r_0 + r_2).$$

Из уравнения (13) выразим $R'_1(z)$ и с учетом (15) получим

$$R_1(z) = \frac{\lambda + \kappa_1}{r_0 + r_2} R_0 \int_0^z (1 - B(x)) dx.$$

В пределе при $z \rightarrow \infty$ найдем

$$(16) \quad R_1 = R_0 b(\lambda + \kappa_1)/(r_0 + r_2),$$

где $b = \int_0^\infty (1 - B(x)) dx$ – среднее время обслуживания заявок. Учитывая, что $R_0 + R_1 = 1$, получим стационарные вероятности того, что прибор находится в состоянии 0 или 1:

$$R_0 = \frac{r_0 + r_2}{r_0 + r_2 + (\lambda + \kappa_1)b}; \quad R_1 = \frac{(\lambda + \kappa_1)b}{r_0 + r_2 + (\lambda + \kappa_1)b}.$$

Из уравнения (14), с учетом выражений (15) и (16), получим выражения для κ_1 (5).

Согласно (10) получаем асимптотические функции

$$F_0(w) = \frac{r_0 + r_2}{r_0 + r_2 + (\lambda + \kappa_1)b} \exp\{j\kappa_1 w\},$$

$$F_1(w) = \frac{(\lambda + \kappa_1)b}{r_0 + r_2 + (\lambda + \kappa_1)b} \exp\{j\kappa_1 w\},$$

тогда аппроксимация первого порядка характеристической функции числа заявок на орбите имеет вид $H(u) \approx F_0(w) + F_1(w) = \exp\{j\kappa_1 u/\sigma\}$ и определяет асимптотическое среднее числа заявок на орбите.

Теорема 2 (Асимптотика второго порядка). Пусть $i(t)$ – число заявок на орбите в системе $M^{[n]}/GI/1$ с обратной связью, рекуррентным обслуживанием, тогда имеет место равенство $\lim_{\sigma \rightarrow 0} M\{\exp\{jw\sqrt{\sigma}(i(t) - \kappa_1/\sigma)\}\} = \exp\{\kappa_2(jw)^2/2\}$, где

$$(17) \quad \kappa_2 = \left(\lambda \nu_2 r_0 (r_0 + r_2) + \lambda \bar{\nu} r_0 (r_2 - r_0) + 2(\lambda \bar{\nu})^2 r_0 b + \right. \\ \left. + (\lambda \bar{\nu})^3 (r_0 + r_2) (b_2 - 2b^2) \right) / \left(2(r_0 - \lambda \bar{\nu} b)^2 \right), \\ \nu_2 = \sum_{\nu=1}^{\infty} \nu^2 q_{\nu}, b_2 = \int_0^{\infty} x^2 dB(x).$$

Доказательство. В системе (3), (4) выполним замену $H_n(u) = H_n^{(2)}(u)e^{ju\kappa_1/\sigma}$, $n = 0, 1$, $\tilde{H}_1(u, z) = \tilde{H}_1^{(2)}(u, z)e^{ju\kappa_1/\sigma}$. Здесь $H_n^{(2)}(u)$ – частичные характеристические функции случайной величины $i(t) - \kappa_1/\sigma$. Тогда получим систему уравнений для $H_n^{(2)}(u)$ в виде

$$-\lambda H_0^{(2)}(u) + (r_0 + r_2 e^{ju}) \frac{\partial \tilde{H}_1^{(2)}(u, 0)}{\partial z} + j\sigma \frac{\partial H_0^{(2)}(u)}{\partial u} - \kappa_1 H_0^{(2)}(u) = 0;$$

$$\lambda B(z) e^{-ju} h(u) H_0^{(2)}(u) + \lambda (h(u) - 1) \tilde{H}_1^{(2)}(u, z) + \frac{\partial \tilde{H}_1^{(2)}(u, z)}{\partial z} + \\ + (r_1 B(z) - 1) \frac{\partial \tilde{H}_1^{(2)}(u, 0)}{\partial z} - j e^{-ju} B(z) \sigma \frac{\partial H_0^{(2)}(u)}{\partial u} + \\ + \kappa_1 e^{-ju} B(z) H_0^{(2)}(u) = 0;$$

$$\begin{aligned} & \lambda [e^{-ju}h(u) - 1] H_0^{(2)}(u) + \lambda (h(u) - 1) H_1^{(2)}(u) + \\ & + r_2 (e^{ju} - 1) \frac{\partial \tilde{H}_1^{(2)}(u, 0)}{\partial z} + j\sigma (1 - e^{-ju}) \frac{\partial H_0^{(2)}(u)}{\partial u} - \\ & - \kappa_1 (1 - e^{-ju}) H_0^{(2)}(u) = 0. \end{aligned}$$

Обозначим $\sigma = \varepsilon^2$ и введем замену $u = \varepsilon w$, $H_n^{(2)}(u) = F_n^{(2)}(w, \varepsilon)$, $n = 0, 1$, $\tilde{H}_1^{(2)}(u, z) = \tilde{F}_1^{(2)}(w, z, \varepsilon)$.

Получим систему

$$\begin{aligned} (18) \quad & -(\lambda + \kappa_1) F_0^{(2)}(w, \varepsilon) + \\ & + (r_0 + r_2 e^{j\varepsilon w}) \frac{\partial \tilde{F}_1^{(2)}(w, 0, \varepsilon)}{\partial z} + j\varepsilon \frac{\partial F_0^{(2)}(w, \varepsilon)}{\partial w} = 0; \\ & B(z) e^{-j\varepsilon w} (\lambda h(w, \varepsilon) + \kappa_1) F_0^{(2)}(w, \varepsilon) + \\ & + \lambda (h(w, \varepsilon) - 1) \tilde{F}_1^{(2)}(w, z, \varepsilon) + \frac{\partial \tilde{F}_1^{(2)}(w, z, \varepsilon)}{\partial z} + \\ & + (r_1 B(z) - 1) \frac{\partial \tilde{F}_1^{(2)}(w, 0, \varepsilon)}{\partial z} - j\varepsilon e^{-j\varepsilon w} B(z) \frac{\partial F_0^{(2)}(w, \varepsilon)}{\partial w} = 0. \\ & (\lambda e^{-j\varepsilon w} h(w, \varepsilon) - \lambda - \kappa_1 + \kappa_1 e^{-j\varepsilon w}) F_0^{(2)}(w, \varepsilon) + \\ & + \lambda (h(w, \varepsilon) - 1) F_1^{(2)}(w, \varepsilon) + \\ & + r_2 (e^{j\varepsilon w} - 1) \frac{\partial \tilde{F}_1^{(2)}(w, 0, \varepsilon)}{\partial z} + j\varepsilon (1 - e^{-j\varepsilon w}) \frac{\partial F_0^{(2)}(w, \varepsilon)}{\partial w} = 0. \end{aligned}$$

Решение системы запишем в виде

$$\begin{aligned} (19) \quad & F_n^{(2)}(w, \varepsilon) = \Phi_2(w)(R_n + j\varepsilon w f_n) + O(\varepsilon^2), n = 0, 1 \\ & \tilde{F}_1^{(2)}(w, z, \varepsilon) = \Phi_2(w)(R_1(z) + j\varepsilon w f_1(z)) + O(\varepsilon^2) \end{aligned}$$

Подставим разложения (19) в систему (18)

$$\begin{aligned} (20) \quad & -(\lambda + \kappa_1) (R_0 + j\varepsilon w f_0) + \\ & + (r_0 + r_2 e^{j\varepsilon w}) (R_1'(0) + j\varepsilon w f_1'(0)) + j\varepsilon R_0 \frac{\Phi_2'(w)}{\Phi_2(w)} = O(\varepsilon^2); \end{aligned}$$

$$\begin{aligned}
 & B(z)e^{-j\varepsilon w} (\lambda h(w, \varepsilon) + \kappa_1) (R_0 + j\varepsilon w f_0) + \lambda (h(w, \varepsilon) - 1) \times \\
 & \times (R_1(z) + j\varepsilon w f_1(z)) + R_1'(z) + j\varepsilon w f_1'(z) + (r_1 B(z) - 1) \times \\
 & \times (R_1'(0) + j\varepsilon w f_1'(0)) - j\varepsilon B(z)e^{-j\varepsilon w} \frac{\Phi_2'(w)}{\Phi_2(w)} R_0 = O(\varepsilon^2); \\
 & ((\lambda h(w, \varepsilon) + \kappa_1)e^{-j\varepsilon w} - \lambda - \kappa_1) (R_0 + j\varepsilon w f_0) + \lambda (h(w, \varepsilon) - 1) \times \\
 & \times (R_1 + j\varepsilon w f_1) + r_2 (e^{j\varepsilon w} - 1) (R_1'(0) + j\varepsilon w f_1'(0)) + \\
 & + j\varepsilon (1 - e^{-j\varepsilon w}) \frac{\Phi_2'(w)}{\Phi_2(w)} (R_0 + j\varepsilon w f_0) = O(\varepsilon^2).
 \end{aligned}$$

В первых двух уравнениях системы (20) используем разложения $e^{\pm jw\varepsilon} = 1 \pm jw\varepsilon + O(\varepsilon^2)$ и $h(w, \varepsilon) = 1 + j\varepsilon w\bar{\nu} + O(\varepsilon^2)$, а в третьем уравнении – разложения $e^{\pm jw\varepsilon} = 1 \pm jw\varepsilon + (jw\varepsilon)^2/2 + O(\varepsilon^3)$ и $h(w, \varepsilon) = 1 + j\varepsilon w\bar{\nu} + (j\varepsilon w)^2\nu_2/2 + O(\varepsilon^3)$, где $\nu_2 = \sum_{\nu=1}^{\infty} \nu^2 q_\nu$, тогда с учетом (12)–(14) систему (20) можно переписать

$$(21) \quad -(\lambda + \kappa_1) f_0 + r_2 R_1'(0) + (r_0 + r_2) f_1'(0) + \frac{\Phi_2'(w)}{w\Phi_2(w)} R_0 = 0;$$

$$(22) \quad (\lambda\bar{\nu} - \lambda - \kappa_1) B(z)R_0 + (\lambda + \kappa_1) B(z)f_0 + \lambda\bar{\nu}R_1(z) + \\ + f_1'(z) + (r_1 B(z) - 1) f_1'(0) - \frac{\Phi_2'(w)}{w\Phi_2(w)} B(z)R_0 = 0;$$

$$(23) \quad \left(\frac{\lambda\nu_2}{2} - \lambda\bar{\nu} + \frac{1}{2}\lambda + \frac{1}{2}\kappa_1 \right) R_0 + (\lambda\bar{\nu} - \lambda - \kappa_1) f_0 + \lambda\bar{\nu}f_1 + \\ + r_2 f_1'(0) + \frac{\lambda\nu_2}{2} R_1 + \frac{1}{2}r_2 R_1'(0) + \frac{\Phi_2'(w)}{w\Phi_2(w)} R_0 = 0.$$

Из системы (21)–(23) видно, что при $z \rightarrow \infty$ уравнения (21) и (22) совпадают, а $\Phi_2'(w)/[w\Phi_2(w)]$ не зависит от w . Можно обозначить $\Phi_2'(w)/[w\Phi_2(w)] = -\kappa_2$ и записать $\Phi_2(w) = \exp\{\kappa_2(jw)^2/2\}$. Из уравнения (21)

$$(24) \quad f_1'(0) = \frac{(\lambda + \kappa_1) f_0 - r_2 R_1'(0) + \kappa_2 R_0}{r_0 + r_2},$$

тогда из уравнения (22) с учетом (14) получим

$$f_1(z) = \frac{(\lambda + \kappa_1) f_0 - r_2 R_1'(0) + \kappa_2 R_0}{r_0 + r_2} \int_0^z (1 - B(x)) dx + \\ + \lambda \bar{\nu} \int_0^z (R_1 B(x) - R_1(x)) dx.$$

В пределе при $z \rightarrow \infty$

$$(25) \quad f_1 = \frac{(\lambda + \kappa_1) f_0 - r_2 R_1'(0) + \kappa_2 R_0}{r_0 + r_2} b + \\ + \lambda \bar{\nu} \left(-R_1 b + R_0 \frac{\lambda + \kappa_1}{r_0 + r_2} \frac{b_2}{2} \right),$$

где b_2 – второй начальный момент длительности обслуживания поступающих заявок. Перепишем (25) с учетом (15):

$$(26) \quad f_1 = \frac{\lambda + \kappa_1}{r_0 + r_2} b f_0 + \frac{\lambda \bar{\nu}}{2} \frac{\lambda + \kappa_1}{r_0 + r_2} b_2 R_0 + \\ + \left(-r_2 \frac{\lambda + \kappa_1}{(r_0 + r_2)^2} + \frac{\kappa_2}{r_0 + r_2} - \lambda \bar{\nu} b \frac{\lambda + \kappa_1}{r_0 + r_2} \right) b R_0.$$

В уравнение (23) подставим (24), (26), (15), (16), и так как слагаемые, содержащие f_0 , взаимно уничтожаются, получим κ_2 в явном виде (17).

4. Численные результаты

Теоремы 1 и 2 показывают, что асимптотическая характеристическая функция числа заявок на орбите является характеристической функцией гауссовской случайной величины, тогда для распределения $P(i)$ построим аппроксимацию

$$(27) \quad P_{appr}(i) = \frac{L(i + 0,5) - L(i - 0,5)}{1 - L(-0,5)},$$

где $L(x)$ – функция нормального распределения с параметрами κ_1/σ и $\sqrt{\kappa_2/\sigma}$.

В случае, когда $B(x)$ – функция экспоненциального распределения, первый и второй начальные моменты имеют вид

$b = 1/\mu, b_2 = 2/\mu^2$, а выражения для параметров распределения, полученные в теоремах 1 и 2, имеют вид

$$\kappa_1 = \lambda \bar{\nu} \mu \frac{r_0 + r_2}{\mu r_0 - \lambda \bar{\nu}} - \lambda,$$

$$\kappa_2 = \frac{\lambda \mu r_0}{2(\mu r_0 - \lambda \bar{\nu})^2} (2\lambda \bar{\nu}^2 + \mu \nu_2 (r_0 + r_2) + \mu \bar{\nu} (r_2 - r_0))$$

и совпадают с выражениями для κ_1 и κ_2 , полученными в [15] для системы $M^{[n]}/M/1$. Для такой системы можно найти область применимости асимптотических результатов, сравнив распределение, полученное допредельным аналитическим методом, с асимптотическим распределением (27). Численные эксперименты показывают, что расстояние Колмогорова уменьшается при уменьшении σ и становится меньше 0,05 при $\sigma < 0,5$ для средней загрузки системы $\rho = \lambda \nu / \mu r_0$, при $\sigma < 0,1$ для высокой загрузки системы $\rho > 0,8$.

Рассмотрим систему с параметрами $\lambda = 1, r_0 = 0,5, r_1 = 0,3, r_2 = 0,2$. Число заявок в пачке имеет геометрическое распределение с параметром $p = 0,5$. Длительность обслуживания имеет гамма-распределение с параметрами α и β .

На рис. 2а представлено распределение вероятностей числа заявок на орбите, полученное по формуле (27) при параметрах $\alpha = 10, \beta = 2$ и различных значениях параметра σ , который определяет длительность задержки заявок на орбите. В этой системе вероятность того, что прибор занят, $R_1 = 0,8$.

На рис. 2б представлено распределение вероятностей при параметрах $\alpha = 10, \sigma = 0,1$ и различных значениях параметра β , который влияет на длительность обслуживания. В этой системе вероятность того, что прибор занят R_1 изменяется в зависимости от параметра β .

5. Заключение

В работе исследована RQ-система с неординарным пуассоновским входящим потоком, рекуррентным обслуживанием и об-

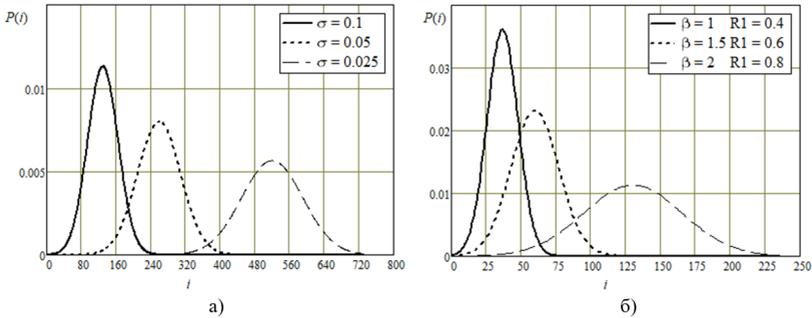


Рис. 2. Асимптотическое распределение вероятностей числа заявок на орбите

ратной связи. Система уравнений Колмогорова решена с помощью метода асимптотического анализа в предельном условии большой задержки заявок на орбите при введении дополнительной переменной – остаточного времени обслуживания. Показано, что асимптотическая характеристическая функция числа заявок на орбите является характеристической функцией гауссовской случайной величины, а также найдены параметры гауссовского распределения. В частном случае экспоненциально распределенного времени обслуживания полученное распределение совпадает с распределением, полученным [15]. В этом случае применение асимптотического анализа целесообразно при $\sigma < 0,1$.

Литература

1. АЛИЕВА С.Г. Численное исследование моделей систем массового обслуживания с отсроченными обратными связями // Вестн. Том. гос. ун-та. Управление, вычислительная техника и информатика. – 2020. – №51.
2. ИВАНОВ А.С., ЛЯХОВ А.И., ХОРОВ Е.М. Математическая модель передачи неординарного потока с помощью периодических резервирований и блочного квитирования в канале с коррелированными помехами // Автоматика и телемеханика. – 2017. – №11. – С. 48–63.

3. КОЖАНОВ Ю.Ф. *Теория телетрафика* : учебное пособие. – Санкт-Петербург: СПбГУТ им. М.А. Бонч-Бруевича, 2020. – 203 с.
4. МОИСЕЕВА Е.А. *Исследование RQ-системы M|GI|I в допредельной ситуации* // Математическое и программное обеспечение информационных, технических и экономических систем. Материалы II Всероссийской молодежной научной конф. Сер. "Труды ТГУ. Серия физико-математическая". – 2013. – С. 116–121.
5. МОНСИК В.Б., СКРЫННИКОВ А.А., ФЕДОТОВ А.Ю. *Система массового обслуживания с групповым обслуживанием неординарного потока требований* // Научный вестник МГТУ ГА. – 2010. – №157.
6. НАЗАРОВ А.А., КВАЧ А.С. *Сравнение методов остаточного и истекшего времени обслуживания для исследования замкнутой RQ-системы M/GI/1/N с конфликтами заявок и ненадежным прибором* // Информационные технологии и математическое моделирование (ИТММ-2017): Материалы XVI Международной конференции имени А.Ф. Терпугова, Казань, 29 сентября – 3 октября 2017 года. Часть 1. – Казань: Изд-во науч.-техн. лит-ры, 2017. – С. 142–149.
7. НАЗАРОВ А.А., РОЖКОВА С.В., ТИТАРЕНКО Е.Ю. *Исследование системы с обратной связью, рекуррентным обслуживанием и неординарным пуассоновским входящим потоком* // Информационные технологии и математическое моделирование (ИТММ-2020): материалы XIX Международной конференции имени А. Ф. Терпугова, 2–5 декабря 2020 г. – Томск: Изд-во НТЛ, 2021. – С. 223–227.
8. РЫЖИКОВ Ю.И. *Расчет систем обслуживания с групповым поступлением заявок* // Информационно-управляющие системы. – 2007. – №2.
9. ARTALEJO J.R., GOMEZ-CORRAL A. *Retrial queueing systems. A Computational Approach*. – Springer, 2008. – 309 p.

10. ARTALEJO J.R., ATENCIA I. *On the single server retrieval queue with batch arrivals* // Sankhya. – 2004. – Vol. 66. – P. 140–158.
11. KALYANARAMAN R. *A Feedback Retrieval Queueing System with Two Types of Batch Arrivals* // Int. Journal of Stochastic Analysis. – 2012. – Vol. 2012. – DOI: <https://doi.org/10.1155/2012/673642>.
12. KALYANARAMAN R. *A Retrieval Queueing System with Two Types of Batch Arrivals and with Feedback to Orbit* // Int. Journal of Science and Research (IJSR). – 2022. – Vol. 11, Iss. 5. – P. 1710–1717.
13. MELIKOV A., ALIYEVA S., NAIR SS., KUMAR BK. *Retrieval Queueing-Inventory Systems with Delayed Feedback and Instantaneous Damaging of Items* // Axioms. – 2022. – Vol. 11(5). – P. 241. – DOI: <https://doi.org/10.3390/axioms11050241>.
14. MELIKOV A., ALIYEVA S., SZTRIK J. *Retrieval Queues with Unreliable Servers and Delayed Feedback* // Mathematics. 2021. – Vol. 9(19). – P. 2415. – DOI: <https://doi.org/10.3390/math9192415>.
15. NAZAROV A.A., ROZHKOVA S.V., TITARENKO E.Y. *Asymptotic analysis of RQ-system with feedback and batch Poisson arrival under the condition of increasing average waiting time in orbit* // Distributed Computer and Communication Networks: Control, Computation, Communications. DCCN 2020. Communications in Computer and Information Science. – Vol 1337. – Springer, Cham. – DOI: https://doi.org/10.1007/978-3-030-66242-4_26.
16. PHUNG-DUC T. *Retrieval Queueing Models: A Survey on Theory and Applications* // arXiv:1906.09560. – 2017. – DOI: <https://doi.org/10.48550/arXiv.1906.09560>.

17. SARAVANAN V., VENUGOPAL P., GODHANDARAMAN P. *Performance Analysis of a Retrial Queueing System with Optional Service, Unreliable Server, Balking and Feedback* // Int. Journal of Mathematical, Engineering and Management Sciences. – 2023. – Vol. 8. – P. 769–786. – DOI: 10.33889/IJMEMS.2023.8.4.044.
18. KEERTHIGA S., KANDAIYAN I. *Two phase of service in M/G/1 queueing system with retrial customers* // The Journal of Analysis. – 2023. – DOI: 10.1007/s41478-023-00635-x.
19. TAKACS L. *A single-server queue with feedback* // Bell System Technical Journal. – 1963. – Vol. 42. – P. 505–519.
20. TAKACS L. *A queueing model with feedback* // Operations Research. – 1977. – Vol. 11. – P. 345–354.

ASYMPTOTIC ANALYSIS OF THE $M^{[N]}/GI/1$ SYSTEM WITH THE REMAINING SERVICE TIME

Anatoly Nazarov, Tomsk State University, Tomsk, Doctor of Science, professor (nazarov.tsu@gmail.com).

Svetlana Rozhkova, Tomsk Polytechnic University, Tomsk, Doctor of Science, professor (rozhkova@tpu.ru).

Ekaterina Titarenko, Tomsk Polytechnic University, Tomsk, lecturer (teu@tpu.ru).

Abstract: A single server queueing system with Poisson batch incoming stream, repeated calls, instant and delayed feedbacks is considered. It is assumed that service time is distributed according to an arbitrary law, and the service durations are independent of each other. When the server is busy, incoming customers are sent into orbit. The problem is to investigate a random process of the number of customers in orbit. When compiling the Kolmogorov equations for the system, an additional variable is used - the remaining service time. The resulting system of equations is solved by the method of asymptotic analysis under the condition of a large delay of customers in orbit. As a result, a stationary probability distribution for the number of customers in orbit was found. The resulting asymptotic distribution is compared with the distribution found in previous papers for the case of an exponentially distributed service time. A numerical example is considered for a system in which the service duration has a gamma distribution with different parameters.

Keywords: retrial queue system, feedback, arbitrary distributed service time, remaining time.

УДК 519.872

ББК 22.17

DOI: 10.25728/ubs.2024.108.2

*Статья представлена к публикации
членом редакционной коллегии А.С. Манделем.*

Поступила в редакцию 24.10.2023.

Дата опубликования 31.03.2024.

ПРИМЕНЕНИЕ ОТРИЦАТЕЛЬНОГО БИНОМИАЛЬНОГО РАСПРЕДЕЛЕНИЯ ДЛЯ АППРОКСИМАЦИИ СТАЦИОНАРНОГО РАСПРЕДЕЛЕНИЯ ЧИСЛА ЗАЯВОК В СМО С ВХОДЯЩИМ МАР-ПОТОКОМ, ИНТЕНСИВНОСТЬ КОТОРОГО ЗАВИСИТ ОТ СОСТОЯНИЯ СИСТЕМЫ

Полин Е. П.¹, Моисеева С. П.², Моисеев А. Н.³

*(Национальный исследовательский Томский
государственный университет, Томск)*

Рассматривается математическая модель бесконечнолинейной системы массового обслуживания с входящим МАР-поток с интенсивностью, зависящей от числа занятых приборов. Параметры входящего потока, а именно его условные интенсивности, меняются каждый раз, когда изменяется состояние системы, то есть появляется новая заявка либо одна из заявок завершает обслуживание. Дисциплина обслуживания определяется тем, что заявка занимает любой из свободных приборов в системе, в котором выполняется ее обслуживание в течение случайного времени, распределенного по экспоненциальному закону. Для данной модели получение стационарного распределения вероятностей числа заявок в системе аналитическим путем не представляется возможным, поэтому в данной работе предлагается эвристический подход, а именно – использование отрицательного биномиального распределения как аппроксимации для искомого распределения. Предлагается два подхода такой аппроксимации, для которых выполнен численный анализ точности на основе сравнения с результатами имитационного моделирования. Первый подход основан на вычислении параметров отрицательного биномиального распределения с помощью точных значений математического ожидания и дисперсии числа заявок в рассматриваемой системе, а второй – на том факте, что интенсивность поступающих заявок определяется управляющей потоком цепью Маркова. Было получено, что первый способ аппроксимации дает более точные результаты, однако при большой загрузке системы обе аппроксимации имеют большую погрешность.

Ключевые слова: гетерогенная система массового обслуживания, переменная интенсивность, отрицательное биномиальное распределение.

¹ Евгений Павлович Полин, аспирант (polin_evgeny@mail.ru).

² Светлана Петровна Моисеева, д.ф.-м.н., профессор (smoiseeva@mail.ru).

³ Александр Николаевич Моисеев, д.ф.-м.н., доцент (moiseev.tsu@gmail.com).

1. Введение

Развитие техники, телефонии, сетей связи привело к необходимости использования более адекватных математических моделей процессов передачи и обработки данных, в которых интенсивность потока и параметры обслуживания могут изменяться случайным образом.

В литературе по теории массового обслуживания рассматриваются модели, в которых параметры, описывающие интенсивность прибытия и/или пропускную способность обслуживания, обладают одной из следующих характеристик: параметры однородны во времени; параметры не являются постоянными, а меняются во времени, однако их временная зависимость является исходной величиной модели; параметры меняются в зависимости от состояния воздействующих факторов [11].

Следует отметить, что указанные системы массового обслуживания (СМО) относятся к классу систем, функционирующих в случайной среде. Как правило, в качестве случайного процесса, описывающего среду, рассматривается число занятых приборов в системе. Примеры технических задач, приводящих к анализу таких СМО, можно найти в [1, 4-6, 14].

Поскольку в реальных жизненных ситуациях системы обслуживания имеют случайное время между моментами поступления заявок и случайное время их обслуживания, а среды обычно недетерминированы, СМО в случайных средах сегодня являются областью активных исследований в прикладной теории вероятностей. Довольно часто такие системы применяются для исследования систем с запасами Inventory Queue [12-16].

Возникающие сегодня сложные технологические и логистические системы возродили интерес к таким моделям. Особый интерес к этим системам представляют, например, услуги различного качества, которые предоставляются отдельным клиентам, или набор данных (сообщения в сетях) при ограниченных ресурсах и при внешних ограничениях [3, 10].

Примером модели, где параметры входящего потока зависят от числа находящихся на обслуживании заявок, может служить ситуация с просмотром контента (например, Youtube), в которой клиенты, в настоящее время просматривающие видео,

могут поделиться ссылками со своими знакомыми, вследствие чего возникает зависимость интенсивности входящего потока от общего числа клиентов. Необходимость понимания поведения таких систем обосновывает актуальность проведения настоящего исследования.

В настоящей работе рассматривается математическая модель бесконечнолинейной системы массового обслуживания с входящим МАР-потокком с интенсивностью, зависящей от числа занятых приборов. Эта модель является обобщением модели СМО, рассмотренной в работе [8], где проводилось исследование системы с входящим пуассоновским потоком с переменной интенсивностью вида $\lambda(i(t)) = a + b \cdot i(t)$. Здесь $i(t)$ – число занятых приборов в системе в момент времени t . В этой же работе было получено, что производящая функция рассматриваемого случайного процесса имеет вид производящей функции случайной величины, имеющей отрицательное биномиальное распределение (ОБР) [2] с параметрами $r = a/b$ и $p = 1 - b/\mu$, где μ – параметр экспоненциального времени обслуживания заявок на приборах. Целью данной работы является нахождение стационарного распределения вероятностей числа заявок в системе с входящим МАР-потокком с интенсивностью, зависящей от числа занятых приборов. Однако для данной модели получение результатов аналитическим путем не представляется возможным, поэтому в данной работе предлагается эвристический подход, а именно, использование отрицательного биномиального распределения как аппроксимации для искомого стационарного распределения вероятностей числа заявок в системе по аналогии с результатом [8]. В дальнейшем мы будем называть это распределение *ОБР-аппроксимацией*.

Поскольку для модели с входящим МАР-потокком параметры, аналогичные a и b , являются матрицами, то основной задачей становится разработка подхода или подходов к вычислению параметров ОБР-аппроксимации. В настоящей работе предлагается два таких подхода: один основан на вычислении параметров ОБР с помощью точных значений математического ожидания и дисперсии числа заявок в рассматриваемой системе, полученных по формулам, представленным в работе [9], а другой – на том факте, что интенсивность поступающих заявок определя-

ется управляющей потоком цепью Маркова $k(t)$. Кроме того, выполнен численный анализ точности построенных таким образом ОБР-аппроксимаций на основе сравнения с результатами имитационного моделирования.

2. Математическая модель

2.1 ОПИСАНИЕ МОДЕЛИ

Рассматривается система массового обслуживания (рис. 1), на вход которой поступает МАР-поток заявок. Управляющая цепь входящего потока $k(t) = 1, 2, \dots, K$ задана матрицей $\mathbf{Q} = [q_{k_1 k_2}]$ инфинитезимальных характеристик. Матрица условных интенсивностей имеет вид $\mathbf{\Lambda} = \mathbf{A} + \mathbf{B}i$, где \mathbf{A} – диагональная матрица с элементами a_k по главной диагонали, \mathbf{B} – диагональная матрица с элементами b_k по главной диагонали, i – число занятых приборов (число заявок) в системе. Обозначим вероятности появления заявок в системе в момент смены состояния МАР-потока с k_1 на k_2 как $d_{k_1 k_2}$, $k_1 \neq k_2$. Элементы $d_{k_1 k_2}$ образуют матрицу $\mathbf{D} = [d_{k_1 k_2}]$, $k_1, k_2 = 1, 2, \dots, K$ (диагональные элементы $d_{kk} = 0$). Таким образом, параметры входящего МАР-потока, а именно его условные интенсивности, меняются каждый раз, когда изменяется состояние системы (появляется новая заявка либо одна из заявок завершает обслуживание).

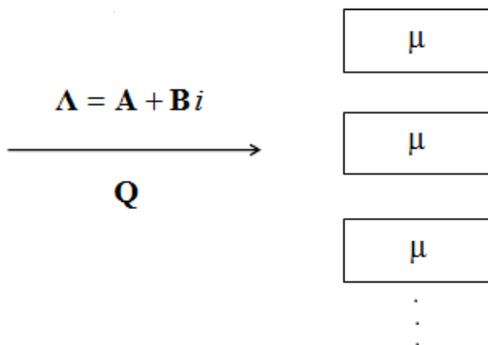


Рис. 1. СМО с интенсивностью входящего МАР-потока, зависящей от числа занятых приборов в системе

Дисциплина обслуживания определяется тем, что поступившая заявка занимает любой из свободных приборов в системе, на котором выполняется ее обслуживание в течение случайного времени, распределенного по экспоненциальному закону с параметром μ . Процесс $i(t)$ характеризует число занятых приборов в системе в момент времени t . Ставится задача нахождения распределения числа заявок в системе в стационарном режиме функционирования. Для существования стационарного режима необходимо, чтобы выполнялось условие $b_k < \mu$ для всех $k = 1, 2, \dots, K$.

2.2 ОТРИЦАТЕЛЬНОЕ БИНОМИАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Отрицательное биномиальное распределение (ОБР) – это распределение дискретной случайной величины ξ , равной количеству неудач, которые произошли в последовательности испытаний Бернулли с вероятностью успеха p до наступления r -го успеха. В оригинале закон распределения выглядит следующим образом:

$$P\{\xi = k\} = \frac{(k+r-1)!}{k!(r-1)!} p^r (1-p)^k,$$

для случаев действительных положительных значений параметра r его можно записать с использованием гамма-функции Эйлера в виде

$$(1) \quad P\{\xi = k\} = \frac{\Gamma(r+k)}{k!\Gamma(r)} p^r (1-p)^k.$$

Производящая функция вероятностей случайной величины с отрицательным биномиальным распределением равна

$$G(z) = \left(\frac{p}{1-(1-p)z} \right)^r, \quad |z| < \frac{1}{1-p},$$

математическое ожидание и дисперсия случайной величины ξ вычисляются по формулам

$$(2) \quad M\{\xi\} = \frac{r(1-p)}{p}, \quad D\{\xi\} = \frac{r(1-p)}{p^2}.$$

Параметр r в «классическом» ОБР может принимать только целые значения, однако для построения аппроксимации более

удобно использовать «расширенный» вариант ОБР с непрерывными значениями этого параметра (распределение Полюа). В случае действительных положительных значений параметра r закон распределения случайной величины ξ определяется формулой (1). Характеристическая функция имеет следующий вид:

$$(3) \quad h(u) = \left(\frac{p}{1 - (1-p)e^{ju}} \right)^r,$$

где j – мнимая единица. Если параметры ОБР известны, распределение вероятностей числа занятых приборов в системе может быть найдено с помощью обратного преобразования Фурье

$$P_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-juk} h(u) du, \quad k = 1, 2, \dots, N,$$

либо по формуле (1).

2.3 ОБР-АППРОКСИМАЦИЯ

Для использования ОБР в качестве аппроксимации необходимо определить его параметры r и p так, чтобы полученное распределение было максимально близко к настоящему распределению числа заявок в системе. Для получения таких результатов можно использовать формулы для математического ожидания и дисперсии (2), откуда получаем следующие выражения для вычисления параметров ОБР-аппроксимации:

$$(4) \quad p = \frac{M}{D}, \quad r = \frac{M \cdot p}{1 - p}.$$

Таким образом, задача сводится к нахождению математического ожидания и дисперсии для искомого распределения. Для случая стационарного распределения числа заявок в СМО с входящим МАР-поток, интенсивность которого зависит от состояния системы, мы предлагаем два подхода к решению данной задачи, которые представлены в следующем разделе.

3. ОБР-аппроксимация стационарного распределения числа заявок в СМО с входящим МАР-поток, интенсивность которого зависит от состояния системы

3.1 АППРОКСИМАЦИЯ НА ОСНОВЕ ТОЧНЫХ МОМЕНТОВ

В качестве первого метода аппроксимации предлагается вычисление параметров ОБР по формулам (4), где значения математического ожидания M и дисперсии D получены в результате исследования рассматриваемой системы методом моментов [9] и вычисляются следующим образом:

$$M = m_1 = Mi(t) = \mathbf{m}_1 \mathbf{e} = \mathbf{r}(\mathbf{A} + \mathbf{C})(\mu \mathbf{I} - \mathbf{B} - \mathbf{Q})^{-1} \mathbf{e},$$

$$(5) \quad m_2 = Mi^2(t) = \mathbf{m}_2 \mathbf{e} =$$

$$= \mathbf{r}(\mathbf{A} + \mathbf{C})\{(\mathbf{B} + \mathbf{Q} - \mu \mathbf{I})^{-1}(2(\mathbf{A} + \mathbf{C}) + \mathbf{B} + \mu \mathbf{I}) - \mathbf{I}\}(2\mathbf{B} + \mathbf{Q} - 2\mu \mathbf{I})^{-1} \mathbf{e},$$

$$D = m_2 - (m_1)^2,$$

где \mathbf{C} – матрица с элементами $d_{vk}q_{vk}$ и нулевыми элементами по главной диагонали; \mathbf{r} – вектор стационарного распределения вероятностей состояний цепи Маркова $k(t)$; \mathbf{I} – единичная матрица; \mathbf{e} – единичный вектор-столбец.

3.2 ЭВРИСТИЧЕСКАЯ АППРОКСИМАЦИЯ

Второй подход основан на том факте, что интенсивность поступающих заявок определяется управляющей потоком цепью Маркова $k(t)$. Если матрица \mathbf{B} является нулевой, то на вход поступает МАР-поток, интенсивность которого определяется как $a = \mathbf{r}(\mathbf{A} + \mathbf{C})\mathbf{e}$ [7], что соответствует параметру a в системе, рассматриваемой в работе [8]. Так как матрица \mathbf{B} также определяет интенсивность поступления дополнительных заявок, то определим параметр b аналогичным образом: $b = \mathbf{r}\mathbf{B}\mathbf{e}$. Параметры ОБР в данном случае вычисляем согласно результатам, полученным в [8]:

$$(6) \quad r = \frac{a}{b}, \quad p = 1 - \frac{b}{\mu}.$$

4. Численный анализ

В настоящем разделе проведено исследование области применимости предложенных аппроксимаций при различных значениях параметров модели.

4.1 ВЫРОЖДЕННЫЙ СЛУЧАЙ. ПОГРЕШНОСТЬ ИМИТАЦИОННОЙ МОДЕЛИ

Очевидно, что если взять в матрице **A** одинаковые диагональные элементы, то же самое проделать с матрицей **B**, а элементы матрицы **D** положить равными нулю, то входящий поток получится пуассоновским с переменной интенсивностью, как в работе [8]. В этом случае мы ожидаем, что результаты, полученные с помощью обоих видов аппроксимации, должны совпадать между собой и с результатом [8]. Проведенные численные эксперименты полностью это подтверждают.

Кроме того, используя такой вид модели можно оценить погрешность результатов имитационного моделирования соответствующих систем при разных объемах выборки. Для оценки погрешности мы используем расстояние Колмогорова, которое для дискретных распределений имеет вид

$$(7) \quad \Delta = \max_{n \geq 0} \left| \sum_{i=0}^n [P_1(i) - P_2(i)] \right|.$$

Здесь $P_1(i)$ и $P_2(i)$ – сравниваемые распределения вероятностей. В нашем случае одно из них будет результатом имитационного моделирования, другое – распределение, полученное аналитически. Результат оценки погрешности при разных объемах выборки (событий изменения состояния в имитационной модели) представлен в таблице 1.

Таблица 1. Погрешность (7) результатов имитационного моделирования при различных объемах выборки

Объем выборки	50 тыс.	100 тыс.	500 тыс.	1 млн	2 млн	5 млн
Расстояние Колмогорова	0,004	0,0038	0,0019	0,0011	0,0010	0,0009

При оценке качества аппроксимаций будем считать, что результат $\Delta \leq 0,05$ является приемлемой погрешностью. В дальнейшем будем брать объем выборки 2 миллиона событий.

4.2 ВЛИЯНИЕ МАТРИЦ **D** И **Q**

Далее установим, как параметры входящего потока влияют на точность аппроксимаций, построенных разными способами. Таких параметров несколько. Во-первых, это элементы матрицы **D**. Во-вторых, это диагональные элементы матриц **A** и **B**, а точнее – различие между элементами в каждой матрице: чем больше разница между условными интенсивностями МАР-потока, тем больше его дисперсия (точнее – дисперсия длин интервалов между событиями). Рассмотрим влияние этих трех факторов (матриц **D**, **A**, **B**) подробнее.

Влияние матрицы **D**. На самом деле влияние элементов матрицы **D** на близость аппроксимаций следует рассматривать в совокупности с матрицей **Q**, поскольку появление заявок в моменты изменения состояния вложенной цепи зависит не только от вероятностей этого появления (матрица **D**), но и от частоты смены состояний (матрица **Q**). Фактически мы имеем зависимость от поэлементного произведения этих матриц.

Итак, для начала возьмем матрицу **A**, у которой на главной диагонали стоят одинаковые элементы, равные 1, и матрицу **B** с диагональными элементами, равными 0,1.

В последующих экспериментах будем использовать матрицу **D** следующего вида:

$$(8) \quad \mathbf{D} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix},$$

а вариацию параметров будем производить посредством изменения элементов матрицы **Q**. Для этого будем задавать матрицу **Q** в виде

$$(9) \quad \mathbf{Q} = N \cdot \mathbf{Q}_0,$$

где \mathbf{Q}_0 – некоторая выбранная (одинаковая для всех экспериментов) матрица с соответствующими свойствами

генератора цепи Маркова, а параметр $N > 0$ будем варьировать, чтобы выяснить свойства аппроксимаций.

Если выбрать матрицу \mathbf{Q}_0 такой, что все ее строки будут одинаковы, то обе аппроксимации (5) и (6) дают одинаковый результат, при этом их погрешность составляет величину близкую к 0,007, что является очень хорошим результатом.

Возьмем матрицу \mathbf{Q}_0 с достаточно сильно различающимися строками, например

$$(10) \mathbf{Q}_0 = \begin{pmatrix} -1 & 0,2 & 0,8 \\ 4 & -5 & 1 \\ 2 & 8 & -10 \end{pmatrix},$$

при этом сохраним те же значения для элементов матриц \mathbf{A} , \mathbf{B} и \mathbf{D} , что и ранее. Результаты сравнения аппроксимаций с эмпирическим распределением, полученным в результате имитационного моделирования, представлены в таблице 2.

Таблица 2. Расстояния Колмогорова Δ_1 и Δ_2 для аппроксимаций первого и второго вида для разных значений параметра N при равных элементах на диагоналях матриц \mathbf{A} и \mathbf{B}

N	1	10	20	50	100	200
Δ_1	0,0095	0,0076	0,0037	0,0045	0,0038	0,0049
Δ_2	0,0415	0,0475	0,0501	0,0441	0,0504	0,0607

Здесь через Δ_1 и Δ_2 обозначены расстояния Колмогорова (7) для аппроксимаций первого (по точным моментам) и второго (эвристическая) вида по отношению к эмпирическому распределению. Из таблицы видно, что изменение параметра N слабо влияет на величину точности аппроксимаций, при этом ОБР-аппроксимация, основанная на точных моментах, имеет значительно (на порядок) меньшую погрешность по сравнению с эвристической аппроксимацией. На рис. 2, можно увидеть графики соответствующих законов распределения при $N = 20$ (обозначения: sim – эмпирическое распределение, approx – ОБР-аппроксимация, основанная на точных моментах, heurist – эвристическая ОБР-аппроксимация. Из рисунка понятно, что эвристическая ОБР-аппроксимация даёт существенно неточный

результат по дисперсии при достаточно точной оценке математического ожидания (это подтверждается числовыми данными).

