

Институт проблем управления
им. В.А. Трапезникова РАН

УПРАВЛЕНИЕ БОЛЬШИМИ СИСТЕМАМИ

Выпуск 115
Май 2025

**СБОРНИК
ТРУДОВ**

ISSN 1819-2467

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

Москва – 2025

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

**УПРАВЛЕНИЕ
БОЛЬШИМИ
СИСТЕМАМИ**

СБОРНИК ТРУДОВ

Выпуск 115

Москва – 2025

КООРДИНАЦИОННЫЙ СОВЕТ

Академики: Васильев С.Н., Желтов С.Ю., Каляев И.А., Куржанский А.Б., Попков Ю.С., Черноусько Ф.Л.; члены-корреспонденты РАН: Галяев А.А., Савватеев А.В.; д-ра техн. наук: Кузнецов О.П. (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. физ.-мат. наук Горбунова А.В. **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. Сараев 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

Гежа В.Н., Козицин И.В.

Аналитическое решение задачи управления мнениями в SCARDO-модели для случая двухэлементной азбуки мнений с одним типом нативных агентов..... 6

Постнова О.С., Тананко И.Е., Рогачко Е.С.

Приближенный анализ длительности пребывания требований в сети массового обслуживания с делением и слиянием требований 33

Салимзянов Р.Р., Моисеев А.Н., Сущенко С.П.

Исследование сети массового обслуживания с неограниченным числом приборов в узлах и прерываниями обслуживания 52

Тали Д.И., Финько О.А.

Модели электронного документа в рамках парадигмы «Индустрия 4.0»..... 66

Математическая теория управления

Подвальный С.Л., Васильев Е.М.

Эволюционная методология многоальтернативного управления сложными системами 100

Юрченков И.А.

Эффективный метод численного решения объемных интегральных уравнений Фредгольма для задач моделирования распространения акустических волн . 138

Анализ и синтез систем управления

Романенко В.А.

Оптимизация распределения воздушных судов по авиалиниям при нечётких исходных данных 156

Информационные технологии в управлении

Русаков К.Д.

Байесовская модель объединения признаков представлений в задаче реидентификации личности 183

Сетевые модели в управлении

Гасанов И.И.

*Исследование аукционной схемы распределения
заказа внутри ассоциации производителей 203*

Управление в социально-экономических системах

**Бузиков М.Э., Петелина Ю.А., Красоткин С.А.,
Рыжов М.С., Козицин И.В.**

*Прогнозирование динамики общественного мнения
по лонгитюдным данным высокой гранулярности:
модель Абельсона, регрессионные модели и ансамбли
моделей..... 220*

**Овчинников П.В., Ткачев А.Н.,
Мирославская М.Д.**

*Использование иерархической многофакторной
регрессионной модели с целью анализа и
прогнозирования сельскохозяйственных показателей 241*

Управление техническими системами и технологическими процессами

Вытовтов К.А., Гладких Т.Я.

*Алгоритм идентификации загрязнений поверхности
водоемов для геоинформационных систем 263*

Фёдорова Е.А., Тюленина Я.А., Мелошникова Н.П.

*Математическое моделирование гибридных
энергетических систем с помощью теории массового
обслуживания 279*

Программы и системы моделирования объектов, средств и систем управления

Тисленко Т.И., Семенова Д.В.

*Разработка адаптивной системы управления
светофорными объектами с использованием
марковских процессов принятия решений..... 298*

АНАЛИТИЧЕСКОЕ РЕШЕНИЕ ЗАДАЧИ УПРАВЛЕНИЯ МНЕНИЯМИ В SCARDO-МОДЕЛИ ДЛЯ СЛУЧАЯ ДВУХЭЛЕМЕНТНОЙ АЗБУКИ МНЕНИЙ С ОДНИМ ТИПОМ НАТИВНЫХ АГЕНТОВ¹

Гежа В. Н.²

(Московский физико-технический институт, Москва)

Козицин И. В.³

*(ФГБУН Институт проблем управления
им. В.А. Трапезникова РАН, Москва, Московский
физико-технический институт, Москва)*

Рассматривается задача управления мнениями агентов, сформулированная в терминах приближения среднего поля для SCARDO-модели, построенного для случая социальной системы с модулярной сетевой структурой, гетерогенными уровнями активности агентов и персонализацией контента. В рамках модели управления специальная группа агентов, имитирующая ботов в реальных социальных медиа, воздействует на остальных агентов, с тем чтобы изменить их мнения в соответствии с интересами управляющего субъекта. Изучается наиболее простой для анализа случай двухэлементной азбуки мнений, в рамках которой существует только два возможных мнения (интерпретация – выборы в двухпартийной системе) и один тип агентов-неботов. Строится аналитическое решение задачи управления. Демонстрируется, что в зависимости от значений параметров модели может быть либо одна оптимальная стратегия управления, либо бесконечно много. При этом в случае, когда оптимальное управление единственно, оно должно быть устроено следующим образом: все социальные боты должны транслировать одно и то же мнение, а переключение распространяемого мнения может произойти не более одного раза лишь для некоторых конфигураций параметров. Полученные результаты сопровождаются примерами с реальными данными из онлайн платформы ВКонтакте.

Ключевые слова: модели формирования мнений, SCARDO-модель, приближение среднего поля, оптимальное управление, модулярные сети, алгоритмы персонализации.

¹ Работа выполнена при финансовой поддержке РФФ, грант №24-71-00070.

² Владислав Николаевич Гежа, аспирант (gezha.vn@phystech.edu).

³ Иван Владимирович Козицин, к.ф.-м.н. (kozitsin.ivan@mail.ru).

1. Введение

Разработка эффективных стратегий управления активными сетевыми структурами является чрезвычайно актуальной задачей современной теории управления [1, 2, 7]. С их помощью можно будет противостоять распространению деструктивного контента в социальных медиа, продвигать научно-обоснованную повестку, а также способствовать снижению социальной напряженности в обществе [3, 4, 10]. При этом исследования в данной области затруднены тем, что ученые до сих пор не знают, как устроено социальное влияние и как меняются мнения людей при информационных взаимодействиях [12]. В частности, активные дебаты ведутся касательно существования диссимилиятивной компоненты социального влияния, которая заставляет мнения людей отдаляться друг от друга в результате общения [6]. Вместе с тем последние эмпирические исследования указывают на легитимность данного механизма влияния [11, 13]. В свою очередь в работе [18] была высказана мысль, что эффект социального влияния в процессах формирования мнений может существенно зависеть от эмпирического контекста и в разных условиях могут работать различные механизмы влияния, такие как ассимилятивное влияние, ограниченное доверие или диссимилиятивное влияние. В связи с этим модели формирования мнений и основанные на них модели управления должны быть достаточно адаптивны.

В данной работе будет рассмотрена модель формирования мнений, предложенная в работе [12] (далее – SCARDO-модель). Данная агентная модель описывает взаимодействия агентов как последовательные парные контакты со стохастическим протоколом изменения мнения. Данный протокол задается при помощи распределения, которое является функцией текущих мнений и, возможно, других параметров агентов. В итоге SCARDO-модель обладает высокой степенью гибкости и позволяет описывать основные механизмы социального влияния как качественно, так и количественно.

В работе [14] на основе приближения среднего поля для

SCARDO-модели была сформулирована задача управления мнениями агентов, найдены необходимые условия оптимальности и разработаны численные методы для ее решения. Целью настоящей работы является аналитическое решение данной задачи управления для наиболее простого случая, когда число возможных мнений в системе равно двум (интерпретация – политические выборы с участием двух кандидатов/партий), а число возможных типов агентов равно одному. Ранее в работе [9] было построено расширение SCARDO-модели, учитывающее гетерогенность активности агентов [16], модулярность структуры взаимодействий, проявляющуюся в неоднородности плотности социальной сети и наличии плотных кластеров-сообществ [21], а также работу алгоритмов персонализации, которые влияют на информационное окружение пользователей в реальных социальных медиа-платформах, фильтруя контент на основании интересов и действий пользователей [5]. В настоящей работе задача управления сначала будет решена для более простой постановки, рассматриваемой в работе [14]. После этого полученные результаты будут расширены на обобщенную SCARDO-модель [9].

Данная работа является существенно расширенной версией доклада, представленного на XX Всероссийской школе-конференции молодых ученых «Управление большими системами».

2. Описание модели

Рассматривается система из N агентов, соединенных графом $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, где $\mathcal{V} = \{1, \dots, V\}$ – множество агентов, а $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ – ребра между ними. Каждый агент описывается мнением, принимающим значения из множества $Z = \{Z_1, \dots, Z_m\}$, и типом, принимающим значения из множества $\Xi = \{\Xi_1, \dots, \Xi_M\}$. Предполагаем, что мнения агентов могут обновляться в результате взаимодействий, в то время как типы агентов неизменны.

В каждый такт t (считаем, что время дискретно) случайным образом на основании наперед заданного на множестве агентов распределения активностей $\{\pi_1, \dots, \pi_N\}$ выбираются агент

(пусть i) и один из его соседей по социальной сети (пусть j). Считаем, что $\pi_1, \dots, \pi_N > 0$ и вероятность выбора агента i есть $\pi_1/(\pi_1 + \dots + \pi_N)$. В свою очередь, вероятность выбора агента j описывается формулой $\pi_j / \sum_{k \in \mathcal{V}_i} \pi_k$, где $\mathcal{V}_i \subseteq \mathcal{V}$ – множество друзей агента i ($j \in \mathcal{V}_i$).

После этого агент j влияет на агента i . Пусть мнения агентов i и j есть $x_i(t) = Z_s, x_j(t) = Z_l$, а типы равны $\xi_i = \Xi_f, \xi_j = \Xi_r$ соответственно. Мнение агента i трансформируется в соответствии с вероятностным распределением $\{p_{s,l,1}^{f,r}, \dots, p_{s,l,m}^{f,r}\}$, где $p_{s,l,k}^{f,r}$ – вероятность события, что мнение агента i примет значение Z_k . У данного параметра верхние индексы f и r отвечают за типы объекта влияния и субъекта влияния соответственно, а нижние индексы s и l согласованы с мнениями объекта влияния и субъекта влияния. Третий нижний индекс k соответствует мнению объекта влияния после взаимодействия. При такой системе обозначений неравенство $p_{s,l,s}^{f,r} > 0$ означает ненулевую вероятность того, что мнение останется неизменным. Выполнено условие нормировки:

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

Вероятности $p_{s,l,k}^{f,r}$ сгруппируем в последовательность $\{P^{f,r}\}_{f \in [M], r \in [M+1]}$, объекты $P^{f,r} = [p_{s,l,k}^{f,r}]_{s,l,k \in [m]}$ будем именовать таблицами переходов. Таблицы переходов удобно представлять как последовательности строчно-стохастических матриц размера $m \times m$: $P_1^{f,r}, \dots, P_m^{f,r}$, где матрица $P_s^{f,r} = [p_{s,l,k}^{f,r}]_{l,k \in [m]}$ определяет то, каким образом реагируют на влияние со стороны агентов с типом Ξ_r агенты с мнением Z_s и типом Ξ_f .