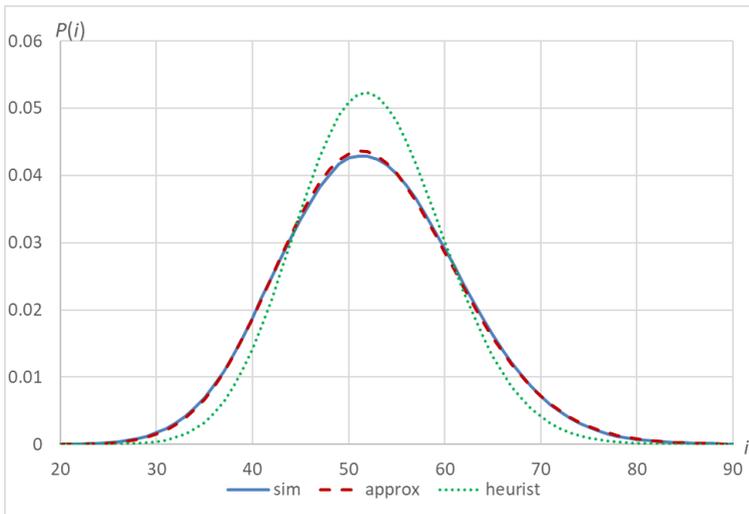


Рис. 2. Графики законов распределения для $N = 20$

4.3 ВЛИЯНИЕ МАТРИЦ \mathbf{A} И \mathbf{B}

Для проведения анализа влияния матриц \mathbf{A} и \mathbf{B} на точность предложенных аппроксимаций возьмем матрицу \mathbf{D} , как в (8), и матрицу $\mathbf{Q} = \mathbf{Q}_0$ из (10). Матрицы \mathbf{A} и \mathbf{B} будем брать в виде

$$(11) \quad \mathbf{A} = N_A \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 8 \end{pmatrix}, \quad \mathbf{B} = N_B \cdot \begin{pmatrix} 0,1 & 0 & 0 \\ 0 & 0,2 & 0 \\ 0 & 0 & 0,8 \end{pmatrix},$$

где $N_A, N_B > 0$ – параметры, которые мы будем варьировать. При таких значениях матриц \mathbf{A} и \mathbf{B} при увеличении параметров N_A и/или N_B будут увеличиваться средняя интенсивность входящего потока и дисперсия числа поступающих заявок в единицу времени, а вместе с ними – среднее и дисперсия результирующего стационарного распределения числа заявок в системе.

Влияние матрицы \mathbf{A} . Для первой серии экспериментов возьмем $N_B = 1$ и будем менять только N_A . В таблице 3

приведены результаты (расстояния Колмогорова) для этой серии экспериментов.

Таблица 3. Расстояния Колмогорова Δ_1 и Δ_2 для аппроксимаций первого и второго вида при разных значениях параметра N_A

N_A	0,5	1	2	5	10	20	50	100
Δ_1	0,007	0,0064	0,0020	0,0132	0,0222	0,0327	0,0435	0,0525
Δ_2	0,074	0,0936	0,1143	0,1664	0,2302	0,2903	0,3695	0,4115

Из таблицы видно, что ОБР-аппроксимация, основанная на точных моментах, дает достаточно небольшую погрешность в широком спектре значений параметра N_A , в то время как эвристическая ОБР-аппроксимация всегда дает неудовлетворительный результат для рассматриваемых случаев (при разных элементах на главных диагоналях матриц **A** и **B**). Графики соответствующих законов распределения можно увидеть на рис. 3. Из графиков видно, что эвристическая ОБР-аппроксимация имеет совсем другой характер, нежели истинное распределение вероятностей. При этом ОБР-аппроксимация, основанная на точных моментах, дает неплохие результаты, однако с ростом значения параметра N_A становится заметно, что левые части графиков существенно отличаются.

Влияние матрицы **B**. Зафиксируем теперь $N_A = 1$ и будем менять только N_B в допустимых пределах. В таблице 4 приведены расстояния Колмогорова для этой серии экспериментов.

Таблица 4. Расстояния Колмогорова Δ_1 и Δ_2 для аппроксимаций первого и второго вида при разных значениях параметра N_B

N_B	0,5	1	1,5	2	2,5	3
Δ_1	0,0113	0,0064	0,0035	0,0181	0,0422	0,1051
Δ_2	0,0938	0,0936	0,0940	0,1083	0,1294	0,1668

Как видим, ОБР-аппроксимация, основанная на точных моментах, дает хорошие результаты при росте параметра N_B вплоть до значения $N_B = 2,5$. После этого характер истинного распределения меняется (см. рис. 3), в результате чего ОБР-аппроксимация становится неприменимой. Эвристическая ОБР-

аппроксимация имеет большую погрешность при всех значениях параметра N_B .

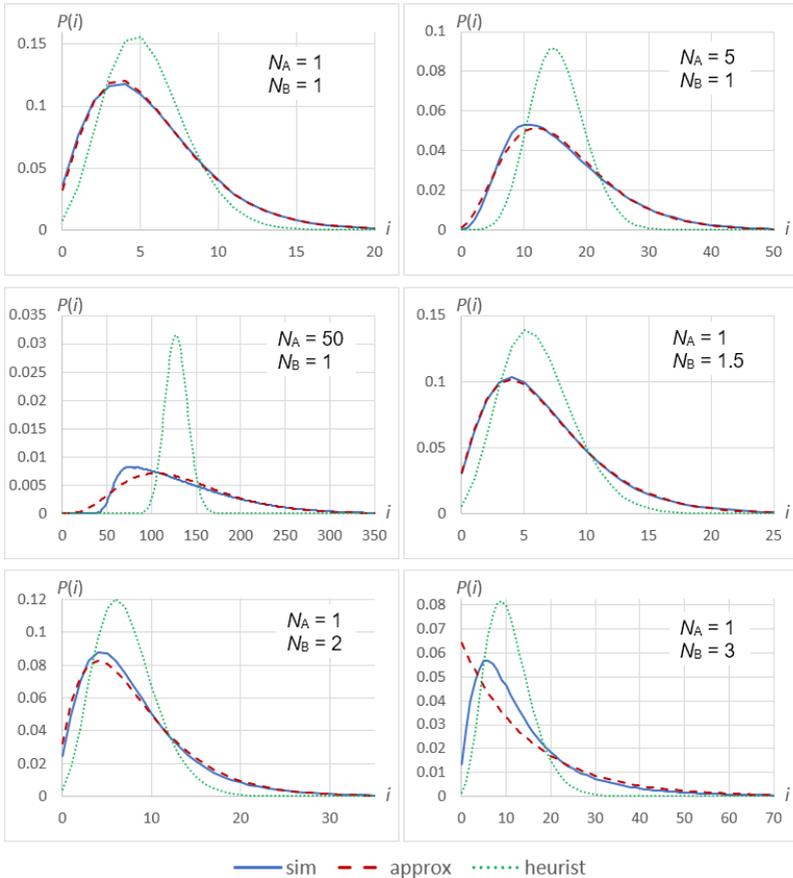


Рис. 3. Графики законов распределения для разных значений N_A и N_B

5. Заключение

Таким образом, в данной статье было представлено два подхода аппроксимации искомого распределения вероятностей числа заявок в системе с входящим МАР-поток с интенсивно-

стью, зависящей от числа занятых приборов. Аппроксимация проводилась с помощью отрицательного биномиального распределения. Был проведен численный анализ точности аппроксимации при различных значениях параметров исследуемой системы. Следует отметить, что при большой загрузке системы обе аппроксимации имеют большую погрешность, поэтому возникает необходимость в применении метода асимптотического анализа для исследования рассматриваемой СМО.

Литература

1. БАШАРИН Г.П., КОКОТУШКИН В.А., НАУМОВ В.А. *Метод эквивалентных замен в теории телетрафика*. – М.: Электросвязь. – 1980. – Т. 2. – С. 82–122.
2. ВАДЗИНСКИЙ Р.Н. *Справочник по вероятностным распределениям*. – СПб.: Наука. – 2001. – 295 с.
3. ГОРБУНОВА А.В., ЛЕБЕДЕВ А.В. *Эффекты стохастической нетранзитивности в системах массового обслуживания // Управление большими системами*. – 2020. – Вып. 85. – С. 23–50. – DOI: <https://doi.org/10.25728/ubs.2020.85.2>.
4. ДУДИН А.Н. *Модель процесса передачи данных в интегральных цифровых сетях связи с адаптивной коммутацией // Вычислительные сети коммутации пакетов*. – Рига: ИЗВТ. – 1987. – Т. I. – С. 121–124.
5. ДУДИН А.Н. *Об обслуживающей системе с переменным режимом работы // Автоматика и вычислительная техника*. – 1985. – №2. – С. 27–29.
6. ЖОЖИКАШВИЛИ В.А., ВИШНЕВСКИЙ В.М. *Сети массового обслуживания. Теория и применение к сетям ЭВМ*. – М.: Радио и связь. – 1988. – 192 с.
7. КАРЛЫХАНОВА Т.А., МОИСЕЕВА С.П., НАЗАРОВ А.А. *Исследование системы MAP/M/∞ методом моментов // Вестник Томского государственного университета*. – 2006. – № 293. – С. 99–104.
8. ПОЛИН Е.П., МОИСЕЕВА С.П., МОИСЕЕВ А.Н. *Исследование бесконечнолинейной СМО с интенсивностью входящего потока, зависящей от состояния системы // Материалы XXIII Международной научной конференции*

- «Распределенные компьютерные и телекоммуникационные сети: управление, вычисление, связь» (DCCN-2020), Distributed computer and communication networks: control, computation, communications (DCCN-2020), 14–18 сентября 2020 г., Москва. – М.: ИПУ РАН, 2020. – С. 638–644.
9. ПОЛИН Е.П., МОИСЕЕВА С.П. *Исследование математической модели бесконечнолинейной СМО с входящим МАР-потокм переменной интенсивности* // Проблемы оптимизации сложных систем: сборник докладов Двенадцатой международной азиатской школы-семинара. Новосибирск, Академгородок, 12–16 декабря 2016 г. Новосибирск. – 2016. – С. 380-384. – URL: <http://conf.nsc.ru/opcs2016/ru/proceedings>.
 10. ТУКУБАЕВ З.Б. *Моделирование и исследование алгоритмов динамического управления потоками и очередями сообщений в компьютерных сетях* // Управление большими системами. – 2009. – Вып. 26. – 2009. – С. 38–46.
 11. BOUCHENTOUF A.A., GUENDOUZI A. *Cost optimization analysis for an $M^X/M/C$ vacation queueing system with waiting servers and impatient customers* // SeMA Journal: Bulletin of the Spanish Society of Applied Mathematics. – 2019. – Vol. 76. – P. 309–341. – DOI: <https://doi.org/10.1007/s40324-018-0180-2>.
 12. БАЕК J.W., MOON S.K. *A production-inventory system with a Markovian service queue and lost sales* // J. Korean Stat. Soc. 45. – 2016. – P. 14–24. – DOI: <https://doi.org/10.1016/j.jkss.2015.05.002>
 13. DEEPAK T.G., KRISHNAMOORTHY A., NARAYANAN V.C. *Inventory with service time and transfer of customers and inventory* // Ann. Oper. Res. – 2008. – Vol. 160. – P. 191–213. – DOI: <https://doi.org/10.1007/s10479-007-0304-z>.
 14. FRALIX B., ADAN I. *An infinite-server queue influenced by a semi-Markovian environment* // Queueing Systems. – 2009. – Vol. 61. – P. 65–84.
 15. MATHEW N., JOSHUA V.C., KRISHNAMOORTHY A. *A production inventory model with server breakdown and customer impatience* // Ann Oper Res. – 2023. – DOI: <https://doi.org/10.1007/s10479-023-05659-x>.

16. MELIKOV A., KRISHNAMOORTHY A., SHAHMALIYEV M. *Perishable Queuing Inventory Systems with Delayed Feedback* // In: Information Technologies and Mathematical Modelling. Queuing Theory and Applications / Eds.: Dudin A., Nazarov A., Moiseev A. – ITMM WRQ Communications in Computer and Information Science. – 2018. – Vol. 912. – Springer, Cham. – DOI: https://doi.org/10.1007/978-3-319-97595-5_5.

APPLICATION OF NEGATIVE BINOMIAL DISTRIBUTION TO APPROXIMATE THE STATIONARY DISTRIBUTION OF THE NUMBER OF ARRIVALS IN A QS WITH AN INCOMING MAP, THE INTENSITY OF WHICH DEPENDS ON THE STATE OF THE SYSTEM

Evgeny Polin, National Research Tomsk State University, Tomsk, graduate student (polin_evgeny@mail.ru).

Svetlana Moiseeva, National Research Tomsk State University, Tomsk, Doctor of Science, professor (smoiseeva@mail.ru).

Alexander Moiseev, National Research Tomsk State University, Tomsk, Doctor of Science, professor (moiseev.tsu@gmail.com).

Abstract: This paper considers a mathematical model of an infinite-linear queuing system with an incoming MAP with an intensity depending on the number of busy servers. The parameters of the incoming process, namely its conditional intensities, change every time the state of the system changes, that is, a new request appears or one of the requests completes servicing. The service discipline is determined by the fact that the request occupies any of the free devices in the system on which its service is performed for a random time distributed according to an exponential distribution. For this model, obtaining a stationary probability distribution of the number of applications in the system by analytical means is not possible, so this paper proposes a heuristic approach, namely, the use of a negative binomial distribution as an approximation for the desired distribution. Two approaches to such approximation are proposed, for which a numerical analysis of the accuracy is performed based on comparison with the results of simulation modeling. The first approach is based on calculating the parameters of the negative binomial distribution using the exact values of the expected value and dispersion of the number of applications in the system under consideration, and the second is based on the fact that the intensity of incoming applications is determined by the Markov chain controlling the incoming process. It was found that the first approximation method gives more accurate results, however, when the system is heavily loaded, both approximations have a large error.

Keywords: heterogeneous queuing system, variable intensity, negative binomial distribution.

УДК 519.2

ББК 22.171

DOI: 10.25728/ubs.2024.108.3

*Статья представлена к публикации
членом редакционной коллегии А.В. Горбуновой.*

Поступила в редакцию 06.12.2023.

Опубликована 31.03.2024.

МОДЕЛИ И МЕТОДИКИ ОБОСНОВАНИЯ ПЛОТНОСТЕЙ ПОГРАНИЧНЫХ СИЛ, ОБЕСПЕЧИВАЮЩИХ НАДЕЖНУЮ ОХРАНУ ГОСУДАРСТВЕННОЙ ГРАНИЦЫ

Шумов В. В.¹

(Международный научно-исследовательский институт проблем управления, Москва)

На основе методов теории управления и учения о пограничной безопасности сформулирован подход к обоснованию плотностей пограничных сил, обеспечивающих надежную охрану государственных границ. С учетом положений военной науки и данных пограничной статистики назначены требуемые вероятности задержания нарушителей для трех уровней охраны (высокий, средний, минимальный). Для нарушителей границы с экономическими мотивами обоснование требуемой плотности охраны выполнено с использованием моделей пограничного сдерживания и агрегированной функции задержания. На примере рассмотрена методика обоснования средних плотностей охраны. Для нарушителей границы с неэкономическими мотивами (диверсионные и террористические группы) их сдерживание обеспечивается достижением требуемой вероятности их задержания (нейтрализации). В этом случае требуемые плотности охраны рассчитываются с использованием функций пограничных сил и средств, аналитических и имитационных моделей. Представлена упрощенная методика расчета вероятностей задержания нарушителей. Численные значения параметров моделей оценены на достаточном объеме данных пограничной статистики. Перспективным направлением исследований является формирование перспективного облика пограничных подразделений.

Ключевые слова: пограничная безопасность, охрана государственной границы, уровни охраны, агрегированная функция задержания, пограничное сдерживание, плотность охраны, плотность пограничных сил, модель.

1. Введение

Одним из семи существующих чудес света является Великая Китайская стена, построенная в III веке до н.э. (династия Цинь) на месте отдельных пограничных сооружений китайских царств; впоследствии, в течение более тысячелетия, она неоднократно

¹ Владислав Вячеславович Шумов, д.т.н., профессор (v.v.shumov@yandex.ru).

достраивалась и реконструировалась. Стена использовалась не только в целях обороны, но и для других целей: пограничного контроля и регулирования торговли с иноземцами. Протяженность Китайской стены составляет около 9 000 км, а со всеми ответвлениями – более 21 000 км. Китайское «го» в слове «государство» изображается иероглифом: ограда и человек с копьем.

На границах Римской империи в I–II веках н.э. возводились лимесы¹ (валы): Германский лимес, Дунайский лимес, вал Адриана, Африканский лимес и др. Они служили как оборонительными сооружениями, так и средствами пограничного и таможенного контроля. В Римской империи численность пограничных легионов составляла до 2/3 от общей численности вооруженных сил. При этом в III веке (212 и 284 гг.) плотность защиты и охраны границы (количество пограничников на километр границы) составляла от 2 чел./км (Африка, Мавритания) до 15–20 чел./км (Сирия, Германия и др.) [1]. В более поздний период Римской империи происходит дальнейшее увеличение плотности защиты и охраны границы. Примерно десятая часть границы была защищена пограничными сооружениями со стенами или валами и около двух десятых – системой опорных пунктов с крепостями, наблюдательными постами и другими элементами предохранения. Предпринятые меры по оборудованию границы и ее охране позволили отразить множество вторжений иноземных племен на территорию Римской (Византийской) империи.

Во времена правления Ивана IV и позже на территории современных Тульской, Рязанской и Калужской областей России существовала Большая засечная черта протяженностью около 400 километров. Историки сравнивают ее с Адриановым валом в Англии, Римским Лимесом и другими пограничными заграждениями².

¹ Лимес (от лат. *Limes* – дорога, граничная тропа, граница) – укрепленный рубеж со сторожевыми башнями.

² Ученые нашли фрагмент Большой русской засечной стены. РИА Новости, 20.10.2023. – URL: <https://ria.ru/20231020/arkheologiya-1903929435.html> (дата обращения: 20.10.2023).

В наше время хорошо укрепленные и охраняемые разделительные стены функционируют в Азии (Пакистан, Индия, Израиль и другие страны), в Америке (США), в Африке и Европе. Например, на границе США с Мексикой плотность охраны границы составляет 6,3 чел./км. По периметру границы расположены бетонные заборы, инфракрасные камеры, датчики, постоянно используются беспилотные летательные аппараты, свыше 20 тыс. американских пограничников обеспечивают безопасность границы [12].

Существует множество причин военно-политического, этнического, экономического и иного характера, побуждающих государства пересматривать соотношение между барьерной и контактной функцией границ, наращивать усилия по укреплению их охраны или, наоборот, превращать их в почти прозрачные для пересечения лиц и грузов.

Проблемы обеспечения безопасности в пограничном пространстве исследуются рядом дисциплин, включая погранологию, лимологию и исследования границ. В таблице 1 показано число употреблений терминов, связанных с исследованиями границ, пограничной политикой и охраной пограничных пространств (по состоянию на 23.01.2023).

Таблица 1. Количество статей и поисковых результатов

Термин	Количество статей, зарегистрированных в Google Scholar	Количество результатов в поисковой системе Яндекс
Погранология	108	2 000
Лимология	507	2 000
Пограничная безопасность	20 000	7 000
Пограничная политика	25 500	386 000
Border security	3 910 000	213 000
Border policy	4 970 000	76 000
Border studies	5 530 000	10 000
Border research	6 500 000	58 000

В докладе [3] отмечается, что процессы глобализации сопровождаются увеличением числа признанных и непризнанных государств, ростом сепаратизма в ряде стран, т.е. фрагментацией мирового политического пространства. Выявление причин этих сдвигов – фундаментальная научная проблема.

Не менее значимой научной и практической проблемой является обоснование требуемых плотностей охраны границы¹ исходя из уровней существующих или ожидаемых угроз безопасности. С обзором моделей охраны границы можно ознакомиться в статьях [8, 10]. Вместе с тем следует отметить, что вопросы обоснования требуемых плотностей охраны границы мало исследованы. Из зарубежных работ в этой области можно назвать статью «Analyzing the Homeland Security of the U. S.-Mexican Border» [14].

Целью настоящей статьи является разработка моделей и методик обоснования плотностей пограничных сил, обеспечивающих надежную охрану государственной границы. Структурно работа организована так.

Во втором разделе для трех категорий нарушителей границы и трех уровней охраны (высокой, средней и минимальной) определены требуемые вероятности задержания нарушителей.

В третьем разделе с использованием агрегированной функции задержания для нарушителей с мотивами экономического характера разработаны модель и методика обоснования средних плотностей охраны.

В четвертом разделе обоснованы требования к построению охраны границы и разработана методика расчета вероятности задержания нарушителей границы.

2. Требуемые уровни охраны границы

Анализ отечественного и зарубежного опыта развития пограничных сил и средств позволяет сформулировать принципы

¹ *Охрана государственной границы возлагается на пограничные ведомства (силы). Поэтому далее термины «плотность охраны границы» и «плотность пограничных сил» полагаются тождественными.*

назначения требуемых уровней охраны границы и обоснования перспективного облика пограничных подразделений:

1. Системный и комплексный подходы, прогнозирование национальной и пограничной безопасности, выявление уязвимостей в охране границы.

2. На основе анализа угроз формирование перечня ожидаемых категорий нарушителей границы, оценка рисков и возможностей нарушителей по прорыву границы и нанесению ущерба.

3. Недопущение нарушений границы реализуется двумя основными способами, допускающими количественную оценку: во-первых, созданием условий и рисков, вынуждающих потенциальных нарушителей отказываться от своих намерений (результат пограничного сдерживания), во-вторых, их задержанием (нейтрализацией) в случае попытки нарушения границы.

4. Разработка вариантов системы пограничных мер с учетом существующих и перспективных образцов вооружения, военной и специальной техники, сетецентрических технологий и автоматизированных систем.

5. Сочетание мобильных и стационарных сил и средств охраны, наличие резерва сил и средств.

6. Применение комплекса моделей (аналитических, имитационных) для оценки эффективности системы пограничных мер, оценка параметров моделей по результатам служебной и служебно-боевой деятельности, подготовка и проведение полигонных и войсковых испытаний.

Положения военной науки, пограничной статистики и опыт служебно-боевых действий дают основания для назначения требуемых уровней охраны границы (таблица 2).

Таблица 2. Требуемые уровни охраны границы

Уровень охраны	Категория нарушителей, требуемая вероятность задержания (нейтрализации)		
	1. Невооруженные нарушители (без пособников)	2. Контрабанда оружия (наркотиков)	3. Диверсионные (террористические) группы
Высокий	0,5–0,7	0,6–0,8	0,8–0,9
Средний	0,3–0,5	0,4–0,6	0,6–0,8
Минимальный	Меньше 0,3	Меньше 0,4	Меньше 0,6

Значения требуемой вероятности задержания (нейтрализации) нарушителей определяются двумя факторами:

- степенью опасности нарушителей для государства и общества (ожидаемым ущербом и его последствиями);
- возможностями пограничного сдерживания нарушителей с экономическими мотивами.

Из трех представленных в таблице категорий нарушителей первые две, как правило, имеют мотивы экономического характера (поиск лучших условий жизни, нарушения с хозяйственно-бытовыми целями, контрабанда и т.д.). Обоснование требуемых вероятностей их задержания выполняется с использованием моделей пограничного сдерживания [9].

Третья группа нарушителей является наиболее опасной по двум причинам: во-первых, эта группа несет угрозу жизни гражданам, представителям власти, общественным деятелям и малыми силами может нанести непоправимый ущерб; во-вторых, террористы и диверсанты являются хорошо подготовленными и мотивированными и их (или их организаторов) может остановить только угроза быть уничтоженными при попытке нарушения границы или на пути следования к объектам атак.

Поэтому для третьей группы высокий уровень охраны означает высокую гарантию их уничтожения (нейтрализации) при попытке пересечения границы. А наличие дополнительных сил и средств других ведомств и организаций приведет почти гарантированно к срыву поставленной задачи.

На протяженных участках границы высокий уровень ее охраны потребует привлечения значительных сил и средств, что может оказаться неэффективным с точки зрения государства. Тогда может оказаться разумным назначить средний уровень охраны границы, при котором окажутся высокими возможности по своевременному обнаружению нарушителей и их последующему задержанию во взаимодействии с другими ведомствами.

Далее рассмотрим задачу обоснования требуемых вероятностей задержания нарушителей с экономическими мотивами.

3. Моделирование пограничного сдерживания

В ряде государств пограничное сдерживание («оптимальное сдерживание», см. [13]) является основным критерием охраны границы и важнейшим принципом пограничной деятельности в силу следующих обстоятельств: во-первых, имеется хорошо разработанный научный аппарат по прямой и косвенной оценке сдерживания нарушителей (пограничная статистика, экономическая теория преступности, теория дискретного выбора и т.д.), во-вторых, эффективность пограничного сдерживания оценена по результатам служебной деятельности (границы СССР, граница США с Мексикой).

3.1. АГРЕГИРОВАННАЯ ФУНКЦИЯ ЗАДЕРЖАНИЯ НАРУШИТЕЛЕЙ

На основе анализа результатов охраны границы СССР в 1960–1980 гг. и охраны американо-мексиканской границы в 1980–2011 гг. (см. [2, 12]) получена статистически значимая агрегированная функция задержания. В соответствии с ней вероятность задержания нарушителей границы равна

$$(1) \quad p_z = 1 - \exp(-\lambda x),$$

где λ – параметр; x – плотность охраны границы (количество личного состава подразделений охраны, включая резервы и обеспечивающие подразделения, на 1 км границы).

Содержательно параметр λ отражает: 1) искусство руководителей применять пограничные силы и средства; 2) возможности типовых нарушителей границы; 3) физико-географические и иные условия. Значения параметра λ оценены методом наименьших квадратов (таблица 3).

Таблица 3. Значения параметра

Категория нарушителей	Неподготовленные нарушители	Подготовленные нелегальные мигранты	Подготовленные контрабандисты
Значение параметра λ	0,3–0,4	0,22–0,25	0,18–0,22

На рис. 1 для трех категорий нарушителей показаны вероятности их задержания в зависимости от плотности охраны границы.

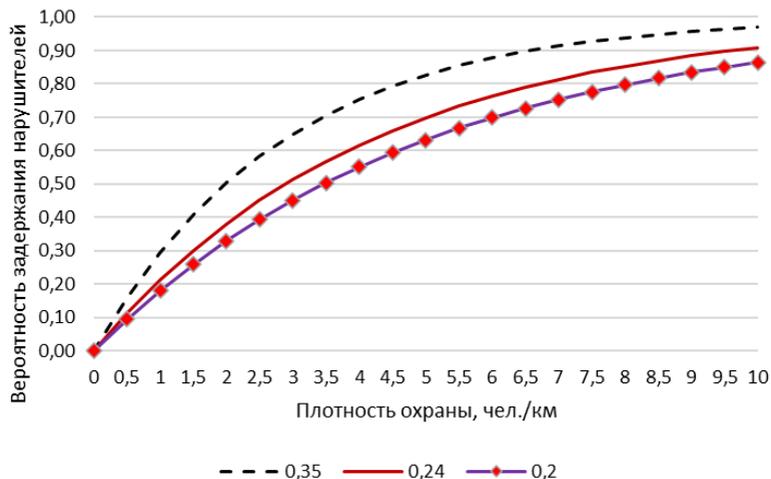


Рис. 1. Зависимость вероятностей задержания нарушителей от плотности охраны границы

Из выражения (1) находим формулу для вычисления требуемой плотности X_r охраны границы:

$$(2) \quad X_r = -\frac{1}{\lambda} \ln(1 - P_r),$$

где P_r – требуемая вероятность задержания нарушителей границы.

Рассмотрим обоснование требуемых вероятностей задержания нарушителей с экономическими мотивами (нелегальные мигранты, контрабандисты и т.д.). Полезность u_1 незаконной деятельности¹ таких нарушителей равна [11]:

$$(3) \quad u_1 = B(p_z)(D - S) + (1 - B(p_z))D,$$

¹ Известно множество моделей ожидаемой полезности (см., например, [7]).

где p_z – вероятность задержания нарушителей; $B(\cdot)$ – функция представления нарушителей о вероятности их задержания (зависит от информации, получаемой нарушителями от пособников, из СМИ и социальных сетей); D – ожидаемая прибыль от незаконной деятельности; S – денежный эквивалент наказания (включая конфискацию имущества, судебные и другие издержки).

Содержательно модель (3) можно трактовать так: прибыль D от преступления – это сила, которая побуждает человека совершить правонарушение. Тяжесть S и неотвратимость p_z наказания – сила, удерживающая его от этого. Если полезность u_1 незаконной деятельности положительна, правонарушение следует ожидать, иначе – нет.

В простейшем случае функция представления $B(\cdot)$ имеет вид [5]

$$(4) \quad B(y, x, p) = \alpha B_+(y, p) + (1 - \alpha) B_-(x, p),$$

$$B_+(y, p) = \frac{p \exp(y)}{1 - p + p \exp(y)},$$

$$B_-(x, p) = \frac{p \exp(-x)}{1 - p + p \exp(-x)},$$

где $0 < \alpha < 1$ – параметр оптимизма в оценке вероятности p ; y – количество (доля) сообщений, направленных на увеличение представления о вероятности p ; x – количество (доля) сообщений, направленных на понижение представления о вероятности p . При $\alpha > 0,5$ индивид склонен к лучшему восприятию (доверию) сообщений (информации) об увеличении вероятности p , при $\alpha < 0,5$ он больше доверяет сообщениям о низких значениях этой вероятности.

При $y = x = 0,5$ и при трех значениях параметра оптимизма (0,25, 0,5 и 0,75) на рис. 2 показан график зависимости представления от вероятности.

Пессимисты (индивиды, стремящиеся к риску, $\alpha = 0,25$) преуменьшают значение вероятности (синяя линия с маркерами), тогда как оптимисты (индивиды, стремящиеся избежать риска, $\alpha = 0,75$) преувеличивают вероятность (красная пунктирная линия). Рисконейтралы ($\alpha = 0,5$, сплошная черная линия) почти точно оценивают вероятность.

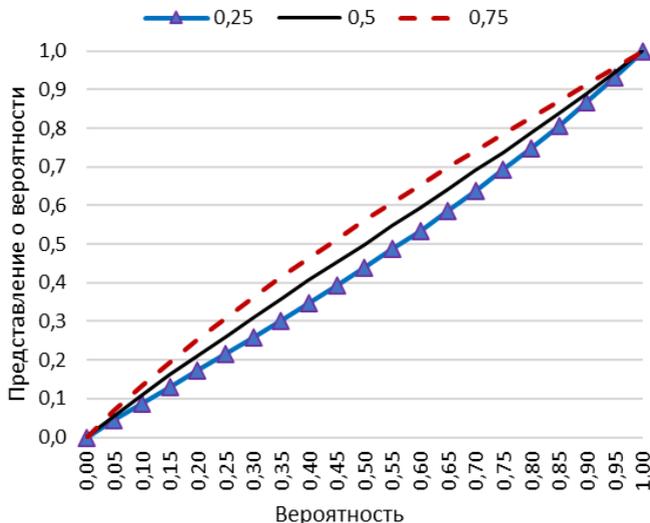


Рис. 2. Зависимость вероятностей задержания нарушителей от плотности охраны границы

Зная по каждой категории нарушителей с экономическими мотивами ожидаемые значения их прибыли D и наказания S , из выражения (3) и условия $u_1 < 0$ можно найти требуемое значение представления о вероятности для сдерживания нарушителей:

$$(5) \quad B(p_z) > \frac{D}{S}.$$

Пример 1. Нарушитель планирует незаконно пересечь государственную границу с целью получения прибыли в размере $D = 200\,000$ руб. и через полгода незаконно пересечь границу в обратном направлении. Рассчитать требуемую вероятность его задержания, при которой ему окажется невыгодной незаконная деятельность. Полагается, что он осведомлен о действиях по охране границы (через пособников или лиц, уже незаконно пересекавших границу).

Решение. В условиях осведомленности нарушителей о системе охраны границы и при отсутствии целенаправленных ин-

формационных воздействий полагаем, что представление о вероятности задержания совпадает с этой вероятностью, т.е. $B(p_z) = p_z$.

В соответствии со статьей 322 УК РФ и сложившейся практикой наказания примем $S = 250\,000$ руб.

Нарушитель будет дважды нарушать границу ($n = 2$). Для привлечения его к ответственности достаточно одного задержания, т.е. модель (3) будет иметь вид (условие пограничного сдерживания):

$$(6) \quad u_1 = (1 - (1 - p_z)^n)(D - S) + (1 - p_z)^n D < 0,$$

откуда находим:

$$(D - S) + (1 - p_z)^n S < 0, \quad p_z > 1 - \sqrt[n]{\frac{S - D}{S}} \approx 0,55.$$

Таким образом, в условиях примера для частичной реализации пограничного сдерживания конкретной категории нарушителей (их отказа от попыток нарушения границы) достаточно обеспечить вероятность задержания нарушителей не ниже 0,55. Отметим, что если нарушители достигают своей цели при однократном нарушении границы ($n = 1$), то потребуется обеспечить вероятность задержания не ниже 0,8:

$$p_z > 1 - \frac{S - D}{S} = 0,8.$$

Важно подчеркнуть, что при $S < D$ (денежный эквивалент наказания меньше ожидаемой прибыли) пограничное сдерживание невозможно обеспечить при любых сколь угодно высоких затратах на охрану границы. Поэтому наряду с реализацией мер по охране границы необходимо вести мониторинг по ожидаемым доходам нарушителей и предпринимать меры по снижению их дохода и/или увеличению меры наказания.

Нарушители оценивают систему охраны границы с ошибками, поэтому для оценки вероятности отказа нарушителей от незаконной деятельности можно воспользоваться моделью бинарного выбора:

$$(7) \quad s_0 = \frac{\exp(\theta)}{\exp(\theta) + \exp(\theta u_1 / u_0)},$$

где $\theta \approx 2-4$ – степень знания нарушителями системы охраны границы; u_0 – полезность законной деятельности.

Пример 2. Полезность законной деятельности потенциальных нарушителей равна $u_0 = 100$, степень знания системы охраны границы $\theta = 3$. Построить график зависимости вероятности отказа нарушителей от незаконной деятельности при изменении ожидаемой ее полезности в интервале $u_1 = (0; 300)$.

Результаты расчета показаны на рис. 3. Из рисунка видно, что при малых значениях ожидаемой полезности незаконной деятельности вероятность отказа s_0 близка к единице, при $U_1 = U_0$ эта вероятность равна 0,5 и с ростом полезности U_1 вероятность стремится к нулю.

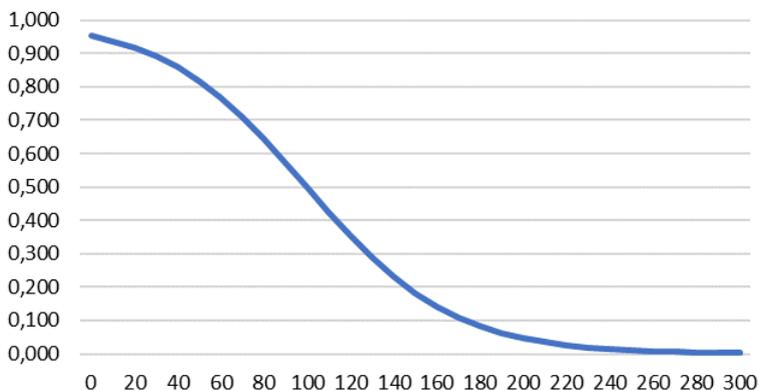


Рис. 3. График зависимости вероятности отказа s_0 от ожидаемой полезности u_1 незаконной деятельности

Представленные математические модели позволяют сформулировать методику обоснования средних плотностей охраны границы, при которых реализуется сдерживание нарушителей с экономическими мотивами.

3.2. МЕТОДИКА ОБОСНОВАНИЯ СРЕДНИХ ПЛОТНОСТЕЙ ОХРАНЫ ГРАНИЦЫ

Реализацию методики рассмотрим на **примере 3**.

Шаг 1. Уточнение (прогноз) сведений о массовых потенциальных нарушителях границы с экономическими мотивами. По каждой категории нарушителей оцениваются данные:

- степень знания нарушителями системы охраны границы θ (при $\theta \approx 3$ нарушители пользуются услугами проводников, при $\theta \approx 2$ нарушители самостоятельно изучают систему охраны границы, при $\theta \approx 4$ имеется организованная преступная группировка);

- ожидаемая прибыль D от незаконной деятельности (доход минус издержки);

- денежный эквивалент наказания S ;

- средний доход (зарплата) в регионе u_0 ;

- количество n нарушений границы для реализации прибыли.

Шаг 2. Назначается вероятность s_0 , при которой реализуется пограничное сдерживание (массовый отказ от попыток нарушений границы данной категорией лиц). Обычно полагается $s_0 = 0,7-0,8$.

Из выражения (7) находим

$$(8) \quad u_1 = u_0 + \frac{u_0}{\theta} \ln \frac{1-s_0}{s_0}.$$

При $\theta = 3$, $s_0 = 0,75$, $u_0 = 30\,000$ руб. получим: $u_1 \approx 19\,014$ руб.

Шаг 3. Из выражения (при $n = 1$ или $n = 2$)

$$u_1 = (1 - (1 - p_z)^n)(D - S) + (1 - p_z)^n D$$

находим требуемую вероятность задержания нарушителей.

Пусть $n = 1$, $D = 200\,000$ руб., $S = 250\,000$ руб. Тогда

$$p_z = \frac{D - u_1}{S} \approx \frac{200000 - 19014}{250000} \approx 0,72.$$

Шаг 4. Зная категорию нарушителей, находим значение параметра λ и по формуле (2) вычисляем требуемую плотность охраны границы:

$$X_r = -\frac{1}{\lambda} \ln(1 - P_r) = -\frac{1}{0,24} \ln(1 - 0,72) \approx 5,4 \text{ чел./км.}$$

Шаг 5. Повторяем вычисления для других значимых категорий нарушителей (текущих и прогнозных) и принимаем в качестве требуемой плотности охраны границы максимальную величину.

Отметим, что нами рассчитана средняя плотность охраны границы (пограничных сил) в регионе. Для оптимального назначения плотностей на конкретных участках ответственности необходимо использовать аналитические модели.

4. Обоснование требований к построению охраны границы

Концептуальная схема аналитического обоснования требований к построению охраны границы основана на функциях пограничных сил и средств [2], математических моделях и показана на рис. 4.

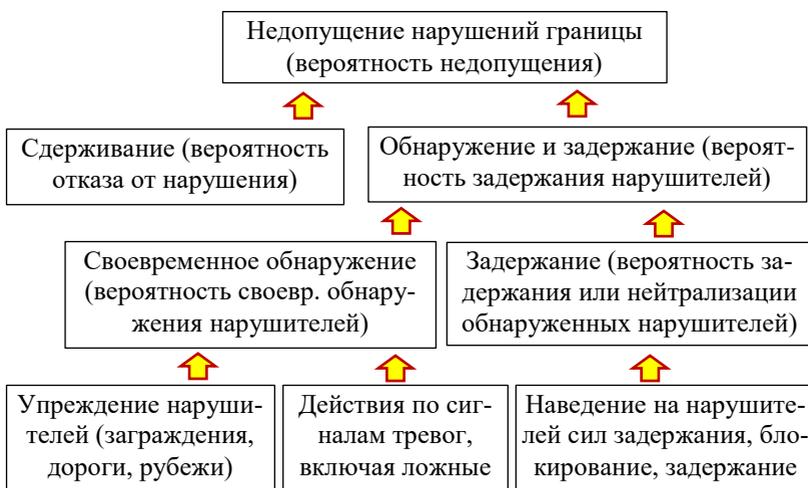


Рис. 4. Концептуальная схема обоснования требований к построению охраны границы

Модели пограничного сдерживания разработаны в основном для нарушителей с экономическими мотивами. Сдерживание диверсионно-разведывательных и террористических групп обеспечивается за счет создания высоких плотностей охраны границы, а следовательно, и рисков быть задержанными или уничтоженными при попытке пересечения границы.

Вероятность задержания нарушителей вычисляется по формуле:

$$(9) \quad P_z = P_{so}P_{mz},$$

где p_{so} – вероятность своевременного обнаружения нарушителей и качественных действий по сигналам тревог; p_{mz} – вероятность задержания (нейтрализации) нарушителей при условии их обнаружения.

При наличии аналитических дифференцируемых функций $p_{so}(x)$ и $p_{mz}(y)$, где x и y – количество сил и средств, выделяемых для решения задач своевременного обнаружения нарушителей и их задержания, условие оптимального распределения между названными задачами имеет вид

$$\frac{p'_{so}(x)}{p_{so}(x)} = \frac{p'_{mz}(y)}{p_{mz}(y)}.$$

Если вероятность p_z задана нормативно и, в частном случае, когда издержки на реализацию функций своевременного обнаружения нарушителей и их задержания примерно равны, то необходимо обеспечить следующие значения вероятностей p_{so} и p_{mz} :

$$(10) \quad p_{so} \approx p_{mz} \approx \sqrt{p_z}.$$

В таблице 4 показаны необходимые уровни вероятностей обнаружения и задержания обнаруженных нарушителей для обеспечения требуемой вероятности задержания.

Таблица 4. Необходимые уровни вероятностей

Требуемая вероятность задержания	0,9	0,8	0,7	0,6	0,5	0,4	0,3
Необходимые уровни вероятностей p_{so} и p_{mz}	0,95	0,89	0,84	0,77	0,71	0,63	0,55

Для получения точных результатов расчетов необходимо использовать данные о местности на участках ответственности и положении пограничных сил и средств, включая резервы.

Кратко опишем методику расчета вероятности задержания нарушителей.

4.1. МЕТОДИКА РАСЧЕТА ВЕРОЯТНОСТИ ЗАДЕРЖАНИЯ НАРУШИТЕЛЕЙ

Шаг 1. Расчет зон упреждения нарушителей резервами (заслонами) и пограничными нарядами (рис. 5).

Время t_y упреждения нарушителей рассчитывается по формуле

$$(11) t_y = t_{ng} - t_{zn},$$

где t_{ng} – время движения нарушителя с момента его обнаружения до выхода на рубеж (к границе, в тыл); t_{zn} – время движения заслона (резерва, погн) с учетом времени сбора).

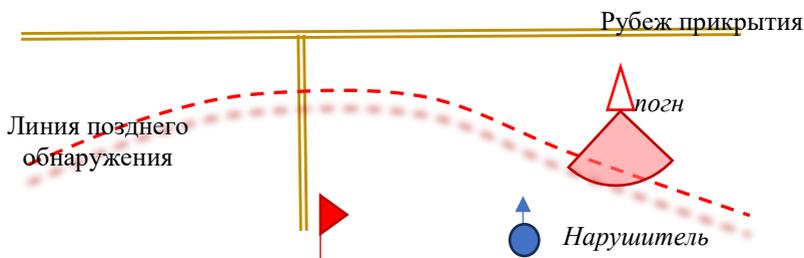


Рис. 5. Зоны упреждения нарушителей

Полагается, что нарушитель движется по кратчайшему проходимому маршруту в сторону границы или в наш тыл (к объекту интереса).

На рис. 5 показана линия позднего обнаружения (пунктирная линия) заслоном и зона упреждения нарядом (сектор). Если нарушитель будет обнаружен выше линии позднего обнаружения и вне зоны упреждения, то он не будет задержан. То есть вероятность его своевременного обнаружения равна нулю. Иначе вероятность своевременного обнаружения полагается равной вероятности p_o его обнаружения:

$$(12) p_{so} = \begin{cases} p_o, & t_y \geq 0, \\ 0, & t_y < 0. \end{cases}$$

Расчеты показывают, что для обеспечения упреждения нарушителей необходимо выполнить минимально необходимые требования (таблица 5).

Таблица 5. Требования по обеспечению упреждения нарушителей

Требование	Уровень охраны границы	
	Средний	Высокий
Количество подготовленных рубежей	Не менее двух	Не менее трех
Наличие заграждений, воспрепятствующих использованию нарушителями транспортных средств	Желательно	Обязательно
Режим пограничной полосы	Желательно	Обязательно

Шаг 2. Расчет вероятности приборного обнаружения. Вероятность обнаружения нарушителей рассчитывается как доля участка и времени суток, при которых возможно обнаружение нарушителей (пространственно-временная схема прикрытия средствами обнаружения [6], рис. 6).



Рис. 6. Зона обнаружения нарушителей

Шаг 3 (учет сигналов ложных тревог). Под вероятностью обнаружения нарушителей понимается приборная вероятность обнаружения с учетом своевременных действий по сигналам тревог (одноканальная система массового обслуживания с ограниченным временем ожидания):

$$(13) P_o = P_{пр} P_{смo},$$

$$(14) P_{смo} = \begin{cases} 1 - \frac{\lambda_t}{\mu} e^{-(\mu - \lambda_t)t_y}, & \mu > \lambda_t, \\ 0; & \mu \leq \lambda_t, \end{cases}$$

где $P_{пр}$ – приборная вероятность обнаружения (плотность прикрытия участка средствами обнаружения); $P_{смo}$ – вероятность

своевременных и качественных действий по сигналам тревог; λ_t – интенсивность сигналов тревог (включая ложные); μ – интенсивность обслуживания сигналов тревог.

Интенсивность обслуживания μ вычисляется по формуле

$$(15) \quad \mu = \begin{cases} \frac{1}{t_d}; & \text{1-й способ,} \\ \frac{1}{t_d + \Delta t}; & \text{2-й способ,} \end{cases}$$

где t_d – среднее время сбора, выезда на участок, обнаружения признаков нарушения границы, непосредственно действий по задержанию нарушителей и возвращения; $\Delta t \approx 12\text{--}15$ ч – время, ежедневно необходимое личному составу на сон, прием пищи, занятия, обслуживание техники и т.д.

При первом способе действий в подразделении назначается резерв личного состава, предназначенный для действий по сигналам тревог, причем час времени нахождения в резерве засчитывается как полчаса несения службы. При втором способе действий в резерв назначается весь свободный от службы личный состав.

Шаг 4 (расчет условной вероятности задержания). Условная вероятность p_{mz} вычисляется по формуле¹ [4]:

$$(16) \quad p_{mz} = \frac{(\beta x)^\alpha}{(\beta x)^\alpha + y^\alpha}, \quad \beta = \sqrt[4]{\beta_s \beta_r \beta_w \beta_m},$$

где x – численность пограничников, участвующих в задержании (нейтрализации) нарушителей; y – численность нарушителей; β – параметр боевого превосходства пограничников над нарушителями; α – параметр масштаба; β_s (β_r , β_w , β_m) – параметр превосходства над противником в согласованности действий (обнаружении, поражении и маневренности).

Параметр масштаба равен:

$\alpha = 0,5$ при отсутствии средств непрерывного сопровождения нарушителей и наведения на них нарядов и групп;

$\alpha = 1$ в противном случае.

¹ Допустимым средним в шкале отношений является среднее геометрическое.

5. Заключение

Таким образом, нами рассмотрены модели и методики обоснования плотностей пограничных сил, обеспечивающих надежную охрану государственной границы.

Перспективными направлениями исследований являются: во-первых, разработка моделей и методик формирования перспективного облика пограничных подразделений, во-вторых, разработка аналитических и имитационных моделей для детального учета тактико-технических характеристик вооружения и специальной техники, особенностей местности.

Автор выражает признательность Д.А. Новикову и А.Г. Чхартишвили за идеи, консультирование и руководство работ в области моделирования пограничной безопасности, а участникам научного семинара по теории управления организационными системами – за советы, критику и рекомендации.

Литература

1. ГИРНИК Е.С. *Реконструкция пограничной статистики Римской империи* // Вопросы безопасности. – 2017. – №5. – С. 33–54.
2. ГИРНИК Е.С., МАНИЛОВ А.Л., ШУМОВ В.В. *Пограничная статистика и ее применение для оценки уровня пограничной безопасности на внешних границах государств – участников СНГ* / Под общ. ред. В.В. Шумова. – М.: КЖИ «Граница», 2022. – 232 с.
3. КОЛОСОВ В.А., СЕБЕНЦОВ А.Б. *Глобализация, суверенитет и фрагментация мирового политического пространства* // Социально-экономическая география: теория, методология и практика преподавания. Материалы Всероссийской научно-практической конференции «Пятые Макасовские чтения», Москва, 24–25.09.2020. – С. 221–235.
4. КОРЕПАНОВ В.О., ЧХАРТИШВИЛИ А.Г., ШУМОВ В.В. *Базовые модели боевых действий* // Управление большими системами. – 2023. – Вып. 103. – С. 40–77.

5. МАКАРЕНКО А.В., ЧХАРТИШВИЛИ А.Г., ШУМОВ В.В. *Системный анализ и прогнозирование безопасности* / Под общ. ред. В.В. Шумова. – М.: ЛЕНАНД, 2022. – 216 с.
6. МАНИЛОВ А.Л., САВЕНКО В.Н., ШУМОВ В.В. *Моделирование деятельности пограничных ведомств государств – участников Содружества Независимых Государств: учебное пособие* / Под ред. В.А. Дмитриева. – М.: Граница, 2014. – 608 с.
7. ШУМЕЙКЕР П. *Модель ожидаемой полезности: разновидности, подходы, результаты и пределы возможностей* // THESIS. – 1994. – №5. – С. 29–80.
8. ШУМОВ В.В. *Классификация и обзор погранометрических моделей* // Труды ИСА РАН. – 2012. – Т. 62.1. – С. 26–39.
9. ШУМОВ В.В. *Модели пограничного сдерживания: монография*. – М.: ЛЕНАНД, 2012. – 200 с.
10. ШУМОВ В.В. *Иерархические и матричные модели пограничной безопасности* // Математическое моделирование. – 2014. – Т. 26. – №3. – С. 137–148.
11. ШУМОВ В.В. *Модель социального влияния и ее применение при анализе пограничной безопасности государства* // Управление большими системами. – 2014. – Вып. 47. – С. 125–166.
12. ШУМОВ В.В. *Модель обоснования направлений сосредоточения усилий пограничной охраны на уровне государства* // Компьютерные исследования и моделирование. – 2019. – Т. 11, №1. – С. 187–196.
13. HADDAL C.C. *Border Security: The Role of the U.S. Border Patrol*. – Washington D.C.: Library of Congress. Congressional Research Service, 2010. – 40 p.
14. WEIN L. M., LIU Y., MOTSKIN A. *Analyzing the Homeland Security of the U. S.-Mexican Border* // Risk Analysis. – 2009. – Vol. 29, No 5. – P. 699–713.

MODELS AND METHODS OF JUSTIFICATION DENSITY OF BOUNDARY FORCES, PROVIDING RELIABLE SECURITY STATE BORDER

Vladislav Shumov, International Research Institute for Advanced Systems, Moscow, Doctor of Sciences, Professor (v.v.shumov@yandex.ru).

Abstract: Based on the methods of management theory and the doctrine of border security, an approach has been formulated to substantiate the density of border forces that ensure reliable protection of state borders. Taking into account the provisions of military science and border statistics data, the required probabilities of detaining violators were assigned for three levels of security (high, medium, minimum). For economically motivated border violators, the justification for the required security density is made using models of border deterrence and the aggregated detention function. Using an example, the methodology for substantiating average security densities is considered. For border violators with non-economic motives (sabotage and terrorist groups), their deterrence is ensured by achieving the required probability of their detention (neutralization). In this case, the required security densities are calculated using the functions of border forces and means, analytical and simulation models. A simplified method for calculating the probabilities of apprehending violators is presented. The numerical values of the model parameters are estimated using a sufficient amount of border statistics data. A promising area of research is the formation of a promising image of border units.

Keywords: border security, state border protection, security levels, aggregate detention function, border deterrence, security density, density of border forces, model.

УДК 519.8

ББК 22.18

DOI: 10.25728/ubs.2024.108.4

Статья представлена к публикации членом редакционной коллегии В.Н. Бурковым.

Поступила в редакцию 02.11.2023.

Опубликована 31.03.2024.

МЕТОДЫ ВЫЧИСЛЕНИЯ СОСТОЯНИЯ ЗАРЯДА ВАНАДИЕВЫХ ПРОТОЧНЫХ АККУМУЛЯТОРОВ: АНАЛИЗ ВЗАИМОСВЯЗЕЙ¹

Парсегов С. Э.²

(Московский физико-технический институт, Сколковский институт науки и технологий, ФГБУН Институт проблем управления им. В.А. Трапезникова РАН, Москва)

Пугач М. А.³

(Сколковский институт науки и технологий, Москва)

Ерофеева В. А.⁴

*(ФГБУН Институт проблем машиноведения РАН,
Санкт-Петербург)*

Ожидается, что в будущем системы накопления энергии будут играть ключевую роль в переходе к энергосистемам с низким уровнем выбросов углекислого газа. Увеличение темпов внедрения возобновляемых источников энергии требует интеграции аккумуляторов, обеспечивающих устойчивость и безопасность энергосистемы и сглаживающих нестабильное поведение возобновляемых источников энергии. Важна также эффективная эксплуатация самих батарей: она продлевает срок их службы, снижая эксплуатационные расходы. Одной из ключевых задач, помогающих управлять аккумуляторными батареями, является мониторинг состояния заряда. В статье проводится анализ существующих явных методов расчета состояния ванадиевых проточных аккумуляторов: метода на основе напряжений холостого хода и метода кулоновского счетчика. Исследуются взаимосвязь между ними и возможность их совместного использования для достижения более надежного и точного отслеживания состояния заряда. В отличие от существующих работ мы получаем аналитическое выражение для общего состояния заряда, учитывающее оба основных компонента батареи, а именно ее стек и резервуары. Анализ их вклада позволяет выявить некоторые недостатки существующих подходов, широко используемых для расчета и мониторинга состояния заряда проточных аккумуляторов.

Ключевые слова: аккумуляторные батареи, вычисление состояния заряда, ванадиевые проточные аккумуляторы.

¹ Исследование выполнено за счет гранта Российского научного фонда (проект № 23-29-00807).

² Сергей Эрнестович Парсегов, к.ф.-м.н., в.н.с. (s.e.parsegov@gmail.com).

³ Михаил Александрович Пугач, Ph.D, с.н.с. (m.pugach@skoltech.ru).

⁴ Виктория Александровна Ерофеева, к.ф.-м.н., с.н.с. (eva@ipme.ru).

1. Введение

Аккумуляторные системы накопления энергии являются одной из ключевых технологий, обеспечивающих переход к углеродно-нейтральным энергосистемам за счет повсеместной интеграции возобновляемых источников энергии (ВИЭ). Как известно, наиболее распространенные ВИЭ, такие как солнечные и ветровые, характеризуются значительными колебаниями и непостоянным режимом работы, что может повлиять на устойчивость и безопасность энергосистемы в целом. В связи с этим особое внимание уделяется сервисам, которые могут предоставлять энергосистеме аккумуляторные системы хранения энергии. Значительное количество современных исследований посвящено сервисам «аккумулятор – сеть», обеспечивающим устойчивость сети и способствующим снижению пиковой нагрузки, см. например, работы [1, 16] и ссылки, приведенные в них.

Ванадиевые проточные аккумуляторы (ВПА) считаются одной из наиболее перспективных технологий для таких приложений благодаря длительному сроку службы, низким эксплуатационным рискам, возможности глубокого разряда и высокой энергоэффективности [8]. Технология ВПА была предложена М. Скиллас-Казакос и ее исследовательской группой в 1980-х годах [14]. С тех пор в ряде исследовательских работ изучались вопросы, связанные с химией, электрохимией и управлением аккумуляторами, например, [2]. Несмотря на растущий интерес к технологии ВПА, она все еще находится на стадии доработки и оптимизации. В этой связи необходимо предлагать новые подходы, направленные на совершенствование существующих решений.

Системы управления аккумуляторами позволяют эффективно эксплуатировать их путем оценки в реальном времени таких важных характеристик, как температура, расход электролита, напряжение на клеммах, состояние заряда (англ. state of charge, SoC) и состояние здоровья (англ. state of health, SoH). Уникальной особенностью ВПА является способность работать практически

без снижения емкости в течение всего периода эксплуатации. Это связано с тем, что для ребалансировки требуется обнаружение дисбаланса электролита в реальном времени, что может быть реализовано с помощью мониторинга состояния заряда (СЗ). Поэтому для ВПА-систем, рассчитанных на длительную эксплуатацию, весьма желательны методы онлайн-мониторинга СЗ, простые в использовании и обладающие высокой точностью.

Методы, пригодные для мониторинга состояния заряда ВПА, можно разделить на две основные группы: (а) оценка на базе моделей, основанных на данных; (б) вычисление в явном виде. В общем случае подход, основанный на моделировании, состоит из трех этапов. Сначала необходимо построить математическую модель, с достаточной точностью описывающую внутреннюю динамику системы ВПА. Затем параметры модели должны быть определены на основе экспериментальных данных с целью получения репрезентативной модели. И наконец, необходимо применить метод оценки состояния, обрабатывающий собранные измерения и определяющий необходимые параметры и состояния батареи.

Вопросам оценивания на базе фильтра Калмана посвящено множество работ. Например, в статье [13] рассмотрен метод оценки СЗ с расширенным фильтром Калмана (ЕКФ) улучшен с точки зрения скорости сходимости, точности и робастности. Далее, в [18] авторы описывают онлайн-идентификацию параметров эквивалентной схемы с помощью рекурсивного метода наименьших квадратов и прогнозирование СЗ на основе «unscented» фильтрации Калмана (UKF). В другом исследовании предлагается комбинация модели эквивалентной схемы с настройкой параметров на основе метода рекурсивного МНК с адаптивным коэффициентом забывания и последующей оценкой СЗ с помощью ЕКФ [5]. В [19] предложенный алгоритм объединяет две оценки СЗ, полученные с помощью различных фильтров (расширенный фильтр Калмана и адаптивный расширенный фильтр Калмана), что, по словам авторов, позволяет достичь более высокой робастности. Некоторые работы посвящены вычислительным вопросам, связанным с моделированием ВПА, например, [7] описывает ал-

горитм оценивания параметров эквивалентной схемы, который позволяет сократить время вычислений, необходимое для оценки модели.

Другое направление исследований связано с применением наблюдателей, работающих в скользящем режиме, см. например, [3, 17]. Этот тип наблюдателей минимизирует ошибку моделирования, а также влияние неопределенностей и сходится за конечное время. Следует отметить, что такие способы также увеличивают и вычислительные затраты для получения оценок. В [19] авторы описывают онлайн-мониторинг СЗ и потерь мощности на основе авторегрессионного моделирования экзогенных факторов и наблюдателя с использованием H_∞ -теории. В [4] рассматривается адаптивный нелинейный наблюдатель, учитывающий эффект кроссовера в модели.

В отличие от методов с использованием математических моделей динамики системы ВПА, кулоновский счетчик отличается простотой реализации и не требует дополнительной измерительной аппаратуры. Кроме того, этот метод не требует настройки параметров модели, которую необходимо должным образом отлаживать в режиме реального времени. Тем не менее существенной проблемой данного метода является накопление погрешностей, возникающих в результате измерений тока, а также кроссовера и нежелательных побочных реакций [10].

В связи с этим такой подход обычно не используется в качестве единственного метода для длительного мониторинга СЗ, поэтому для преодоления этого ограничения зачастую требуются дополнительные измерения. Наиболее распространенным подходом является включение измерений напряжения холостого хода (НХХ) (англ. open-circuit voltage, OCV). Однако этот метод также имеет недостатки, поскольку значения НХХ постоянно дрейфуют из-за кроссовера, что приводит к накоплению дополнительной погрешности расчета СЗ при длительной работе батареи. Кроме того, для измерения НХХ требуется специальное оборудование, которое необходимо бережно эксплуатировать, что неминуемо приводит к увеличению затрат [12].

Следует отметить, что состояние заряда ВПА напрямую зависит от концентрации ионов ванадия, участвующих в электрохимических реакциях в ячейках и содержащихся в резервуарах (баках). Многие существующие подходы к мониторингу определяют СЗ с учетом концентрации ионов исключительно в резервуарах. Поэтому при увеличении соотношения объемов бака и стека в расчет СЗ вносятся дополнительные неточности.

В работе проводится критический анализ способов прямого мониторинга СЗ ВПА, таких как НХХ-подход и метод кулоновского счетчика, а также изучается их взаимосвязь. Начиная с модели ВПА восьмого порядка, формулируется ряд предположений, что сводит исходную модель к модели второго порядка. Учитывая свойства редуцированной модели, выявляется зависимость, связывающая напряжения холостого хода и измерения тока нагрузки. Полученное соотношение может помочь в разработке более надежных средств оценивания и мониторинга СЗ, поскольку оно учитывает возможность восстановления недостающих измерений либо получение более точных оценок при зашумленных измерениях/неисправных измерительных приборах.

Остальная часть работы устроена следующим образом. В разделе 2 приводится общее описание технологии ВПА, затем формулируется задача мониторинга СЗ и приводится динамическая модель ВПА в пространстве состояний. В разделе 3 представлен анализ модели ВПА второго порядка и показана связь между различными стратегиями измерения СЗ. Раздел 4 раскрывает некоторые полезные свойства полученных теоретических результатов, дополненных имитационным моделированием. Раздел 5 завершает работу.

2. Постановка задачи

2.1. ОСНОВНЫЕ ПОНЯТИЯ ВПА

Рассмотрим структуру системы ВПА, изображенную на рис. 1. Как правило, система состоит из двух баков, заполненных по отдельности отрицательным и положительным электролитами, электрохимического стека, вырабатывающего электроэнер-

гию, и гидравлической подсистемы, обеспечивающей переток электролита между основными компонентами. На входе и/или выходе системы могут быть установлены дополнительные измерительные приборы, например, ячейки для измерения напряжений холостого хода. Электроэнергия вырабатывается в результате электрохимической реакции между ионами ванадия. В «отрицательном» резервуаре находятся ионы V^{2+} и V^{3+} , в «положительном» – ионы V^{4+} и V^{5+} в виде оксидов VO_2^+ и VO^{2+} соответственно:



Количество энергии, вырабатываемой системой, напрямую связано как с объемом электролита, хранящегося в резервуарах, так и с концентрациями ионов ванадия. Пусть V_{tk} обозначает объем каждого резервуара, а $c_i^{tk} > 0, i \in \{2, \dots, 5\}$ – концентрация i -го иона в резервуаре. Стек состоит из n_c ячеек, каждая из которых имеет соответствующий объем, обозначаемый через V_c . Концентрация i -го иона в стеке обозначается $c_i^{st} > 0, i \in \{2, \dots, 5\}$.

Таким образом, общий вектор состояния ВПА-батареи имеет вид

$$\begin{aligned} \xi &= [\xi_1, \xi_2, \dots, \xi_8]^T \in \mathbb{R}^8, \\ \xi_1 &= c_2^{tk}, \quad \xi_2 = c_3^{tk}, \quad \xi_3 = c_4^{tk}, \quad \xi_4 = c_5^{tk}, \\ \xi_5 &= c_2^{st}, \quad \xi_6 = c_3^{st}, \quad \xi_7 = c_4^{st}, \quad \xi_8 = c_5^{st}. \end{aligned}$$

Если система не подключена к нагрузке (источнику) или оснащена специальными измерительными ячейками холостого хода, то доступны измерения напряжений холостого хода. Выходной сигнал системы y имеет вид

$$(1) \quad y = \begin{bmatrix} y_{in} \\ y_{out} \end{bmatrix} = h(\xi).$$

Здесь функция $y = h : \mathbb{R}^8 \rightarrow \mathbb{R}^2$ является векторной формой уравнения Нернста

$$(2) \quad h(\xi) = \begin{bmatrix} k_0 + k_1 \ln \frac{\xi_1 \xi_4}{\xi_2 \xi_3} \\ k_0 + k_1 \ln \frac{\xi_5 \xi_8}{\xi_6 \xi_7} \end{bmatrix}$$

с известными константами $k_0, k_1 > 0$: k_0 представляет собой стандартный потенциал, а $k_1 = \frac{RT}{F}$, где $R = 8,314 \frac{\text{Дж}}{\text{К} \cdot \text{моль}}$ – газовая постоянная, $F = 96485,332 \frac{\text{Кл}}{\text{моль}}$ – постоянная Фарадея, а $T = 298 \text{ К}$ – температура. Выходы y_{in} и y_{out} соответствуют напряжениям холостого хода на входе (OCV_{in}) и выходе (OCV_{out}) ячейки (стека) соответственно (см. рис. 1).

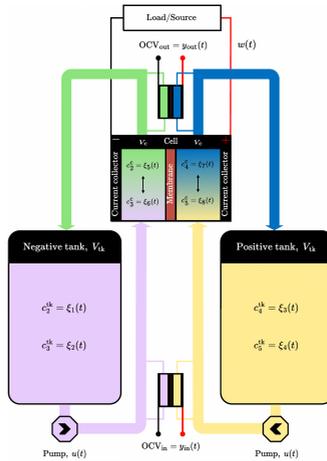


Рис. 1. Структура ВПА, оснащенного входными/выходными ячейками для измерения напряжений холостого хода

2.2. МЕТОД КУЛОНОВСКОГО СЧЕТЧИКА ДЛЯ ОПРЕДЕЛЕНИЯ СОСТОЯНИЯ ЗАРЯДА

Состояние заряда аккумулятора обычно выражается как

$$SoC = \frac{Q_{act}}{Q_{max}},$$

где Q_{act} – имеющееся в данный момент количество электрического заряда, Q_{max} – максимальный электрический заряд батареи.

Среди различных известных методов оценки СЗ наиболее простым и понятным является кулоновский счетчик. Он основан

на учете перенесенного заряда

$$\Delta Q = \int_{t_0}^{t_1} w(\tau) d\tau,$$

рассчитываемого как интеграл, взятый на интервале времени $[t_0; t_1]$ от измеряемого внешнего электрического тока $w(t)$, который положителен при заряде и отрицателен при разряде.

Таким образом, оценка СЗ на основе кулоновского счетчика может быть описана следующим образом:

$$(3) \quad SoC = SoC_0 + \frac{\Delta Q}{Q_{max}},$$

где SoC_0 – известное базовое значение.

Недостатком этого метода является накопление погрешности на интеграторе из-за неточного измерения тока. Например, если измерение тока имеет систематическую погрешность. В этом случае будет иметь место линейный рост ошибки. Этот недостаток метода широко известен, поэтому в таких приложениях точность амперметра является критичной. Кроме того, этот метод может давать большую погрешность при усилении побочных реакций, на которые уходит часть энтропийной энергии, но это не отражается на кулоновском счетчике, который дает только интегральное значение. В связи с этим данный метод не используется на больших временных интервалах и часто дополняется другими методами оценки состояния заряда. Обсуждение и анализ этого и других недостатков данного метода можно найти в [9].

В данной работе показано, что возможное накопление ошибок – не единственная особенность метода кулоновского счетчика при его использовании в ВПА. Учитывая специфику строения и работы проточного аккумулятора, накопление (расходование) заряда происходит в различных элементах накопителя. На основе анализа модели динамики такой системы мы показываем, как общее состояние заряда связано с зарядами в стеке и баке. Для анализа изменения состояния заряда рассмотрим сначала типовую модель с сосредоточенными параметрами, описывающую динамику концентраций в ВПА.

2.3. ДИНАМИКА ВПА

В общем случае динамика состояния аккумулятора ξ может быть описана системой дифференциальных уравнений восьмого порядка

$$(4) \quad \dot{\xi} = u\mathcal{A}\xi + \mathcal{J}\xi + \mathcal{D}w$$

с двумя алгебраическими ограничениями:

$$(5) \quad \xi_1 + \xi_2 + \xi_3 + \xi_4 = 2c_t,$$

$$(6) \quad \xi_5 + \xi_6 + \xi_7 + \xi_8 = 2c_t,$$

где $u \in \mathbb{R}$ обозначает скорость потока электролита, $w \in \mathbb{R}$ – измеряемое возмущение (ток), а c_t – постоянная общая концентрация, которая известна. Соответствующие матрицы имеют вид

$$(7) \quad \mathcal{A} = \begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix} \otimes \mathbf{I}_4,$$

$$(8) \quad \mathcal{J} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} -J_2 & 0 & -J_4 & -2J_5 \\ 0 & -J_3 & 2J_4 & 3J_5 \\ 3J_2 & 2J_3 & -J_4 & 0 \\ -2J_2 & -J_3 & 0 & -J_5 \end{bmatrix},$$

$$(9) \quad \mathcal{D} = [0 \ 0 \ 0 \ 0 \ d_5 \ d_6 \ d_7 \ d_8]^\top,$$

где $d_5 = d_8 = -d_6 = -d_7 = d$.

Здесь $d = \frac{1}{FV_c}$, $\alpha = \frac{1}{V_{tk}}$, $\beta = \frac{1}{n_c V_c}$.

Матрица \mathcal{J} отражает динамику кроссовера. Каждая константа J_i , $i \in \{2, \dots, 5\}$, представляет собой коэффициент диффузии i -го иона ванадия. Как правило, эти коэффициенты имеют достаточно малый порядок, например, 10^{-11} , и ими можно пренебречь на интервале от нескольких циклов до десятков циклов заряда/разряда.

3. Основной результат

3.1. МОДЕЛЬ ВТОРОГО ПОРЯДКА

При пренебрежении вторым членом (4) можно видеть, что система становится симметричной. Это означает, что концентрации положительной и отрицательной сторон аккумулятора имеют одинаковую динамику.

Алгебраические ограничения в данном случае примут вид

$$(10) \quad \begin{aligned} \xi_1 + \xi_2 = \xi_3 + \xi_4 = \xi_5 + \xi_6 = \xi_7 + \xi_8 = c_t, \\ \xi_1 = \xi_4, \quad \xi_5 = \xi_8. \end{aligned}$$

Таким образом, переписав динамику системы с учетом ее алгебраической части (10), получим:

$$(11) \quad \dot{x} = uAx + Dw,$$

где $x_1 = \xi_1, x_2 = \xi_5$,

$$A = \begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix}, \quad D = [0 \quad d]^\top.$$

Отметим, что в простейшем случае, когда скорость потока u постоянна, а внешнее возмущение $w \equiv 0$, система (11) имеет обычную консенсусную (усредняющую) динамику с матрицей Лапласа $L = -uA$: асимптотически концентрации в стеке и резервуаре стремятся достичь одного значения.

Выход системы (11) примет вид

$$(12) \quad \begin{bmatrix} y_{in} \\ y_{out} \end{bmatrix} = \begin{bmatrix} k_0 + 2k_1 \ln \frac{x_1}{c_t - x_1} \\ k_0 + 2k_1 \ln \frac{x_2}{c_t - x_2} \end{bmatrix}.$$

Очевидно, что с помощью следующих выражений можно в явном виде вычислить состояния системы:

$$(13) \quad x_1 = \frac{c_t e^{\frac{y_{in} - k_0}{2k_1}}}{1 + e^{\frac{y_{in} - k_0}{2k_1}}}, \quad x_2 = \frac{c_t e^{\frac{y_{out} - k_0}{2k_1}}}{1 + e^{\frac{y_{out} - k_0}{2k_1}}}.$$

Из (13) становится понятно, что, имея измерения обоих НХХ, можно легко найти концентрации, причем динамическую часть модели рассматривать не нужно.

Далее будет показана связь метода кулоновского счетчика с изменением концентраций.

3.2. ОБЩЕЕ СОСТОЯНИЕ ЗАРЯДА ВПА

Предположим, что измерение тока доступно и точно совпадает с реальным сигналом $w(t)$. Тогда для батареи с динамикой концентрации (11) и измеряемыми выходами (12) справедливо следующее утверждение.

Утверждение 1. Пусть в начальный момент времени $t = 0$ значения концентраций ионов равны $x_1(0), x_2(0)$. Тогда справедливо следующее равенство:

$$(14) \quad V_{tk}x_1 + n_c V_c x_2 = V_{tk}x_1(0) + n_c V_c x_2(0) + \frac{n_c}{F} \int_0^t w(\tau) d\tau.$$

Доказательство. Система (11) может быть переписана с учетом свойств матрицы A . Собственные значения A и соответствующие им собственные векторы имеют вид

$$(15) \quad \lambda_1 = 0, \quad \nu_1 = [1, 1]^\top,$$

$$(16) \quad \lambda_2 = -(\alpha + \beta), \quad \nu_2 = \left[1, -\frac{\beta}{\alpha}\right]^\top.$$

Введем новые переменные

$$(17) \quad \tilde{x} = S^{-1}x,$$

где S – новый базис из собственных векторов

$$(18) \quad S = \begin{bmatrix} 1 & 1 \\ 1 & -\frac{\beta}{\alpha} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -\frac{V_{tk}}{n_c V_c} \end{bmatrix}.$$

Тогда в новых координатах динамика системы примет вид

$$(19) \quad \dot{\tilde{x}} = u\Lambda\tilde{x} + \tilde{D}w,$$

где

$$(20) \quad \Lambda = \begin{bmatrix} 0 & 0 \\ 0 & -(\alpha + \beta) \end{bmatrix},$$

$$(21) \quad \tilde{D} = [\tilde{d} \quad -\tilde{d}]^\top$$

и

$$\tilde{d} = \frac{\alpha d}{\alpha + \beta} = \frac{n_c}{F(n_c V_c + V_{tk})}.$$

Проанализируем проведенную замену координат. Подставляя (18) в (17), получим

$$(22) \quad \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{bmatrix} = \frac{1}{V_{tk} + n_c V_c} \begin{bmatrix} V_{tk} & n_c V_c \\ n_c V_c & -n_c V_c \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Динамика новых переменных имеет вид

$$(23) \quad \begin{bmatrix} \dot{\tilde{x}}_1 \\ \dot{\tilde{x}}_2 \end{bmatrix} = u \begin{bmatrix} 0 & 0 \\ 0 & -(\frac{1}{V_{tk}} + \frac{1}{n_c V_c}) \end{bmatrix} \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{bmatrix} + \\ + \begin{bmatrix} \frac{n_c}{F(V_{tk} + n_c V_c)} \\ -\frac{n_c}{F(V_{tk} + n_c V_c)} \end{bmatrix} w.$$

Легко видеть, что полученная система содержит две независимые части.

Переменная \tilde{x}_1 описывается динамикой простого интегратора, зависящей от входного сигнала w и начального условия $\tilde{x}_1(0)$:

$$(24) \quad \tilde{x}_1 = \tilde{x}_1(0) + \frac{n_c}{F(V_{tk} + n_c V_c)} \int_0^t w(\tau) d\tau.$$

Перепишав последнее уравнение в исходных координатах, получаем (14). На этом доказательство завершено.

Следствие 1. *Полученный результат отражает следующий факт: общее СЗ аккумулятора является выпуклой комбинацией состояний заряда баков SoC_{tk} и стека SoC_{st} . При этом выполняется следующее равенство:*

$$(25) \quad \begin{aligned} SoC &= k_{tk} SoC_{tk}(t) + k_{st} SoC_{st}(t) = \\ &= k_{tk} SoC_{tk}(0) + k_{st} SoC_{st}(0) + \frac{n_c \int_0^t w(\tau) d\tau}{F_{ct}(V_{tk} + n_c V_c)}, \end{aligned}$$

где веса имеют вид

$$k_{tk} = \frac{V_{tk}}{V_{tk} + n_c V_c}, \quad k_{st} = \frac{n_c V_c}{V_{tk} + n_c V_c}$$

и удовлетворяют $k_{tk} + k_{st} = 1$.

Замечание 1. При наличии паразитной динамики, разбалансировки батареи, деградации, потерь различной природы, а также неточности измерений, рассмотренные уравнения, равенства и преобразования перестают быть верными. Тем не менее подобный анализ системы (11)–(12) проливает свет на взаимосвязь между сигналами, связанными с ее динамикой.

4. Анализ результатов и моделирование

В данном разделе излагаются выводы, сделанные на основе (25), анализируются возможные приложения к оценке состояния и приводятся некоторые результаты моделирования.

Влияние соотношения объемов стека и бака.

Сначала продемонстрируем важность соотношения объемов стека и бака. Оба состояния заряда, а именно SoC_{tk} и SoC_{st} , вносят свой вклад в общее состояние, и ни одним из них нельзя пренебрегать. Если объем стека не очень мал по сравнению с объемом резервуара, то его вклад в общий заряд может быть значительным. Например, такая ситуация характерна для лабораторных стендов. В таблице 1 приводятся параметры ВПА некоторых реальных экспериментальных накопителей. Здесь «Сколтех-1» и «Сколтех-2» обозначают батареи малого и среднего размера, установленные в институте Сколтех, Москва. Накопитель «Padova» установлен в университете Падуи, а «UNSW» – в университете Нового Южного Уэльса соответственно. Параметр $\mu = \frac{n_c V_c}{V_{tk}} \cdot 100\%$ обозначает процентное соотношение объемов стека и бака.

Таблица 1. Параметры экспериментальных установок ВПА

Установка ВПА	Параметры				Ссылки
	$V_c, \text{ м}^3$	$V_{tk}, \text{ м}^3$	n_c	$\mu, \%$	
Сколтех-1	7,50e-06	4,00e-04	10	18,75	[11]
Сколтех-2	1,84e-04	1,00e-01	40	7,36	[11]
Padova	3,42e-04	5,50e-01	40	2,49	[6]
UNSW	4,50e-04	0,2	40	9	[15]

Влияние соотношения объемов стека и бака можно проанализировать с помощью коэффициента k_{st} , определяемого как $k_{st}(\mu) = \frac{\mu}{\mu+100}$.

Изменение $k_{st}(\mu)$ в зависимости от μ показано на рис. 2.

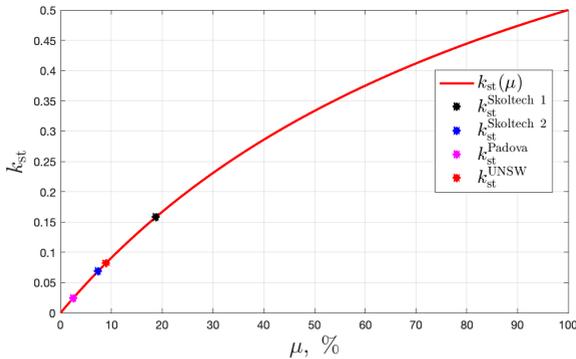


Рис. 2. Вес k_{st} вклада состояния заряда в стеке SoC_{st} как функция отношения объемов стека и бака μ

Звездочки на кривой соответствуют лабораторным батареям из таблицы. Как видно из рисунка, стеки трех установок дают вклад до 7–15 % в общий заряд батареи. В некоторых работах, например, см. [4], не делается различия между общим состоянием заряда SoC и состоянием заряда стека SoC_{tk} . Как будет показано в примере, такая интерпретация состояния заряда может привести к некорректной работе аккумулятора, см. рис. 4.

Отсутствующие измерения.

Состояние заряда, найденное с помощью метода кулоновского счетчика, может служить заменой отсутствующему измерению напряжения холостого хода. Очевидно, что одна концентрация (или состояние заряда) может быть выражена через другую, как следует из (14), (25). Если ток измеряется с достаточно высокой точностью, то довольно легко вычислить недостающую концентрацию (или состояние заряда). Это можно проиллюстрировать на следующем примере. В численном эксперименте аккумулятор проходит цикл заряда от 10% до 75%. Кроме того, имеется десятисекундный интервал, на котором батарея резко переходит в режим разряда. Именно на этом интервале удобно оценить разницу между различными состояниями заряда системы.

Пусть параметры системы составляют $u = 3 \cdot 10^{-5} \text{ м}^3/\text{с}$,

$V_{tk} = 4 \cdot 10^{-4} \text{ м}^3$, $V_c = 3,6 \cdot 10^{-6} \text{ м}^3$, $n_c = 10$, $SoC_{tk}(0) = SoC_{st}(0) = 0,1$. Измерения НХХ и тока содержат нормально распределенную случайную ошибку с нулевым средним. Тогда, согласно (25), мы имеем два способа вычисления общего состояния заряда:

$$(26) \quad SoC_1 = k_{tk}SoC_{tk}(t) + k_{st}SoC_{st}(t),$$

$$(27) \quad SoC_2 = k_{tk}SoC_{tk}(0) + k_{st}SoC_{st}(0) + \frac{n_c \int_0^t w(\tau) d\tau}{Fc_t(V_{tk} + n_c V_c)}.$$

Как видно из графиков состояний заряда, изображенных на рис. 3, они практически совпадают.

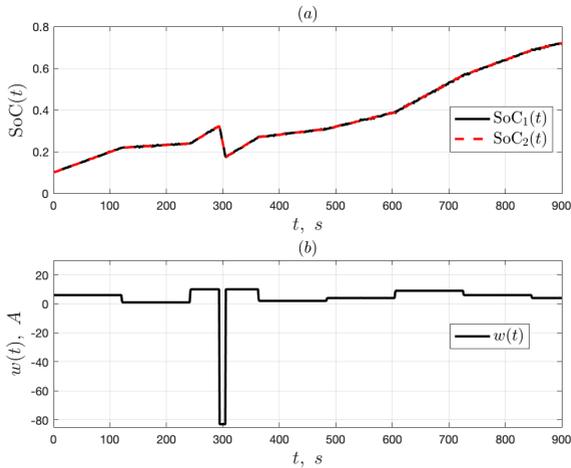


Рис. 3. Изменение состояний заряда, вычисленных при помощи (26), (27) (а), а также измеренный ток (б)

Резкий переход батареи из режима заряда в режим разряда с высокой нагрузкой свидетельствует о значительном изменении состояния заряда в стеке, батарея по сути близка к выходу из строя, см. рис.4. В то же время состояние заряда в баке выше общего, более «правильного» состояния заряда.

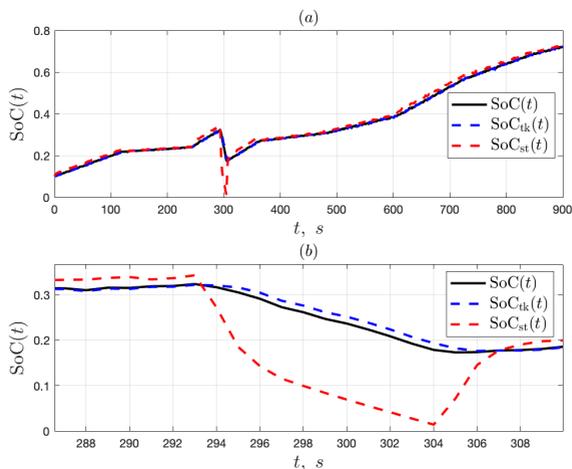


Рис. 4. Изменение состояний заряда SoC , SoC_{tk} , SoC_{st} (a); увеличенный фрагмент, соответствующий резкому скачку тока нагрузки (b)

Более робастные способы оценки состояния заряда.

Точность оценки состояния заряда может быть повышена за счет увеличения избыточности измерительной системы. Если доступны как измерения напряжений холостого хода, так и тока нагрузки, то избыточность налицо. В то же время они по-прежнему содержат шумы: избыточность измерений может быть использована для получения более точных оценок концентраций ионов при использовании статических оценивающих устройств, а также наблюдателей/фильтров Калмана. Это может повысить точность получаемых оценок.

5. Заключение

Определение состояния заряда является важной задачей для обеспечения правильной эксплуатации ванадиевых проточных аккумуляторов (ВПА), так как оно указывает на способность к накоплению энергии. Из-за технических различий между ВПА-системами и обычными батареями широко используемые методы

определения/оценки состояния заряда могут давать большие погрешности. Более того, эти погрешности могут накапливаться в процессе длительной эксплуатации и должны быть обязательно учтены.

В данной работе проведен подробный анализ существующих методов явного определения состояния заряда ВПА. В результате анализа выявлена связь между традиционным подходом, использующего технику кулоновского счетчика, и вычислением состояний заряда в стеке и резервуарах с помощью измерений напряжений холостого хода как на входе в стек, так и на его выходе. Результаты исследования показали, что:

– при отсутствии потерь и паразитной динамики общее состояние заряда представляет собой взвешенную сумму состояний заряда баков и стека. Весовыми коэффициентами в формуле являются относительные объемы. Анализ вклада стека и бака в общее состояние заряда показывает, что величиной накопленного заряда в стеке пренебрегать нельзя;

– связь концентраций (или состояний заряда) в стеке и баке с зарядом, оцененным методом кулоновского счета, позволяет: 1) выразить одни состояния в терминах других в случае отсутствия измерений; 2) увеличить избыточность и, следовательно, точность оценок при измерении как тока, так и напряжений холостого хода.

Дальнейшие планы включают в себя а) экспериментальную проверку полученных результатов, б) анализ применения метода к более сложной динамике с кроссовером и в) применение кулоновского счетчика в сочетании с измерениями напряжения к задаче синтеза наблюдателей состояния.

Литература

1. BARSALI S., GIGLIOLI R., LUTZEMBERGER G. et al. *Optimised operation of storage systems integrated with mv photovoltaic plants, considering the impact on the battery lifetime* // Journal of Energy Storage. – 2017. – Vol. 12. – P. 178–185.

2. BARELLI L., BIDINI G., CIUPAGEANU D.-A. et al. *Electrical performance analysis of an innovative vanadium redox flow battery stack for enhanced power density applications* // IEEE Madrid PowerTech. – 2021. – P. 1–6.
3. CLEMENTE A., MONTIEL M., BARRERAS F. et al. *Vanadium redox flow battery state of charge estimation using a concentration model and a sliding mode observer* // IEEE Access. – 2021. – Vol. 9. – P. 72368–72376.
4. CLEMENTE A., CECILIA A., COSTA-CASTELLO R. *Soc and diffusion rate estimation in redox flow batteries: An i&i- based high-gain observer approach* // European Control Conference (ECC). – 2021. – P. 1640–1644.
5. DAWEI Q., ZIXUAN L., FAN Y. et al. *State of charge estimation for the vanadium redox flow battery based on extended kalman filter using modified parameter identification* // Energy Sources, Part A: Recovery, Utilization, and Environmental Effects. – 2022. – Vol. 44, No. 4. – P. 9747–9763.
6. GUARNIERI M., TROVÒ A., MARINI G. et al. *High current polarization tests on a 9 kw vanadium redox flow battery* // Journal of Power Sources. – 2019. – Vol. 431. – P. 239–249.
7. KHAKI B., DAS P. *Fast and simplified algorithms for soc and soh estimation of vanadium redox flow batteries* // IEEE Green Technologies Conference (GreenTech). – 2021. – P. 494–501.
8. LUCAS A., CHONDROGIANNIS S. *Smart grid energy storage controller for frequency regulation and peak shaving, using a vanadium redox flow battery* // Int. Journal of Electrical Power & Energy Systems. – 2016. – Vol. 80. – P. 26–36.
9. MOVASSAGH K., RAIHAN A., BALASINGAM B. et al. *A critical look at coulomb counting approach for state of charge estimation in batteries* // Energies. – 2021. – Vol. 14, No. 14. – P. 40–74.
10. MISYRIS G.S., TENGNER T., MARINOPOULOS A.G. et al. *Battery energy storage systems modeling for online applications* // IEEE Manchester PowerTech. – 2017. – P. 1–6.

11. PARSEGOV S., PUGACH M., POLYAKOV A. et al. *Analysis of flow factor control strategy in vanadium redox flow batteries* // IFAC-PapersOnLine. – 2022. – Vol. 55, No. 9. – P. 187–192.
12. PULESTON T., CLEMENTE A., COSTA-CASTELLO R. et al. *Modelling and estimation of vanadium redox flow batteries: A review* // Batteries. – 2022. – Vol. 8, No. 9. – P. 1–21.
13. QIU Y., LI X., CHEN W. et al. *State of charge estimation of vanadium redox battery based on improved extended kalman filter* // ISA Trans. – 2019. – Vol. 94. – P. 326–337.
14. SKYLLAS-KAZACOS M., KASHERMAN D., HONG D. et al. *Characteristics and performance of 1 kW UNSW vanadium redox battery* // Journal of Power Sources. – 1991. – Vol. 35, No. 4. – P. 399–404.
15. TANG A., BAO J., SKYLLAS-KAZACOS M. *Studies on pressure losses and flow rate optimization in vanadium redox flow battery* // Journal of power sources. – 2014. – Vol. 248. – P. 154–162.
16. UDDIN M. , ROMLIE M. F., ABDULLAH M. F. et al. *A review on peak load shaving strategies* // Renewable and Sustainable Energy Reviews. – 2018. – Vol. 82. – P. 3323–3332.
17. XIONG B., ZHAO J. , SU Y. et al. *State of charge estimation of vanadium redox flow battery based on sliding mode observer and dynamic model including capacity fading factor* // IEEE Trans. on Sustainable Energy. – 2017. – Vol. 8, No. 4. – P. 1658–1667.
18. ZHENG C., TIAN X., NIE G. et al. *State of power and state of charge estimation of vanadium redox flow battery based on an online equivalent circuit model* // IEEE 18th Int. Conf. on Industrial Informatics (INDIN-2020). – 2020. – Vol. 1. – P. 633–638.
19. ZHAO X., KIM K., JUNG S. *State-of-charge estimation using data fusion for vanadium redox flow battery* // Journal of Energy Storage. – 2022. – Vol. 52. – P. 104852.

METHODS FOR CALCULATING THE STATE OF CHARGE OF VANADIUM REDOX FLOW BATTERIES: AN ANALYSIS OF THE RELATIONSHIPS

Sergei Parsegov, Moscow Institute of Physics and Technology, Skolkovo Institute of Science and Technology, Institute of Control Sciences of RAS, Moscow, Cand.Sc., leading research scientist (s.e.parsegov@gmail.com).