Далее будем предполагать, что параметры активности агентов π_1, \dots, π_N однозначно определяются их типами.

Также будем считать, что социальная сеть определяется стохастичной блочной моделью [15]. В рамках данной модели вершины делятся на блоки и для каждой пары блоков задаются вероятности возникновения связи между вершинами из этих блоков. Аналогично задаются вероятности для внутриблочных связей (для каждого из блоков отдельно). Для каждой пары вершин

связь образуется независимо. Мы будем считать, что блоки соответствуют типам агентов. Пусть $f \in [M]$ и $r \in [M]$ – индексы блоков. Тогда параметр $\rho_{f,r} \in [0, 1]$ задает вероятность формирования связи между двумя произвольными вершинами с типами Ξ_f и Ξ_r . Отметим, что при $f = r$ параметр $\rho_{f,f}$ описывает вероятность образования связи между вершинами одного типа (Ξ_f).

Дополнительно определим $(M + 1)$ -й тип агентов [14]. Будем считать, что агенты данного типа невосприимчивы к социальному влиянию и характеризуются параметром активности $\pi_{M+1} > 0$. Для таких агентов имеет место $p_{s,l,k}^{M+1,r} = 0$ при $s \neq k$. Содержательная интерпретация агентов с типом $M + 1$ – боты, аккаунты СМИ или рекламные сообщения в социальных медиа. Далее будем называть их агентами-ботами или просто ботами. Остальных агентов будем именовать нативными агентами. Агенты-боты разбиваются на группы (когорты), каждая когорта воздействует со своим типом нативных агентов и только с ними. Такая постановка позволяет учесть таргетированность реальных ботов или рекламных сообщений в социальных медиа. Для формализации такой структуры взаимодействий также будем использовать стохастичную блочную модель, в которой блоками выступают типы нативных агентов, а также «нацеленные» на них когорты ботов (которые будем индексировать сообразно типам нативных агентов). Более точно, вероятность формирования связи между нативным агентом с типом Ξ_r и ботом из соответствующей когорты будем описывать параметром $\rho^r \in [0, 1]$.

Будем предполагать, что связи между нативными агентами статичны, в то время как составы когорт могут меняться со временем, как и структура связей между ботами и нативными агентами (эта динамика операционализируется при помощи изменения величин ρ^r). Иными словами, предполагается, что боты адаптивны и могут менять тактику влияния в зависимости от текущей ситуации. Предполагаем, что мнения ботов, составы их когорт и структура связей с нативными агентами задаются экзогенно.

Также считаем, что коммуникации агентов модулируются алгоритмом персонализации Δ , который запрещает контакт между

двумя агентами i и j (пусть второй влияет на первого) с мнениями $x_i(t) = Z_s, x_j(t) = Z_l$ и типами $\xi_i = \Xi_f, \xi_j = \Xi_r$ с заданной вероятностью $\Delta_{s,l}^{f,r}$.

3. Приближение среднего поля

Для представленной выше модели в работе [9] было построено приближение среднего поля, описывающее динамику численностей сторонников различных мнений с учетом их типов в предположении, что число агентов в системе велико ($N \rightarrow \infty$). Пусть τ – масштабированное время, определяемое следующим образом: $\tau = t/N, \delta\tau = 1/N$. Пусть $y_{a,f}(\tau)$ – доля (относительно N) adeptов мнения Z_a с типом Ξ_f для $f \in [M]$ (где $[M]$ обозначает множество натуральных чисел от 1 до M) в момент времени τ . Пусть $u_l^f(\tau)$ – доля (относительно N) ботов из соответствующей когорты f с мнением Z_l в момент времени τ . Пусть

$$A = \sum_{f \in [M]} n_f \pi_f + u \pi_{M+1},$$

где n_f – доля нативных агентов с типом Ξ_f , а u – доля ботов. Пусть

$$B^f(\tau) = \sum_{r \in [M]} n_r \pi_r \rho_{f,r} + u^f(\tau) \pi_{M+1} \rho^f,$$

где $u^f(\tau) = u_1^f(\tau) + \dots + u_m^f(\tau)$ – доля когорты с индексом f .