Mikhail Pugach, Skolkovo Institute of Science and Technology, Moscow, Ph.D, senior research scientist (m.pugach@skoltech.ru).

Victoria Erofeeva, Institute for Problems in Mechanical Engineering of RAS, Saint Petersburg, Cand.Sc., senior research scientist (eva@ipme.ru).

Abstract: In the future, energy storage systems are expected to play a key role in the transition to low-carbon energy systems. Increasing renewable energy deployment rates require the integration of batteries that ensure the stability and security of the power system and mitigate the unstable behavior of renewables. Efficient operation of the batteries themselves is also important: it extends their life, reducing operating costs. Monitoring of the state of charge is one of the key tasks to help battery operation. This paper studies the existing explicit methods for calculation of the state of charge of vanadium flow batteries: the open-circuit voltage-based method and the Coulomb counter method. The relationship between them and the possibility of using them together to achieve more reliable and accurate charge state monitoring is investigated. In contrast to existing works, we derive an analytical expression for the overall state of charge that takes into account both main components of the battery, namely its stack and tanks. Analyzing their contributions reveals some shortcomings of existing approaches widely used to calculate and monitor the state of charge of flow batteries.

Keywords: battery energy storages, state of charge calculation, vanadium redox flow batteries.

УДК 681.5

ББК 31.251

DOI: 10.25728/ubs.2024.108.5

Статья представлена к публикации членом редакционной коллегии П.С. Щербаковым.

Поступила в редакцию 07.12.2023.

Дата опубликования 31.03.2024.

АЛГОРИТМ АНАЛИЗА МУЛЬТИСПЕКТРАЛЬНЫХ АЭРОФОТОСНИМКОВ ДЛЯ ИДЕНТИФИКАЦИИ ЗАГРЯЗНЕНИЙ ВОДОЕМОВ С ИСПОЛЬЗОВАНИЕМ АНАЛИТИЧЕСКИХ МЕТОДОВ И НЕЙРОСЕТЕВЫХ ПОДХОДОВ¹

Диане С. А. К.², Вытовтов К. А.³, Барабанова Е. А.⁴
(ФГБУН Институт проблем управления
им. В.А. Трапезникова РАН, Москва)

Статья посвящена разработке алгоритмов анализа загрязнений на поверхности водоемов по визуальной информации, полученной с использованием мультиспектральной камеры, закрепленной на корпусе БПЛА. Предложена структура алгоритмического комплекса для анализа мультиспектральных аэрофотоснимков. В рамках развиваемого подхода каждое из анализируемых изображений проходит процедуру предобработки, обеспечивающую выравнивание и совмещение его спектральных каналов в единый многомерный растр. Разработанный аналитический алгоритм позволяет осуществлять обработку и свертку каналов мультиспектрального изображения с применением трех математических операторов: полосовой фильтрации, изменения контраста и изменения яркости. При этом выбор параметров для выделения загрязнений на поверхности водоемов основан на предварительном этапе, связанном с максимизацией показателя превышения контраста для эталонной области. Предложенный нейросетевой алгоритм анализа загрязнений основывается на применении метода скользящего окна в сочетании со сверточной архитектурой нейросетевого классификатора для анализа фрагментов изображения, расположенных по прямоугольной сетке. На основе программной реализации предложенных алгоритмов и графического интерфейса пользователя проведены экспериментальные исследования, которые подтвердили эффективность каждого из рассмотренных подходов и показали, что нейросетевой алгоритм выигрывает в точности, а аналитический подход легче поддается интерпретации с точки зрения эксперта.

Ключевые слова: аэрофотоснимок, аналитический метод, нейросетевой подход.

¹ Исследование выполнено при финансовой поддержке Российского научного фонда №23-29-00795, <https://rscf.ru/project/23-29-00795>.

² Секу Абдель Кадер Диане, к.т.н., с.н.с. (diane1990@yandex.ru).

³ Константин Анатольевич Вытовтов, д.т.н., в.н.с. (vytovtov_konstan@mail.ru).

⁴ Елизавета Александровна Барабанова, д.т.н., в.н.с. (elizavetaalex@yandex.ru).

1. Введение

Экологический мониторинг местности, лесных массивов, посевов сельхозкультур, анализ загрязнений водоемов по материалам аэрофотосъемки являются актуальными научными и прикладными задачами. Экосистема отдельно взятого участка природной среды, рассматриваемая как большая система (с позиций теории автоматического управления) нуждается в контуре обратной связи для поддержания ее состояния в пределах допустимых диапазонов. Данный контур может быть представлен робототехнической системой, оснащенной цифровой фото- или видеокамерой, в сочетании с системой комплексной обработки информации и поддержки принятия решений.

На сегодняшний день существует ряд подходов к анализу поверхности Земли и водоемов по изображениям, полученным с беспилотных летательных аппаратов и спутников. При этом в ряде случаев использование БПЛА представляется наиболее целесообразным решением с точки зрения экономических затрат, точности определения загрязнений и оперативности получения результатов [13].

Методология анализа аэрофотоснимков, получаемых БПЛА по результатам мониторинга местности, основывается на пяти взаимодополняющих подходах: метод анализа коэффициентов спектральной яркости [11, 12], метод гистограммного анализа сопоставления изображений [2, 5], методы сопоставления коэффициентов вейвлет преобразований изображений [4, 8, 9], методы сопоставления дескрипторов ключевых точек изображений [6], нейросетевые методы визуальной детекции объектов [3, 7, 14].

Так, в работе [11] исследованы возможности применения коэффициентов спектральной яркости для анализа аэрофотоснимков земной поверхности. Получены коэффициенты спектральной яркости для различных видов поверхности и различных углов места. В [12] представлен метод идентификации загрязнений водной поверхности с использованием коэффициентов спектральной яркости. В работе также приведены результаты обработки мультиспектральных изображений, полученных

с БПЛА. Достоинством этой группы методов является возможность нахождения и анализа коэффициентов спектральной яркости в явной аналитической форме. Недостатком является сильная зависимость результатов мониторинга от погодных условий.

Ряд исследований посвящен анализу аэрофотоснимков и спутниковых фотографий на основе сопоставления гистограмм участков изображений. В частности, в работе [2] представлен пример использования квадрокоптера в сельскохозяйственной задаче визуального обнаружения плодов растений. В исследовании [5] также применен гистограммный подход, но для анализа более масштабных участков местности по спутниковым снимкам на предмет изменений в текстуре поверхности, произошедших за определенный интервал времени.

В работах [4, 8] применены частотные методы анализа изображений. Так, в работе [4] для оценки площади рисового поля предложено использовать индекс растительности, основанный на расчете энергии вейвлет-преобразования в пределах рассматриваемого участка аэрофотоснимка. В работе [8] рассмотрены методы повышения четкости аэрофотоснимков за счет применения линейного преобразования к их вейвлет-спектрограммам.

Авторы работ [6, 9] перечисляют ряд алгоритмов, основанных на выделении ключевых точек на изображении и формировании визуальных дескрипторов. Исследования в данном направлении позволяют, с одной стороны, осуществить распознавание целевых участков местности и более надежную локализацию БПЛА, а с другой – производить совмещение аэрофотоснимков, полученных с нескольких ракурсов для формирования единой карты наблюдаемого участка Земли.

Улучшенный подход к выравниванию аэрофотоснимков представлен в работе [14]. Авторы предлагают использовать нейронную сеть для многомасштабной оценки взаимных смещений аэрофотоснимков.

Перспективность нейросетевого анализа аэрофотоснимков подтверждается и в исследовании [7], где данная технология используется для фенотипирования пшеницы с применением

беспилотного квадрокоптера, оснащенного гиперспектральной камерой.

Следует отметить, что в ряде случаев анализ аэрофотоснимков может выполняться и вручную на основе экспертного разбиения аэрофотоснимков на фрагменты, подсчета объектов целевого интереса и вычисления обобщенных характеристик местности, таких как, например, индексы растительности и городской застройки [3].

Вне зависимости от выбранного метода анализа изображений наилучшая точность визуальной детекции целевых объектов на аэрофотоснимках достигается при использовании мультиспектрального формата регистрации и хранения соответствующей растровой информации.

2. Постановка задачи и исходные данные

Целью данной работы является разработка алгоритмического обеспечения и программного комплекса для автоматизированного анализа снимков мультиспектральной камеры, полученных с БПЛА, в задаче идентификации загрязнений водных поверхностей. К числу таких загрязнений относятся нефтяные и масляные пятна, а также органические выбросы.

Основным требованием к программно-алгоритмическому обеспечению является возможность загрузки подготовленных изображений, их коррекции и обработки с применением аналитических и нейросетевых методов. На выходе программного комплекса должны формироваться растровые карты, отражающие загрязненность исследуемого участка местности.

В качестве анализируемых изображений в данной работе использованы снимки мультиспектральной камеры с пятиканальным форматом изображений {NIR, RE, R, G, B}, сочетающим в себе следующие каналы: ближний инфракрасный (NIR – 840 нм), красный край (RE – 717 нм), красный (R – 668 нм), зеленый (G – 560 нм) и синий (B – 475 нм). Угол обзора камеры 62,7°, фокусное расстояние – 5,74 мм (40 мм – эквивалент формата 35 мм), автофокус установлен на ∞ , диафрагма f/2.2. Камера содержит шесть 1/2,9-дюймовых матриц CMOS, включая

одну матрицу видимого излучения RGB и пять монохромных. Число эффективных пикселей каждой матрицы: 2,08 млн (общее число пикселей – 2,12 млн). Высота съемки составляет 25 метров.

3. Описание алгоритмического обеспечения для обработки мультиспектральных снимков

Обобщенный алгоритм анализа мультиспектральных аэрофотоснимков для идентификации загрязнений водоемов должен основываться на комплексном применении как аналитических методов, так и нейросетевых подходов. При этом существенным требованием является участие эксперта как на этапе разметки аэрофотоснимков, так и на этапе верификации и обобщения результатов, получаемых алгоритмическим способом.

Разработанное авторами программно-алгоритмическое обеспечение состоит из трех подсистем (рис. 1).

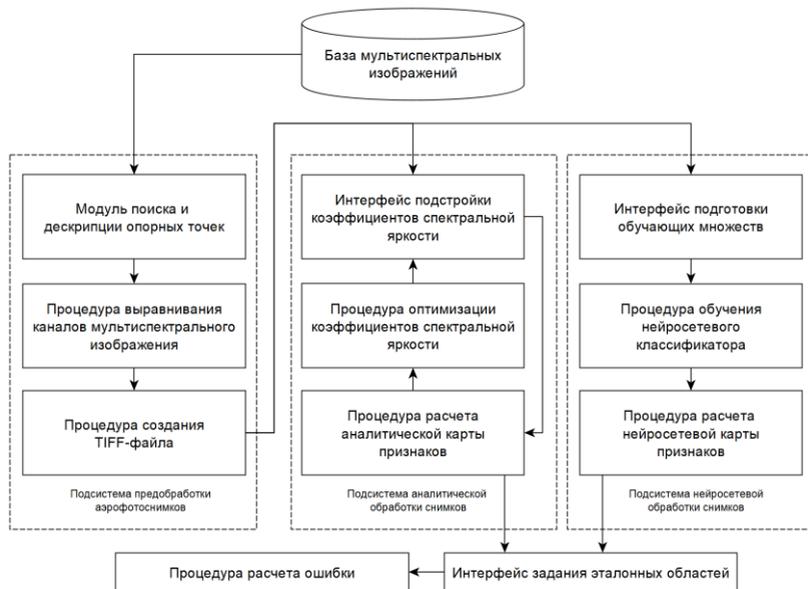


Рис. 1. Структура алгоритмического комплекса для анализа мультиспектральных аэрофотоснимков

Первая подгруппа программных модулей и процедур отвечает за предобработку аэрофотоснимков. Основной решаемой задачей при этом является выравнивание растров многоканального изображения. Потребность в данной процедуре обусловлена тем фактом, что объективы мультиспектральных камер располагаются с некоторым смещением друг относительно друга, и простое наложение каналов без дополнительного выравнивания приводит к размытости результирующего аэрофотоснимка.

Вторая подгруппа реализует аналитическую обработку изображений (подстройку яркости в ручном и автоматическом режимах, расчет результирующей аналитической карты признаков).

Третья подгруппа программных компонент обеспечивает нейросетевую обработку изображений. При этом обеспечивается как возможность выбора требуемой архитектуры сверточной нейронной сети, так и ее настройки по выбранным экспертом обучающим примерам. Как и в случае аналитической обработки, в результате работы нейросетевого алгоритма формируется карта признаков, отражающая наличие или отсутствие загрязненностей на выбранном участке Земли.

Отметим, что оценка результатов аналитической и нейросетевой обработки аэрофотоснимков производится с учетом эталонных областей загрязнений, заданных экспертом на тестовом подмножестве снимков.

Предполагаемое местоположение загрязнения на рассматриваемом участке изображения формально может быть описано совокупностью типовых геометрических фигур. В частности, удобными способами задания областей являются полигональный формат, когда целевая зона задается в виде замкнутого многоугольника, и формат набора кругов, пересекающих целевую область. В данном исследовании был использован последний подход, поскольку он вычислительно прост и удобен с точки зрения пользователя (рис. 2).

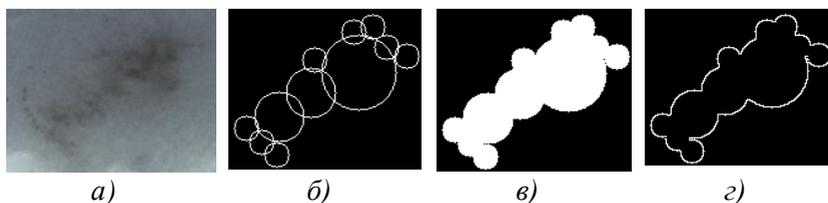


Рис. 2. Процесс интерактивного описания области загрязнения: а) фрагмент аэрофотоснимка; б) совокупность окружностей, сформированных экспертом; в) целевая область, состоящая из объединения кругов; г) ограничивающий контур области

4. Аналитические алгоритмы обработки аэрофотоснимков мест загрязнений

Каждый аэрофотоснимок, получаемый мультиспектральной камерой, может быть представлен в формате $A = \{I_1, I_2, I_3, I_4, I_5\}$, где $I_k \in \mathbb{R}^{W \times H}$, $k = 1, \dots, 5$, – растр отдельно взятого канала из перечня $\{NIR, RE, R, G, B\}$, W – ширина изображения, H – высота изображения.

Первоочередным этапом обработки аэрофотоснимка является совмещение его каналов, которое может быть сведено к попарному выравниванию растров I_k, I_{k+1} , $k = 1, \dots, 4$. Причем, поскольку смещения объективов мультиспектральной камеры постоянны, то данная операция может быть выполнена однократно. Одним из эффективных способов для решения данной задачи является расстановка трех опорных точек на каждом из растров и расчет матриц проективного преобразования (гомографии) $H_k \in \mathbb{R}^{3 \times 3}$ для каждой последовательной пары растров [15]. Матрицы данного типа хранят информацию о взаимном смещении, повороте и перспективном искажении растров, которые обусловлены несовпадением положений и оптических осей светочувствительных матриц в составе мультиспектральной камеры. Расчет матриц гомографии возможен по трем опорным точкам, задаваемым на каждом из совмещаемых растров. Расстановка данных точек может быть выполнена вручную или с применением алгоритмов детекции и дескрипции ключевых точек на изображении [10].

Выравнивание каналов сводится к смещению точек одного из двух растров по формуле

$$(1) \quad p_{x,y} = H_k \cdot p_{x,y}^*,$$

где $p_{x,y} = (x, y, 1)^T$ – координаты точки исходного растра (служащего источником значения яркости); $p_{x,y}^* = (x^*, y^*, 1)^T$ – координаты смещенных точек, на модифицированном растре многоканального изображения. Данная формула позволяет скомпенсировать не только линейное смещение точек изображения и их поворот, но и перспективное искажение, обусловленное вариациями в наклоне объективов мультиспектральной камеры.

Результирующая карта признаков на выходе аналитического алгоритма анализа аэрофотоснимков представляет собой монохромное изображение в формате $J \in \mathbb{R}^{W \times H}$. При этом взаимосвязь между формируемой картой признаков и растрами мультиспектрального изображения имеет вид

$$(2) \quad J = \sum_{k=1}^5 w_k \cdot \Psi_k(I_k),$$

где Ψ_k – преобразование k -го растра, являющееся суперпозицией операторов фильтрации яркости F , коррекции яркости B и изменения контраста C , примененных в порядке перечисления; $w_k \in \mathbb{R}$ – весовые коэффициенты для сбалансированного учета отдельных каналов в формируемой карте признаков.

Первичная фильтрация яркости, рассматриваемая как процедура выделения подмножества точек изображения, близких по яркости к загрязнению, заключается в амплитудном преобразовании значений $v_{x,y}$ для выбранного растра I_k по формуле

$$(3) \quad F(v_{x,y}) = \max(0, v_{x,y} \cdot (1 - |v_{x,y} - c|/R)),$$

где c – центральная яркость фильтра, R – полуширина фильтра. По существу данное преобразование приводит к нулю яркости точек растра, отличающихся от выбранного центрального значения на величину, большую R .

Формула изменения яркости для каждой точки растра по выбранному каналу имеет следующий вид:

$$(4) \quad B(v_{x,y}) = v_{x,y} + b,$$

где $b \in [-255, 255] \subset \mathbb{R}$ – смещение яркости.

Формула изменения контраста для каждой точки растра по выбранному каналу имеет следующий вид:

$$(5) \quad C(v_{x,y}) = v_{x,y} + k \cdot (v_{x,y} - v_{avg}),$$

где v_{avg} – усредненная яркость растра, k – степень изменения контраста.

Отметим, что преобразования растров по формулам (3)–(5) производятся в пространстве вещественных чисел. Лишь на финальном этапе цветокоррекции растр преобразуется в целочисленный формат с применением кодировки 8 бит на пиксель.

Для автоматизации подбора параметров яркости и контраста отдельных каналов в составе мультиспектрального изображения использованы методы дискретной оптимизации.

Вектор настраиваемых параметров $P_k \in \mathbb{R}^5$ для k -го канала содержит 5 параметров:

$$(6) \quad P_k = \{c, R, b, k, w\},$$

где в порядке перечисления указаны параметры центральной яркости фильтра, полуширины фильтра, смещения яркости, степени изменения контраста, весового коэффициента для каждого канала.

Таким образом, общий вектор параметров для 5 каналов включает 25 значений.

Критерием оптимальности решения является максимизация функционала четкости Q загрязнения, наблюдаемого экспертом.

Для пояснения принципа расчета данного функционала рассмотрим схематичное изображение зоны загрязнения S_2 внутри свободной от загрязнения части аэрофотоснимка S_1 (рис. 3). Отметим, что выполняется условие непересечения двух областей: $S_1 \cap S_2 = \emptyset$. При этом область S_2 задается либо многоугольником, либо объединением типовых геометрических фигур.

Отчетливость наблюдаемого загрязнения определяется различием в статистических характеристиках точек растра (x, y) , расположенных внутри и снаружи от пятна.

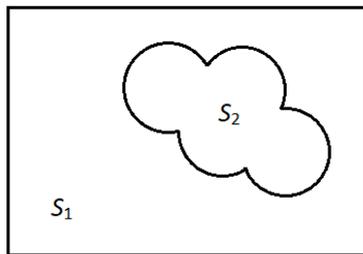


Рис. 3. Схема расположения участка загрязнения в границах аэрофотоснимка

Матожидание и смещенная оценка дисперсии яркости $v(x, y)$ на некотором участке S раstra задаются формулами

$$(7) \quad M(v)|_{v \in S} = \frac{1}{|S|} \cdot \sum_{(x,y) \in S} v_{x,y},$$

$$(8) \quad D(v)|_{v \in S} = \frac{1}{|S|} \cdot \sum_{(x,y) \in S} (v_{x,y} - M(v)|_{v \in S})^2.$$

Степень контрастности области раstra S определяется как стандартное отклонение яркости на подмножестве точек из данной области:

$$(9) \quad \gamma(S) = \sqrt{D(v)|_{v \in S}}.$$

Тогда в качестве критерия оптимальности при выборе определенных параметров цветовой фильтрации можно использовать показатель превышения контраста целевой области над контрастом фоновой области:

$$(10) \quad Q(J, P) = \frac{\gamma(S_2)}{\gamma(S_1)}.$$

При этом следует исходить из предположения, что контраст пятна в области S_2 должен быть выше, чем контраст элементов изображения в области S_1 (рис. 3). Дело в том, что контур области S_2 , заданный конечным числом геометрических примитивов, на практике никогда не аппроксимирует достаточно точно границы загрязнения, поскольку на практике они имеют характер, близкий к фрактальному. Гетерогенность рассматриваемого участка S_2 (содержащего как само загрязнение, так и вкрапления

окружающей среды) позволяет говорить о ненулевом контрасте γ_2 данной области. Вместе с тем необходимо добиться существенного превышения контраста γ_2 над аналогичным показателем γ_1 для внешней среды. Выполнение данного требования возможно за счет выбора параметров P коррекции многоканального изображения A с применением методов дискретной оптимизации.

Однократно выполненная максимизация показателя превышения контраста позволяет выделить зону загрязнения не только на калибровочном аэрофотоснимке, но и на других изображениях, полученных в режиме штатной эксплуатации системы экологического мониторинга.

Обобщенный алгоритм поиска загрязнения на основе предлагаемого аналитического подхода представлен на рис. 4.

Ключевыми этапами алгоритма являются: перебор дискретизированных векторов параметров фильтрации, расчет монохромных карт признаков по формуле (2) и расчет критерия оптимальности по формуле (9). Выходом алгоритма является карта признаков, наиболее четкая с точки зрения обнаружения загрязнений.

Отметим, что для реализации перебора параметров фильтрации эксперту необходимо задать набор дискретных значений для каждой из пяти переменных в формуле (6). Так, например, при выборе n опорных значений для каждого из параметров общее количество перебираемых вариантов будет определяться как $|P| = n^5$. Сокращение числа рассматриваемых вариантов возможно за счет перехода от метода полного перебора к методу покоординатного спуска, однако подобный подход в большей степени подвержен попаданию в локальные максимумы целевого функционала.

В качестве модельного примера, иллюстрирующего работу аналитического алгоритма поиска загрязнения, рассмотрим аэрофотоснимок, представленный на рис. 5. Эталонный участок, заданный экспертом в интерактивном режиме, показан красной линией.

Результат расчета карты признаков по совокупности выбранных параметров показан на рис. 6.

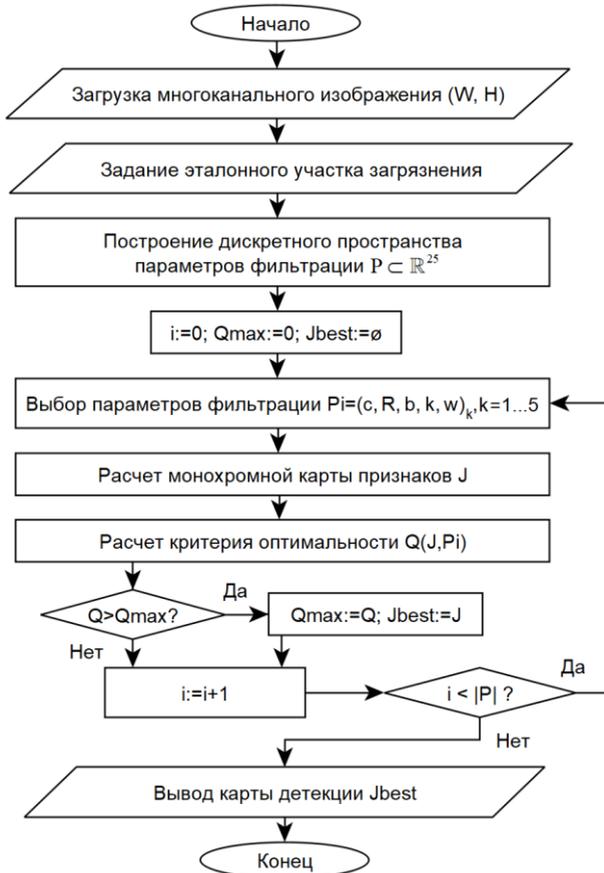


Рис. 4. Аналитический алгоритм обработки аэрофотоснимка

Значения параметров фильтрации, смещения яркости, контрастирования и взвешенного суммирования каналов перечислены ниже в таблице 1.



Рис. 5. Оригинальный аэрофотоснимок с нанесенным поверх контуром предполагаемого расположения загрязнения

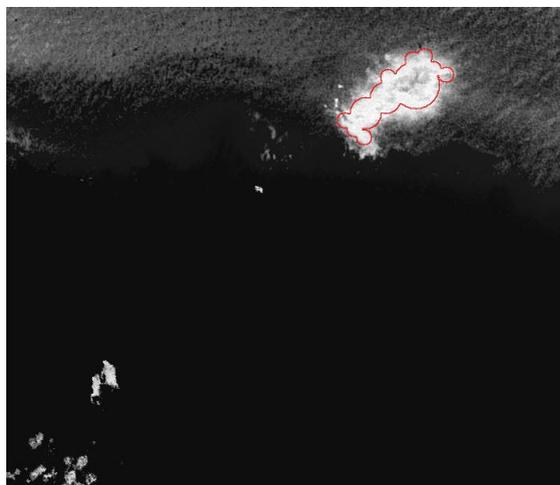


Рис. 6. Результат аналитической обработки мультиспектрального аэрофотоснимка

Таблица 1. Результирующие значения коэффициентов для аналитического подхода

Тип канала	c	R	b	k	w
NIR	15	20	100	1,5	0,25
RE	15	20	100	1,5	-0,25
R	20	20	100	1,5	1
G	10	20	100	1,5	-0,5
B	15	20	100	1,5	1

На полученном растре довольно отчетливо видно загрязнение, несколько превышающее область, изначально заданную экспертом. Однако в качестве недостатка подхода можно отметить наличие ложных срабатываний на объектах в левой нижней части снимка.

5. Нейросетевой алгоритм поиска загрязнений на поверхности водоемов

Предобработка входных изображений для нейросетевого анализа мультиспектральных аэрофотоснимков также предполагает выполнение операций выбора опорных точек и взаимного выравнивания каналов и верификации результатов детекции с привязкой к априорно заданным эталонным областям.

Основное отличие нейросетевого алгоритма заключается в способе построения карты признаков. Если для аналитического способа карта признаков строится поточечно на основе расчета суперпозиции описанных выше функций, то при нейросетевом подходе применяется классификация фрагментов $G_{x,y} \in \mathbb{R}^{30 \times 30 \times 5}$ аэрофотоснимка A в сочетании с методом скользящего окна.

Преимущество такого подхода заключается в том, что размер анализируемых фрагментов достаточен для выявления не только лишь яркости наблюдаемой местности, но и ее текстурных особенностей. Это позволяет, в частности, осуществлять классификацию 4 классов поверхности: {песок, трава, вода, загрязнение}.

Одной из основных сложностей при этом является оперативное формирование обучающего множества для настройки нейронной сети с учетом особенностей требуемой географической зоны.

Для решения этой проблемы разработано алгоритмическое обеспечение, которое включает интерфейс интерактивного формирования множества обучающих примеров. После загрузки аэрофотоснимка пользователь выбирает номер класса объектов, по которым формируется обучающий пример, позиционирует курсор в области целевой точки местности (x, y) и нажимает на клавишу компьютерной мыши. После этого в заданную папку на диске сохраняется изображение фрагмента G с именем файла, отражающим принадлежность к выбранному классу.

Многократное повторение данной процедуры позволяет сформировать первичное обучающее множество. При этом дополнительно применяется аугментация изображений для увеличения количества тренировочных примеров.

Не менее важной задачей является выбор структуры нейросетевого классификатора. С одной стороны, она должна быть вычислительно простой, а с другой – совмещать в себе сверточные, субдискретизирующие и полносвязные слои для комплексной обработки входного изображения. Архитектура использованного классификатора приведена в таблице 2. Выходом сети являются 4 оцениваемых вероятности принадлежности входного образа одному из возможных классов поверхности.

Таблица 2. Архитектура нейросетевого классификатора

№	Название слоя	Размер слоя	Ядро подвыборки
1.	Входное изображение	$30 \times 30 \times 5$	–
2.	Сверточный слой	$26 \times 26 \times 64$	5×5
3.	Слой субдискретизации	$13 \times 13 \times 64$	2×2
4.	Полносвязный слой	64	–
5.	Выходной слой softmax-активации	4	–

Обучение выбранной нейронной сети производится алгоритмом обратного распространения ошибки на основе метода градиентного спуска [1].

После обучения классификатора возможен запуск процедуры формирования нейросетевой карты признаков за счет многократного расчета классификатора на различных фрагментах аэрофотоснимка с использованием метода скользящего окна. При этом шаги сдвига скользящего окна в горизонтальном и вертикальном направлениях могут быть выбраны равными половине размера анализируемого изображения: $stepX = 15$ пикселей, $stepY = 15$ пикселей.

Обобщенный алгоритм нейросетевой обработки мультиспектрального аэрофотоснимка представлен на рис. 7.

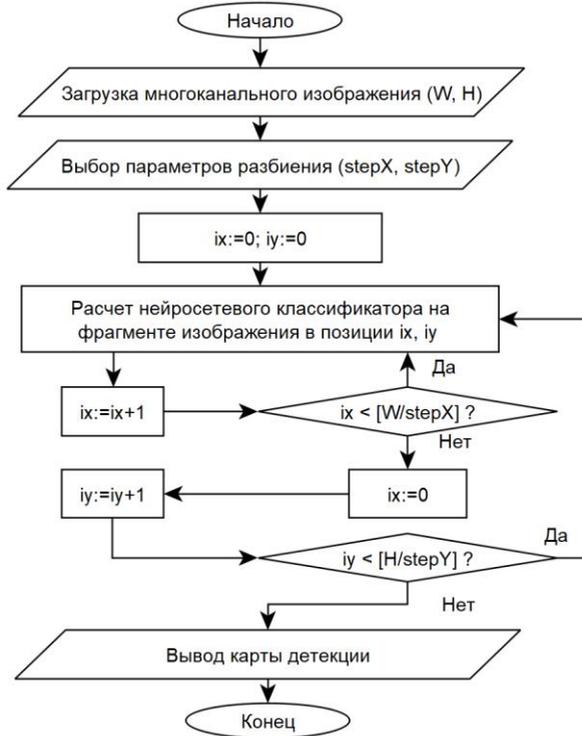


Рис. 7. Нейросетевой алгоритм обработки аэрофотоснимка

Расположение анализируемых фрагментов вдоль строк и столбцов исходного изображения обуславливает наличие двух циклов в структуре алгоритма. Ключевой этап алгоритма, связанный с расчетом нейросетевого классификатора, опирается на данные таблицы 2.

Визуализация результата расчета нейросетевой карты признаков для одного из аэрофотоснимков представлена на рис. 8. Фиолетовым цветом показана найденная область загрязнения на поверхности реки.

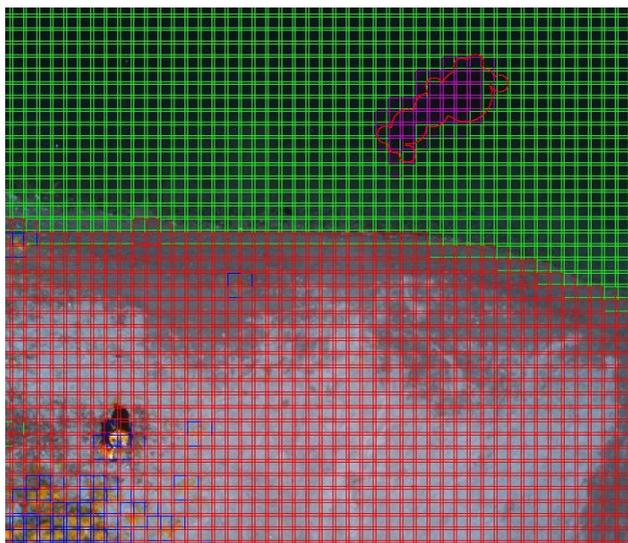


Рис. 8. Результат нейросетевой детекции загрязнения на поверхности реки

6. Программная реализация алгоритмов обработки аэрофотоснимков и оценка эффективности

Программная реализация предложенных алгоритмов потребовала разработки программного комплекса в соответствии с архитектурой, представленной на рис. 1. Описание входящих в состав программного комплекса модулей и процедур приведено в таблице 3.

Таблица 3. Описание компонент программно-алгоритмического комплекса для анализа мультиспектральных аэрофотоснимков

№	Название	Описание
1.	Модуль поиска и дескрипции опорных точек	Группа программных функций для выделения на изображении контрастных участков с привязкой их центральных точек к системе координат фотоснимка. При этом вектор дескрипции данных участков инвариантен к изменению цветовой гаммы каналов мультиспектрального изображения.
2.	Процедура выравнивания каналов мультиспектрального изображения	Программная функция, осуществляющая сдвиг каналов мультиспектрального изображения по осям X и Y, такой, что достигается максимум критерия взаимной корреляции каналов
3.	Процедура создания TIFF-файла	Программная функция, обеспечивающая объединение нескольких одноканальных растровых изображений в единый TIFF-файл
4.	Интерфейс подстройки коэффициентов спектральной яркости	Группа функций и графических элементов взаимодействия с пользователем для фильтрации диапазона значений яркости на изображении, корректировки яркости и контраста каналов TIFF изображения.
5.	Модуль расчета аналитической карты признаков	Совокупность программных функций, реализующая аналитическое определение областей вероятного расположения загрязнения на наблюдаемом участке местности
6.	Процедура оптимизации коэффициентов спектральной яркости	Программно реализованный алгоритм выбора наилучших значений параметров фильтрации, яркости и контраста для каждого из каналов мультиспектрального аэрофотоснимка

Таблица 3 (продолжение)

7.	Интерфейс подготовки обучающих множеств	Совокупность программных функций и графических элементов взаимодействия с пользователем, позволяющая выбрать и сохранить на диск множество фрагментов изображения, относящихся к одному из типовых классов (водная поверхность, пятно загрязнения, песок, трава и т.д.)
8.	Процедура обучения нейросетевого классификатора	Программно реализованный алгоритм подстройки весовых коэффициентов нейронной сети на базе метода градиентного спуска (алгоритм обратного распространения ошибки)
9.	Модуль расчета нейросетевой карты признаков	Программный модуль, основанный на применении метода скользящего окна в сочетании с расчетом нейросетевого классификатора изображений
10.	Интерфейс задания эталонных областей	Совокупность программных функций и графических элементов взаимодействия с пользователем, позволяющая эксперту очертить многоугольник целевого интереса вокруг обнаруженного загрязнения на поверхности воды
11.	Процедура расчета ошибки	Программная функция для оценки точности обнаружения загрязнения аналитическим или нейросетевым методом на основе сравнения с экспертной оценкой

Отдельного пояснения требует процедура расчета ошибки детекции загрязнения в пределах рассматриваемого участка местности S_i . Как показано на рис. 9, эталонная область загрязнения S_e и детектируемые области загрязнения S_n , S_a могут не совпадать.

И для аналитического, и для нейросетевого детекторов оценки точности рассчитываются с применением операций пересечения к некоторым из множеств точек, представленным на рис. 9.

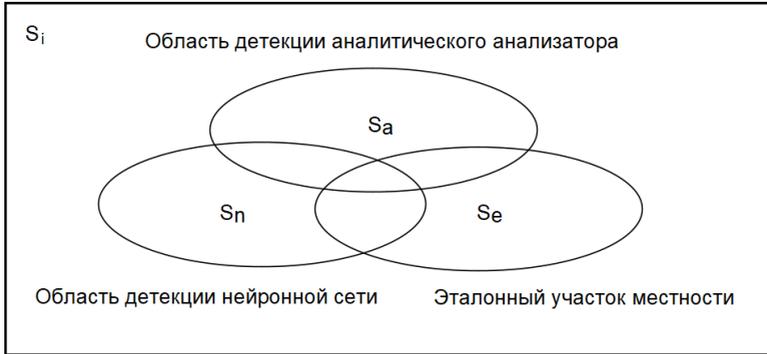


Рис. 9. Принцип оценки точности обнаружения загрязнений

При этом целесообразен расчет трех основных оценок, применяемых в машинном обучении: точности P , полноты R и обобщенной метрики $F1$. В частности, формулы для оценки данных показателей для аналитического детектора имеют следующий вид:

$$(11) P_a = \frac{|S_a \cap S_e|}{|S_a|},$$

$$(12) R_a = \frac{|S_a \cap S_e|}{|S_e|},$$

$$(13) F1_a = \frac{2P_a R_a}{P_a + R_a}.$$

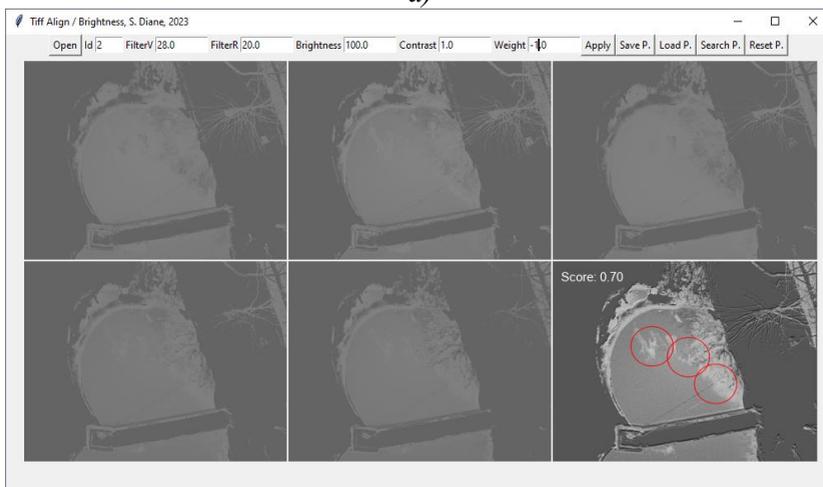
Аналогичным образом рассчитываются данные показатели и для нейросетевого детектора.

Для реализации программно-алгоритмического обеспечения был выбран язык программирования Python, в том числе входящие в его состав библиотеки OpenCV, TKInter, PyGame.

Интерфейс разработанного программного комплекса представлен на рис. 10. В выведенном на экран графическом окне представлено 5 каналов мультиспектрального аэрофотоснимка побережья реки (рис. 10а).



а)



б)

Рис. 10. Интерфейс программного комплекса для анализа мультиспектральных аэрофотоснимков

К каждому из каналов применен вектор параметров фильтрации, подстройки яркости и контраста, полученный в результате запуска процедуры оптимизации коэффициентов спек-

тральной яркости. Объединение данных каналов в монохромное изображение карты признаков представлено в области 6 графического окна.

Аналогичным образом разработанное программное обеспечение может быть применено и для анализа других природных объектов. Так, на рис. 10б показан пример использования разработанного программного обеспечения для анализа поверхности водоема, расположенного на территории ИПУ РАН. Отчетливость результирующей карты признаков в данном случае существенно снижена, что обусловлено текстурными особенностями окружающей среды и наличием нависающих над загрязнением ветвей деревьев.

Красными окружностями на картах признаков показаны целевые участки местности, заданные экспертом и служащие для расчета показателя превышения контраста (10) в процессе подстройки коэффициентов спектральной яркости и метрики $F1$ (13) в процессе верификации алгоритма.

Предварительные экспериментальные исследования по оценке точности аналитического и нейросетевого детекторов проводились на тестовом множестве аэрофотоснимков, включающем 50 изображений, размеченных экспертом. Точность аналитического способа обнаружения загрязнений в проведенных экспериментах составила 78%. Точность нейронной сети составила 93%.

Запланированная серия дополнительных исследований включает расширение тренировочного и тестового набора примеров для настройки аналитического и нейросетевого детекторов, проработку методики отсека заведомо нерелевантных участков местности при обнаружении загрязнений водной поверхности.

7. Заключение

В данной статье представлены алгоритмы и программный комплекс автоматизированного анализа снимков мультиспектральной камеры БПЛА для идентификации загрязнений водных поверхностей.

Предлагаемый подход является комплексным и основан на использовании математического аппарата цифровой обработки информации, методов оптимизации, технологий искусственного интеллекта, элементов теории множеств.

Основными требованиями к программно-алгоритмическому обеспечению являлись возможность загрузки подготовленных изображений, их коррекции и обработки с применением аналитических и нейросетевых методов. На выходе программного комплекса необходимо было формировать растровые карты, отражающие загрязненность исследуемого участка местности.

Разработанный аналитический алгоритм, основанный на взвешенном суммировании нескольких каналов мультиспектрального снимка, достаточно прост в интерпретации экспертом и, кроме того, не требует больших вычислительных мощностей. В свою очередь, преимуществом использования разработанного нейросетевого алгоритма является возможность более эффективного учета текстурных составляющих на фрагментах изображений.

Отметим, что выбранный подход к использованию метода оконного сканирования для расчета нейронной сети не может считаться оптимальным с точки зрения требований к вычислительным ресурсам. В этой связи дальнейшие перспективы развития предложенного алгоритмического комплекса связаны с рассмотрением сегментирующих нейронных сетей, принимающих на вход весь аэрофотоснимок целиком без разбиения его на фрагменты и потенциально способных достичь более высокой производительности и точности при определении загрязнений.

Литература

1. ХАЙКИН С. *Нейронные сети: полный курс*. – М.: Вильямс, 2008. – 1103 с.
2. ALI N.M., KARIS M.S., SOBRAN N.M.M., BAHAR M.B. et al. *Detection of multiple mangoes using histogram of oriented gradient technique in aerial monitoring // ARPN Journal of*

- Engineering and Applied Sciences. – 2017. – Vol. 12 (8). – P. 2730–2736.
3. BARRIENTOS Z., NÁJERA J.M., ZÚÑIGA M. *A Satellite and Ground Evaluation of Urban Vegetation and Infrastructure in the Landscape of a Tropical City: Heredia, Costa Rica* // Cities and the environment. – 2013.
 4. CONG Z., YAN G., SHENGHUI F., KAILI Y. et al. *Combining spectral and wavelet texture features for unmanned aerial vehicles remote estimation of rice leaf area index* // Frontiers in Plant Science. – 2022. – Vol. 13. – DOI: 10.3389/fpls.2022.957870.
 5. KANTH P.C., GUPTA N., ARI S. *Change Detection in Multi-spectral Satellite Images using Histogram based Thresholding Technique* // Proc. of 2019 IEEE Int. Conf. on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore, India. – 2019. – P. 1–5. – DOI: 10.1109/ICECCT.2019.8869500.
 6. MENG L., ZHOU J., LIU S., DING L. et al. *Investigation and evaluation of algorithms for unmanned aerial vehicle multi-spectral image registration, International* // Journal of Applied Earth Observation and Geoinformation. – 2021. – Vol. 102, 102403. – DOI: 10.1016/j.jag.2021.102403.
 7. MOGHIMI A., YANG C., ANDERSON J.A. *Aerial hyperspectral imagery and deep neural networks for high-throughput yield phenotyping in wheat* // Computers and Electronics in Agriculture. – 2020. – Vol. 172. – DOI: 10.1016/j.compag.2020.105299.
 8. PYKA K. *Wavelet-Based Local Contrast Enhancement for Satellite, Aerial and Close Range Images* // Remote Sens. – 2017. – Vol. 9. – P. 25. – DOI: 10.3390/rs9010025.
 9. RODRIGUES R.B., PELLEGRINO S., PISTORI H. *Combining Color and Haar Wavelet Responses for Aerial Image Classification* // In: Artificial Intelligence and Soft Computing / Eds. – Springer Berlin Heidelberg, 2012. – Vol. 7267. – P. 583–591. – DOI: 10.1007/978-3-642-29347-4_68.

10. RUBLEE E., RABAUD V., KONOLIGE K., BRADSKI G. *ORB: an efficient alternative to SIFT or SURF* // Proc. of the IEEE Int. Conf. on Computer Vision. – 2011. – P. 2564–2571.
11. VASILYEV A.V., MELNIKOVA I.N., POBEROVSKAYA L.N., TOVSTENKO I.A. *Spectral brightness coefficient of natural ground surfaces in spectral ranges 0.35 to 0.85 μm based on airborne measurements* // Proc. SPIE 3237, 23rd European Meeting on Atmospheric Studies by Optical Methods. – 1997. – DOI: 10.1117/12.284763.
12. VYTOVTOV K.A., BARABANOVA E.A., GLADKIKH T.Y., KULINA A.L. et al. *Remote Monitoring of Water Pollution With Oil Products in the Visible Range by Using UAV Multi-spectral Camera* // Proc. 2022 Int. Conf. on Information, Control, and Communication Technologies (ICCT-2022), Astrakhan, Russian Federation, 2022. – P. 1–5. – DOI: 10.1109/ICCT56057.2022.9976826.
13. XIANG T.-Z., XIA G.-S., ZHANG L. *Mini-Unmanned Aerial Vehicle-Based Remote Sensing: Techniques, applications, and prospects* // IEEE Geoscience and Remote Sensing Magazine. – 2019. – Vol. 7, No. 3. – P. 29–63. – DOI: 10.1109/MGRS.2019.2918840.
14. ZAMPIERI A., CHARPIAT G., GIRARD N., TARABALKA Y. *Multimodal image alignment through a multiscale chain of neural networks with application to remote sensing* // The European Conference on Computer Vision (ECCV). – 2018. – P. 657–673. – DOI: 10.1007/978-3-030-01270-0_40.
15. ZHANG Z., HANSON A.R. *3D Reconstruction based on homography mapping* // Proc. ARPA96. – 1996. – P. 1007–1012.

ALGORITHM FOR ANALYSIS OF MULTISPECTRAL AERIAL IMAGES FROM UAV FOR IDENTIFICATION OF WATER POLLUTION USING ANALYTICAL METHODS AND NEURAL NETWORK APPROACHES

Sekou Diane, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, Candidate of Science, (diane1990@yandex.ru)

Konstantin Vytovtov, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, Doctor of Science, assistant professor (vytovtov_konstan@mail.ru)

Elizaveta Barabanova, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, Doctor of Science, assistant professor (elizavetaalex@yandex.ru)

Abstract: The article is devoted to the development of algorithms for the analysis of pollution on the surface of water bodies based on visual information obtained using a multispectral camera mounted on the body of a UAV. The structure of the algorithmic complex for the analysis of multispectral aerial photographs is proposed. Within the framework of the developed approach, each of the analyzed images undergoes a preprocessing procedure that ensures the alignment and alignment of its spectral channels into a single multidimensional raster. The developed analytical algorithm makes it possible to process and convolve the channels of a multispectral image using three mathematical operators - bandpass filtering, contrast change, and brightness change. At the same time, the choice of parameters for identifying pollution on the surface of water bodies is based on a preliminary stage associated with maximizing the contrast excess index for the reference area. The proposed neural network pollution analysis algorithm is based on the application of the sliding window method in combination with the convolutional architecture of the neural network classifier for the analysis of image fragments located on a rectangular grid. The software implementation of these algorithms, as well as the development of a graphical user interface, made it possible to confirm the assumption about the effectiveness of each of the considered approaches. Experimental studies have shown that the neural network algorithm wins in accuracy, and the analytical approach is easier to interpret from the point of view of an expert.

Keywords: aerial photograph, analytical method, neural network approach.

УДК 519.7 + 62

ББК 22.18+40

DOI: 10.25728/ubs.2024.108.6

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

Поступила в редакцию 07.07.2023.

Опубликована 31.03.2024.

ПОСТРОЕНИЕ ПРОГНОЗА ДИНАМИКИ ОБЩЕСТВЕННОГО МНЕНИЯ ПРИ ПОМОЩИ SCARDO-МОДЕЛИ¹

Козицин И. В.²

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

Последние 20 лет теория агент-ориентированных моделей социального влияния активно развивается, что связано с необходимостью описания процессов формирования мнений в условиях перехода каналов общения в цифровую среду и интенсификации информационных потоков. Вместе с тем практическая сторона данной теории остается слабо изученной. Причиной этому в первую очередь являются трудности в калибровке параметров моделей и построении эмпирической базы. В настоящей работе проводится валидация SCARDO-модели формирования мнений на эмпирических лонгитюдных данных из социальной сети ВКонтакте, включающих три снимка мнений крупномасштабной выборки пользователей, а также снимок дружеских связей между ними. Параметры модели калибруются на первых двух снимках, а третий используется для проверки точности прогноза модели (объект прогноза – численности сторонников различных взглядов). В качестве ориентира выступает модель постоянного тренда. Проведенный анализ показывает, что в зависимости от способа калибровки параметров прогноз SCARDO-модели может быть более или менее точным, чем предсказание модели постоянного тренда. Вместе с тем изменения общественного мнения в рассматриваемом датасете (несмотря на то, что достаточны для калибровки параметров модели) малы и в результате характерная величина ошибки прогноза не превышает одного процента «головов».

Ключевые слова: модели формирования мнений, SCARDO-модель, социальные сети, валидация.

1. Введение

Противодействие распространению деструктивного контента, локализация и купирование информационных пузырей, а также деполяризация общества требуют понимания того, какие механизмы отвечают за распространение информации и формирование мнений в социальных сетях [3, 7, 13]. Основным инстру-

¹ Работа выполнена при финансовой поддержке РФФ, грант № 22-71-00075.

² Иван Владимирович Козицин, к.ф.-м.н. (kozitsin.ivan@mail.ru).

ментом изучения такого рода процессов в структурированных социальных системах являются агент-ориентированные модели социального влияния (их также называют моделями формирования мнений) [2]. В этих моделях агенты, имитирующие людей, наделены определенными характеристиками (в частности, мнениями) и взаимодействуют друг с другом на основании некоторых наперед заданных правил.

Несмотря на теоретическую проработанность данной теории, ее прикладной аспект требует значительно большего внимания со стороны научного сообщества [5]. Основная проблема кроется в сложности построения эмпирической базы для проведения процедуры валидации моделей [11]. Подготовка датасетов, описывающих динамику мнений людей и структуру/историю их взаимодействий, требует значительно более тесной кооперации специалистов из различных областей (социальная психология, математическое моделирование, машинное обучение, сетевой анализ), чем та, которая имеет место в настоящий момент [10, 12]. Отметим, что еще шесть лет назад консенсусной точкой зрения было то, что именно ассимилятивный механизм влияния в линейной форме (являющийся основной, к примеру, модели Де-Гроота [4]) наилучшим образом соответствует тому, что наблюдается в эмпирике. Однако последние исследования, проведенные в рамках полевых экспериментов в онлайн-социальных сетях, показали легитимность других механизмов влияния, в том числе ассимилятивного влияния в умеренной форме и диссимилятивного влияния; также были выявлены нетривиальные паттерны их сосуществования [1, 7, 9].

Ранее в работе [8] была построена агент-ориентированная модель формирования мнений, предлагающая табличный подход к моделированию процессов влияния, в рамках которого мнение агента в результате воздействия на него изменяется стохастически в соответствии с распределением, которое является функцией от характеристик взаимодействующих агентов. Данная функция может быть эффективно представлена при помощи серии таблиц. Такой способ кодировки дает возможность воспроизводить ши-

рокий спектр механизмов социального влияния как качественно, так и количественно, не используя при этом традиционные разностные выражения [7].

Преыдушие исследования были посвящены дескриптивным аспектам SCARDO-модели [6, 8]. В данной работе будет изучена ее предиктивная способность. Для этого будут использованы эмпирические данные, полученные из социальной сети ВКонтакте. Эти данные представляют собой три снимка мнений выборки пользователей, а также снимок их дружеских связей, сделанный в середине промежутка наблюдения за системой. Это позволяет калибровать параметры SCARDO-модели на первых двух снимках мнений, а третий использовать для ее валидации. Отметим, что одно из важных свойств SCARDO-модели заключается в ее толерантности к формату данных и относительной простоте процедуры идентификации ее параметров.

2. Модель

В данном разделе будет представлена SCARDO-модель [8], расширение которой для описания агентов с различными уровнями влиятельности и восприимчивости ко влиянию было предложено в работе [6]. Далее мы будем рассматривать базовую версию SCARDO-модели [8].

Рассматривается система из N агентов, каждый из которых характеризуется мнением $x_i(t) \in Z = \{z_1, \dots, z_m\}$. Множество Z будем называть азбукой мнений, ее элементы могут соответствовать номинальным альтернативам (за какую из партий вы будете голосовать на выборах?), а могут быть упорядочены для того, чтобы описывать степень доверия к той или иной сущности (как вы относитесь к закону о ношении оружия по шкале от 1 до 10?):

$$(1) \quad z_1 \prec \dots \prec z_m.$$

Агенты связаны сетью $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, где \mathcal{V} – множество вершин, $\#\mathcal{V} = N$, и $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ – множество ребер. Данный граф описывает дружеские связи между агентами. Граф по умолчанию является простым, связным, неориентированным и невзвешенным.

Множество друзей агента i будем обозначать

$$\mathcal{V}_i = \{j \in \mathcal{V} \mid (i, j) \in \mathcal{E}\}.$$

Время дискретно. В каждый момент времени t случайным образом (по умолчанию равномерно) выбирается агент i и – также случайно (по умолчанию равномерно) – один из его друзей $j \in \mathcal{V}_i$. Далее агент j (субъект влияния) влияет на агента i (объект влияния). Отметим, что это не противоречит ненаправленности графа: направление действия влияния появляется в результате случайного выбора вершин. В результате мнение агента i обновляется, принимая одно из m возможных значений: z_1, \dots, z_m . Исход влияния определяется стохастически.

Зададим распределение, которое описывает эту процедуру. Будем считать, что оно зависит только от мнений взаимодействующих агентов. Пусть объект влияния характеризуется мнением $x_i(t) = z_s$, а субъект влияния – мнением $x_j(t) = z_l$. Далее к этим элементам азбуки мнений будем обращаться через их индексы s и l . Итоговое мнение агента i будет определяться распределением

$$P_{s,l} = \left(p_{s,l,1}, \dots, p_{s,l,m} \right),$$

в котором величина $p_{s,l,k} \in [0, 1]$ есть вероятность изменения мнения $z_s \rightarrow z_k$ объекта влияния. Третий нижний индекс у $p_{s,l,k}$ синхронизирован с итоговым мнением, а первые два нижних индекса согласованы с индексами мнений взаимодействующих агентов. При этом выполнено

$$p_{s,l,1} + \dots + p_{s,l,m} = 1.$$

Все величины $p_{s,l,k}$ сгруппируем в трехмерный объект

$$\mathcal{P} = [p_{s,l,k}]_{s,l,k \in [m]},$$

который будем называть таблицей переходов. Таблица переходов \mathcal{P} может быть представлена как последовательность квадратных $m \times m$ строчно-стохастических матриц:

$$\mathcal{P} = [P_1, \dots, P_m],$$

$$P_1 = \begin{bmatrix} p_{1,1,1} & \cdots & p_{1,1,m} \\ \cdots & \cdots & \cdots \\ p_{1,m,1} & \cdots & p_{1,m,m} \end{bmatrix}, \dots, P_m = \begin{bmatrix} p_{m,1,1} & \cdots & p_{m,1,m} \\ \cdots & \cdots & \cdots \\ p_{m,m,1} & \cdots & p_{m,m,m} \end{bmatrix},$$

где матрица P_s кодирует то, каким образом агенты с мнением z_s реагируют на все возможные воздействующие мнения. Именно такой способ представления будет далее использоваться при описании таблиц переходов (см. рис. 1 и 2 ниже).

3. Численные эксперименты

Для изучения предиктивной способности SCARDO-модели были проведены численные эксперименты, в рамках которых значения параметров модели калибровались на эмпирических данных. После этого строился прогноз, который затем сопоставлялся с данными, которые не участвовали в калибровке параметров модели. Объектом прогноза выступали величины

$$y_s(t) = \frac{\#\{i \mid x_i(t) = z_s\}}{N},$$

описывающие нормированные численности фракций сторонников различных взглядов. В качестве целевой метрики выступала суммарная абсолютная ошибка. Для построения базового прогноза была использована модель постоянного тренда.

3.1. Эмпирические данные

Рассматривался лонгитюдный датасет, включающий три снимка мнений выборки из $N = 1,660,927$ пользователей социальной сети ВКонтакте относительно политической повестки, сделанные в феврале (T_1), июле (T_2) и декабре (T_3) 2018 года, а также граф социальных (дружеских) связей, полученный в июле 2018 года. Мнения пользователей оценивались по непрерывной шкале от 0 до 1, где 0 соответствует оппозиционной позиции, а 1 – провластной. Подробное описание данных можно найти в работах [8, 9]. Предварительно из графа была выделена гигантская компонента связности, состоящая из $N = 1,648,829$

вершин (примерно 99% от их исходного числа вершин), а несвязные вершины были удалены из рассмотрения.

3.2. Идентификация параметров модели

Мнения агентов инициализировались на основании эмпирических данных. Поскольку азбука мнений SCARDO-модели является дискретной конструкцией, то к эмпирическим мнениям, принадлежащим промежутку $[0, 1]$, применялись различные дискретизирующие отображения. Рассматривались четыре дискретизирующих отображения:

$$\left[0, \frac{1}{2}\right) \rightarrow z_1, \left[\frac{1}{2}, 1\right] \rightarrow z_2, \quad (m = 2),$$

$$\left[0, \frac{1}{3}\right) \rightarrow z_1, \left[\frac{1}{3}, \frac{2}{3}\right) \rightarrow z_2, \left[\frac{2}{3}, 1\right] \rightarrow z_3, \quad (m = 3),$$

$$\left[0, \frac{1}{5}\right) \rightarrow z_1, \left[\frac{1}{5}, \frac{2}{5}\right) \rightarrow z_2, \dots, \left[\frac{4}{5}, 1\right] \rightarrow z_5, \quad (m = 5).$$

$$\left[0, \frac{1}{10}\right) \rightarrow z_1, \left[\frac{1}{10}, \frac{2}{10}\right) \rightarrow z_2, \dots, \left[\frac{9}{10}, 1\right] \rightarrow z_{10}, \quad (m = 10),$$

порождающие четыре различных азбуки мнений. В каждой из них элементы были упорядочены в соответствии с (1).

Сеть \mathcal{G} определялась на основе графа дружеских связей.

Для идентификации таблиц переходов был использован алгоритм, предложенный в работе [8]. Для каждого пользователя была построена оценка воздействующего на него мнения, операционализируемая как среднее мнение его друзей (переменная $x \in [0, 1]$ описывает эмпирические мнения):

$$x_{-i}(T_1) = \frac{\sum_{j \in \mathcal{V}_i} x_j(T_1)}{\#\mathcal{V}_i}.$$

После этого величины $x_i(T_1)$, $x_{-i}(T_1)$ и $x_i(T_2)$ были дискретизированы, а компоненты $p_{s,l,k}$ рассчитаны по следующей формуле:

$$p_{s,l,k} = \frac{\#\{i \mid \mathcal{D}(x_i(T_1)) = z_s, \mathcal{D}(x_{-i}(T_1)) = z_l, \mathcal{D}(x_i(T_2)) = z_k\}}{\#\{i \mid \mathcal{D}(x_i(T_1)) = z_s, \mathcal{D}(x_{-i}(T_1)) = z_l\}},$$

где \mathcal{D} – дискретизирующее отображение. При этом из рассмотрения были вычеркнуты пользователи, для которых

$$|x_{-i}(T_2) - x_{-i}(T_1)| > 0,05,$$

что позволило отчасти контролировать эффект влияния третьего лица.

На рис. 1 и 2 представлены рассчитанные таблицы переходов для случаев $m = 2$, $m = 3$ и $m = 5$ (таблица переходов для случая $m = 10$ слишком габаритна).

В качестве начального состояния системы рассматривался второй снимок датасета (сюда относятся и значения мнений агентов, и структура связей).

3.3. Дизайн экспериментов

Для каждого значения m было проведено 30 независимых испытаний, каждое из которых длилось $T = N$ итераций (один шаг Монте-Карло), что соответствовало промежутку времени в реальном мире между вторым и третьим снимками датасета (полгода).

В качестве целевого ориентира выступала модель постоянного тренда, согласно которой численности фракций к моменту времени T_3 определяются как

$$(2) \quad y_{pred}^{const}(T_3) = y(T_2) + y(T_2) - y(T_1).$$

Для оценивания качества прогноза y_{pred} использовалась суммарная абсолютная ошибка (AE):

$$(3) \quad AE = \|y(T_3) - y_{pred}\|_1.$$

0.975	0.025
0.952	0.048

0.066	0.934
0.049	0.951

0.96	0.04	0
0.942	0.057	0.001
0.907	0.091	0.002

0.04	0.952	0.008
0.021	0.969	0.01
0.02	0.944	0.036

0.001	0.082	0.917
0.001	0.07	0.929
0.001	0.054	0.945

Рис. 1. Таблицы переходов для $t = 2$ (панели а, б) и $t = 3$ (панели в–д). Все величины представлены с точностью до трех значащих цифр после запятой. Компонента $p_{1,1,3}$ матрицы P_1 (панель в) была положительна до округления

0.947	0.044	0.008	0.001	0
0.954	0.04	0.006	0	0
0.938	0.051	0.01	0.001	0
0.911	0.07	0.016	0.003	0
0.919	0.063	0.009	0.009	0

0.055	0.872	0.071	0.002	0
0.046	0.898	0.055	0.001	0
0.032	0.893	0.073	0.002	0
0.04	0.859	0.097	0.004	0
0.047	0.858	0.092	0.003	0

0.002	0.057	0.919	0.022	0
0.002	0.052	0.935	0.011	0
0.001	0.038	0.943	0.018	0
0.001	0.039	0.913	0.046	0.001
0.001	0.043	0.896	0.059	0.001

0.001	0.002	0.102	0.876	0.019
0.001	0.005	0.093	0.886	0.015
0	0.002	0.078	0.905	0.015
0.001	0.002	0.068	0.898	0.031
0.001	0.003	0.069	0.884	0.043

0	0	0.011	0.097	0.892
0	0.003	0.009	0.081	0.907
0.001	0.001	0.006	0.072	0.92
0	0.001	0.006	0.069	0.924
0.001	0	0.005	0.071	0.923

Рис. 2. Таблица переходов для $t = 5$. Все величины представлены с точностью до трех значащих цифр после запятой. Компоненты, равные нулю до округления, выделены серым цветом

4. Результаты

Для каждого испытания SCARDO-модели было рассчитано значение (3). Статистические тесты Д'Агостино – Пирсона ($P = 0,24$, $P = 0,41$, $P = 0,55$, $P = 0,14$ для $m = 2$, $m = 3$, $m = 5$ и $m = 10$ соответственно) и Шапиро – Вилка ($P = 0,5$, $P = 0,81$, $P = 0,45$, $P = 0,57$ для $m = 2$, $m = 3$, $m = 5$ и $m = 10$ соответственно) показали, что нулевая гипотеза о том, что данные распределения являются нормальными, не может быть отвергнута.

На рис. 3 представлены средние значения данных распределений в сравнении с прогнозами модели (2).

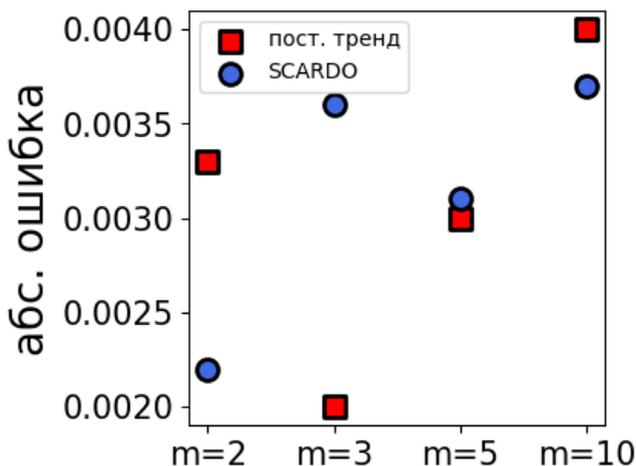


Рис. 3. Сравнение качества прогнозов модели постоянного тренда и SCARDO-модели.

При этом для $m = 2$, $m = 3$ и $m = 10$ разницы с прогнозами модели постоянного тренда оказались статистически значимы (одновыборочный t -критерий: $P < 0,001$ для всех трех случаев). Из рис. 3 можно сделать вывод, что при $m = 2$ и $m = 10$ SCARDO-модель показывает бóльшую точность по сравнению с моделью постоянного тренда, в то время как для $m = 3$ послед-

няя оказывается более точной. При $m = 5$ статистически значимых различий не наблюдается (одновыборочный t -критерий: $P = 0,18$).

Стоит отметить, что характерные значения ошибки прогноза оказались достаточно малы и не превышали 0,004, что составляет менее одного процента «голосов».

5. Обсуждение результатов и заключение

В данной работе была изучена прогностическая способность SCARDO-модели. Для этого были использованы эмпирические лонгитюдные данные, состоящие из трех снимков мнений выборки пользователей социальной сети ВКонтакте. Для калибровки параметров модели использовались первые два снимка мнений, а также снимок дружеских связей, в то время как третий снимок мнений использовался для оценки качества прогноза.

Проведенные эксперименты показали, что точность прогноза SCARDO-модели существенно зависит от выбора дискретизирующего отображения, переводящего эмпирические мнения (принадлежащие непрерывному спектру) в модельные (дискретные), поскольку от этого зависит структура таблицы переходов. Слишком маленькие значения m (число элементов азбуки мнений) приводят к тому, что многие эффекты, связанные с изменениями мнений, «затираются» в результате агрегирования и не находят отражения в таблице переходов.

Было показано, что при $m = 2$ и $m = 10$ точность прогноза SCARDO-модели выше, чем у модели постоянного тренда, в то время как для $m = 3$ ситуация обратная. Для случая $m = 5$ статистически значимых различий выявлено не было.

Отметим, что характерная величина ошибки прогноза оказалась достаточно мала: ее значения не превышали 0,004, что составляет менее одного процента «голосов». По-видимому, это объясняется спецификой рассматриваемого фрагмента данных, в котором, согласно проведенным ранее исследованиям [9], мнения пользователей изменяются относительно редко: к примеру было показано, что вероятность изменения мнения за полгода

составляет порядка 0,1 – всего один пользователь из десяти изменил свое мнение. В связи с этим в дальнейших исследованиях необходимо уделить внимание сбору эмпирических данных в рамках большего временного промежутка. Однако необходимо учитывать тот факт, что таблица переходов – ключевой элемент SCARDO-модели, описывающий то, каким образом агенты реагируют на внешнее влияние, – может быть подвержена изменениям под действием событий, происходящих в обществе.

Литература

1. BAIL C.A., ARGYLE L.P., BROWN T.W., BUMPUS J.P. et al. *Exposure to opposing views on social media can increase political polarization* // Proc. of the National Academy of Sciences. – 2018. – Vol. 115. No. 37. – P. 9216–9221.
2. CHATTOE-BROWN E. *Why questions like ‘do networks matter?’ matter to methodology: How Agent-Based Modelling makes it possible to answer them* // Int. Journal of Social Research Methodology. – 2021. – Vol. 24. No. 4. – P. 429–442.
3. COMBS A., TIERNEY G., GUAY B., MERHOUT F. et al. *Reducing political polarization in the United States with a mobile chat platform* // Nature human behaviour. – 2023. – Vol. 7. No. 9. – P. 1454–1461.
4. DEGROOT M.H. *Reaching a consensus* // Journal of the American Statistical association. – 1974. – Vol. 69. No. 345. – P. 118–121.
5. FLACHE A., MAS M., FELICIANI T., CHATTOE-BROWN E. et al. *Models of social influence: Towards the next frontiers* // Journal of Artificial Societies and Social Simulation. – 2017. – Vol. 20. No. 4.
6. GEZHA V.N., KOZITSIN I.V. *The Effects of Individuals’ Opinion and Non-Opinion Characteristics on the Organization of Influence Networks in the Online Domain* // Computers. – 2023. – Vol. 12. No. 6. – P. 116.

7. KEIJZER M., MAS M., FLACHE A. *Polarization on social media: Micro-level evidence and macro-level implications* // Journal of Artificial Societies and Social Simulation. – 2024. – Vol. 27. No. 1. – P. 1–7.
8. KOZITSIN I.V. *A general framework to link theory and empirics in opinion formation models* // Scientific reports. – 2022. – Vol. 12. No. 1. – P. 5543.
9. KOZITSIN I.V. *Opinion dynamics of online social network users: a micro-level analysis* // The Journal of Mathematical Sociology. – 2023. – Vol. 47. No. 1. – P. 1–41.
10. MASTROENI L., VELLUCCI P., NALDI M. *WAgent-based models for opinion formation: A bibliographic survey* // IEEE Access. – 2019. – Vol. 7. – P. 58836–58848.
11. MAS M., BISCHOFBERGER L. *Will the personalization of online social networks foster opinion polarization?* // Available at SSRN 2553436. – 2015.
12. PROSKURNIKOV A.V., TEMPO R. *A tutorial on modeling and analysis of dynamic social networks. Part II* // Annual Reviews in Control. – 2018. – Vol. 45. – P. 166–190.
13. ROBERTSON R.E., GREEN J., RUCK D.J., OGNANOVA K. et al. *Users choose to engage with more partisan news than they are exposed to on Google Search* // Nature. – 2023. – P. 1–7.

PREDICTING PUBLIC OPINION DYNAMICS WITH THE SCARDO-MODEL

Ivan Kozitsin, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, Cand.Sc. (kozitsin.ivan@mail.ru).

Abstract: Over the past 20 years, the theory of agent-based social influence models has been actively developing, a trend which is associated with the need to explain opinion formation processes in the context of the digitalization of communication channels and the intensification of information exchange processes. However, the practical side of this theory remains poorly studied. The key reason for this is the difficulties in calibrating model parameters and thus constructing an empirical foundation. This paper validates the SCARDO-model of opinion formation using empirical longitudinal data from the social network VKontakte. The data include three opinion snapshots of a large-scale sample of VKontakte users and a snapshot of their friendship connections. The model parameters are calibrated on the first two snapshots, whereas the third one is used to check the accuracy of the model's forecast regarding the populations of opinion fractions at the next time moment. The constant trend model serves as a benchmark. The analysis performed shows that, depending on the method of parameter calibration, the prediction of the SCARDO-model can be more or less accurate than those of the constant trend model. At the same time, changes in public opinion in the dataset at hand (despite being sufficient to calibrate the model parameters) are small from the macro-scale point of view and, as a result, the typical value of the forecast error does not exceed one percent of «votes».

Keywords: opinion formation models, SCARDO-model, social networks, validation.

УДК 519.7

ББК 22.18

DOI: 10.25728/ubs.2024.108.7

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

Поступила в редакцию 28.02.2024.

Дата опубликования 31.03.2024.

ЛОГИСТИЧЕСКИЕ МОДЕЛИ ЖИЗНЕННОГО ЦИКЛА ТЕХНОЛОГИЙ КАК ИНСТРУМЕНТ ОЦЕНКИ ЭФФЕКТИВНОСТИ ЗАТРАТ НА НИОКР

Нижегородцев Р. М.¹, Рослякова Н. А.², Горидько Н. П.³
(ФГБУН Институт проблем управления
им. В.А. Трапезникова РАН, Москва)

Исследование жизненных циклов технологий, их квантификация и определение переломных точек является актуальной научной задачей. Наиболее обоснованной теоретической конструкцией исследования динамики жизненного цикла технология является логистическая кривая. Основой является сопоставление динамических рядов затрат и эффектов и выявление их взаимной связи, допускающей аппроксимацию при помощи логистической кривой. Данная статья посвящена расчету логистических трендов, выражающих связь годовых данных валовой выручки и затрат на НИОКР для компании Яндекс в 2009–2021 гг. На основе аппроксимации, проведенной методами нелинейного регрессионного анализа, рассчитываются значения максимальной интегральной эффективности и максимальной дифференциальной (точечной) эффективности затрат на НИОКР по каждому из рассматриваемых временных промежутков. Исследование логистических трендов и представленные инструментарий и результаты позволяют выявить периоды доминирования той или иной технологической (или организационно-управленческой) парадигмы в жизни определенной высокотехнологичной компании на основе сопоставления совокупной и/или мгновенной эффективности для разных периодов развития компании. Кроме того, предложенные результаты актуальны для оценки перспектив технологических сдвигов в развитии высокотехнологичной компании, а именно, определения уровня технологического или стоимостного верхнего предела, выражаемых верхней горизонтальной асимптотой соответствующей логисты.

Ключевые слова: логистические модели, экономическая эффективность, высокотехнологичные компании, нелинейный регрессионный анализ.

1. Введение

Жизненный цикл технологии можно описать при помощи различных количественных моделей, выражающих связь между затратами и результатами инновационной деятельности. Очень

¹ Роберт Михайлович Нижегородцев, д.э.н. (bell44@rambler.ru).

² Наталья Андреевна Рослякова, к.э.н. (na@roslyakova24.ru).

³ Нина Павловна Горидько, к.э.н. (horidko@mail.ru).

распространенной (можно сказать традиционной) является группа логистических моделей, подразумевающих, что зависимость эффекта от затрат выражается логистической кривой.

При этом характер логистической связи не зависит от того, в каких единицах выражены эффект и затраты: это могут быть стоимостные единицы или натуральные. Кроме того, это могут быть полные, совокупные затраты и эффекты, а могут быть частичные (затраты какого-либо отдельного ресурса, выражаемые не обязательно в денежных единицах, а, например, в человеко-часах или тонно-километрах). Масштабы единиц измерения по координатным осям при этом могут быть совершенно различны, но логистический вид соответствующей зависимости между затратами и эффектом не претерпит изменений [5].

При условии, что затраты и эффект являются количественно измеримыми при помощи статистических наблюдений, составляющих динамические ряды случайных величин, можно построить соответствующие логистические кривые, используя инструмент регрессионных моделей. На основе построенных логистических трендов (при условии, что они выражают статистически значимые связи и все входящие в уравнение регрессоры значимы с надлежащей степенью доверия) можно оценить максимальный уровень совокупной (интегральной) эффективности затрат на рассматриваемом временном промежутке и максимальный уровень мгновенной (краткосрочной, дифференциальной) эффективности на том же промежутке. Таким образом, цель данной статьи заключается в том, чтобы предложить метод трендовой аппроксимации для оценки эффективности затрат на НИОКР и апробировать этот метод на примере одной из высокотехнологичных российских компаний.

2. Постановка задачи

Пусть логистическая кривая выражает связь между затратами и результатами хозяйственной деятельности агента (фирмы) на определенном временном промежутке (рис. 1). В таком случае различные прямые, проходящие через начало координат и пересекающие данную логисту, имеют различные углы наклона к оси

абсцисс. Максимальную интегральную эффективность характеризует касательная q_1 , имеющая максимальный угол наклона α_1 к оси абсцисс, и тангенс этого угла равен максимально возможному значению интегральной эффективности усилий агента на данном периоде времени.

Максимальная мгновенная эффективность достигается в точке, где логистическая кривая растет максимально быстро, – это точка перегиба, корень второй производной соответствующей функции. На рис. 1 этому соответствует точка A_2 . Касательная q_2 к логистической кривой, проходящая через данную точку, наклонена к оси абсцисс под углом α_2 , тангенс которого и равен максимальной мгновенной эффективности усилий агента на данном временном промежутке.

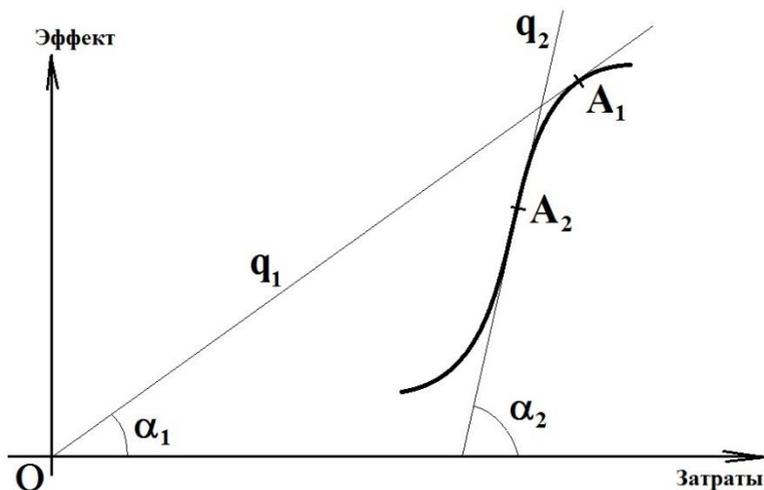


Рис. 1. Интегральная и дифференциальная эффективность инновационных процессов

Если исходить из того, что на рис. 1 представлено стоимостное выражение затрат и эффекта, то положение верхней горизонтальной асимптоты логистической кривой зависит не только от технологических параметров, но и от емкости рынка, на который выносятся инновационный продукт.

3. Аппроксимационные методы оценки эффективности затрат

Высокотехнологичный сектор экономики традиционно характеризуется капиталоемкостью и значительным временным горизонтом, который необходим для реализации инноваций в практической плоскости. Всё это обуславливает существенную неопределённость процесса реализации затрат на НИОКР и ставит вопрос поиска методов оценки их эффективности. На определение этих методов имеют влияние многие характеристики компаний, например, масштаб осуществляемых инвестиций и получаемых результатов. Так, в работах [7, 15] авторы указывают, что крупные высокотехнологичные компании имеют больший потенциал привлечения финансовых и человеческих ресурсов по сравнению с небольшими компаниями. По этой причине малые компании чаще прибегают усилиям по поиску внешних источников открытого доступа для снятия своих ограничений. Однако такие решения далеко не всегда являются наиболее эффективными, что проявляется в недоступности всего необходимого функционала, недостаточной адаптивности под требования и характеристики конкретной компании и пр. Это в свою очередь накладывает ограничения и на достижимые для малых высокотехнологичных компаний параметров эффективности. Помимо этого большое значение имеет и работа в конкретной отрасли или с определённой технологией, что также определяет вариацию в уровне отдачи. Так в работе [14] автор доказывает, что если компания концентрируется на развитии одной инновационной технологии, то использование различных источников затрат ведёт к падению отдачи и эффективности в случае сервисной специализации компании. Для производственных компаний, напротив, если компания концентрируется на развитии одной инновационной технологии, то использование различных источников затрат ведёт к возрастанию отдачи и эффективности, что, вероятно, можно объяснить большей предрасположенностью производственных инноваций к масштабированию и получению соответствующего эффекта.

Наиболее часто такие исследования проводятся на данных отдельных компаний, собранных в массивы на страновом или отраслевом уровне. Примерами таких работ являются [10, 16–17,

19–20]. Традиционной теоретической рамкой таких исследований является анализ жизненных циклов технологий, выраженных через затраты на НИОКР. Трансформируясь в успешные инновации, они увеличивают экономическую отдачу (выручку или прибыль) компании со значительным ускорением, что обычно моделируется через логистическую кривую (S-образную кривую, S-curve) [2, 13, 18]. Она позволяет представить динамику процесса накопления результатов процесса развития инноваций. Однако существенным ограничением всех этих исследований является необходимость набора определённого массива данных.

Во многих случаях подход, связанный с поточечным оцениванием и формированием массива данных, является невозможным. Для высокотехнологичных компаний часто характерно доминирование на рынке, которое выражается через значительный масштаб деятельности, несопоставимый с оборотом ближайших конкурентов, и технологическое превосходство, что не позволяет оценивать другие компании наравне с ними. Примерами таких компаний являются: Яндекс в России, Tencent в Китае, Alphabet, Amazon, Nvidia в США, Samsung в Южной Корее. Традиционным инструментарием работы с такими исследовательскими объектами являются имитационные модели, основанные на симуляционной аппроксимации, реализуемой на базе инструментария DEA (data envelopment analysis, оболочечный анализ данных – анализ среды функционирования), который использует непараметрические методы, позволяющие определить границу эффективных состояний системы (оценка оптимальных уровней для рассматриваемого набора затратных и результирующих параметров), примерами таких работ являются [9, 21]. Альтернативой DEA и наибольшим приближением к предлагаемому нами методическому подходу является параметрический метод оценивания стохастической границы (SFA, stochastic frontier analysis), основывающийся на производственной функции и исходящий из предположения, что случайная ошибка распределена симметрично, а составляющие неэффективного поведения компании распределены несимметрично, примеры работы с использованием данного инструментария [3–4, 8, 12].

Предложенный нами подход оценки эффективности также относится к аппроксимационным, с той разницей, что DEA и SFA

базируются на имитационной аппроксимации, тогда как в нашей работе мы имеем дело с трендовым методом аппроксимации. При этом в отличие от метода SFA, который также относится к параметрическим, где исходные условия задаются случайным образом, предложенный нами подход основывается на предположении об определенном экзогенно задаваемом виде тренда, выражающего связь между затратами и эффектом наукоемких видов деятельности. В соответствии с теорией инноваций, в роли трендов такого рода могут выступать различные виды логистических кривых, параметры которых подлежат количественной оценке на основе регрессионного анализа наблюдаемых рядов статистических данных.

4. Методика расчётов

Аппроксимационный тренд ищется в виде регрессионного уравнения стандартной логистической кривой, при помощи которой традиционно описываются жизненные циклы технологий:

$$(1) \quad y(x) = a + \frac{b}{c + e^{-wx}}.$$

Калибруя данную модель методом наименьших квадратов и находя ее параметры, мы проследим взаимосвязь между расходами на НИОКР и валовой выручкой одной из широко известных высокотехнологичных компаний, а именно – компании Яндекс.

Некоторая смысловая некорректность заключается в том, что рассматриваемая модель улавливает зависимость валовой выручки компании исключительно от объема ее расходов на НИОКР, тогда как на самом деле эта выручка выступает итогом синфазного действия большого числа факторов, не сводимых к затратам на НИОКР. Данная некорректность является в каком-то смысле традиционной для кейнсианской школы, которая рассматривает приращение ВВП макросистемы как прямое следствие прироста инвестиций, тогда как разогрев любой составной части совокупного спроса приводит к росту ВВП, и поэтому каждая часть совокупного спроса имеет собственный мультипликатор, и существуют вполне корректные методики совместного расчета этих мультипликаторов (см., например, [1]).

В то же время отмеченная смысловая некорректность не критична с точки зрения получаемых результатов, поскольку влияние различных факторов, отличных от затрат на НИОКР, может быть учтено в модели (1) в форме констант, причем как аддитивным, так и мультипликативным образом: путем включения в состав параметра a различных факторов, не коллинеарных с динамическим рядом x , и посредством синхронного умножения входящих в данную формулу коэффициентов a и b на один и тот же сомножитель.

4.1. ПОСТРОЕНИЕ ЛОГИСТИЧЕСКИХ ТРЕНДОВ

Взятая нами в качестве примера компания Яндекс представлена годовыми данными с 2009 по 2021 гг., которые взяты из годовых финансовых отчетов [6] и приведены на основании дефляторов ВВП, представленных на сайте Мирового банка [11], к уровню цен 2021 года (таблица 1).

Таблица 1. Данные о затратах на НИОКР и выручке компании Яндекс в 2009–2021 гг.

Год	Расходы на НИОКР (R&D), млрд. руб. в ценах 2021 г. x	Выручка, млрд. руб. в ценах 2021 г. y
2009	3,299	17,939
2010	4,032	24,001
2011	5,194	33,510
2012	6,822	45,688
2013	8,893	60,561
2014	12,833	74,082
2015	18,570	82,871
2016	21,446	103,023
2017	24,513	122,693
2018	27,208	153,736
2019	34,200	205,437
2020	42,278	254,248
2021	48,500	356,200

В результате график, отображающий взаимосвязь указанных величин, имеет вид, представленный на рис. 2.



Рис. 2. Соотношение расходов на НИОКР и выручки компании Яндекс, 2009–2021 гг.

На показанном выше рисунке мы предполагаем наличие двух разных логистических кривых: одна – с 2009 по 2015 гг., а вторая – примерно с 2014 по 2020 гг. Точка, соответствующая 2021 году, скорее всего, является началом следующей логисты.

Исходя из этих предположений и пользуясь инструментами ППП Statistica, мы построили два разных логистических тренда и, соответственно, нашли значения переменных формулы (1). При этом количество наблюдений в каждой из кусочных выборок небольшое и не позволяет эффективно (т.е. с минимальной вероятностью ошибки – ее показывает уровень p -level) определить значение всех четырех параметров. Так, для периода 2009–2015 гг. с помощью пользовательской функции нелинейного оценивания в Statistica мы получили вырожденную функцию. Поэтому некоторые значения коэффициентов мы подобрали путем предварительного анализа эмпирических данных; так, с целью сокращения числа оцениваемых параметров была предварительно построена линейная функция:

$$(2) \quad y(x) = 11,472 + 4,332x.$$

Она указывает на возможность оценки параметров линейной связи между расходами на НИОКР и выручкой компании Яндекс, причем, исходя из значения коэффициента детерминации, 91% изменения выручки объясняется вариацией расходов на НИОКР. Полученная модель адекватно описывает исходные данные, и ее параметры значимы: регрессор – на уровне значимости 0,01,

а свободный член – на уровне значимости 0,15. Тем не менее именно свободный член нас интересует как точка пересечения аппроксимирующей функции с осью ординат, в нашем случае – это параметр a из формулы (1). Остальные коэффициенты регрессии определяются с помощью нелинейной функции методом Гаусса – Ньютона; после 42 итераций получена следующая формула:

$$(3) \quad y(x) = 11,472 + \frac{2,24}{0,032 + e^{-0,489x}}$$

Для нее $R^2 = 99,85\%$, F -критерий значим на уровне значимости 0,01, все параметры значимы на уровне погрешности 0,05 (см. таблицу 2).

Таблица 2. Статистические характеристики модели (2)

Модель: $y = 11,472 + b/(c + \exp((-w*x)))$						
Уровень значимости: 95,0%						
	Оценка	Стандарт ошибка	t -знач. $cc = 4$	p -знач.	Ниж. дов. предел	Верх. дов. предел
b	2,24023	0,758369	2,954014	0,041802	0,134663	4,345802
c	0,03236	0,010558	3,064661	0,037488	0,003043	0,061669
w	0,48926	0,054305	9,009560	0,000840	0,338489	0,640039

Логистическая кривая, выражаемая уравнением (3), изображена на рис. 3.

Предполагая, что в период 2014–2015 гг. произошел переход на другую логистическую кривую, построим следующую логисту по наблюдениям с 2014 г., при этом попытка включить в данную выборку наблюдение за 2021 г. успеха не имела: параметры такой модели оказались незначимыми, и это косвенно подтверждает предположение о дальнейшем технологическом сдвиге: о том, что наблюдения за 2019–2021 годы характеризуют начало новой, следующей логистической кривой.

Итак, логистическая функция вида (1) по данным 2014–2020 гг., построенная с помощью нелинейной функции методом Гаусса – Ньютона, в результате 25 итераций имеет вид

$$(4) \quad y(x) = 58,806 + \frac{1,38}{0,006 + e^{-0,173x}}$$

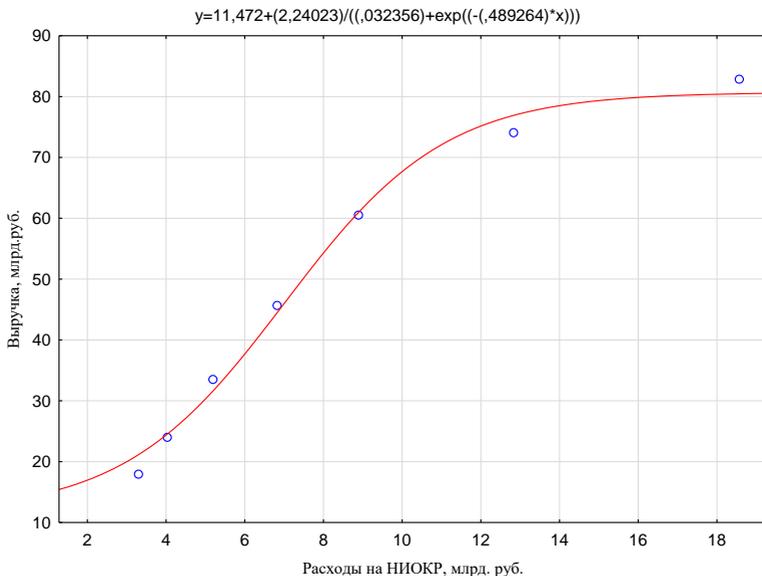


Рис. 3. График функции (3), 2009–2015 гг.