Пусть

$$D_a^f(\tau) = \sum_{l \in [m]} \sum_{r \in [M]} y_{l,r}(\tau) \pi_r \rho_{f,r} \Delta_{a,l}^{f,r} + \sum_{l \in [m]} u_l^f(\tau) \pi_{M+1} \rho^f \Delta_{a,l}^{f,M+1}.$$

Пусть, кроме того,

$$C_{s,l,k}^f(\tau) = \sum_{r \in [M]} y_{l,r}(\tau) \pi_r \rho_{f,r} \Delta_{s,l}^{f,r} p_{s,l,k}^{f,r} + u_l^f(\tau) \pi_{M+1} \rho^f \Delta_{s,l}^{f,M+1} p_{s,l,k}^{f,M+1}.$$

Имеет место следующая система дифференциальных уравнений:

$$(1) \quad \frac{dy_{a,f}(\tau)}{d\tau} = \frac{1}{A \cdot B^f(\tau)} \left[\sum_{s \in [m]} \sum_{l \in [m]} y_{s,f}(\tau) \pi_f C_{s,l,a}^f(\tau) - y_{a,f}(\tau) \pi_f D_a^f(\tau) \right], \quad a \in [m], f \in [M].$$

Система (1) описывает динамику переменных $y_{a,f}(\tau)$ с учетом заданного начального условия

$$(2) \quad y_{a,f}(\tau_0) = q_{a,f}, \quad a \in [m], f \in [M].$$

Далее будем считать, что величины $q_{a,f}$ в (2) удовлетворяют

$$q_{a,f} \geq 0, \quad \sum_{a \in [m]} q_{a,f} = n_f.$$

Величины A и $B^f(\tau)$ играют нормировочную роль для распределений вероятностей, описывающих выбор объекта влияния и субъекта влияния соответственно. Для более глубокого понимания можно ознакомиться с выводом уравнения (1) в работе [9] (см. леммы 1 и 2). В свою очередь, величину $C_{s,l,k}^f(\tau)$ можно интерпретировать как агрегированную интенсивность принятия агентами с мнением Z_s и типом Ξ_f мнения Z_k в результате взаимодействия с мнением Z_l . При этом усреднение ведется по типу субъекта влияния: рассматриваются и нативные агенты (первое слагаемое в выражении для $C_{s,l,k}^f(\tau)$), и боты (второе слагаемое). Величина $D_a^f(\tau)$ трудно интерпретируема, ее появление обусловлено введенной гетерогенностью активности агентов и структуры сети.

4. Задача управления мнениями

Рассмотрим задачу управления мнениями, основанную на модели среднего поля (1). Будем считать, что управляющий субъект (далее – УС) определяет мнения ботов, а также составы когорт с тем, чтобы изменить распределение мнений в системе. При этом предполагаем, что УС никак не влияет на интенсивность контактов между ботами и нативными агентами – величины заданы ρ^f и постоянны во времени.

$$(3) \quad \underset{u(\cdot)}{\text{minimize}} \quad J(u(\cdot)),$$

$$\frac{dy_{a,f}(\tau)}{d\tau} = \frac{1}{A \cdot B^f(\tau)} \left[\sum_{s \in [m]} \sum_{l \in [m]} y_{s,f}(\tau) \pi_f C_{s,l,a}^f(\tau) - y_{a,f}(\tau) - \pi_f D_a^f(\tau) \right],$$

$$y_{a,f}(\tau_0) = q_{a,f}, \quad a \in [m], f \in [M],$$

$$u_l^f(\tau) \geq 0, \quad l \in [m], f \in [M],$$

$$\sum_{l \in [m], f \in [M]} u_l^f(\tau) = u.$$

В задаче управления (3) функционал J предполагается линейным и определяется следующим образом:

$$(4) \quad J = K \int_{\tau_0}^{\tau_T} \sum_{\alpha \in [m]} v_\alpha y_\alpha(\tau) d\tau + \sum_{\alpha \in [m]} v_\alpha y_\alpha(\tau_1),$$

где τ_T определяет горизонт планирования, $y_\alpha(\tau) = \sum_{f \in [M]} y_{\alpha,f}(\tau)$ – доля нативных агентов, имеющих мнение Z_α в момент времени τ , а $v = (v_1, \dots, v_m)$ – заданный вектор весов оптимизации для конкретных мнений, отражающий интересы УС. Чем больше значение v_α , тем менее желательно представительство мнения Z_α . При этом в работе [14] было показано, что без ограничения общности задачи можно ограничиться неотрицательными векторами весов. В (4) параметр $K > 0$ отвечает за относительную важность интегрального слагаемого функционала.

5. Простейший случай: $m = 2, M = 1$, система гомогенна.

Далее будем рассматривать задачу (3) в предположении, что $m = 2$ (два возможных значения мнения) и $M = 1$ (один тип нативных агентов, а значит и всего одна когорта ботов). Без уменьшения общности, пусть $\tau_0 = 0$ и $\tau_1 = T$. Также будем считать, что все параметры стохастической блочной модели равны друг другу, как и параметры активности агентов. В результате они сокращаются в числителе и знаменателе (1). Также будем считать, что алгоритм персонализации «отключен»: все $\Delta_{s,l}^{f,r} = 1$

Запишем одно из уравнений системы (1) для описываемого случая (второе рассматривать не имеет смысла, так как оно элементарно первому в силу $y_{1,1} + y_{2,1} = 1 - u$). Для краткости опустим аргумент времени τ :

$$(5) \quad \frac{dy_{1,1}}{d\tau} = \sum_{s=1}^2 \sum_{l=1}^2 y_{s,1} \left[y_{l,1} p_{s,l,1}^{1,1} + u_l^1 p_{s,l,1}^{1,2} \right] - y_{1,1} =$$

$$= \sum_{s=1}^2 y_{s,1} \left[y_{1,1} p_{s,1,1}^{1,1} + u_1^1 p_{s,1,1}^{1,2} + y_{2,1} p_{s,2,1}^{1,1} + u_2^1 p_{s,2,1}^{1,2} \right] - y_{1,1} =$$

$$= y_{1,1} \left[y_{1,1} p_{1,1,1}^{1,1} + u_1^1 p_{1,1,1}^{1,2} + y_{2,1} p_{1,2,1}^{1,1} + u_2^1 p_{1,2,1}^{1,2} \right] + \\ + y_{2,1} \left[y_{1,1} p_{2,1,1}^{1,1} + u_1^1 p_{2,1,1}^{1,2} + y_{2,1} p_{2,2,1}^{1,1} + u_2^1 p_{2,2,1}^{1,2} \right] - y_{1,1}.$$

Также для краткости будем игнорировать индексы типов и когорт (так как всего один тип нативных агентов и одна когорта ботов), используя обозначения $y_1 = y_{1,1}$, $y_2 = y_{2,1}$, $u_1 = u_1^1$, $u_2 = u_2^1$:

$$\frac{dy_1}{d\tau} = y_1 \left[y_1 p_{1,1,1}^{1,1} + u_1 p_{1,1,1}^{1,2} + y_2 p_{1,2,1}^{1,1} + u_2 p_{1,2,1}^{1,2} \right] + \\ + y_2 \left[y_1 p_{2,1,1}^{1,1} + u_1 p_{2,1,1}^{1,2} + y_2 p_{2,2,1}^{1,1} + u_2 p_{2,2,1}^{1,2} \right] - y_1.$$

При сделанных предположениях целевой функционал (4) принимает вид

$$(6) \quad J = K \int_0^T (v_1 y_1(\tau) + v_1 y_2(\tau)) d\tau + v_1 y_{1,1}(T) + v_1 y_{2,1}(T).$$

При этом содержательно рассматривать только два вектора весов: $v = (0, 1)$ или $v = (1, 0)$. Действительно, пусть, без ограничения общности, $v_2 > v_1$. Тогда, представляя $v_2 = v_1 + v'$, где $v' > 0$, можно вынести v_1 из интегрального и терминального слагаемых (6), после чего воспользоваться тем, что $y_1 + y_2 = 1 - u$. Таким образом из функционала выносятся константа и остаётся

$$J_{new} = K \int_0^T v' \cdot y_2(\tau) d\tau + v' y_2(T).$$

Также не имеет смысла рассматривать ситуацию $v = (0, v')$ для $v' \neq 1$, так как данная константа входит в функционал мультипликативно.

Далее будем рассматривать случай, когда $v = (0, 1)$, т.е. УС стремится минимизировать распространенность мнения Z_2 в системе. В таком случае

$$J = K \int_0^T y_2(\tau) d\tau + y_2(T).$$

Распишем J как сумму интегралов с шагом $\varepsilon > 0$, стремящимся к нулю:

$$J = K \left(\sum_{i=0}^{\frac{T}{\varepsilon}-1} \int_{i\varepsilon}^{(i+1)\varepsilon} y_2(\tau) d\tau \right) + y_2(T).$$

Далее приблизим каждый из этих интегралов суммой

$$\int_{i\varepsilon}^{(i+1)\varepsilon} y_2(\tau) d\tau = \frac{y_2(i\varepsilon) + y_2((i+1)\varepsilon)}{2}.$$

Получаем

$$J = \frac{K}{2} y_2(0) + K \sum_{i=1}^{\frac{T}{\varepsilon}-1} y_2(i\varepsilon) + \left(\frac{K}{2} + 1 \right) y_2(T).$$

С учётом $y_1 + y_2 = 1 - u$ минимизация J эквивалентна максимизации функционала

$$(7) \quad J^* = \frac{K}{2} y_1(0) + K \sum_{i=1}^{\frac{T}{\varepsilon}-1} y_1(i\varepsilon) + \left(\frac{K}{2} + 1 \right) y_1(T).$$

Для удобства параметризуем таблицы переходов (для рассматриваемого случая необходимо определить две таблицы переходов – внутренних взаимодействий (нативных) агентов $P^{1,1}$ и влияния ботов на нативных агентов $P^{1,2}$) следующим образом:

$$(8) \quad P_1^{1,i} = \begin{bmatrix} 1 - \alpha_{1,i} & \alpha_{1,i} \\ 1 - \beta_{1,i} & \beta_{1,i} \end{bmatrix}, P_2^{1,i} = \begin{bmatrix} \gamma_{1,i} & 1 - \gamma_{1,i} \\ \delta_{1,i} & 1 - \delta_{1,i} \end{bmatrix}.$$

Поскольку в рамках данной работы будет рассматриваться только случай $M = 1$, то в представлении (8) можно отказаться от фрагмента «1,» нижнего индекса у компонент таблиц переходов:

$$P_1^{1,i} = \begin{bmatrix} 1 - \alpha_i & \alpha_i \\ 1 - \beta_i & \beta_i \end{bmatrix}, P_2^{1,i} = \begin{bmatrix} \gamma_i & 1 - \gamma_i \\ \delta_i & 1 - \delta_i \end{bmatrix}.$$

Сформулируем основополагающий результат данной работы.