Анализируя значимость ее параметров, мы пришли к выводу, что коэффициенты a и w значимы на уровне значимости 0,05, b и c оказались незначимыми. Соответственно, мы попробовали зафиксировать параметр a таким же образом, как и для первой серии наблюдений, вследствие чего получили модель

$$(5) \quad y(x) = 58,81 + \frac{1,381}{0,006 + e^{-0,172x}}$$

Как видим, по сравнению с формулой (4) в функции (5) немного уточнены коэффициенты регрессии (впрочем, крайне незначительно), а сама она характеризуется высокой объясняющей способностью ($R^2 = 99,94\%$), адекватна исходным данным, и ее параметры значимы как минимум на уровне погрешности 0,1, см. таблицу 3.

График функции (5) представлен на рис. 4.

Если наложить два построенных логистических тренда на один график, получим изображение, представленное на рис. 5.

Таблица 3. Статистические характеристики модели (5)

Модель: $y = 58,81 + b / (c + \exp((-w*x)))$						
Уровень значимости: 95,0%						
	Оценка	Стандарт ошибка	t -знач. $ss=4$	p -знач.	Ниж. дов. предел	Верх. дов. предел
b	1,380780	0,517163	2,66991	0,055811	-0,05509	2,816655
c	0,006432	0,002183	2,94714	0,042089	0,00037	0,012492
w	0,172494	0,015428	11,18062	0,000364	0,12966	0,215329

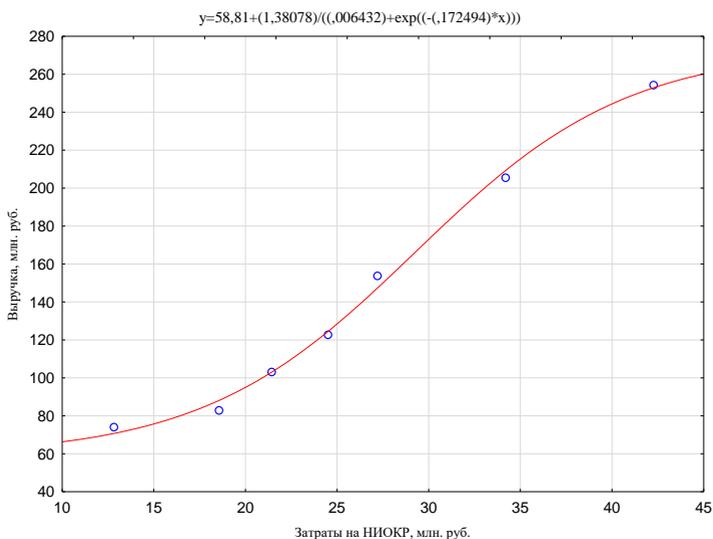


Рис. 4. График функции (5), 2014–2020 гг.

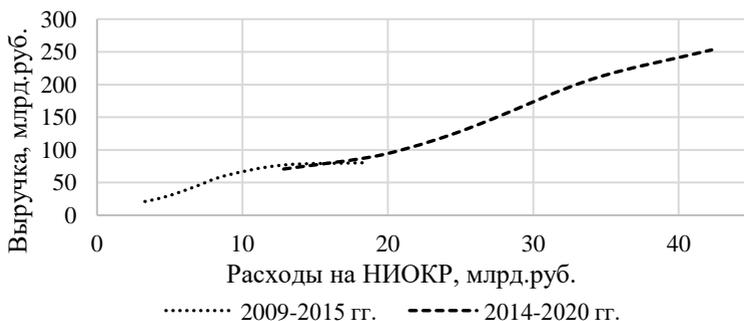


Рис. 5. Логистические кривые соотношения затрат на НИОКР и выручки компании Яндекс, 2009–2020 гг.

4.2. ПОСТРОЕНИЕ ЛОГИСТИЧЕСКИХ ТРЕНДОВ

На следующем этапе нами была поставлена задача поиска параметров дифференциальной и интегральной эффективности инновационных процессов, в соответствии с логикой рис. 1.

Поскольку линия q_1 на рис. 1 представляет собой прямую, проходящую через начало координат, то в ее уравнении свободный член равен нулю. Ее вид представлен в правой части первого равенства в системе уравнений (6). Первая производная будет представлять моментное изменение функции, описывающей прямую q_1 (второе равенство в системе уравнений (6)):

$$(6) \quad \begin{cases} y(x) = a + \frac{b}{c+e^{(-\omega x)}} = k_1 x, \\ y'(x) = \frac{\omega b e^{-\omega x}}{(c+e^{-\omega x})^2} = k_1. \end{cases}$$

Решив систему уравнений, мы получаем параметр k_1 , который представляет собой коэффициент максимальной совокупной эффективности инвестиций в НИОКР за рассматриваемый период. Для каждого логистического тренда имеются две касательные, проходящие через начало координат: одна касается логисты в верхней ее части и характеризуется максимальным значением угла наклона α_1 (см. рис. 1), другая касается той же логисты в ее нижней части и характеризуется минимальным значением угла наклона к оси абсцисс.

Таким образом, решив систему уравнений (6) методом приближенных вычислений для двух построенных логистических трендов (2009–2015 гг. и 2014–2020 гг.), мы получаем по два значения каждого параметра (таблица 4), одно из которых представляет точку минимума функции совокупной эффективности, другое – точку ее максимума.

Таблица 4. Результаты решения системы уравнений (6) для компании Яндекс

Решение системы уравнений (6) для компании Яндекс в 2009–2015 гг.		Решение системы уравнений (6) для компании Яндекс в 2014–2020 гг.	
$k_1 = 6,030$	$k_1 = 6,857$	$k_1 = 4,740$	$k_1 = 6,183$
$x = 4,562$	$x = \mathbf{8,924}$	$x = 19,232$	$x = \mathbf{36,872}$
$y = 27,512$	$y = \mathbf{61,192}$	$y = 91,155$	$y = \mathbf{227,990}$
$k_1 = 6,030$	$k_1 = 6,857$	$k_1 = 4,740$	$k_1 = 6,183$

Тем самым интегральная эффективность затрат на НИОКР за период 2009–2015 гг. колеблется в пределах 6,03–6,86, а за период 2014–2020 гг. – в пределах 4,74–6,18. Координаты точек, соответствующих максимальной интегральной эффективности затрат на НИОКР за рассматриваемые периоды, выделены полужирным шрифтом в таблице 4.

Уравнение касательной к логистическому тренду, построенному по данным периода 2009–2015 гг., с максимальным углом наклона к оси абсцисс, таково: $y = 6,857x$.

Уравнение касательной к логистическому тренду, построенному по данным периода 2014–2020 гг., с максимальным углом наклона к оси абсцисс, таково: $y = 6,183x$.

4.3. РАСЧЕТ ДИФФЕРЕНЦИАЛЬНОЙ ЭФФЕКТИВНОСТИ

Для получения параметров дифференциальной эффективности необходимо перейти к решению системы уравнений со второй производной, а именно:

$$(7) \begin{cases} y(x) = a + \frac{b}{c+e^{-\omega x}} = k_2 x + p, \\ y'(x) = \frac{\omega b e^{-\omega x}}{(c+e^{-\omega x})^2} = k_2, \\ y''(x) = \frac{-b\omega^2 e^{-\omega x} + \frac{2b\omega^2 e^{-\omega x}}{c+e^{-\omega x}}}{(c+e^{-\omega x})^2} = 0. \end{cases}$$

Решив систему уравнений, мы получаем параметр k_2 , который представляет собой коэффициент максимальной (за рассматриваемый период) дифференциальной эффективности, а x характеризует объем затрат на НИОКР, при котором достигается максимальная скорость получения результата (т.е. максимально быстрый прирост выручки y).

Решив систему уравнений (7) методом приближенных вычислений для двух найденных логистических трендов (2009–2015 гг. и 2014–2020 гг.), мы получаем следующие значения каждого параметра (таблица 5).

Таблица 5. Результаты решения системы уравнений (7) для компании Яндекс

Решение системы уравнений (7) для компании Яндекс в 2009–2015 гг.	Решение системы уравнений (7) для компании Яндекс в 2014–2020 гг.
$k_2 = 7,528$	$k_2 = 8,229$
$p = -5,824$	$p = -71,882$
$x = 8,429$	$x = 33,274$
$y = 57,630$	$y = 201,928$

Уравнение прямой, касающейся логистической кривой, построенной по данным периода 2009–2015 гг., в точке ее перегиба, таково: $y = 7,528x - 5,824$.

Уравнение прямой, касающейся логистической кривой, построенной по данным периода 2014–2020 гг., в точке ее перегиба, таково: $y = 8,229x - 71,882$.

5. Заключение: выводы и обсуждение результатов

Еще раз подчеркнем, что найденные коэффициенты k_1 и k_2 представляют собой тангенсы углов соответственно α_1 и α_2 на рис. 1. Чем более пологим является логистический тренд в фазе своего максимально быстрого роста, тем меньше количественное различие между максимальной мгновенной эффективностью, достигаемой на соответствующем промежутке времени, и максимальной совокупной эффективностью на том же промежутке.

Итак, для временного периода 2009–2015 гг. значение максимальной интегральной эффективности затрат на НИОКР для компании Яндекс составило 6,86 и было достигнуто в 2013 году (этот факт вытекает из годовых значений динамического ряда x , приведенных в таблице 1), значение максимальной мгновенной эффективности затрат на НИОКР составило 7,53 и приходилось на промежуток 2012–2013 гг.

Для временного периода 2014–2020 гг. значение максимальной интегральной эффективности затрат на НИОКР составило 6,18 и приходилось на 2019–2020 гг., значение максимальной мгновенной эффективности составило 8,23 и приходилось на промежуток 2018–2019 гг.

Для того чтобы просто оценить операционную эффективность (отдачу от затрат определенного вида) той или иной компании и выявить, в какие годы она была выше или ниже, никакие трендовые модели (логистические в том числе), собственно, не нужны. Достаточно подготовить годовые данные затрат и эффекта, приведенные к сопоставимым ценам одного и того же года, и разделить эффект на затраты – получаем динамический ряд показателя, характеризующего интегральную эффективность усилий компании в том или ином направлении, по которому рассматриваются затраты. Решая ту же задачу в годовых приращениях, получаем динамический ряд показателя дифференциальной (точечной, мгновенной) эффективности действий этой компании.

Логистические тренды нужны для того, чтобы выявить периоды доминирования той или иной технологической (или организационно-управленческой) парадигмы в жизни определенной высокотехнологичной компании и иметь возможность сопоставлять между собой соответствующие временные промежутки, находя в рамках каждого из них точки максимальной за данный период совокупной и/или мгновенной эффективности.

Кроме того, применение поточечных методов оценки оправдано в тех случаях, когда заведомо можно абстрагироваться от наличия ошибок измерений и вычислений в динамических рядах статистических наблюдений. В задачах экономического анализа это чаще всего не так. Трендовые методы оценки хороши, в частности, тем, что они содержат инструмент сглаживания выбросов различного происхождения и потому позволяют получить надежную оценку как средней, так и предельной эффективности затрат.

Логистические модели позволяют также оценить перспективы технологических сдвигов в развитии определенной высокотехнологичной компании, в том числе высоту технологических либо стоимостных верхних пределов, выражаемых верхней горизонтальной асимптотой соответствующей логисты. Построив логистический тренд по первым нескольким точкам, можно также спрогнозировать максимально возможную интегральную и мгновенную отдачу, которой исследуемая компания вправе ожидать от внедрения соответствующей технологии.

Для других компаний эффективность внедрения той же самой технологии может оказаться совершенно иной. Это зависит, в частности, от того, в какой мере оргструктура конкретной компании и существующие в ней алгоритмы принятия решений способны выступить проводниками возможностей, предоставляемых наличием данной технологии.

Литература

1. ГОРИДЬКО Н.П., НИЖЕГОРОДЦЕВ Р.М. *Регрессионная оценка мультипликаторов совокупного спроса по агрегатам* // Анализ, моделирование, управление, развитие социально-экономических систем: Сб. науч. трудов XIV Всеросс. с междунар. уч-ем шк.-симп. АМУР-2020. – Симферополь, 2020. – С. 109–110.
2. ДХАКАЛ Т., МИН К.С. *Макроанализ и прогноз перспектив распространения электромобилей* // Форсайт – 2021. – Т. 15, №1. – С. 67–73.
3. ИЛБИНА Е.А., САРАЕВ Л.А. *Стохастические модели динамики максимальной и оптимальной прибыли производственного предприятия, внедряющего технологические инновации* // Вестник Самарского университета. Экономика и управление. – 2023. – Т. 14, №2. – С. 197–213.
4. ИПАТОВА И.Б., ПЕРЕСЕЦКИЙ А.А. *Техническая эффективность предприятий отрасли производства резиновых и пластмассовых изделий* // Прикладная эконометрика. – 2013. – Т. 32, №4. – С. 71–92.
5. НИЖЕГОРОДЦЕВ Р.М. *Экономика инноваций: Учебное пособие*. – М.: РУСАЙНС, 2016. – 154 с.
6. *Яндекс: годовые финансовые отчеты* / Смартлаб [Электронный ресурс]. – Режим доступа: <https://smartlab.ru/q/YNDX/f/y/GAAP/>.
7. D'ANGELO A., BARONCELLI A. *An Investigation Over Inbound Open Innovation in SMEs: Insights from an Italian Manufacturing Sample* // Technology Analysis & Strategic Management. – 2020. – Vol. 32, No. 5. – P. 542–560.

8. DOAN A.-T., KHAN A., HOLMES S., TRAN T. *SMEs' efficiency in a transitional economy: does innovation and public support schemes matter?* // Journal of the Asia Pacific Economy. – 2023. – Vol. 28, No. 3. – P. 1029–1060.
9. DOBRZANSKI P. *The efficiency of spending on R&D in Latin America region* // Applied Economics. – 2020. – Vol. 52, No. 46. – P. 5020–5034.
10. FORÉS B., CAMISÓN C. *Does Incremental and Radical Innovation Performance Depend on Different Types of Knowledge Accumulation Capabilities and Organizational Size?* // Journal of Business Research. – 2016. – Vol. 69. – P. 831–848.
11. *Indicators* / WorldBank. [Электронный ресурс]. – Режим доступа: https://data.worldbank.org/indicator/NY.GDP.DEFL.KD.ZG?end=2021&name_desc=true&start=2000&view=map&year=1961.
12. KIM K., LEE J., LEE C. *Which innovation type is better for production efficiency? A comparison between product/service, process, organisational and marketing innovations using stochastic frontier and meta-frontier analysis* // Technology Analysis & Strategic Management. – 2023. – Vol. 35, No. 1. – P. 59–72.
13. MIAMO C.W., BERTIN C.P.K. *The effect of external knowledge on innovation capacity of SMES: Does the source of knowledge matter?* // African Journal of Science, Technology, Innovation and Development. – 2022. – Vol. 14, No. 6. – P. 1655–1666.
14. RADICIC D. *Breadth of external knowledge search in service sectors* // Business Process Management Journal. – 2021. – Vol. 27, No. 1. – P. 230–252.
15. RADICIC D., ALKARAAN F. *Relative effectiveness of open innovation strategies in single and complex SME innovators* // Technology Analysis & Strategic Management. – 2022. – DOI: 10.1080/09537325.2022.2130042.
16. RODRÍGUEZ-POSE A., DI CATALDO M. *Quality of Government and Innovative Performance in the Regions of Europe* // Journal of Economic Geography. – 2015. – Vol. 15, No. 4. – P. 673–706.
17. SERRANO-BEDIA A.M., LÓPEZ-FERNÁNDEZ M., GARCÍA-PIQUERES G. *Complementarity between innovation knowledge sources: Does the innovation performance measure*

- matter?* // BRQ Business Research Quarterly. – 2018. – Vol. 21, No. 1. – P. 53–67.
18. UGUR M., VIVARELLI M. *Innovation, firm survival and productivity: the state of the art* // Economics of Innovation and New Technology. – 2021. – Vol. 30, No. 5. – P. 433–467.
 19. WANG P., CEN C. *Does digital economy development promote innovation efficiency? A spatial econometric approach for Chinese regions* // Technology Analysis & Strategic Management. – 2022. – DOI: 10.1080/09537325.2022.2065980.
 20. YEH M.-L., CHU H.-P., SHER P. J., CHIU Y.-C. *R&D intensity, firm performance and the identification of the threshold: fresh evidence from the panel threshold regression model* // Applied Economics. – 2010. – Vol. 42, No. 3. – P. 389–401.
 21. ZHU Y., WANG Z., YANG J., ZHANG ZH. *Evaluating performance of innovation resource allocation in industrial enterprises: an improved two-stage DEA model* // Technology Analysis & Strategic Management. – 2022. – DOI: 10.1080/09537325.2022.2157254.

LOGISTIC MODELS OF THE TECHNOLOGY LIFE CYCLE AS A TOOL FOR ASSESSING THE EFFICIENCY OF R&D EXPENDITURES FOR KNOWLEDGE INTENSIVE COMPANIES

Robert Nizhegorodtsev, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, Doctor of Science, professor (bell44@rambler.ru).

Natalia Roslyakova, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, Cand.Sc. (na@roslyakova24.ru).

Nina Goridko, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, Cand.Sc. (horidko@mail.ru).

Abstract: The study of the life cycles of technologies, their quantifications and the definition of breakpoints is an urgent scientific task. The most well-founded theoretical construction of the technology life cycle dynamics study is the logistic curve. The basis is a comparison of the dynamic series of costs and effects. The paper deals with the calculation of logistics trends expressing the relationship between annual data of gross revenue (effects) and R&D expenditures for Yandex in 2009-2021 (costs). Based on the approximation carried out by methods of nonlinear regression analysis, the values of maximum integral efficiency and maximum differential (point) efficiency

of R&D expenditures for each of the considered time intervals are calculated. The study of logistics trends and the presented tools and results allow us to reveal the periods of dominance of one or another technological (or organizational and managerial) paradigm in the life of a certain high-tech company based on a comparison of aggregate and/or instantaneous efficiency for different periods of the company's development. In addition, the proposed results are relevant for assessing the prospects of technological shifts in the development of a high-tech company, namely, determining the level of technological or cost upper limit, expressed by the upper horizontal asymptote of the corresponding logistics.

Keywords: logistics models, economic efficiency, knowledge intensive companies, nonlinear regression analysis.

УДК 330.341 + 330.43

ББК 65.011.4

DOI: 10.25728/ubs.2024.108.8

*Статья представлена к публикации
членом редакционной коллегии В.В. Ключковым.*

Поступила в редакцию 27.10.2023.

Опубликована 31.03.2024.

ИСПОЛЬЗОВАНИЕ ВИРТУАЛЬНЫХ СТРУКТУР ПРИ ФОРМИРОВАНИИ СЦЕНАРНО-КОГНИТИВНЫХ МОДЕЛЕЙ НА ОСНОВЕ ИСПОЛЬЗОВАНИЯ ЭКСПЕРТНЫХ ЗНАНИЙ

Шульц В. Л.¹

(Центр исследования проблем безопасности РАН, Москва)

Чернов И. В.²

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

Работа посвящена исследованию проблемы повышения адекватности формирования сценарно-когнитивных моделей на основе экспертных знаний в ограниченном пространстве факторов. Одной из важнейших задач при формировании сценарно-когнитивной модели на основе экспертных знаний является проблема учет совокупного влияния внешней среды, т.е. тех факторов, которые остались за пределами структуры модели, однако оказывающие влияние на достижение необходимой точности результатов моделирования. При построении сценарно-когнитивных моделей сложных социально-экономических и политических систем обычно приходится применять существенное упрощение, которое заключается в том, чтобы все многообразие факторов и связей между ними заключить в относительно простой и обозримой модели. Качество модели, построенной на основе объединения экспертных данных, должно определяться адекватностью представлению образа реального объекта или обстановки. Следовательно, при формировании моделей с использованием экспертных знаний необходимо также «экспертно замыкать» структуру модели некоторыми виртуальными подструктурами, которые способны генерировать определенные сигналы, отражающие влияние внешней среды. Приведены типовые сигналы, имитирующие воздействия внешней среды. Представлены типовые структуры экспертной идентификации влияния внешней среды на факторы сценарных моделей. Приведена общая схема сценарно-когнитивной модели, сформированная на основе экспертных знаний и состоящая из множества реальных факторов сложной системы и виртуальных структур.

Ключевые слова: сценарно-когнитивное моделирование, адекватность, сценарная модель, сценарный анализ, внешняя среда, виртуальные факторы.

¹ Владимир Леопольдович Шульц, д.фил.наук, профессор, член-корр. РАН (9380752@mail.ru).

² Игорь Викторович Чернов, к.т.н., в.н.с. (ichernov@gmail.com).

1. Введение

Сценарный подход предполагает исследование процессов развития крупномасштабных систем, изучение сложных многоаспектных проблем, синтез и комплексный анализ сценариев рационального в рамках заданных целевых установок развития таких систем, обеспечивая тем самым эффективную поддержку процессов подготовки, принятия и реализации управленческих решений различного уровня.

Сценарии для лица принимающего решения (ЛПР), а также аналитиков являются эффективным и наглядным способом представления прогнозной информации об альтернативных направлениях развития социально-экономических систем, способствующим выработке эффективных управленческих решений в условиях рисков и неопределенностей [3].

Важнейшим качеством сценарной модели, как и прочих моделей, является адекватное представление образа реального объекта или обстановки. Когда мы говорим о сложных социально-экономических, политических, информационных системах, а чаще всего о комплексных сложных системах, включающих все эти и прочие аспекты, наиболее трудной задачей является выделение множества факторов модели.

В работе рассматривается проблема повышения адекватности сценарных моделей с применением методов сценарно-когнитивного моделирования [3] в ограниченном множестве факторов.

2. Общая схема построения сценарно-когнитивных моделей на основе экспертных данных

Одной из важнейших задач при формировании сценарной модели на основе экспертных знаний является проблема включения в нее ограниченного множества факторов, без ущерба для точности и адекватности. При построении моделей подобных систем обычно приходится применять существенное упрощение, которое заключается в том, чтобы все многообразие факторов и связей между ними заключить в относительно простой и обобщимой модели.

Очевидно, что далеко не всегда модель, построенная при помощи такой группы факторов, способна адекватно описывать поведение сложной системы или развитие обстановки, поскольку сложная система с точки зрения формирования модели является открытой, вследствие чего существует сложная сеть взаимовлияний между факторами как включёнными в модель, так и не включёнными. Однако невозможно расширять модель бесконечно, включая в нее все новые и новые факторы внешней среды, привлекая все возрастающее количество экспертов в различных областях. Внешняя среда сложных систем может оказаться неисчерпаемой. Второй стороной процесса расширения множества факторов является, как ни странно, потеря адекватности модели, поскольку все труднее согласовывать экспертные знания, устанавливать причинно-следственные и иные связи и отслеживать их корректность, например дублирование. Следовательно, важнейшей проблемой моделирования является вопрос, как построить для подобной сложной системы понятную и практичную модель, включающую ограниченное число факторов, но как можно более адекватно отражающую происходящие в этой системе динамические процессы.

С этой целью в сценарную модель вводятся ряд виртуальных факторов и структур, отражающих возможные направления влияния внешней среды. На рис. 1 представлена общая схема построения сценарных моделей на основе экспертных знаний.

Описательная модель проблемы представляет собой базу данных накапливаемой и изменяемой информации о существующих и новых ситуациях в моделируемой сфере, в которой отражается информация по следующим взаимосвязанным компонентам:

- основные тенденции развития моделируемого объекта и окружения;
- интересы сторон, регионов страны и социальных групп и т.д.;
- уязвимости и угрозы реализации интересов сторон;
- ресурсы, которые могут быть использованы для реализации интересов;

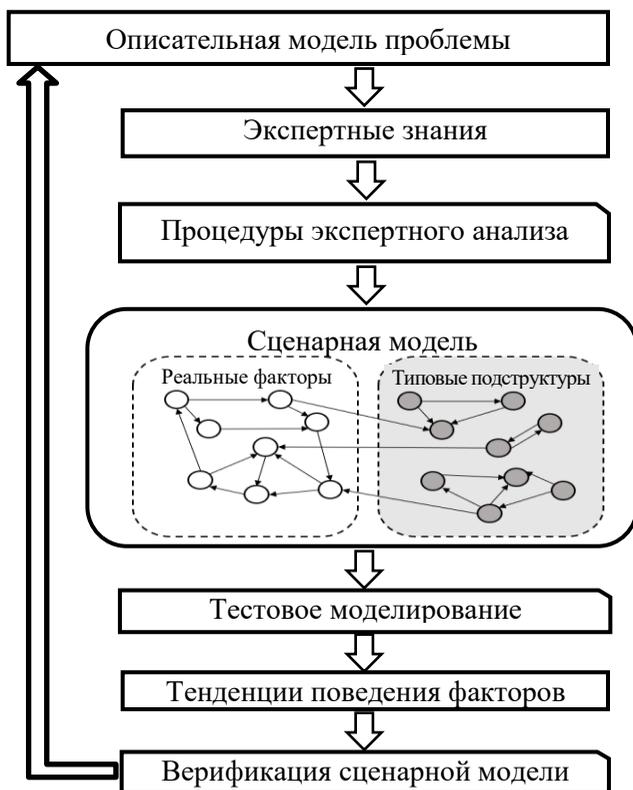


Рис. 1. Общая схема построения сценарных моделей

- внутренние и внешние конфликты, кризисные ситуации и возможные последствия их реализации;
- недопустимые сценарии развития ситуации для основных сторон, параметры состояний факторов, характеризующих недопустимые сценарии развития ситуации;
- механизмы ограничения параметров недопустимых состояний факторов (внутреннее законодательство, международные структуры, институты, договоры и т.п.) и т.п.

При формировании моделей сложной многоуровневой многоаспектной системы часто бывает затруднительно или невозможно собрать численные данные, характеризующих динамику

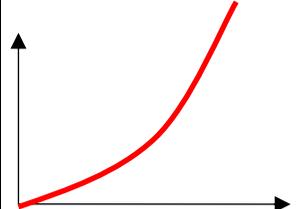
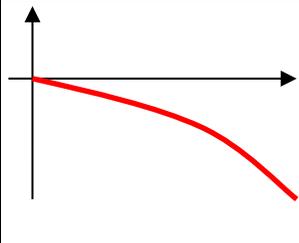
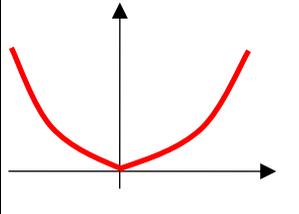
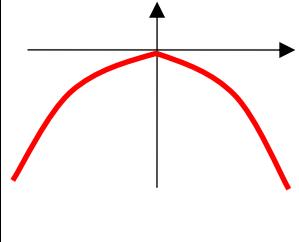
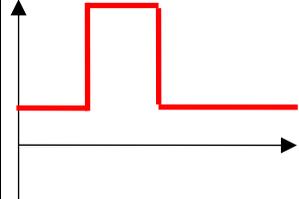
поведения значимых факторов, влияющих на обстановку. Особенно такая ситуация характерна для моделей таких сложных систем как страна, регион, отрасль и т.п. Исходя из этого особую ценность при выделении множества факторов и определения связей между ними приобретают механизмы и методы получения и согласования экспертных оценок [1,4,7].

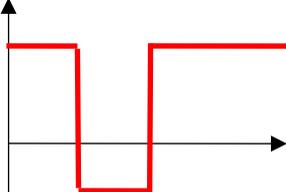
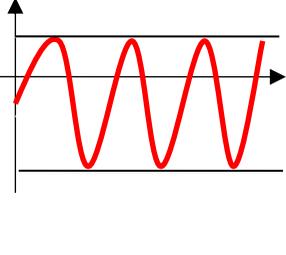
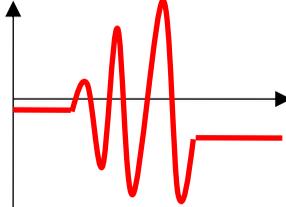
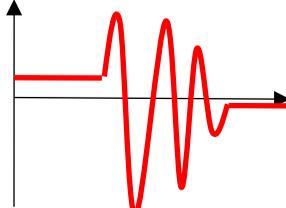
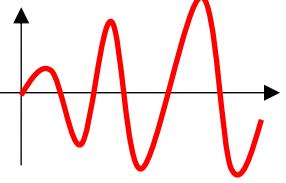
Модели сложных систем могут охватывать различные сферы деятельности: социальную, экономическую, промышленную, политическую и др. Объединение знаний экспертов в различных сферах происходит на этапах формирования сценарной модели [2]. Построенная модель служит прежде всего в качестве средства объединения и согласования знаний, полученных из разных источников, об исследуемой ситуации или сложной системе. Одним из важнейших преимуществ построения подобных моделей является «отчуждение» экспертных знаний от их носителей. Экспертные знания уже заложены в модель и могут многократно использоваться в ситуациях, когда необходимые эксперты недоступны оператору-аналитику, а также при жестких ограничениях на время анализа ситуации или принятия решения.

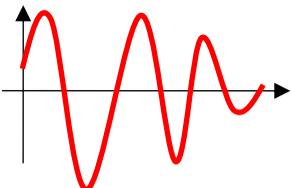
3. Типовые сигналы имитирующие влияния внешней среды на факторы сценарных моделей

Качество сценарной модели, построенной на основе объединения экспертных данных, должно определяться адекватностью представлению образа реального объекта или обстановки. С этой целью при построении сценарно-когнитивных моделей с использованием экспертных знаний необходимо также «экспертно замыкать» структуру модели некоторыми виртуальными подструктурами, генерирующими определенные сигналы, отражающие влияние внешней среды. При этом влияние внешних факторов, непосредственно не включенных в модель, имитируется типовыми импульсными воздействиями, направленными на факторы модели. В таблице 1. представлены основные типы внешних воздействий на факторы исследуемой сценарной модели.

Таблица 1. Типы сигналов, имитирующих внешние воздействия

Тип внешнего воздействия	Характер внешнего воздействия	Изменения в структуре сценарной модели
F_{\uparrow} 	Усиление существующего в сценарной модели взаимодействия между выделенными факторами	Формирование дублирующей взаимосвязи между факторами, приводящей к усилению значимости выделенных факторов сценарной модели
F_{\downarrow} 	Подавление существующего в сценарной модели взаимодействия между выделенными факторами	Формирование дублирующей взаимосвязи между факторами, приводящей к снижению значимости выделенных факторов сценарной модели или к их противоположному влиянию
F_{\perp} 	Усиление существующего в сценарной модели взаимодействия между выделенными факторами по модулю	Формирование дублирующей избирательной взаимосвязи между факторами, приводящей к усилению значимости выделенных факторов сценарной модели независимо от их динамики
F_{\top} 	Ослабление системного взаимодействия по модулю	Формирование дублирующей избирательной взаимосвязи между факторами, приводящей к ослаблению значимости выделенных факторов сценарной модели независимо от их динамики
F_{+} 	Одинокое усиливающее воздействие на выбранный фактор сценарной модели	Неконтролируемое (возможно, временно) субъектом управления кратковременное усиление фактора модели

F_{\downarrow}		<p>Одинокое ослабляющее воздействие на выбранный фактор сценарной модели</p>	<p>Неконтролируемое (возможно, временно) субъектом управления кратковременное ослабление фактора модели</p>
$F_{=}$		<p>Колебательные процессы, ограниченные по амплитуде</p>	<p>Потенциальная скрытая неустойчивость структуры сценарной модели (уязвимость), которая при активации воздействий F_{\uparrow} и F_{\downarrow} может привести к риску влияния деструктивных процессов на реализацию цели управления</p>
$F_{<}<$		<p>Деструктивный процесс, ограниченный по времени</p>	<p>Использование кратковременного перехвата управления («окна уязвимости») для неадекватной реакции субъекта управления или обнаружения уязвимостей</p>
$F_{>}>$		<p>Процесс затухания дестабилизирующих воздействий, ограниченный по времени</p>	<p>Использование кратковременного перехвата управления («окна уязвимости») для маскировки структурных уязвимостей</p>
$F_{<}$		<p>Деструктивные процессы</p>	<p>Активация постоянных деструктивных воздействий в структуре модели объекта управления</p>

F_2		Стабилизирующие процессы	Выведение структуры модели объекта управления из нестабильного состояния
-------	---	--------------------------	--

Легко заметить, что подобные воздействия также имеют экспертную интерпретацию и хорошо понимаемы при составлении моделей.

Использование сценарных моделей с монотонными и линейными причинно-следственными взаимосвязями практически всегда в процессе исследования приводит к результатам, далеким от действительной ситуации, поскольку невозможно постоянно расширять множество факторов модели. Сложные объекты или ситуации являются с этой точки зрения открытыми системами. Внешение в сценарную модель нелинейности или структур, имитирующих внешние сигналы, позволяет приблизить моделируемые процессы к действительности, описывая влияние внешней среды в ограниченном множестве факторов и типовых подструктур.

Кроме того, в зависимости от характера получаемых в процессе сценарного моделирования результатов зачастую необходимо изменять структуру сценарной модели. Например, нестабильность поведения одного значимого фактора и рост другого фактора в течении определенного времени (шагах моделирования) способен вызвать реализацию дополнительных управленческих воздействий в качестве реакции на критическое с точки зрения ЛПР изменения ситуации. Подобные управленческие воздействия могут быть направлены на изменения характера зависящих от ситуации причинно-следственных связей между факторами, на активацию или деактивацию определенного фактора или подмножества факторов, инициирование внешних импульсов и т.п. Таким образом, для обеспечения большей адекватности модели исследуемому объекту управления или анализируемой ситуации необходимо наличие возможности гибкой перестройки

структуры и состава факторов модели в зависимости от результатов оценивания текущих результатов сценарного моделирования, переходя тем самым к сценарно-когнитивным моделям, основанным на правилах, реализованным на функциональных графах. С этой целью введем в качестве аргументов функций весов дуг графа параметра, значение которого соответствует базовому типу динамики факторов, присвоив, таким образом, каждому i -му фактору идентификатор текущей динамики SC_i (см. таблицу 2).

Таблица 2. Идентификаторы динамик факторов

Идентификатор сценария, SC_i	Тип динамики фактора
0	не определен (не рассчитывается)
1	рост
2	падение
3	постоянно
4	колебания, ограниченные по амплитуде
5	расходящиеся колебания
6	сходящиеся колебания

Алгоритм анализа динамик i -го фактора модели f_i (f_i – логарифм значения i -го фактора) для периода времени с t_k до t_n включает несколько этапов [5, 6]:

1. Выбрать временной диапазон, на котором происходит анализ динамики факторов. На этом диапазоне необходимо определить два множества: локальных минимальных f_i^{min} и максимальных f_i^{max} значений (точек перегиба) анализируемого фактора модели, а также два множества шагов моделирования этих локальных экстремумов: t_i^{min} и t_i^{max} . При формировании этих множеств необходимо исключить из рассмотрения незначительные («ложные») локальные экстремумы.

2. Определить мощности множеств $col_i^{min} = |f_i^{min}|$ и $col_i^{max} = |f_i^{max}|$.

3. Если $col_i^{min} < 2$ или $col_i^{max} < 2$, то переход к шагу 4, в ином случае переход к шагу 8.

4. Производятся дальнейшие расчеты только с минимальными или максимальными значениями, поэтому индексы min

и max не используются. Определить средние значения времени и значений фактора: t_i^{cp} и f_i^{cp} .

5. Рассчитать значение ковариаций и выборочных дисперсий:

$$(1) S(f_i) = cov(f_i, f_i) = \sum_{j=k}^n \frac{(f_i(j) - f_i^{cp})^2}{n-k+1},$$

$$(2) cov(f_i, t_i) = \sum_{j=k}^n \frac{(f_i(j) - f_i^{cp}) * (t(j) - t_i^{cp})}{n-k+1},$$

$$(3) S(t_i) = cov(t_i, t_i) = \sum_{j=k}^n \frac{(t(j) - t_i^{cp})^2}{n-k+1}.$$

6. Рассчитать угол наклона линии регрессии, который иллюстрирует изменение значение минимума или максимума выбранного фактора с ростом шага:

$$(4) \theta_i = \frac{cov(f_i, t_i) cov(f_i, t_i)}{cov(t_i, t_i) cov(t_i, t_i)}.$$

7. Определить качество подобранной регрессии, вычислив коэффициент детерминации R^2 , который определяет, насколько хорошо подобрана регрессия:

$$(5) \text{ если } cov(f_i, t_i) = 0(t_i, t_i) = 0, \text{ то } R^2 = 0,$$

$$(6) \text{ иначе } R^2 = \left(\frac{cov(f_i, t_i)}{\sqrt{cov(t_i, t_i)} * \sqrt{cov(f_i, f_i)}} \right)^2.$$

8. Расчет двух линий регрессии. Определить средние значения для каждого множества f_i^{min} , f_i^{max} , t_i^{min} и t_i^{max} :

$$(7) f_i^{min\ cp} = \sum_{j=1}^{col_i^{min}} \frac{f_i^{min}(j)}{col_i^{min}},$$

$$(8) t_i^{min\ cp} = \sum_{j=1}^{col_i^{min}} \frac{t_i^{min}(j)}{col_i^{min}},$$

$$(9) f_i^{max\ cp} = \sum_{j=1}^{col_i^{max}} \frac{f_i^{max}(j)}{col_i^{max}},$$

$$(10) t_i^{max\ cp} = \sum_{j=1}^{col_i^{max}} \frac{t_i^{max}(j)}{col_i^{max}}.$$

9. Расчет ковариаций и дисперсий для этих множеств:

$$(11) cov(f_i^{max}, f_i^{max}) = \sum_{j=1}^{col_i^{max}} \frac{(f_i^{max}(j) - f_i^{max\ cp})^2}{col_i^{max}},$$

$$(12) cov(t_i^{max}, t_i^{max}) = \sum_{j=1}^{col_i^{max}} \frac{(t_i^{max}(j) - t_i^{max\ cp})^2}{col_i^{max}},$$

$$(13) cov(f_i^{max}, t_i^{max}) = \sum_{j=1}^{col_i^{max}} \frac{(f_i^{max}(j) - f_i^{max\ cp}) * (t_i^{max}(j) - t_i^{max\ cp})}{col_i^{max}},$$

$$(14) \quad cov(f_i^{min}, f_i^{min}) = \sum_{j=1}^{col_i^{min}} \frac{(f_i^{min}(j) - f_i^{min\text{cp}})^2}{col_i^{min}},$$

$$(15) \quad cov(t_i^{min}, t_i^{min}) = \sum_{j=1}^{col_i^{min}} \frac{(t_i^{min}(j) - t_i^{min\text{cp}})^2}{col_i^{min}},$$

$$(16) \quad cov(f_i^{min}, t_i^{min}) = \sum_{j=1}^{col_i^{min}} \frac{(f_i^{min}(j) - f_i^{min\text{cp}})(t_i^{min}(j) - t_i^{min\text{cp}})}{col_i^{min}}.$$

10. Вычислить коэффициент детерминации R^2 (см. шаг 7).

11. Рассчитать углы наклона линий регрессии для двух множеств локальных максимумов и минимумов:

$$(17) \quad \theta_i^{max} = \frac{cov(f_i^{max}, t_i^{max})}{cov(t_i^{max}, t_i^{max})},$$

$$(18) \quad \theta_i^{min} = \frac{cov(f_i^{min}, t_i^{min})}{cov(t_i^{min}, t_i^{min})}.$$

12. Определить тип динамики фактора на основе вычисленных параметров регрессий.

13. Линия регрессии представлена формулой $y = \alpha + \theta x$. Определить оставшиеся параметры линий регрессий для визуального представления результатов расчета. Параметры для верхней и нижней (по точкам перегиба локальных максимумов и минимумов) линии регрессии:

$$(19) \quad \alpha_i^{max} = f_i^{max\text{cp}} - \theta_i^{max} t_i^{max\text{cp}},$$

$$(20) \quad \alpha_i^{min} = f_i^{min\text{cp}} - \theta_i^{min} t_i^{min\text{cp}}.$$

14. Выражения для линий регрессий, по которым находятся координаты линий для визуального представления тенденций и для определения значений SC_i :

$$(21) \quad f_i^{max}(t) = \alpha_i^{max} + \theta_i^{max} t$$

$$(22) \quad f_i^{min}(t) = \alpha_i^{min} + \theta_i^{min} t.$$

4. Типовые структуры экспертной идентификации влияния внешней среды на факторы сценарных моделей

В таблице 3 приведены типовые виртуальные графовые подструктуры, отвечающих за имитацию внешних воздействий на сценарную модель. Используются следующие обозначения: $V_{i,j}$ – вес дуги между факторами (сила взаимосвязи) i и j ; I_k – импульс, исходящий из фактора (вершины графа) k ; $I_{вн}$ – внешнее

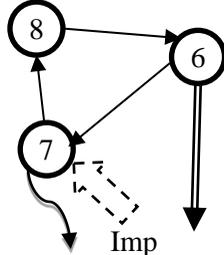
воздействие на выбранный фактор (импульс) – управленческое решение; D_{ij} – вес дуги; Imp – внешний импульс; SC_5 – тип динамики i -го фактора (рост, падение, колебание и т.д.).

Таблица 3. Типовые подструктуры имитационных моделей

Id	Типовая топология	Параметры	Id _s
S ₁		если $D_{1,2}D_{1,3}D_{3,2} > 0$, $Imp > 0$	F_{\uparrow}
		если $D_{1,2}D_{1,3}D_{3,2} > 0$, $Imp < 0$	F_{\downarrow}
		если $D_{1,2}D_{1,3}D_{3,2} < 0$, $Imp \neq 0$	$F_{=}$
S ₂		непосредственное соединение опосредованно связанных факторов (дублирование связи)	F_{\uparrow} F_{\downarrow} F_{\perp} F_{\top}
S ₃		усиливающая нелинейная вза- имосвязь $D_{2,1} = c I_2 $, $c > 0$	F_{\perp}
		ослабляющая нелинейная вза- имосвязь $D_{2,1} = c I_2 $, $c > 0$	F_{\top}
S ₄		внешний положительный им- пульс $Imp > 0$	F_{+}
		ослабляющее внешнее возде- йствиe $Imp < 0$	F_{-}
S ₅		добавленный расходящийся колебательный контур $(D_{3,4}D_{4,5}D_{4,5}) > 1$, $(D_{3,1}D_{1,2}D_{2,5}) < 1$	$F_{<}$
		добавленный сходящийся кон- тур $0 < (D_{3,4}D_{4,5}D_{4,5}) < 1$, $-1 < (D_{3,1}D_{1,2}D_{2,5}) < 0$	$F_{>}$
		добавленный временный рас- ходящийся контур $(D_{3,4}D_{4,5}D_{4,5}) > 1$, $(D_{3,1}D_{1,2}D_{2,5}) < 1$, вершина 3 активна на проме- жутке времени Δt	$F_{<=}$

		<p>добавленный временный сходящийся контур</p> $0 < (D_{3,4}D_{4,5}D_{4,5}) < 1,$ $-1 < (D_{3,1}D_{1,2}D_{2,5}) < 0,$ <p>вершина 3 активна на промежутке времени Δt</p>	$F_{>=}$
S_6		<p>подключаемая подструктура</p> $D_{1,7} = \text{if } [SC_5 = F_{\uparrow}, -5\text{Imp}(1), 0]$ <p>Если динамика фактора D_5 «рост», вес $D_{1,7}$ «-5», если нет – вес $D_{1,7} = 0$ (дуга не существует)</p>	Все
S_7		<p>подключаемая подструктура</p> $D_{1,7} = \text{if } [SC_5 = F_{\uparrow}, -5\text{Imp}(1), 0]$ <p>Если динамика фактора D_5 «рост», вес $D_{1,7}$ «-5», если нет – вес $D_{1,7} = 0$ (дуга не существует)</p>	Все
S_8		<p>межмодельное взаимодействие</p> $D_{1,7} = \text{if } [SC_6 = F_{\downarrow}, -1I(1), 0]$ $D_{6,3} = \text{if } [SC_6 = F_{=}, 5I(6), -1(I(6) + I(8))]$ $D_{10,3} = \text{if } [SC_2 = F_{\uparrow} \ \& \ S_3 = F_{\#}, -1, 0]$	Все

<p>S₉</p>		<p>Межуровневое взаимодействие в иерархической структуре</p> <p>$D_{1,7} = \text{if } (SC_6 = F_1, -II(1), 0)$ $D_{6,3} = \text{if } (SC_6 = F_2, 5I(6), -1(I(6) + I(8)))$ $D_{10,3} = \text{if } (SC_2 = F_1 \& \& S_3 = F_\#, -II(10), 0)$ $D_{8,13} = \text{if } [SC_6 = F_<, -II(10), 0]$ $D_{1,12} = \text{if } [SC_2 = F_2 \parallel S_6 = F_<, -II(1), I(1)]$</p>	<p>Все</p>
<p>S₁₀</p>		<p>Актуальный мониторинг</p> <p>$D_{1,7} = \text{if } (SC_2 = F_1, -II(1), 0)$ $SC_2(t) = f(D_2(t - k), D_2(t - k + 1), \dots D_2(t))$, где k – количество предыдущих шагов, учитываемых при расчете сценария</p> <p>Увеличение запаздывания мониторинга</p> <p>$D_{1,7} = \text{if } (SC_2 = F_1, -II(1), 0)$ $SC_2(t) = f(D_2(t - k - z), D_2(t - k - z + 1), \dots D_2(t - z))$, z – кол-во последних шагов, не учитываемых при расчете сценария</p> <p>меняющаяся частота периодов мониторинга</p> <p>$D_{1,7} = \text{if } (SC_2 = F_1, -II(1), 0)$, $SC_2(t_p) = f[D_2(t_p - k - z), D_2(t_p - k - z + 1), \dots D_2(t_p - z)]$, где p – период обновления определения типа поведения факторов.</p>	<p>Все</p>

<p>S₁₁</p>		<p>Внешнее воздействие в зависимости от текущего сценария. $Imp = if (SC_6 = F_1, 0,7V(8), 0)$, если сценарий фактора D_6 – уменьшение, то величина импульса определяется как 70% значения фактора 8, в противном случае импульс отсутствует</p>
-----------------------	---	--

Следует также отметить, что возможно не только влияние сигналов, исходящих из подобных подструктур на значимые факторы модели, но и обратное влияние. Включение виртуальных подструктур позволит повысить адекватность сценарных моделей, имитировав влияние внешней среды без расширения множества факторов.