Теорема 1. *Справедливы следующие утверждения:*

1. Пусть $\gamma_2 = \delta_2$ и $\beta_2 = \alpha_2 > 0$. Тогда оптимальным будет любое допустимое управление.

2. Пусть $\gamma_2 < \delta_2$ и $\beta_2 - \alpha_2 - \gamma_2 + \delta_2 \leq 0$ или $\gamma_2 = \delta_2$ и $\beta_2 - \alpha_2 < 0$. Тогда оптимальным будет управление $u_1(\tau) \equiv 0$.

3. Пусть $\gamma_2 > \delta_2$ и $\beta_2 - \alpha_2 - \gamma_2 + \delta_2 \geq 0$ или $\gamma_2 = \delta_2$ и $\beta_2 - \alpha_2 > 0$. Тогда оптимальным будет управление $u_1(\tau) \equiv u$.

Доказательство. Запишем дифференциальное уравнение (5) при помощи следующей разностной схемы:

$$\begin{aligned} \frac{y_1((i+1)\varepsilon) - y_1(i\varepsilon)}{\varepsilon} = & y_1(i\varepsilon) \left[y_1(i\varepsilon)(1 - \alpha_1) + u_1(i\varepsilon)(1 - \alpha_2) + \right. \\ & \left. + (1 - u - y_1(i\varepsilon))(1 - \beta_1) + (u - u_1(i\varepsilon))(1 - \beta_2) \right] + \\ & + (1 - u - y_1(i\varepsilon)) \left[y_1(i\varepsilon)\gamma_1 + u_1(i\varepsilon)\gamma_2 + (1 - u - y_1(i\varepsilon))\delta_1 + \right. \\ & \left. + (u - u_1(i\varepsilon))\delta_2 \right] - y_1(i\varepsilon), \quad i = 0, 1, \dots, \frac{T}{\varepsilon} - 1. \end{aligned}$$

Отсюда получаем

$$(9) \quad y_1((i+1)\varepsilon) = \varepsilon \left[\Phi(i\varepsilon) + u_1(i\varepsilon)\Psi(i\varepsilon) \right],$$

где

$$(10) \quad \begin{aligned} \Phi(i\varepsilon) = & y_1^2(i\varepsilon)(1 - \alpha_1) + y_1(i\varepsilon)(1 - u - y_1(i\varepsilon))(1 - \beta_1) + \\ & + u(1 - \beta_2)y_1(i\varepsilon) + (1 - u - y_1(i\varepsilon))y_1(i\varepsilon)\gamma_{1,1} + (1 - u - y_1(i\varepsilon)) \times \\ & \times (1 - u - y_1(i\varepsilon))\delta_1 + (1 - u - y_1(i\varepsilon))u\delta_2 + \left(\frac{1}{\varepsilon} - 1\right)y_1(i\varepsilon) \end{aligned}$$

и

$$(11) \quad \Psi(i\varepsilon) = y_1(i\varepsilon) \left[\beta_2 - \alpha_2 - \gamma_2 + \delta_2 \right] + (1 - u)(\gamma_2 - \delta_2).$$

Заметим, что в силу фазовых ограничений все слагаемые (10) неотрицательны и поэтому имеет место $\Phi(i\varepsilon) \geq 0$. В этом случае максимизация $y_1((i+1)\varepsilon)$ при известном значении $y_1(i\varepsilon)$ относительно переменной управления $u_1(i\varepsilon)$, требуемая функционалом (7), обусловлена лишь выражением (11).

Пусть $i = \frac{T}{\varepsilon} - 1$ и значение $y_1(i\varepsilon)$ задано. Определим управление, оптимальное на промежутке $[T - \varepsilon, T]$. Для этого рассмотрим следующие альтернативы:

1. Пусть $\gamma_2 = \delta_2$. Рассмотрим два возможных случая:

а. Если $y_1(i\varepsilon) = 0$, то $u_1(i\varepsilon)$ – любое из множества допустимых управлений $[0, u]$.

б. Пусть $y_1(i\varepsilon) > 0$. Если $\beta_2 - \alpha_2 > 0$, то тогда $u_1(i\varepsilon)$ должен принимать максимальное значение из множества допустимых управлений, т.е. u . Если $\beta_2 - \alpha_2 < 0$, то $u_1(i\varepsilon)$ должен принимать минимальное значение из множества допустимых управлений, т.е. 0. Если $\beta_2 - \alpha_2 = 0$, то оптимальным будет любое управление.

2. Пусть $\gamma_2 < \delta_2$. Рассмотрим два возможных случая:

а. Если $\beta_2 - \alpha_2 - \gamma_2 + \delta_2 \leq 0$, то неравенство $\Psi(i\varepsilon) < 0$ будет выполнено для любого значения $y_1(i\varepsilon)$. Тогда оптимальным будет управление $u_1(i\varepsilon) = 0$.

б. Пусть $\beta_2 - \alpha_2 - \gamma_2 + \delta_2 > 0$. Неравенство $\Psi(i\varepsilon) > 0$ будет выполнено при

$$(12) \quad y_1(i\varepsilon) > \Theta,$$

где

$$(13) \quad \Theta = \frac{(1-u)(\delta_2 - \gamma_2)}{\beta_2 - \alpha_2 - \gamma_2 + \delta_2}.$$

В таком случае оптимальным будет управление $u_1(i\varepsilon) = u$. Если (12) не выполнено, то тогда оптимальным будет управление $u_1(i\varepsilon) = 0$.

3. Пусть $\gamma_2 > \delta_2$. Рассмотрим два возможных случая:

а. Если $\beta_2 - \alpha_2 - \gamma_2 + \delta_2 \geq 0$, то неравенство $\Psi(i\varepsilon) > 0$ выполнено для любого значения $y_1(i\varepsilon)$. В этом случае оптимальным будет управление $u_1(i\varepsilon) = u$.

б. Пусть $\beta_2 - \alpha_2 - \gamma_2 + \delta_2 < 0$. В таком случае неравенство $\Psi(i\varepsilon) > 0$ выполнено при условии

$$(14) \quad y(i\varepsilon) < \Theta.$$

Тогда оптимальным будет управление $u_1(i\varepsilon) = u$. Если (14) не выполнено, то тогда оптимальным будет управление $u_1(i\varepsilon) = 0$.

Используя данные наблюдения и принцип оптимальности Беллмана, строим оптимальное управление для всего промежутка $[0, T]$ и получаем утверждение теоремы. \square

Отметим, что теорема 1 не исчерпывает все возможные комбинации параметров модели: в ее доказательстве описаны случаи, которые не вошли в итоговую формулировку. Их мы рассмотрим отдельно. Начнем с ситуации, описанной в случае 1а. Тогда $y(T) = (1 - u)^2 \delta_1 + (1 - u)u \delta_2$ и имеет место следующее утверждение.

Утверждение 1. Пусть $\gamma_2 = \delta_2$ и $\delta_1 = 0, \delta_2 = 0$. Пусть для некоторого τ_1 имеет место $y_1(\tau_1) = 0$. Тогда для $\tau > \tau_1$ любое допустимое управление будет оптимальным.

Для случаев 2б и 3б оптимальная стратегия в каждый момент времени определена однозначно, но зависит от текущего состояния системы. Следующие две теоремы показывают, что для этих случаев число возможных переключений не превосходит одного.

Теорема 2. Пусть $\gamma_2 > \delta_2$ и $\beta_2 - \alpha_2 - \gamma_2 + \delta_2 < 0$. Тогда оптимальное управление в момент времени τ зависит от текущего состояния системы $y_1(\tau)$ и имеет следующую структуру:

$$(15) \quad u_1(\tau) = \begin{cases} u, & \text{если } y_1(\tau) < \Theta, \\ 0 & \text{иначе.} \end{cases}$$

При этом если УС применяет управление (15), то число возможных переключений с $u_1(\tau) = 0$ на $u_1(\tau) = u$ не превосходит одного.

Доказательство. Из-за характера уравнения (5) при соблюдении оптимального управления после попадания системы в зону фазового пространства

$$y_1 < \Theta$$

при некотором τ система и далее будет оставаться в этой зоне. Действительно, в силу (9) при $\varepsilon \rightarrow 0, u_1(i\varepsilon) = u$ и $y_1(i\varepsilon) < \Theta$ будет выполнено $y_1((i+1)\varepsilon) \geq y_1(i\varepsilon)$. Это обусловлено последним слагаемым в (10), которое после умножения на ε и предельного перехода становится равным $y_1(i\varepsilon)$. В таком случае для достаточно малого ε будет иметь место $y_1((i+1)\varepsilon) = y_1(i\varepsilon) + \Omega$, где Ω – некоторая небольшая неотрицательная величина. Следовательно, момент переключения оптимального управления, если он будет, произойдет один раз. \square

Теорема 3. Пусть $\gamma_2 < \delta_2$ и $\beta_2 - \alpha_2 - \gamma_2 + \delta_2 > 0$. Тогда оптимальное управление в момент времени τ зависит от текущего состояния системы $y_1(\tau)$ и имеет следующую структуру:

$$(16) \quad u_1(\tau) = \begin{cases} u, & \text{если } y_1(\tau) > \Theta, \\ 0 & \text{иначе.} \end{cases}$$

При этом если УС применяет управление (16), то число возможных переключений с $u_1(\tau) = u$ на $u_1(\tau) = 0$ не превосходит одного.

Доказательство. Доказывается аналогично теореме 2. \square

Возникает вопрос: может ли быть найдено время переключения управления (при условии, что оно принадлежит целевому отрезку $[0, T]$) для ситуаций, рассмотренных в теоремах 2 и 3. Для определенности предположим, что выполнено условие теоремы 2 и $y_1(0) < \Theta$. В таком случае необходимо решить уравнение (5) при $u_1(\tau) \equiv 0$ подобно тому, как это было сделано в работе [8], и после этого определить τ_0 , при котором $y_1(\tau_0) = \Theta$. Для нахождения оптимальной траектории уравнение (5) необходимо решить для нового начального условия $y_1(\tau_0) = \Theta$ при $u_1(\tau) \equiv u$. Может так случиться, что $\tau_0 > T$. В этом случае оптимальное на отрезке $[0, T]$ управление будет иметь вид $u_1(\tau) \equiv 0$.

В конце данного раздела отметим, что случай $v = (1, 0)$ разбирается аналогичным образом. Все результаты будут такими же с точностью до замены u_1 на u_2 и y_1 на y_2 .

6. Основной случай: $m = 2, M = 1$, гетерогенная система.