Таким образом, сформированная на основе экспертных знаний сценарно-когнитивная модель разбивается на части подструктуры, одна из которых отражает обычное взаимовлияние факторов, а другая предназначена для формирования типовых сигналов из внешней среды, тем самым экспертно замыкая сложную модель в ограниченном пространстве факторов (рис. 2).

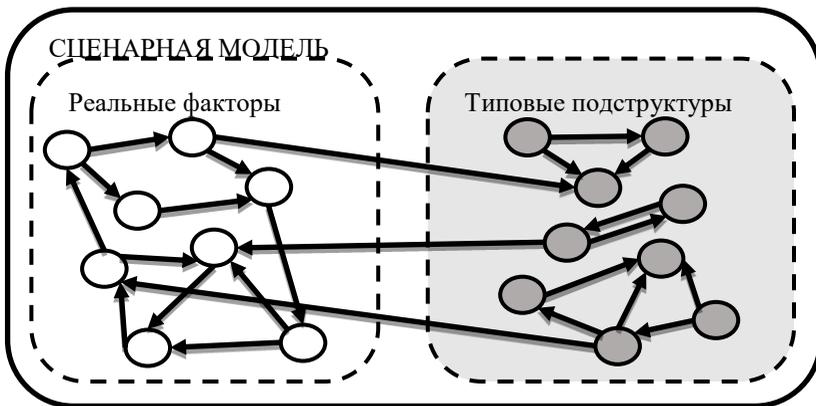


Рис. 2. Сценарная модель, сформированная на основе экспертных знаний и виртуальных факторов

Верификация построенных подобным образом моделей происходит в процессе тестового моделирования. Если в результате такого моделирования полученные тенденции изменения факторов не противоречат логико-математическому описанию собранной экспертной информации, то модель является непротиворечивой.

Для исследования подобных сценарно-когнитивных моделей необходимо создание в соответствующем комплексе сценарного моделирования библиотеки типовых графовых подструктур. Для практического использования подобных типовых подструктур необходимо также предусмотреть программные средства последовательного попарного объединения моделей в единую сценарную модель.

Таким образом, показано, что процедуры формирования моделей на основе экспертных знаний и с использованием типовых подструктур, генерирующих сигналы внешней среды, могут быть использованы для построения сценарных моделей сложным систем и обстановок.

5. Заключение

Таким образом, в работе предложена методика формирования и исследования сценарно-когнитивных моделей, содержащих знания экспертов, на основе логико-математического описания ситуации.

Применение предложенного алгоритма сценарно-событийной идентификации динамики факторов позволяет перейти к сценарно-когнитивному моделированию, основанному на правилах, и реализовать тем самым технологию событийных функциональных нелинейных взаимосвязей между факторами модели, в частности многослойного (в том числе иерархического) представления структуры мультимодели, что обеспечит возможность эффективного применения знаний экспертов в различных ситуациях и предметных областях на стадии разработки адекватных сценарных моделей и при исследовании процессов развития сложных социально-экономических систем и ситуаций.

Разработанные типовые подструктуры модели, имитирующие влияние внешней среды, позволяют также повысить адекватность построения модели на основе экспертных данных и при использовании ограниченного множества факторов и взаимосвязей.

Литература

1. ГУБАНОВ Д.А., КОРГИН Н.А., НОВИКОВ Д.А., РАЙКОВ А.Н. *Сетевая экспертиза* / Под ред. чл.-к. РАН Д.А. Новикова, проф. А.Н. Райкова. – М.: Эгвес, 2010. – 168 с.
2. ДРАНКО О.И., НОВИКОВ Д.А., РАЙКОВ А.Н., ЧЕРНОВ И.В. *Управление развитием региона. Моделирование возможностей*. – М.: URSS, ООО «ЛЕНАНД», 2023. – 432 с.
3. *Модели и методы анализа и синтеза сценариев развития социально-экономических систем*: в 2-х кн. / Под ред. В.Л. Шульца, В.В. Кульбы. – М.: Наука, 2012. – Кн. 1 – 304 с., кн. 2. – 358 с.
4. ОРЛОВ А.И. *Искусственный интеллект: экспертные оценки* : учебник. – М.: Ай Пи Ар Медиа, 2022. – 436 с.
5. SEN A., SRIVASTAVA M. *Regression Analysis: Theory, Methods, and Applications*. – Springer-Verlag, Berlin, 2011. – 348 p.
6. STULP F., SIGAUD O. *Many Regression Algorithms, One Unified Model: A Review* // *Neural Networks*. – 2015. –Vol. 69. – P. 60–79.
7. *The Delphi Method: Techniques and Applications* / Eds.: H.A. Linstone, M. Turoff. 1975. – 620 p.

THE UTILIZATION OF VIRTUAL STRUCTURES IN THE FORMATION OF SCENARIO-COGNITIVE MODELS BASED ON THE UTILIZATION OF EXPERT KNOWLEDGE

Vladimir Schultz, Institute of Socio-political Research of RAS, Moscow, Doctor of Philosophy, professor, Corresponding Member RAS (9380752@mail.ru).

Igor Chernov, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, Candidate of Technical Sciences, Leading Researcher (chernov@ipu.ru).

Abstract: This study is dedicated to investigating the problem of enhancing the adequacy of scenario-cognitive models based on expert knowledge within a limited set of factors. One of the most important tasks in the formation of a scenario-cognitive model based on expert knowledge is the problem of taking into account the total influence of the external environment, i.e., those factors that remain outside the structure of the model, but influence the achievement of the required accuracy of modeling results. When constructing scenario-cognitive models of complex socio-economic and political systems, it is usually necessary to apply a significant simplification, which consists in concluding all the diversity of factors and connections between them in a relatively simple and understandable model. The quality of a model built on the basis of combining expert data should be determined by the adequacy of the image of a real object or situation. Consequently, when forming models using expert knowledge, it is also necessary to “expertly close” the structure of the model with some virtual sub-structures that are capable of generating certain signals reflecting the influence of the external environment. Typical signals simulating the influences of the external environment are presented. Typical structures of expert identification of the impact of the external environment on scenario model factors are introduced. An overall pattern of the scenario-cognitive model is presented, which is formed based on expert knowledge and consists of a multitude of actual factors of a complex system and proxy structures.

Keywords: scenario-cognitive modeling, adequacy, scenario model, scenario analysis, external environment, proxy factors.

УДК 005.7

ББК 13.42

DOI: 10.25728/ubs.2024.108.9

*Статья представлена к публикации
членом редакционной коллегии Р.М. Нижегородцевым.*

*Поступила в редакцию 12.01.2024.
Опубликована 31.03.2024.*

АВТОМАТИЗИРОВАННАЯ ИНФОРМАЦИОННАЯ СИСТЕМА АНАЛИЗА КАЧЕСТВА ПРИРОДНОГО ГАЗА

Брокарев И. А.¹

*(РГУ нефти и газа (НИУ) имени И.М. Губкина, Москва,
ФГБУН Институт проблем управления
им. В.А. Трапезникова РАН, Москва)*

Васьковский С. В.², Фархадов М. П.³

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

Рассматривается и исследуется автоматизированная информационная система определения энергетических параметров природного газа, приведены её основные узлы, показан вариант работы системы на экспериментальных данных. Разработанная архитектура автоматизированной системы состоит из следующих узлов: информационная подсистема, реализующая разработанные алгоритмы; подсистема измерительной информации; подсистема проведения анализа. Рассмотрен метод анализа качества природного газа на основе нейронных сетей. Основные достоинства метода и системы на его основе заключаются в следующем: преимуществом каждой из подсистем является их многофункциональность, позволяющая модифицировать систему на каждом этапе под определенную задачу и осуществлять получение необходимой измерительной информации коммерчески доступным и относительно недорогим измерительным оборудованием. Разработка модели для решения задачи анализа качества природного газа включает в себя ряд этапов: выбор данных для обучения модели; выбор архитектуры модели; выбор метода обучения модели; оценка точности работы модели. Система апробирована с использованием результатов экспериментов, проведенных в лабораторных условиях на данных реальных газовых смесей. Рассчитаны показатели точности определения энергетических параметров, по которым можно сделать вывод о возможности применения исследуемого метода и системы на его основе для анализа качества реальных газовых смесей. Реализованная архитектура автоматизированной информационной системы обеспечивает анализ качества газа с низкими временными затратами.

Ключевые слова: анализ качества природного газа, оценка точности систем анализа, автоматизированные информационные системы.

¹ Иван Андреевич Брокарев, старший преподаватель (brokarev.i@gubkin.ru).

² Сергей Владимирович Васьковский, к.т.н., с.н.с. (vb3v@yandex.ru).

³ Маис Паша оглы Фархадов, д.т.н., г.н.с. (mais@ipu.ru).

1. Введение

В связи с модернизацией нефтегазовой промышленности, а также со сложностью решения промышленных задач традиционными методами в данной отрасли существует большое число примеров применения нейросетей, в частности модели на основе нейронных сетей показали свою эффективность при решении задачи прогнозирования эффективности геолого-технических мероприятий [5, 7, 9], при прогнозировании нефтеотдачи и оценке применимости методов увеличения нефтеотдачи [1, 4, 6], при прогнозировании работы скважин [8, 15]. В работе [2] проводится сравнение моделей на основе нейронной сети и различных методов машинного обучения для вычисления коэффициента сжимаемости природного газа, в статье [3] предложена модель, комбинирующая различные алгоритмы машинного обучения для прогнозирования выбросов оксидов углерода и азота газовых турбин. Одной из перспективных задач, в которых возможно применение искусственных нейронных сетей, является задача анализа качества природного газа. В настоящее время существует большое количество различных методов анализа показателей качества природного газа [14]. В результате проведённого исследования достоинств и недостатков существующих методов и автоматизированных информационных систем определения показателей качества газа сделаны следующие выводы. В настоящее время в промышленности преобладают физико-химические методы анализа качества газа и системы на их основе. Основными методами являются методы газовой хроматографии и калориметрии для задач определения энергетических характеристик природного газа. Физико-химические методы анализа компонентного состава и энергетических параметров газа обладают рядом существенных недостатков, в частности дороговизна и крупные габариты используемого оборудования, большие временные затраты на проведение единичного анализа. В настоящее время разрабатываются альтернативные аналитические методы анализа качества газа и системы на их основе, но широкого распространения они пока не получили. Характерными особенностями систем на основе аналитических методов,

определяющими актуальность их разработки, являются низкие временные затраты на проведение анализа, использование моделей, в частности нейросетевых, для получения необходимых концентраций компонентов, использование относительно недорогих и доступных на рынке измерительных приборов. Разработанная архитектура системы анализа качества газа обладает следующими свойствами: определение компонентного состава модели эквивалентного псевдогаза и энергетических параметров природного газа с низкими временными затратами; использование нейросетевых моделей, в частности нейросетей, для нахождения неизвестных концентраций компонентов модели эквивалентного псевдогаза, зная измеренные свойства газа; измерение физических параметров газа коммерчески доступными и относительно недорогими измерительными приборами. Основными преимуществами системы являются: проведение анализа с низкими временными затратами – до нескольких секунд (для портативных газовых хроматографов анализ занимает десятки минут); высокая точность проведения анализа за счет применения нейросетевых моделей – до третьего класса ($\pm 0,5$ МДж/м³) определения энергетических параметров газа; снижение стоимости измерительного комплекса за счет использования доступных на рынке измерительных приборов до сотен тысяч рублей (стоимость высокоточного промышленного хроматографа превышает миллион рублей); меньшие габариты измерительного комплекса по сравнению с традиционным хроматографом.

Для решения проблемы точного анализа качества природного газа в настоящее время в промышленности используются различные автоматизированные информационные системы (АИС), основанные на физико-химических методах анализа. Следует отметить, что при применении таких АИС обеспечивается точность анализа за счёт использования прямых методов измерения и соответствующих методов обработки информации. Но при этом такие системы имеют следующие недостатки: значительные временные и экономические затраты на проведение анализа, высокая трудоёмкость разработки системы и высокие затраты на обслуживание ее технических средств. Поэтому актуальной является разработка нового метода и средств обработ-

ки информации, а также реализующих их АИС, позволяющих обеспечить высокую скорость проведения анализа качества природного газа, характеризующихся низкой стоимостью разработки и эксплуатации. В настоящее время нейронные сети широко используются в нефтегазовой отрасли в связи с её модернизацией, а также со сложностью решения промышленных задач существующими методами и алгоритмами. Подход с использованием нейросетевых моделей, включая обучение моделей на экспериментальных данных для определения эксплуатационных и стоимостных показателей, применяется в большом числе практических задач нефтегазовой отрасли, таких как контроль давления в газораспределительной сети, прогнозирование вязкости природного газа и т.д. Актуальным является применение нейросетевых технологий и при разработке нового метода и средств обработки информации для интеллектуального анализа качества природного газа.

В качестве практической реализации предлагаемых решений в данной статье рассмотрено экспериментальное подтверждение предложенного метода обработки информации, в частности реализация АИС анализа качества газа. При исследовании системы использовались ранее полученные в работах [10–13] результаты, в частности: использование разработанной нейросетевой модели, применение многокритериальной оценки для выбора входных параметров, использование алгоритма перехода к модели эквивалентного псевдогаза, оценка надёжности с применением вероятностного метода. Архитектура АИС функционирует следующим образом. На технологическом объекте, к которому планируется применить разработанную АИС, в байпасе (обводном трубопроводе технологической установки, применяющемся для транспортировки газа параллельно запорной и регулирующей арматуре) от основной технологической трубы располагаются две измерительные камеры (основная и резервная). В каждой измерительной камере установлены следующие измерительные приборы: прибор TCG3880 компании Xensor Integration для измерения коэффициента теплопроводности, прибор TDS0119 компании Dynament Infrared Gas Sensors для измерения концентрации диоксида углерода, прибор Optim

компании Optel для измерения скорости звука. Данные измерений через концентратор поступают на персональный компьютер. Поступившие от блока входных данных измерения передаются на подсистему алгоритмов, реализованную на ПК в Matlab. Архитектура АИС состоит из трёх основных узлов: информационная подсистема, реализующая разработанные алгоритмы; подсистема получения измерительной информации, включая измерение физических параметров газа; подсистема проведения расчёта энергетических параметров. Архитектура АИС показана на рис. 1. Под алгоритмом нейросетевого анализа понимается выбор архитектуры сети и её обучение.

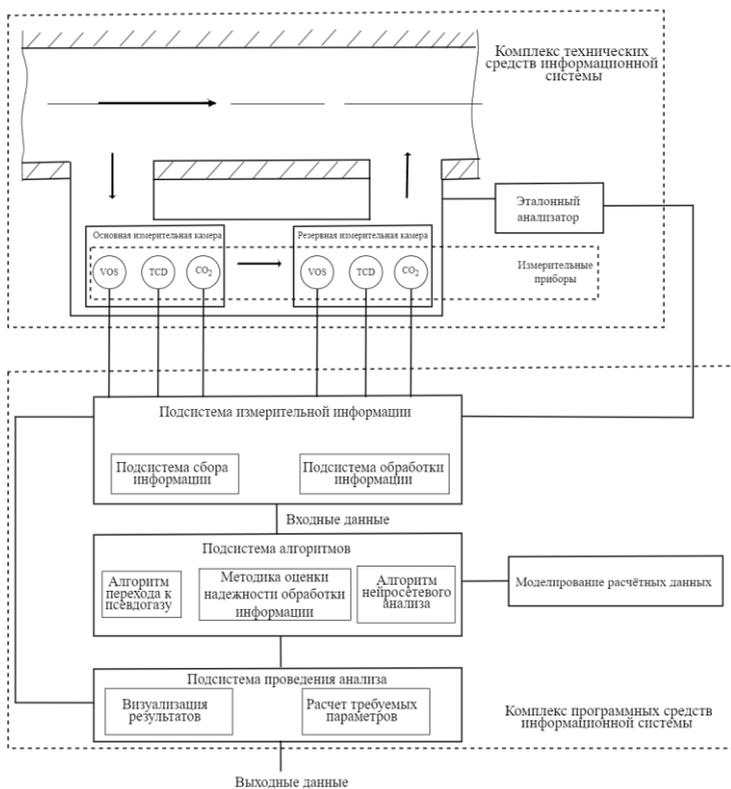


Рис. 1. Архитектура АИС

2. Описание архитектуры АИС

2.1. ПОДСИСТЕМА ИЗМЕРИТЕЛЬНОЙ ИНФОРМАЦИИ

Подсистема измерительной информации состоит из измерительных приборов и аппаратуры, которую возможно использовать в исследуемой задаче для получения значений физических параметров для оценки работы системы в целом. Стоит отметить, что данная подсистема апробировалась на модели, обладающей всеми свойствами предлагаемой системы, в частности возможностью достижения высокой точности анализа за счёт подготовки смесей с использованием регуляторов массового расхода, проведение измерений коммерчески доступными и относительно недорогими измерительными приборами.

Также исследуемая подсистема измерительной информации включает в себя визуализацию основных измеряемых параметров, в частности скорости звука, теплопроводности и концентрации диоксида углерода для наглядного представления процесса измерения. Визуализация реализуется при помощи алгоритмов из соответствующей подсистемы, в частности алгоритма для визуализации различия параметров исходных смесей и модели псевдогазовых смесей. Предлагаемая подсистема получения измерительной информации апробировалась на доступном измерительном оборудовании, её основная функция – предоставление измерительной информации для последующей обработки.

2.2. ПОДСИСТЕМА АЛГОРИТМОВ И ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ

Данная подсистема включает программное обеспечение с реализованными алгоритмами основных функций подсистемы. В предлагаемой реализации архитектуры АИС в качестве программного обеспечения используется пакет прикладных программ для решения задач технических вычислений Matlab 2019b [19] с плагином NIST REFPROP [22] – программы, производящей расчёт различных свойств газовых смесей, как физических параметров, являющихся измеряемыми в исследуемой задаче, так и энергетических параметров, являющихся выходными це-

левыми параметрами в исследуемой задаче. В качестве пакета прикладных программ для реализации алгоритмов и расчёта параметров газа в исследуемой задаче возможно использовать программное обеспечение, функционирующее на современных операционных системах, включая Linux, macOS и Windows, в котором возможно реализовать приведенные алгоритмы, что является преимуществом предлагаемой системы – свойством многофункциональности системы. Подсистема алгоритмов включает в себя следующие алгоритмы: алгоритмы формирования данных, алгоритмы проведения нейросетевого анализа, алгоритмы первичной визуализации данных. Алгоритм формирования компонентного состава газа заключается в задании минимального и максимального (согласно стандартам) значения концентрации каждого компонента с последующим перебором всех возможных вариантов компонентного состава с заданием шага по каждому компоненту. Алгоритм расчёта компонентного состава модели эквивалентного псевдогаза заключается в использовании методик для расчёта по компонентному составу исходного газа компонентного состава модели четырёх-, пятикомпонентного псевдогаза или двух видов псевдогаза одновременно, что можно задать в данном алгоритме. Алгоритм задания температуры и давления позволяет добавлять к компонентному составу значения давления и температуры в задаваемом диапазоне для формирования расчётных данных. Алгоритм расчёта физических параметров и энергетических характеристик для исходного газа и модели эквивалентного псевдогаза заключается в выборе стандарта, по которому будет производиться расчет, выборе параметров, которые будут рассчитаны, единиц измерения, формирования отчёта о возможных ошибках расчёта и формирования массива данных. Входной информацией для алгоритма является количество доступных технических средств информационной системы и сведения о рассматриваемом объекте исследования. В зависимости от соотношения количества доступных технических средств и числа компонентов в объекте расчёт возможен по моделям четырёх- и пятикомпонентного псевдогаза. Выходной информацией в алгоритме является компонентный состав выбранной модели псевдогаза.

Алгоритмы, реализующие нейросетевой анализ, предусматривают формирование данных для модели, разработку самой модели, её обучение и тестирование на массиве данных. Алгоритм импорта данных предназначен для формирования удобного для представления вида данных в зависимости от платформы, из которой осуществляется импорт. Основной алгоритм нейросетевого анализа включает в себя ряд алгоритмов, реализующих следующие функции: возможность пропуска ряда моделей в случае отсутствия необходимости их рассмотрения; многокритериальная оценка входных параметров с последующим выбором наиболее подходящего набора входных данных; разделение данных на тренировочную, валидационную, тестовую выборки с выбором объема выборки, количества разбиений, процентом данных, идущих на каждую из выборок; выбор набора входных и выходных данных, их округление до значений, которые возможно получить посредством измерений приборами; вычисление матрицы коэффициентов корреляции, нормализация и кросс-валидация данных; реализация нейросетевых моделей, для нейросетевых моделей настройка параметров, в том числе архитектуры модели, алгоритма обучения, критерия окончания обучения, количества циклов обучения, расчёта ошибок при обучении для всех моделей, обратной денормализации данных, расчёта показателей точности при обучении и тестировании нейросетевых моделей и выбор модели для последующего тестирования.

Архитектура используемой нейросетевой модели представляет собой простую рекуррентную нейронную сеть (РНС) с одним скрытым слоем. Количество нейронов во входном слое равно трем по количеству входных физических параметров (скорость звука, теплопроводность, концентрация диоксида углерода). Количество нейронов в выходном слое равно трем по количеству выходных концентраций компонентов псевдогаза за исключением концентрации диоксида углерода. Количество нейронов в скрытом слое равно одиннадцати, функциями активации нейронов были выбраны сигмоидальная функция в виде гиперболического тангенса для скрытого слоя и линейная функция активации для выходного слоя. Структура разработан-

ной РНС показана на рис. 2 (n , k , m – количество нейронов во входном, скрытом и выходном слоях).

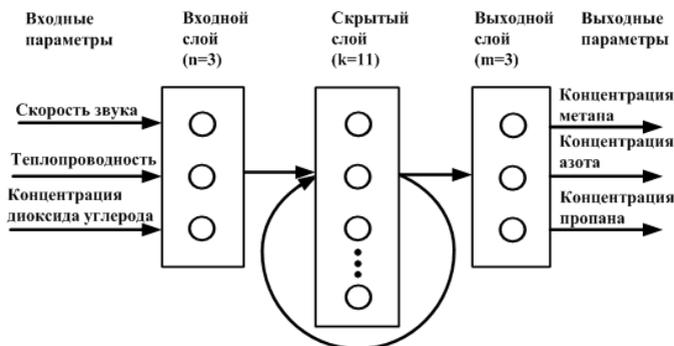


Рис. 2. Нейросетевая модель определения компонентного состава эквивалентного псевдогаза

Для проверки эффективности работы разработанной модели проводится исследование на тестовой выборке данных. Данная выборка включала данные, которые не участвовали в обучении нейросетевой модели. Предварительная тестовая выборка включала 4358 газовых смесей, сформированных аналогично тренировочной выборке, которая включала в себя 96701 газовую смесь.

Подсистема приведённых алгоритмов является важным преимуществом предлагаемой архитектуры АИС анализа качества газа, реализующая алгоритмы предварительной подготовки массива данных и разработки нейросетевой модели. На этом этапе в силу многофункциональности подсистемы возможно пропустить или модифицировать некоторые этапы, подстроив подсистему под определённую задачу. Также в данной подсистеме возможно реализовать предварительное обучение и тестирование модели, многокритериальную оценку входных параметров, переход к модели эквивалентного псевдогаза и другие алгоритмы, позволяющие упростить дальнейшую разработку системы и дающие возможность оттестировать некоторые этапы функционирования системы.

2.3. ПОДСИСТЕМА ПРОВЕДЕНИЯ АНАЛИЗА

Данная подсистема была реализована с использованием того же программного обеспечения, что и подсистема алгоритмов. Для апробации предлагаемой подсистемы был произведён измерительный эксперимент, который заключался в измерении физических параметров газа, в частности, скорости звука, коэффициента теплопроводности и концентрации диоксида углерода для исходных газовых смесей и для соответствующих им четырёх- и пятикомпонентных псевдогазовых смесей. Данные параметры при выбранных термодинамических стандартных условиях измерялись двумя однотипными измерительными приборами для оценки надёжности получаемых измерительных данных. Скорость звука, теплопроводность измерялись для матрицы газовых смесей. Данная матрица строилась путём постепенного увеличения содержания компонентов для покрытия выбранного объекта, а именно российского природного газа. Исследуемая матрица газовых смесей соответствует диапазонам по компонентам российского природного газа, что означает, что в проведенном эксперименте был покрыт весь исследуемый природный газ, что может считаться преимуществом системы по её качеству апробации.

На рис. 3 показаны данные об измерении скорости звука для газовой смеси метан-пропан-азот-диоксид углерода. На рис. 4 показаны результаты измерения для смеси метан 94%, пропан 1%, диоксид углерода 1%, азот 4%. Участки горизонтальных линий на рис. 3 соответствуют определённому компонентному составу, при этом для каждого состава вариация показаний по скорости звука не выходит за пределы 5 см/с. Стоит отметить, что скачки на графиках 3 и 4 объясняются тем, что в начале и конце измерений происходит проверка на чистом метане для контроля достоверности измерений и дальнейшей коррекции данных по температуре и давлению.

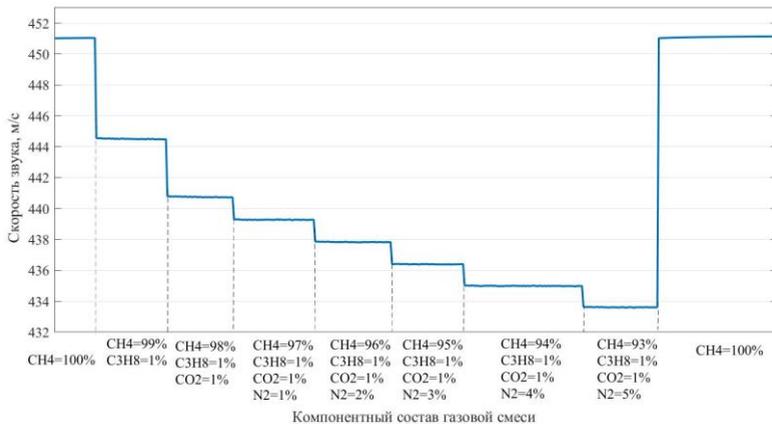


Рис. 3. Результаты измерения скорости звука газовой смеси метан-пропан-азот-диоксид углерода

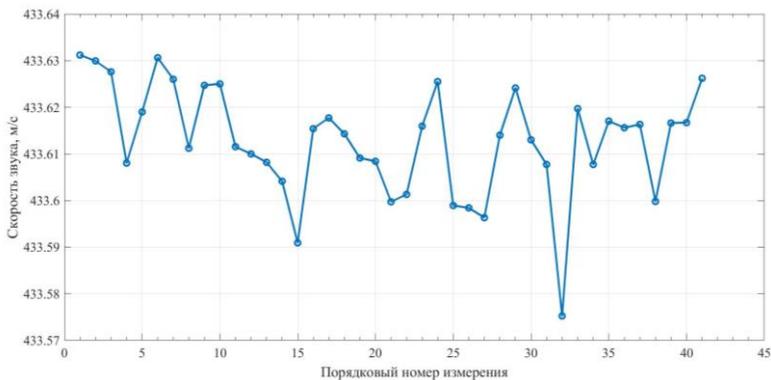


Рис. 4. Результаты измерения скорости звука для газовой смеси метан 94%, пропан 1%, диоксид углерода 1%, азот 4%

На следующем шаге рассчитывались показатели точности определения как компонентного состава модели эквивалентного псевдогаза, так и точность определения энергетических параметров газа. Работа данной подсистемы заключается в тестировании разработанной нейросетевой модели на полученных экспериментальных данных после их коррекции и предварительной обработки. Результаты тестирования (с рассчитанными макси-

мальным (MAO), среднем абсолютным отклонением (CAO) и среднеквадратичным отклонением (СКО) показаны в таблице 1.

Таблица 1. Точность определения компонентного состава псевдогаза разработанной нейросетевой моделью на экспериментальных данных

Компонент	MAO, молярная доля, %	CAO, молярная доля, %	СКО
Метан	0,94	0,56	0,64
Пропан	0,53	0,36	0,32
Азот	0,44	0,25	0,23

В таблице 2 и на рис. 5 и 6 показана точность определения энергетических характеристик газа (низшая объемная теплотворная способность и число Воббе), полученных путем расчета по полученному компонентному составу модели эквивалентного псевдогаза.

Таблица 2. Точность определения энергетических параметров газа по разработанной нейросетевой модели на этапе тестирования

Параметр точности	Теплотворная способность, МДж/м ³	Число Воббе, МДж/м ³
MAO	0,351	0,478
CAO	0,112	0,235

Стоит отметить, что высокое быстродействие системы крайне важно для тех случаев, когда свойства газа могут изменяться быстро и требуют постоянного мониторинга, что делает разработку исследуемой системы крайне актуальной, например, это переработка попутного газа нефтяных месторождений, который раньше шел в факелы, газ, получаемый гидроразрывом пласта и биогаз из различных источников [18, 20, 21, 23, 24].

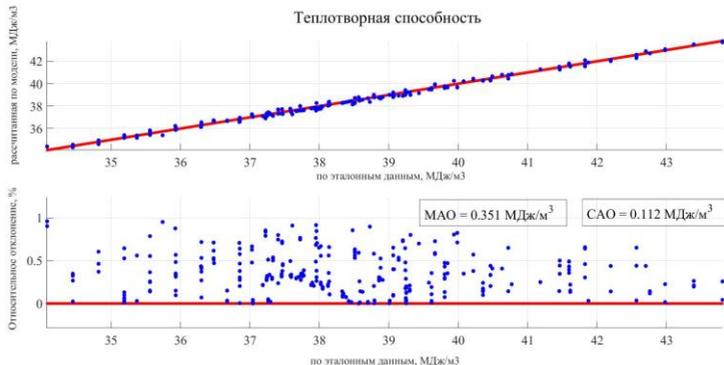


Рис. 5. Точность определения теплотворной способности по разработанной нейросетевой модели на этапе тестирования

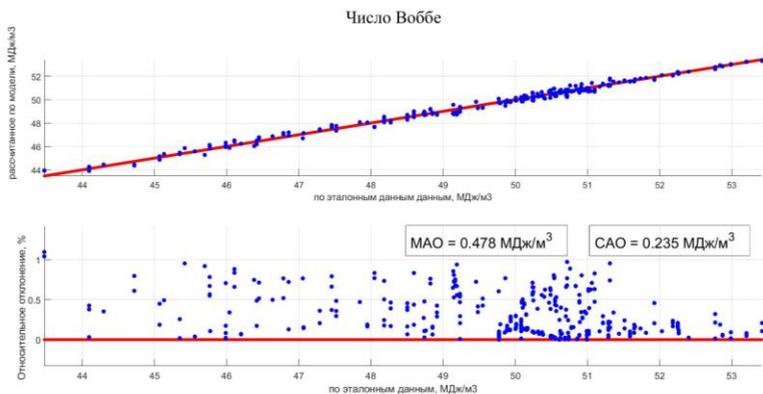


Рис. 6. Точность определения числа Воббе по разработанной нейросетевой модели на этапе тестирования

Реализация в системе ряда алгоритмов, методов и технологий, в том числе использование нейросетевых технологий, многокритериальной оценки входных параметров, алгоритма перехода к модели эквивалентного псевдогаза, оценка надёжности с применением вероятностного метода также являются преимуществами системы.

3. Заключение

Разработана АИС для определения значений энергетических характеристик природного газа по результатам измерения заданного набора его физических параметров. Архитектура АИС состоит из информационной подсистемы, реализующей разработанные алгоритмы, подсистемы измерительной информации и подсистемы проведения анализа. Преимуществом системы является её распределённая структура, заключающаяся в том, что выход из строя одного из узлов не приводит к полной остановке всей системы. Возможность функционирования АИС обосновывается использованием программного обеспечения, выполняющего расчёты, и измерительных приборов, предоставляющих измерительную информацию. Разработанные методы и архитектура системы позволяют проводить исследование и анализ различных современных методов для определения качества природного газа; осуществлять верификацию качества функционирования оборудования; выполнять экспериментальные исследования для оценки точности измерений качества природного газа. По результатам исследований было предложено алгоритмическое решение, на основе которого реализовано программное обеспечение, функционирующее в рамках существующей автоматизированной информационной системы, которое позволяет проводить анализ качества природного газа с меньшими временными и стоимостными затратами.

Литература

1. КЕЛЛЕР Ю.А. *Разработка искусственных нейронных сетей для предсказания технологической эффективности от выравнивания профиля приемистости* // Известия Томского политехнического университета. Информационные технологии. – 2014. – № 5. – С. 60–65.
2. КОЧУЕВА О.Н. *Аппроксимация коэффициента сжимаемости газа на основе генетических алгоритмов* // Автоматизация и информатизация ТЭК. – 2023. – №11(604). – С. 59–68. – DOI: 10.33285/2782-604X-2023-11(604)-59-68.

3. КОЧУЕВА О.Н. *Разработка моделей прогнозирования выбросов оксидов углерода и азота газовых турбин на основе генетических алгоритмов* // Деловой журнал Neftegaz.RU. – 2022. – №5-6(125-126). – С. 14–20.
4. КОЧУЕВА О.Н., СКОРОВ Д.С., ЛЕБЕДЕВА А.Ю. *Оценка применимости методов увеличения нефтеотдачи с использованием искусственной нейронной сети* // Автоматизация, телемеханизация и связь в нефтяной промышленности. – 2020. – №1(558). – С. 8–13. – DOI: 10.33285/0132-2222-2020-1(558)-8-13.
5. ЛЕОНОВ М.Г., ЖИРНОВ Б.С. *Обработка экспериментальных данных процесса коксования с помощью нейронных сетей* // Нефтегазовое дело. – 2014. – №2. – С. 151–165.
6. МАНДРИК И.Э., ШАХВЕРДИЕВ А.Х., СУЛЕЙМАНОВ И.В. *Оценка и прогноз нефтеотдачи на основе моделирования нейронными сетями* // Нефтяное хозяйство. – 2005. – №10. – С. 31–34.
7. ПЧЕЛЬНИКОВ И.В., БОРХОВИЧ С.Ю., НАТАРОВ А.Л. *Перспективы прогнозирования эффективности ГТМ на основе нейросетевого моделирования* // Нефть. Газ. Новации. – 2016. – №4. – С. 37–40.
8. СОЛОМАТИН Г.И., ЗАХАРЯН А.З., АШКАРИН Н.И. *Прогнозирование работы скважин с помощью искусственных нейронных сетей* // Нефтяное хозяйство. – 2002. – №10. – С. 92–96.
9. ASHENA R., THONHAUSER G *Application of Artificial Neural Networks in Geoscience and Petroleum Industry* // Artificial Intelligent Approaches in Petroleum Geosciences. – 2015. – P. 127–166.
10. BROKAREV I.A., FARKHADOV M.P., VASKOVSKII S.V. *Recurrent neural networks to analyze the quality of natural gas* // Vestnik Tomskogo Gosudarstvennogo Universiteta – Управление, Vychislitel'naya Tekhnika i Informatika. – 2021. – No. 55. – P. 11–17.

11. BROKAREV I.A., VASKOVSKII S.V. *Gas Quality Determination Using Neural Network Model-based System* // Proc. of the 2nd Int. Workshop on Stochastic Modeling and Applied Research of Technology (SMARTY 2020). – Petrozavodsk: Institute of Applied Mathematical Research, Karelia Research Centre of the Russian Academy of Sciences, 2020. – Vol. 2792. – P. 113–128.
12. BROKAREV I.A., VASKOVSKII S.V. *Investigation of Reliability for Information System for Natural Gas Quality Analysis* // Advances in Systems Science and Applications. – 2022. – Vol. 22, No. 2. – P. 11–20.
13. BROKAREV I.A., VASKOVSKII S.V. *Multi-criteria estimation of input parameters in natural gas quality analysis* // Advances in Systems Science and Applications. – 2020. – Vol. 20, No. 2. – P. 60-69.
14. DÖRR H., KOTURBASH T., KUTCHEROV V. *Review of impacts of gas qualities with regard to quality determination and energy metering of natural gas* // Measurement Science and Technology. – 2019. – Vol. 30, No. 2. – P. 1–20.
15. JIMENEZ DE LA CRUZ G., RUZ-HERNANDEZ J., SHELOMOV E., SALAZAR-MENDOZA R. *Optimization of an Oil Production System using Neural Networks and Genetic Algorithms* // Proc. of the Joint 2009 Int. Fuzzy Systems Association World Congress and 2009 European Society of Fuzzy Logic and Technology Conference. – 2009. – P. 1815–1820.
16. KAUR J., ADAMCHUK V., WHALEN J., ISMAIL A. *Development of an NDIR CO₂ Sensor-Based System for Assessing Soil Toxicity Using Substrate-Induced Respiration* // Sensors. – 2015. – No. 15. – P. 4734–4748.
17. KOTURBASH T., BICZ A., BICZ W. *New instrument for measuring velocity of sound and quantitative characterization of binary gas mixtures composition* // Measurement Automation Monitoring. – 2016. – P. 254–258.
18. KOTURBASH T., BICZ A., KUTCHEROV V.G. *Real-time quality metering of propanated biomethane*. – 2021. – Vol. 27, No. 1. – P. 8.

19. *Matlab 2019b Software*. – URL: <https://www.mathworks.com> (Дата обращения: 30.05.2023).
20. MIRZAEI-PAIAMAN A., SALAVATI S. *The Application of Artificial Neural Networks for the Prediction of Oil Production Flow Rate // Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*. – 2012. – No. 34:19. – P. 1834–1843.
21. PETCULESCU A. *An acoustic approach to assess natural gas quality in real time // The Journal of the Acoustical Society of America*. – 2017. – No. 142. – P. 2547.
22. *REFPROP Software*. – URL: <https://www.nist.gov/srd/refprop> (Дата обращения: 30.05.2023).
23. WU F., YAN Y., YIN C. *Real-time microseismic monitoring technology for hydraulic fracturing in shale gas reservoirs: A case study from the Southern Sichuan Basin // Natural Gas Industry*. – 2017. – Vol. 4, Iss. 1. – P. 68–71.
24. YI P., LIZHI X., YUANZHONG Z. *Remote real-time monitoring system for oil and gas well based on wireless sensor networks // Int. Conf. on Mechanic Automation and Control Engineering*. – 2010. – P. 2427–2429.

AUTOMATED INFORMATION SYSTEM FOR NATURAL GAS QUALITY ANALYSIS

Ivan Brokarev, National University of Oil and Gas «Gubkin University», Moscow, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, senior lector (brokarev.i@gubkin.ru).

Sergei Vaskovskii, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, Candidate of Sciences, senior researcher (v63v@yandex.ru).

Mais Farhadov, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, Doctor of Sciences, chief researcher (mais@ipu.ru).

Abstract: The article proposes an automated information system for determining the energy parameters of natural gas, shows its main components, and shows a version of the system's operation based on experimental data. The proposed architecture of the automated system consists of the following parts: an information subsystem

that implements the developed algorithms, a measurement information subsystem, and an analysis subsystem. The main advantages of the method and the system based on it are as follows: multifunctionality, which allows to modify the system at each stage for a specific task and obtain the necessary measurement information using commercially available and relatively inexpensive measuring equipment. Development of a model for solving the problem of analyzing the quality of natural gas includes a number of successive stages, namely: selection of data for training the model; choice of model architecture; choosing a model training method; assessing the accuracy of the model. The system was tested using the results of experiments conducted in laboratory conditions using data from real gas mixtures. Indicators of the accuracy of determining energy parameters have been calculated, from which it can be concluded that the method under study and the system based on it can be used to analyze the quality of real gas mixtures. The implemented architecture of the automated information system is information and computing, providing analysis of gas quality with low time costs.

Keywords: analysis of the quality of natural gas, assessment of the accuracy of gas analysis systems, automated information systems.

УДК 519.6

ББК 30.1

DOI: 10.25728/ubs.2024.108.10

*Статья представлена к публикации
членом редакционной коллегии В.Г. Лебедевым.*

Поступила в редакцию 15.02.2024.

Опубликована 31.03.2024.

АДАПТИВНАЯ СИСТЕМА ТРАЕКТОРНОГО УПРАВЛЕНИЯ АНПА НА БАЗЕ НЕЙРОННОЙ СЕТИ ПРЯМОГО РАСПРОСТРАНЕНИЯ

Романова В. Р.¹

*(Крымский федеральный университет
имени В.И. Вернадского, Симферополь)*

Зуев С. В.²

*(Белгородский государственный технологический
университет им. В.Г. Шухова, Белгород)*

Работа посвящена разработке и исследованию интеллектуальной системы адаптивного автоматического управления с заданной целью, основанной на использовании искусственной нейронной сети прямого распространения. Объектом управления является автономный необитаемый подводный аппарат (АНПА). В работе предлагается подать на вход нейронной сети сигналы, получаемые от систем АНПА, и использовать выходной сигнал нейронной сети для управления с целью удержания аппарата на заданной траектории. В качестве результата работы предлагается модель и метод обучения, которые приводят к удержанию АНПА на заданной траектории при внешнем воздействии с естественным ограничением для рассматриваемого мобильного робота. При наличии непрерывной заданной траектории и дискретных сигналов от систем АНПА это позволяет следовать заданной траектории с помощью простой интеллектуальной системы управления, не требующей больших вычислительных мощностей. Предложенный метод обучения системы управления допускает предварительное обучение на численной модели движения аппарата со случайными внешними воздействиями, но не требует такого предварительного обучения при определенных условиях. Показано, в частности, что в случае достаточно большой скорости обучения модель успевает переадаптироваться и реагирует на изменившиеся обстоятельства. Предложенная интеллектуальная система адаптивного автоматического управления может найти применение в тех случаях, когда характерное время изменений в системе имеет порядок времени обучения, а траектория движения удовлетворяет изложенным в работе требованиям.

Ключевые слова: искусственная нейронная сеть, автономный обитаемый подводный аппарат, система адаптивного управления, интеллектуальная система управления, обучение искусственной нейронной сети.

¹ Валерия Романовна Романова, студент (lero4ka2004ro@gmail.com).

² Сергей Валентинович Зуев, к.ф.-м.н., доцент (sergey.zuev@bk.ru).

1. Введение

Задача адаптивного управления является одной из наиболее важных в области автоматического управления. Она заключается в разработке системы управления, способной эффективно адаптироваться к изменениям внешних условий.

Адаптивное управление используется там, где происходит изменение внешних условий, влияющих на работу системы, – это может происходить практически в любом виде деятельности. Внешние условия меняются независимо от состояния рабочей системы и влияют на ее работу по-разному, что с необходимостью приводит к учету в системах управления такого воздействия: предусматривается адаптивное управление, которое компенсирует внешние изменения за счет дополнительного управляющего воздействия. Задача адаптивного управления была поставлена в середине XX века [3], но ее универсальное решение пока отсутствует. Использование для этих целей систем искусственного интеллекта считается перспективным [31].

В этой работе рассматривается адаптивное управление автономным необитаемым подводным аппаратом (АНПА) с помощью искусственной нейронной сети (далее – ИНС) прямого распространения. Такой выбор (АНПА – ИНС прямого распространения) связан с тем, для АНПА характерное время изменений в системе, равное отношению размера аппарата к его скорости, больше, чем у большинства других мобильных роботов. Поэтому задача для такой системы управления допускает более сложные вычисления. То есть предложенная идея адаптивного управления может использоваться и в других системах, в которых имеется достаточно времени или ресурсов для вычислений, но из наиболее распространенных сейчас систем для этой цели хорошо подходит именно АНПА.

Автономные необитаемые подводные аппараты – это подводные мобильные роботы, которые предназначены для получения данных о внешней среде или воздействия на нее в автоматическом режиме, зачастую без информационно-командной связи с человеком-оператором. Применение АНПА является широким и включает в себя (не ограничиваясь этим) обследования мест авиакатастроф, океанографические исследования, разнообраз-

ные исследования морского дна, доставку грузов, производственные операции, добычу полезных ископаемых (например, аппараты компании Impossible Metals, <https://impossiblemetals.com>). Диапазон инновационных применений АНПА также огромен [19, 33].

Указанные задачи часто невозможно выполнить без использования автоматизированных систем управления АНПА. За последние годы было предложено множество различных схем управления. Проблема отслеживания траектории исследовалась в работе [15] с помощью модели сингулярных возмущений. Модельный прогностический контроллер для АНПА был предложен в [23]. Система локализации и планирования маршрута с учетом рельефа для навигации в течение длительного периода времени и определения местоположения АНПА описана в работе [27]. В работе [6] было разработано отказоустойчивое отслеживающее устройство. Координированное управление системой нескольких АНПА рассматривалось в [29].

Отдельно стоит указать системы управления, предложенные в связи с ограничениями вычислительной производительности бортовых вычислительных систем АНПА. Адаптивный не-сингулярный контроллер скользящего режима для АНПА, проявляющих только ответную активность (в норме пассивных), исследовался в работе [24]. В статье [22] модель АНПА была создана с использованием метода Лагранжа, а управление движением было организовано с помощью адаптивного контроллера скользящего режима на основе кватернионов. Ограниченный прогностический контроллер для АНПА был предложен в [21]. Контроллер на нечеткой логике рассмотрен в [12]. Оптимальный робастный контроллер для задачи отслеживания траектории АНПА исследован в [32].

В этой работе имеет существенное значение дискретность получения сигналов, используемых для управления. Поэтому имеет смысл указать и работы, направленные на преодоление ограничений управления в условиях задержек связи. Например, алгоритм управления формированием сигнала был разработан в [15], и был предложен контроллер, основанный на наблюдении расширенного состояния.

Таким образом, в данных работах рассмотрены различные подходы к адаптивному управлению в контексте разных задач управления и ситуаций, использованы различные подходы. Однако интеллектуальные системы управления использовались только в работах [12, 21], т.е. адаптивность к широкому спектру воздействий рассматривалась только в них. Вопрос об адаптивности к произвольным воздействиям по-прежнему открыт и интеллектуальная система на основе ИНС является перспективным кандидатом на его решение. Преодоление ограничений адаптации к произвольным внешним воздействиям особенно ценно при учете малых доступных вычислительных мощностей и задержек в получении данных.

Фактически АНПА представляют собой весьма неопределенные, связанные, изменяющиеся во времени и нелинейные системы [10]. Большинство моделей АНПА чрезвычайно упрощены из-за неточного характера методов и сложности моделирования, а также неизвестных корреляций и взаимосвязей между переменными, что приводит к несовершенству контроллеров для реальных подводных миссий. Немоделируемая динамика является неотъемлемой частью АНПА, впрочем, как и беспилотных летательных аппаратов (БПЛА), но характер управления для этих двух классов мобильных роботов все же существенно различен.

Интеграция искусственной нейронной сети в систему управления АНПА может быть сделана на основе мультиагентной архитектуры системы управления, описанной в [1]. На рис. 1 представлена схема системы управления из [1], в которой ИНС может быть интегрирована в сеть обмена данными в качестве посредника между системой навигации, системами позиционирования с одной стороны, и движительно-рулевым комплексом и системой дифферентовки – с другой стороны.

Обычно считается, что задача оптимального управления нелинейными системами решается как уравнение Гамильтона – Якоби – Беллмана, которое является очень сложным и не может быть решено напрямую для реальных систем. Реальные системы могут быть описаны как сложные, в данном случае – динамически сложные, т.е. их поведение не является детерминированным и, строго говоря, не определяется решением какого-либо урав-

нения. Моделирование таких систем часто делается с помощью машинного обучения. Этот метод использует закономерности, которые есть в системе, не предьявляя их явно.



Рис. 1. Схема системы управления АНПА [17]

Одним из самых распространенных моделей машинного обучения является искусственная нейронная сеть (ИНС). Среди методов обучения ИНС есть обучение с подкреплением [28], которое является подходящим в случае изменчивости среды. Обучение с подкреплением подразумевает обратную связь от системных переменных и постоянно улучшает предсказание работы системы [4, 5, 11]. Для оценки текущей эффективности системы используется обратная связь от управляющего воздействия и последующие шаги управления делаются с учетом этой обратной связи и ее влияния на достижение цели.

Применительно к настоящей работе можно указать следующие современные исследования в области интеллектуального управления с обучением с подкреплением. В [25] был разработан контроллер отслеживания траектории АНПА с помощью обучения с подкреплением. Такой же вид обучения был использован в работе [34] для решения задачи оптимального управления для систем с насыщением и отказами. Проблема адаптивной стратегии обучения с подкреплением была изучена для класса

аффинных нелинейных систем с дискретным временем в работе [13]. Для класса стохастических нелинейных систем в [30] был разработан упрощенный оптимизированный метод управления с использованием обучения с подкреплением. В [20] разработали оптимальный контроллер отслеживания для мультиагентных систем с дискретным временем. Контроллер, основанный на обучении с подкреплением, был предложен в [35] для малоактивных (в норме пассивных) морских надводных судов. Оптимальный контроллер оценки бесконечного горизонта был предложен для системы с неизвестными параметрами дрейфа в [8].

Несмотря на указанные достижения в области контроллеров на основе обучения с подкреплением для нелинейных систем, практически везде имело место насыщение привода. Насыщение привода происходит, когда контроллер генерирует уровень управляющего сигнала, превышающий допустимые пределы работы приводов. Такая ситуация наносит вред как контроллеру, так и транспортному средству и, кроме того, приводит к сбою управления. Чтобы решить эту проблему, необходимо разработать системы управления с защитой от насыщения, что является математически сложной задачей [9]. Одним из путей ее решения является компенсация нелинейности насыщения привода с помощью нейронных сетей.

В нашей работе делается попытка построения простого контроллера, претендующего на преодоление указанных выше проблем накопления ошибок и насыщения с минимальными требованиями к вычислительным мощностям.

2. Постановка задачи

Начальное положение аппарата $p^k(0)$, $k = 1, 2, 3$, и его желаемую трехмерную траекторию движения $x^k(t)$, $t \in [0, T]$, будем считать известными. На основе дополнительных предположений о нахождении внешних воздействий в определенном диапазоне в течение всего времени движения от 0 до T требуется обеспечить движение АНПА по траектории с минимальными отклонениями, используя данные бортовой инерциальной навигационной системы, данные системы позиционирования, обновляющи-

еся с определенными частотами, а также произведенные управляющие воздействия.

Будем считать, что имеется некоторая система позиционирования АНПА, использующая как инерциальные данные, так и возможные данные от реперных точек на дне, которая после первоначального определения положения АНПА периодически отслеживает изменение позиции аппарата и выдает его текущее положение в той же системе координат, в которой задано начальное положение и желаемая траектория АНПА. Таким образом, имеются значения $p^k(t_j)$ в некоторые моменты времени t_j .

Для дальнейшего требуется описать ограничения на допустимые внешние воздействия. Под ними будем подразумевать скорость потока воды в месте нахождения аппарата в момент времени t . Обозначим компоненты этой скорости через $f^k(t)$. Тогда корректная работа контроллера будет возможна только в случае

$$\int_0^T u(t)dt > \int_0^T f(t)dt,$$

где $u(t)$, $f(t)$ – абсолютные моментальные значения скоростей потоков в связанной системе координат и в абсолютной системе координат соответственно. Физически это условие означает, что управляющее воздействие в течение времени T превышает случайные внешние воздействия за то же время. На самом деле для корректной работы контроллера необходимо, чтобы разница в этом неравенстве была хотя бы на порядок величины.

Как уже упоминалось, выбор АНПА в качестве объекта управления связан с тем, что у этого вида мобильных роботов сравнительно большое характерное время. Например, для аппарата длиной 2 м, движущегося с максимальной скоростью 10 м/с, это время равно 200 мс. Процесс обучения ИНС рассматриваемого далее вида занимает около 30 мс, если имеются все данные. Данные, поступающие с наибольшими задержками, – это обычно позиционирование. Будем считать, что даже эти данные не задерживают процесс обучения на время, более чем на порядок превышающее характерное время для объекта. Тогда можно считать, что все изменения за цикл обучения относительно малы.

Управляющее воздействие на АНПА в нашей модели производится одинаковыми двигателями, число которых равно S . Ничто не мешает включить в управляющие элементы еще и систему дифферентовки, но тогда нужно будет описывать преобразования ее сигналов отдельно. Это лишь загромоздит рассмотрение без определенного смысла. Поэтому далее рассматриваем только двигатели. Каждый из них может воспринимать управляющий сигнал m^s , причем минимальный промежуток времени между последовательными командами равен θ_m . Конечно, результат воздействия на режим движения АНПА разных управляющих сигналов зависит нелинейным образом от самого сигнала и от внешних факторов (f), но мы будем считать, что эта зависимость в результате обучения будет содержаться в структуре данных весов и смещений ИНС, поскольку при обучении используется результирующее положение АНПА.

Таким образом, в задаче определены следующие сигналы:

1) $u^l(t)$ – ускорения и угловые скорости АНПА; в каждый момент времени это шестикомпонентный вектор, который предоставляется инерциальной навигационной системой; эти сигналы поступают с периодичностью θ_u и имеют квант η (одинаковый для всех компонент);

2) $p^k(t)$ – позиционирование АНПА, поступающее из подсистемы позиционирования; эти сигналы поступают в систему с периодичностью θ_p и имеют квант ξ (погрешность подсистемы положим постоянной и одинаковой для всех осей);

3) $f^k(t)$ – компоненты скорости внешнего течения (в абсолютной системе координат), не измеряемые и непредсказуемые; не имеет периодичности и установленного кванта;

4) $m^s(t)$ – управляющее воздействие – сигнал, который система подает двигателям, связанный какой-то функцией с их тягой (вид функции не имеет значения); в момент времени t имеются S значений $m^s(t)$, а работа управляющей системы состоит в том, чтобы получить столько же значений $m^s(t + \theta_m)$ и применить их в момент $t + \theta_m$; квантование этих сигналов задается возможностями управления и пусть квант равен μ ;

5) $x^k(t)$ – желаемая траектория; задается последовательностью значений с дискретностью θ_k , которую мы можем установить произвольно.

В этом случае решением задачи будет такая (зависящая от $p^k(t)$) последовательность векторов $m^s(t)$, которая приведет к минимальному отклонению $p^k(t)$ от $x^k(t)$ при любом допустимом (в вышеуказанном смысле) внешнем воздействии.

3. Метод решения с помощью алгоритма прямого распространения

Искусственная нейронная сеть – это компьютерная система, реализующая концепцию машинного обучения, т.е. построения модели процесса на примерах его работы. О нейронных сетях написано огромное число работ и нет смысла делать их полный обзор, но используемые в данном исследовании ИНС прямого распространения хорошо описаны, например, в книгах [2] и [7].

ИНС прямого распространения (FFNN) – это простой тип нейронной сети, в котором информация передается только в одном направлении: от входных нейронов к выходным. Такие ИНС просты в реализации и сравнительно быстро обучаются, имеют хорошую производительность на задачах классификации и регрессии [14, 17, 18]. Недостатками таких ИНС является плохая адаптация к обработке последовательностей или данных, меняющихся со временем. Поэтому, казалось бы, такая архитектура ИНС не очень подходит для поставленной в нашей работе задачи.

Однако если переформулировать задачу так, чтобы в каждый фиксированный сравнительно небольшой промежуток времени система управления обеспечивала уменьшение отклонения аппарата от траектории, то такая задача уже не является обусловленной прошлыми значениями сигналов и может быть успешно решена ИНС прямого распространения. Иными словами, мы предлагаем заново обучать систему сразу после того как получили прямой проход на только что полученных весах и смещениях. С учетом того, что характерное время объекта управления того же порядка величины, что и время обучения и работы ИНС, таким образом получится онлайн обучаемая си-

стема, постоянно воспринимающая новые данные. Мало того, такая система будет реализовывать концепцию обучения с подкреплением, так как регулярно получаемая оценка имеется – это разность выдачи ИНС в канале позиционирования с данными траектории.

В задаче имеется несколько величин, характеризующих временную дискретность: θ_u , θ_p , θ_m , θ_x , причем θ_x можно выбирать произвольно: непрерывную кривую, соединяющую начальное и конечное положение аппарата, можно представить последовательностью с любой дискретностью. Остальные величины зависят от характеристик инерциальной навигационной системы, возможностей системы позиционирования и вычислительных мощностей системы управления – эти величины нельзя менять произвольно.

Для расчетов на ЭВМ все величины так или иначе представляются целыми числами. То есть можно выбрать такую величину кванта времени τ , что все указанные выше величины дискретностей будут целыми положительными числами.

Можно выбрать τ настолько малым, что $T_x = \theta_x / \tau$ будет с достаточной степенью точности наибольшим общим делителем величин $T_u = \theta_u / \tau$, $T_p = \theta_p / \tau$, $T_m = \theta_m / \tau$. Поэтому далее будем использовать именно целочисленную нотацию, предполагая справедливость следующих соотношений:

- (1) $T_u, T_p, T_m \in N$,
- (2) $T_x = \text{НОД}(T_u, T_p, T_m)$.

Иначе говоря, T_u / T_x , T_p / T_x , $T_m / T_x \in N$. В этом случае любой момент времени, в котором значения величин в системе существенны для состояния системы, будет иметь вид $t_j = j\tau T_x$, $j = 0, 1, \dots, \frac{T}{\tau T_x}$, а в начале траектории $t = 0$. Зададим переход

системы из состояния, соответствующего времени t_j , в состояние, соответствующее времени t_{j+1} , в виде модели распространения сигнала в полносвязной ИНС прямого распространения с числом входов, на 3 большим числа нейронов (так как f не предсказывается), см. рис. 2.

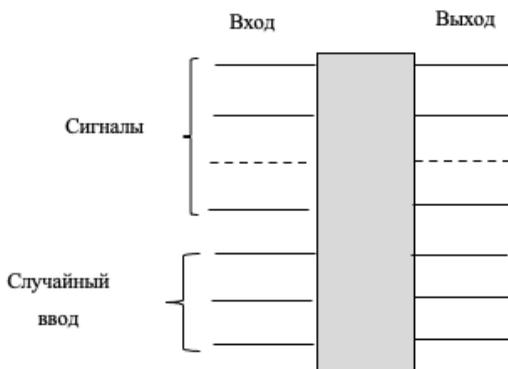


Рис. 2. Первый слой ИНС: переход системы из состояния t_j к состоянию t_{j+1}

Далее будем достраивать слои нейронной сети, сохраняя число сигналов, причем не будем заботиться о вводе внешних воздействий: первый ввод их был случайным, а далее ИНС сама их моделирует для учета в дальнейших оценках позиции и других сигналов, значения которых время от времени становятся известными. Так как сигналы в этой ИНС характеризуют текущее состояние АНПА и управляющие воздействия на него, то можно считать эту систему управления адаптивной: внешние воздействия на нее будут отражаться в сигналах системы позиционирования, а вначале введенное случайное воздействие в конечном счете превратится просто в набор скрытых параметров системы. При должном обучении такая система будет стремиться удерживать АНПА на заданной траектории, если правильно построить целевую функцию.

В начале известны все величины, кроме $f^k(0)$, и движители положим выключенными:

$$(3) \quad u_0^j, p_0^k \equiv x_0^k, m_0^s = 0.$$

Для нахождения весов и смещений ИНС требуются данные для ее обучения. А именно, соответствующие реальности совокупности сигналов u , p , m и меток x такие, что отклонение $|p - x|$ стремится к нулю (уменьшается по сравнению с предыдущим). Иначе говоря, для каких-то значений времени имеются u^k, p^k, m^s

такие, что для более позднего момента времени величина $|p - x|$ меньше, чем для начального.

Ввиду того, что отклонения, вызываемые внешними воздействиями за рассматриваемое время, сравнительно малы, мы положим, что в выходе последнего слоя должно быть $p^k = x^k$ для соответствующего момента времени. Тем самым «усилия» ИНС всегда будут направлены на достижение нужных точек траектории, куда бы вначале ни было направлено судно. И управляющее воздействие будет окончательно определяться именно в последнем слое.

Участок ИНС, в конце которого будут известны все сигналы (кроме трех, изображенных на рис. 1 внизу), определится числом слоев $N = \text{НОК}(T_u, T_p, T_m)$.

В этой N -слойной нейронной сети будут все параметры системы управления в виде ее весов и смещений.

4. Обучение ИНС прямого распространения

Рассмотрим процесс обучения на примере конкретного числа слоев. Пусть $T_u = 2$, $T_p = 4$, $T_m = 3$, $T_x = 1$, $N = 12$. Тогда полносвязная ИНС прямого распространения будет иметь вид, представленный на рис. 3.

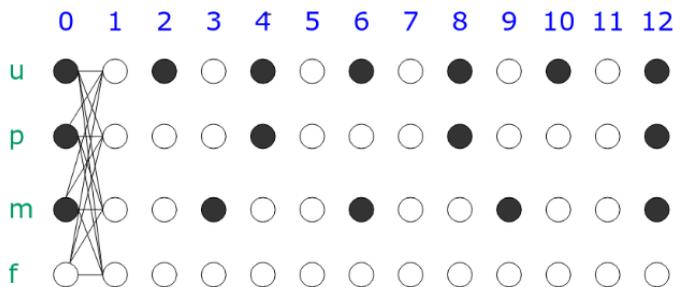


Рис. 3. Частный случай ИНС системы адаптивного управления

Обозначим веса и смещения через W_l^{da} и b_l^n соответственно, где l – номер слоя, n – номер нейрона, d, a – номера нейронов

донора (в предыдущем слое) и акцептора (в текущем слое). Тогда выход любого слоя запишется в виде

$$(4) \quad y_l^a = \delta_l^a \left(\sum_{d=0}^{D-1} W_l^{da} y_{l-1}^d + b_l^a \right),$$

$$(5) \quad l = 1, \dots, N,$$

$$(6) \quad D = J + S + 6,$$

где D – количество признаков, соответствующих разным физическим величинам; J – число моментальных показаний инерциальной навигационной системы; S – количество движителей.

Кодировка каждого слоя будет следующей:

$$(4) \quad y_0^0 = u_0^0, \dots, y_0^{J-1} = u_0^{J-1};$$

$$y_0^J = p_0^0, \dots, y_0^{J+2} = p_0^2;$$

$$y_0^{J+3} = m_0^0, \dots, y_0^{J+2+S} = m_0^{S-1};$$

$$y_0^{J+3+S} = f_0^0, \dots, y_0^{J+5+S} = f_0^2.$$

Инициуруем случайные параметры W_l^{da} и b_l^n . Пройдем ИНС прямо до последнего слоя, получим выходы y_l^a . Размеченные значения выходов будем обозначать через \hat{y} . Начнем обучение с последнего слоя (обратное распространение ошибки).

Коррекция весов последнего слоя ($l = 12$) будет выглядеть так (дельта-правило):

$$(8) \quad \Delta W_l^{da} = -k \sum_a (y_l^a - \hat{y}_l^a) y_l^a (1 - y_l^a) y_{l-1}^d,$$

$$(9) \quad \Delta b_l^a = -k \sum_a (y_l^a - \hat{y}_l^a) y_l^a (1 - y_l^a),$$

где суммирование ведется по всем имеющимся пакетам сигналов.

Каждый пакет сигналов формируется после прохода ИНС. То есть после первого прохода $a = 1$, затем $a = 2$ и т.д. Разница при каждом новом проходе ИНС будет в наборе сигналов на ее входе, а также в наборе меток для обучения в слоях. То есть данные для обучения модель получает прямо из своей работы. Конечно, в начале работы обучение не будет заметно и модель будет вести себя скорее всего неадекватно, но чем дальше будет происходить обучение, тем лучше будет удержание траектории.

Далее увидим это на графиках. В реальности неадекватное поведение не будет заметно, так как длительность этого режима меньше или порядка характерного времени для объекта управления.

Согласно принятому в методе обратного распространения ошибки предположению, разметка слоя $l - 1$ будет иметь вид

$$(10) \hat{y}_{l-1}^d = y_{l-1}^d - \sum_a \delta_l^a W_l^{da},$$

$$(11) \delta_l^a = (y_l^a - \hat{y}_l^a) y_l^a (1 - y_l^a).$$

При обучении ИНС мы будем пользоваться этим правилом только тогда, когда нет реальных данных (в случае прозрачных точек на рис. 2). Если же реальные данные имеются, то будем использовать их вместо соответствующего \hat{y}_{l-1}^d без расчета по обобщенному дельта-правилу. В последнем слое текущего прохода сети в сигналах p всегда будем ставить разметку x . Но при дальнейшем прохождении сети данные для пакета номер $a - 1$ будут содержать реальную позицию p вместо x . Так будет обеспечена подстройка весов сети на большем, чем $N\tau$, промежутке времени. Кроме того, напомним, что для всех пакетов, кроме первого, имеются сигналы и от f (их нет физически, они моделируется нейронной сетью).

В целом сигналы, подаваемые в ИНС, будут иметь следующий вид:

1. Показания инерциальной навигационной системы u^j в количестве J и каждое может принимать ряд значений (их количество равно мощности P_u признака u^j и зависит от кванта и максимального значения).

2. Данные системы позиционирования p^k в виде трех векторов и каждое может принимать ряд значений (их количество равно мощности P_p признака p^j и зависит от кванта и максимального значения).

3. Данные управляющего воздействия m^s в виде допустимых команд на каждый из S движителей.

После этого можно тестировать АНПА: запускать его с разными начальными данными и по ходу его движения он всегда

должен стремиться к желаемой траектории, корректируя направление движителями.

5. Численное моделирование работы системы управления

5.1. ДАННЫЕ И СТРУКТУРЫ ДАННЫХ

Для численного моделирования использовался язык Python 3.10 с библиотеками NumPy (работа с массивами данных) и matplotlib для построения графиков.

Сигналы инерциальной навигационной системы u смоделированы в виде шести значений, лежащих в диапазоне от 0 до 9,9 с квантом 0,1, т.е. 6 признаков мощностью 100 каждый. На каждый момент времени, когда эти сигналы доступны, выдаются все 6 значений. Для удобства хранения и работы входящий сигнал сети записывается в один массив с кодировкой в соответствии с формулами (4). Для процесса обучения таких сигналов должно быть много, т.е. мы будем их накапливать при каждом проходе сети. Поэтому этот массив представляет собой матрицу, в каждом векторе данных которой (строке) первые 6 (от 0 до 5) значений – показания инерциальной навигационной системы. Сигналы, получаемые в движении, хранятся в оперативной памяти в виде словарей. Ключом первого словаря является временная метка (номер слоя), значением – второй словарь. Ключом второго словаря является номер сигнала, а значением – массив значений этого сигнала в это время для каждого экземпляра данных.

Сигналы системы позиционирования p , когда они доступны, представляют собой три значения. Их диапазон и квант в модели приняты равными соответственно $-100, \dots, 99,99$ и $0,01$. То есть это три признака с мощностями по 20000 каждый. Их записываем аналогично: входящие значения – в массив входящих данных признаками 8, 9, 10. Ясно, что ключи первого словаря не будут охватывать все слои нейронной сети (так как данные появляются не на каждом шаге времени). Для каждого пакета, используемого при обучении, имеется своя такая структура данных.

Сигналы управления двигателями m доступны каждые θ_m единиц времени. В модели для визуализации рассматривается 4 двигателя. В стартовой позиции все двигатели отключены: это четыре сигнала, равных 0. Далее они будут принимать значения от $-0,9$ до $1,0$ с квантом $0,1$. Таким образом, у нас будет 4 сигнала с мощностью 20 каждый. Их хранение осуществляется аналогично вышеуказанным сигналам: входящий – в массиве, текущие – в своих словарях для каждого пакета (т.е. для каждого прохода сети).

Параметризация желаемой траектории задается в виде последовательности точек, т.е. в виде матрицы. Структурой данных для неё является массив. В каждый момент времени (в каждом слое ИНС) кроме момента выхода пакета из ИНС будет присутствовать либо предсказанное значение p , либо значение, пришедшее из навигационной системы, и только на выходе будет использоваться траекторное значение x . Поэтому и значения из словарей на этапе первичного обучения никак не будут взаимодействовать с желаемой траекторией [16]. Грубо говоря, система должна сначала описать в параметрах свое окружение, а уже потом должна обучаться движению по траектории.

5.2. ПРОГРАММНАЯ РЕАЛИЗАЦИЯ

В качестве примера создадим нейронную сеть из 12 слоев по 16 нейронов в каждом. 16 сигналов на входе формируются как 6 сигналов инерциальной навигационной системы, 3 сигнала системы позиционирования, 4 сигнала управляющего воздействия и три сигнала генерируются случайно (только для первого пакета!) – это эмуляция внешних течений f , которые далее полагаются задаваемыми точно такими же параметрами, как u : с теми же диапазонами и квантованием.

В программном коде использовались следующие функции:

- 1) генерации весов и смещений сети;
- 2) прохода слоя вперед;
- 3) проход нескольких слоев вперед;
- 4) коррекции весов и смещений последнего слоя (по меткам выхода нейронной сети);

5) коррекции весов и смещений скрытых слоев (с учетом частично размеченных данных).

Изначально генерируем случайные веса и смещения, затем обучаем нейронную сеть.

Внутри функции обучения основной цикл проходит по каждой эпохе обучения, которая соответствует каждому новому пакету данных. То есть эпоха использует все данные, которые накопились до нее и для нее. Для каждой эпохи проходит цикл по каждому слою ИНС. Обучение каждого слоя методом обратного распространения ошибки с использованием градиентного спуска. Обновленные весовые матрицы и векторы смещений сохраняются.

В результате обучения получаем матрицу весов и смещений, которая содержит информацию о закономерностях движения АНПА.

С помощью библиотеки `matplotlib` производим построение графиков абсолютной величины отклонения аппарата от желаемой траектории в каждый момент времени (рис. 4).

5.3. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Моделирование движения аппарата по траектории для испытания работы предложенной системы управления было произведено путем генерации небольших отклонений от заданной траектории, которая, в свою очередь, представляет собой кривую в абсолютной системе координат. В численном моделировании турбулентные эффекты не учитывались, так как предполагалось малое смещение аппарата. Здесь надо сказать, что природа движения объекта для целей управления в данном случае вообще не важна: главное, чтобы во время обучения (менее секунды) и управляющего воздействия (еще меньше) это движение подчинялось каким-то неизменным в этом промежутке времени законам (любым). Воздействие делалось добавками к компонентам скорости, нормально распределенными около 0 со стандартным отклонением 0,1 модуля скорости.

Проведено по три испытания с разными стартовыми сигналами f для систем, обученных на 80 и на 100 эпохах. Как видно на графике на рис. 6, отклонение от траектории в последних 12

шагах после 80 эпох обучения начало уменьшаться с 8-го шага во всех испытаниях.

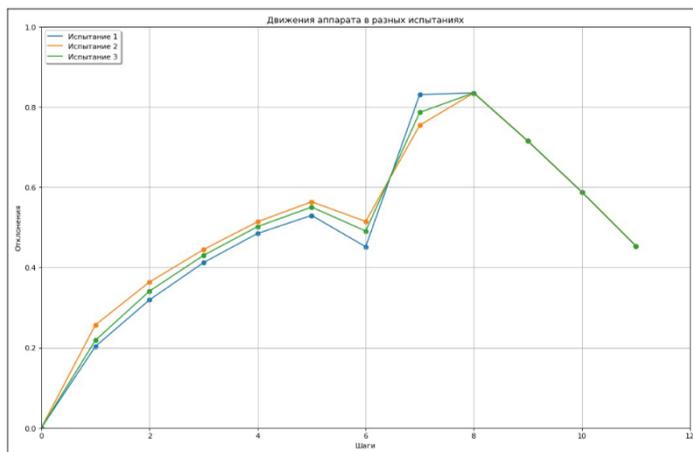


Рис. 4. Абсолютные величины отклонения аппарата от желаемой траектории в каждый момент времени в каждом испытании после 80 эпох обучения

На 3D-графиках рис. 5 и 6 можно увидеть, как смоделированная и обученная соответственно в 80 и 100 эпохах система управления провела аппарат по заданной траектории.

Видно, что на 100 эпохах качество проводки лучше, но в целом и система, обученная в 80 эпохах, установила аппарат на траекторию и удержала даже на значительном искривлении.

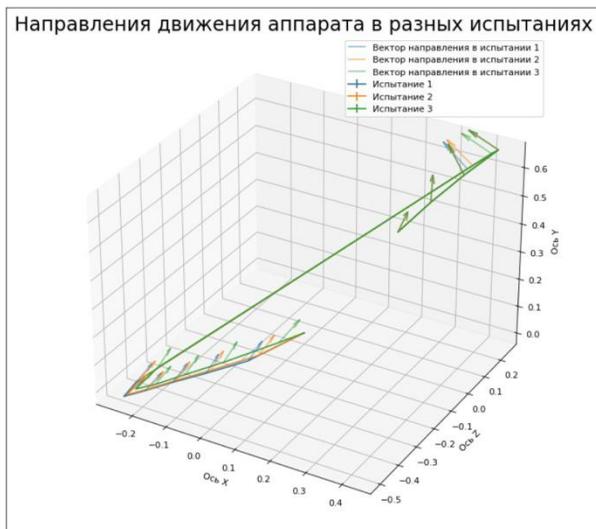


Рис. 5. Траектории движения аппарата в испытаниях с системой, обученной 80 эпохами

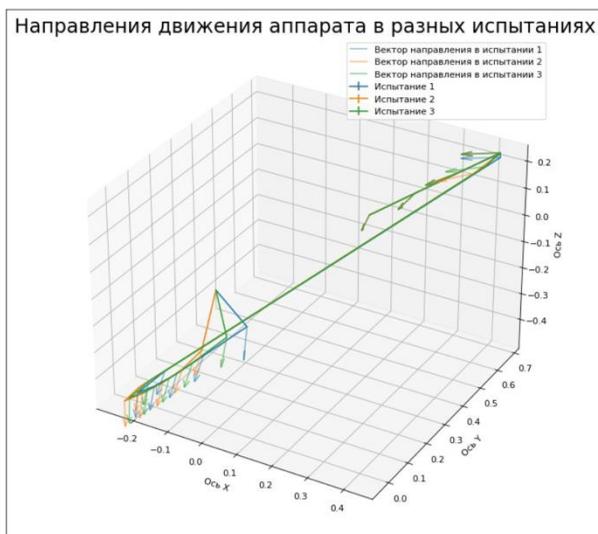


Рис. 6. Траектории движения аппарата в испытаниях с системой, обученной в 100 эпохах

Графики кривых обучения для разных испытаний приведены на рис. 7. Видно, что скорость обучения варьируется в зависимости от стартовых условий, но незначительно. На графике представлены только нейроны, отвечающие за данные инерциальной навигационной системы, внешние и управляющие воздействия, так как позиционирование уже проиллюстрировано выше.

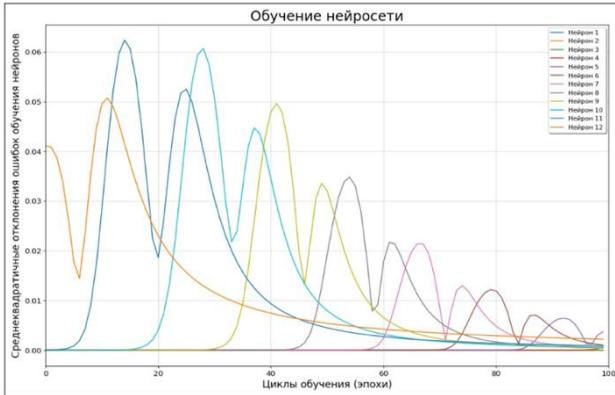


Рис. 7. Процесс обучения ИНС

6. Заключение

В настоящем исследовании была предложена и смоделирована интеллектуальная система адаптивного управления для АНПА с заданной целью – удержание аппарата на заданной траектории.

Ключевой идеей работы было использование простой ИНС, которая эксплуатируется как онлайн-обучаемая система: она выдает требуемую информацию об управляющих воздействиях и обучается одновременно. Для обучения используются накапливающиеся данные о работе АНПА и системы управления.

В результате численного моделирования работы системы управления предложенного типа были получены срезы моделированных и целевых траекторий движения АНПА, представленные на графиках.

Полученный результат показывает, что даже сравнительно простые системы искусственного интеллекта при определенной

стратегии обучения способны достигать хороших результатов в задачах управления. Так, в модели 12-слойной ИНС обучение достигло устойчивых результатов через 100 эпох. То есть понадобилось прохождение 1200 слоев. Поскольку каждый слой эквивалентен промежутку времени минимальной дискретности, который может соответствовать сотым долям секунды, обучение практически было закончено примерно за секунду. Но конкретные значения временных параметров зависят от объекта управления и условий его движения.

Литература

1. БОРЕЙКО А.А., ИНЗАРЦЕВ А.В., МАШОШИН А.И., ПАВИН А.М. и др. *Система управления АНПА большой автономности на базе мультиагентного подхода* // Подводные исследования и робототехника. – 2019. – Т. 28, №2 – С. 23–31.
2. РАШИД Т. *Создаем нейронную сеть*. – СПб. : ООО «Диалектика», 2019. – 272 с.
3. ЦЫПКИН Я.З. *Адаптация и обучение в автоматических системах*. – М.: Гл. ред. физ.-мат. лит-ры изд-ва «Наука», 1968. – 400 с.
4. ABOUSSALAH A.M., LEE C. *Continuous control with Stacked Deep Dynamic Recurrent Reinforcement Learning for portfolio optimization* // Expert Systems with Applications. – 2020. – Vol. 140. – P. 1–11.
5. CARLUCHO I., PAULA M.D., ACOSTA G.G. *Double Q-PID algorithm for mobile robot control* // Expert Systems with Applications. – 2019. – Vol. 137. – P. 292–307.
6. CHE G., YU Z. *Neural-network estimators based fault-tolerant tracking control for AUV via ADP with rudders faults and ocean current disturbance* // Neurocomputing. – 2020. – Vol. 411. – P. 442–454.
7. *Deep Neural Evolution* / Eds.: H. Iba, N. Noman // Springer Singapore. – 2020. – 438 p. – ISBN 978-981-15-3685-4 (electronic).
8. DEPTULA P., BELL Z.I., DOUCETTE E.A., CURTIS J.W. et al. *Data-based reinforcement learning approximate optimal con-*

- trol for an uncertain nonlinear system with control effectiveness faults // Automatica. – 2020. – Vol. 116. – P. 1–10.*
9. ELHAKI O., SHOJAEI K. *A robust neural network approximation-based prescribed performance output-feedback controller for autonomous underwater vehicles with actuators saturation // Engineering Applications of Artificial Intelligence – 2020. – Vol. 88. – P. 1–16.*
 10. ELHAKI O., SHOJAEI K. *Neural network-based target tracking control of underactuated autonomous underwater vehicles with a prescribed performance // Ocean Engineering. – 2018. – Vol. 167. – P. 239–256.*
 11. GLATT R., SILVA F.L., BIANCHI R.A., COSTA A.H. *DECAF: Deep Case-based Policy Inference for knowledge transfer in Reinforcement Learning // Expert Systems with Applications. – 2020. – Vol. 156. – P. 1–13.*
 12. HAN H., WEI Y., YE X., LIU W. *Modeling and fuzzy decoupling control of an underwater vehicle-manipulator system // IEEE Access. – 2020. – Vol. 8. – P. 18962–18983.*
 13. HUANG Z, BAI W., LI T., LONG Y. et al. *Adaptive reinforcement learning optimal tracking control for strict-feedback nonlinear systems with prescribed performance // Information Sciences. – 2023. – Vol. 621. – P. 407–423.*
 14. JAIN A., THOKE A.S., PATEL R.N. *Fault classification of double circuit transmission line using artificial neural network // Int. Journal of Electrical Systems Science and Engineering. – 2008. – Vol. 1. – P. 230–235.*
 15. LEI M. *Nonlinear diving stability and control for an AUV via singular perturbation // Ocean Engineering. – 2020. – Vol. 197. – P. 1–11.*
 16. LEWIS F.L., VRABIE D.L. *Reinforcement learning and adaptive dynamic programming for feedback control // IEEE Circuits and Systems magazine. – 2009. – Vol. 9. – P. 32–50.*
 17. MAHANTY R.N., DUTTA GUPTA P.B. *Comparison of Fault Classification Methods Based on Wavelet Analysis and ANN // Electric Power Components and Systems. – 2006. – Vol. 34 – P. 47–60.*
 18. MONTGOMERY D.C. *Design and Analysis of Experiments. – John Wiley & Sons, Inc, 1997. – P. 757.*

19. NEVES G., RUIZ M., FONTINELE J., OLIVEIRA L. *Rotated object detection with forward-looking sonar in underwater applications* // Expert Systems with Applications. – 2020. – Vol. 140. – P. 1–18.
20. PENG Z., ZHAO Y., HU J., GHOSH B.K. *Data-driven optimal tracking control of discrete-time multi-agent systems with two-stage policy iteration algorithm* // Information Sciences. – 2020. – Vol. 481. – P. 189–202.
21. RASHIDI A.J., KARIMI B., KHODAPARAST A. *A constrained predictive controller for AUV and computational optimization using Laguerre functions in unknown environments* // Int. J. Control Autom. Syst. – 2020. – Vol. 18. – P. 753–767.
22. RODRÍGUEZ J., CASTAÑEDA H., GORDILLO J.L. *Lagrange modeling and navigation based on quaternion for controlling a micro AUV under perturbations* // Robotics and Autonomous Systems. – 2020. – Vol. 124. – P. 1–12.
23. SHEN C., SHI Y. *Distributed implementation of nonlinear model predictive control for AUV trajectory tracking* // Automatica. – 2020. – Vol. 115. – P. 1–9.
24. SU H., ZHANG H., SUN S., CAI Y. *Integral reinforcement learning-based online adaptive event-triggered control for non-zero-sum games of partially unknown nonlinear systems* // Neurocomputing. – 2020. – Vol. 377. – P. 243–255.
25. SUN Y., XIANGRUI R., ZHANG G., XIANGBIN W., XU H. *AUV path following controlled by modified Deep Deterministic Policy Gradient* // Ocean Engineering. – 2020. – Vol. 210. – P. 1–14.
26. SURYENDU C., SUBUDHI B. *Formation control of multiple autonomous underwater vehicles under communication delays* // IEEE Trans. on circuits and systems II: Express briefs. – 2020. – Vol. 67, No. 12. – P. 3182–3186.
27. TENG M., YE L., YUXIN Z., YANQING J. et al. *An AUV localization and path planning algorithm for terrain-aided navigation* // ISA Transactions. – 2020. – Vol. 103. – P. 215–227.
28. VAMVOUDAKIS K.G., LEWIS F.L. *Online actor–critic algorithm to solve the continuous-time infinite horizon optimal control problem* // Automatica. – 2010. – Vol. 46. – P. 878–888.

29. WANG X., ZERR B., THOMAS H., CLEMENT B. et al. *Pattern formation of multi-AUV systems with the optical sensor based on displacement-based formation control* // Int. Journal of Systems Science. – 2020. – Vol.51:2. – P. 348–367.
30. WEN G., CHEN C.L., LI W.N. *Simplified optimized control using reinforcement learning algorithm for a class of stochastic nonlinear systems* // Information Sciences. – 2020. – Vol. 517. – P. 230–243.
31. *What's New in Artificial Intelligence from the 2022 Gartner Hype Cycle* [Электронный ресурс] // Gartner. – URL: <https://tinyurl.com/4z4vrhcy> (дата обращения: 26.06.2023).
32. XIA Y., XU K., WANG W., XU G. et al. *Optimal robust trajectory tracking control of a X-rudder AUV with velocity sensor failures and uncertainties* // Ocean Engineering. – 2020. – Vol. 198. – P. 1–16.
33. YAN X-P., WANG S-W., MA F., LIU Y-C. et al. *A novel path planning approach for smart cargo ships based on anisotropic fast marching* // Expert Systems with Applications. – 2020. – Vol. 159. – P. 1–16.
34. YUAN H, SHUAI C., MA J., FANG Y. *An efficient control allocation algorithm for over-actuated AUVs trajectory tracking with fault-tolerant control* // Ocean Engineering. – 2023. – Vol. 273. – P. 113976.
35. ZHENG Z., RUAN L., ZHU M., GUO X. *Reinforcement learning control for underactuated surface vessel with output error constraints and uncertainties* // Neurocomputing. – 2020. – Vol. 399. – P. 479–490.

ADAPTIVE TRAJECTORY CONTROL SYSTEM AUV BASED ON A DIRECT PROPAGATION NEURAL NETWORK

Valeria Romanova, V.I. Vernadsky Crimean Federal University, Simferopol, student (lero4ka2004ro@gmail.com).

Sergei Zuev, Belgorod State Technological University named after V.G. Shukhov, Belgorod, Candidate of Physical and Mathematical Sciences, Associate Professor (sergey.zuev@bk.ru).

Abstract. The paper is devoted to the development and study of an intelligent system of adaptive automatic control with a given target based on the use of artificial neural network of forward propagation. The control object is an autonomous unmanned underwater vehicle (AUV). In this paper, it is proposed to feed the signals received from the systems of the AUV to the input of the neural network, and use the output signal of the neural network for control to keep the vehicle on a given trajectory. As a result of this work, a model and a learning method are proposed that lead to holding the ANPA on a given trajectory under an external influence with a natural constraint for the considered mobile robot. Given a continuous preset trajectory and discrete signals from the ANPA systems, this allows following the preset trajectory with a simple intelligent control system that does not require large computational power. The proposed method of control system training allows pre-training on a numerical model of vehicle motion with random external influences, but does not require such pre-training under certain conditions. It is shown, in particular, that in the case of a sufficiently large learning rate, the model has time to rearrange itself and reacts to changed circumstances. The proposed intelligent system of adaptive automatic control can find application in those cases when the characteristic time of changes in the system is of the order of the training time, and the trajectory of motion satisfies the requirements stated in the paper.

Keywords: artificial neural network, autonomous unmanned underwater vehicle, adaptive control system, intelligent control system, artificial neural network training.

УДК 004.8

ББК 32.813.5

DOI: 10.25728/ubs.2024.108.11

*Статья представлена к публикации
членом редакционной коллегии Б.Р. Андриевским.*

*Поступила в редакцию 21.10.2023.
Опубликована 31.03.2024.*