Теперь рассмотрим случай $m = 2, M = 1$ в рамках наиболее общей постановки: с гетерогенностью уровней активностей агентов, модулярностью структуры сети и алгоритмом персонализации. Так же, как и в предыдущем разделе, будем использовать обозначения $y_1 = y_{1,1}, y_2 = y_{2,1}, u_1 = u_1^1, u_2 = u_2^1$ и параметризацию таблицы переходов (8).

Для рассматриваемого случая в уравнении (1) возможны только две комбинации индексов типов: $(1, 1)$ и $(1, 2)$. Для удоб-

ства первую комбинацию будем заменять пустым верхним индексом, а вторую – индексом u . Аналогично нижним индексом u будем отмечать переменные активности и сетевой структуры, описывающие агентов управления:

$$\pi = \pi_1, \pi_u = \pi_2, \quad \rho = \rho_{1,1}, \rho^u = \rho_{1,2}$$

Кроме того, при $M = 1$ имеет место $u^1(\tau) = u$, поэтому для краткости будем писать $B = B^1(\tau)$ (напомним, что величина $B^f(\tau)$ была определена в начале раздела 3). Помимо этого введем следующие обозначения:

$$\Pi_{s,l} = \pi \rho \Delta_{s,l}, \quad \Pi_{s,l}^u = \pi_u \rho^u \Delta_{s,l}^u.$$

С учетом вышесказанного получим

$$(17) \quad \frac{dy_1}{d\tau} = \frac{\pi}{AB} \left[\sum_{s=1}^2 \sum_{l=1}^2 y_{s,1} C_{s,l,1} - y_1 D_1(\tau) \right] =$$

$$= \frac{\pi}{AB} \left[y_1 \left(y_1 \Pi_{1,1} p_{1,1,1} + u_1 \cdot \Pi_{1,1}^u p_{1,1,1}^u + y_2 \Pi_{1,2} p_{1,2,1} + u_2 \Pi_{1,2}^u p_{1,2,1}^u \right) + \right.$$

$$\left. + y_2 \left(y_1 \Pi_{2,1} p_{2,1,1} + u_1 \cdot \Pi_{2,1}^u p_{2,1,1}^u + y_2 \Pi_{2,2} p_{2,2,1} + u_2 \Pi_{2,2}^u p_{2,2,1}^u \right) - \right.$$

$$\left. - y_1 D_1(\tau) \right].$$

Теорема 1 из предыдущего раздела допускает естественное обобщение на рассматриваемый в данном разделе гетерогенный случай.

Теорема 4. *Справедливы следующие утверждения:*

1. Пусть $\Delta_{2,1}^u \gamma_2 = \Delta_{2,2}^u \delta_2$ и $\Delta_{1,2}^u \beta_2 = \Delta_{1,1}^u \alpha_2 > 0$. Тогда оптимальным будет любое допустимое управление.

2. Пусть $\Delta_{2,1}^u \gamma_2 < \Delta_{2,2}^u \delta_2$ и $\Delta_{1,2}^u \beta_2 - \Delta_{1,1}^u \alpha_2 - \Delta_{2,1}^u \gamma_2 + \Delta_{2,2}^u \delta_2 \leq 0$ или $\Delta_{2,1}^u \gamma_2 = \Delta_{2,2}^u \delta_2$ и $\Delta_{1,2}^u \beta_2 - \Delta_{1,1}^u \alpha_2 < 0$. Тогда оптимальным будет управление $u_1(\tau) \equiv 0$.

3. Пусть $\Delta_{2,1}^u \gamma_2 > \Delta_{2,2}^u \delta_2$ и $\Delta_{1,2}^u \beta_2 - \Delta_{1,1}^u \alpha_2 - \Delta_{2,1}^u \gamma_2 + \Delta_{2,2}^u \delta_2 \geq 0$ или $\Delta_{2,1}^u \gamma_2 = \Delta_{2,2}^u \delta_2$ и $\Delta_{1,2}^u \beta_2 - \Delta_{1,1}^u \alpha_2 > 0$. Тогда оптимальным будет управление $u_1(\tau) \equiv u$.

Доказательство. Рассматривая разностную схему, аналогичную описанной в предыдущем разделе, применительно

к уравнению (17) получим

$$y_1((i+1)\varepsilon) = \frac{\pi\varepsilon}{AB} \left[\Phi(i\varepsilon) + u_1(i\varepsilon)\Psi(i\varepsilon) \right],$$

где

$$(18) \quad \begin{aligned} \Phi(i\varepsilon) = & y_1^2(i\varepsilon)\Pi_{1,1}(1-\alpha_1) + \\ & + y_1(i\varepsilon)(1-u-y_1(i\varepsilon))\Pi_{1,2}(1-\beta_1) + u(1-\beta_2)\Pi_{1,2}^u y_1(i\varepsilon) + \\ & + (1-u-y_1(i\varepsilon))y_1(i\varepsilon)\Pi_{2,1}p_{2,1,1}\gamma_1 + \\ & + (1-u-y_1(i\varepsilon))(1-u-y_1(i\varepsilon))\Pi_{2,2}\delta_1 + (1-u-y_1(i\varepsilon))u\Pi_{2,2}^u\delta_2 + \\ & + \left(\frac{1}{\pi\varepsilon/(AB)} - Q(i\varepsilon) \right) y_1(i\varepsilon) \end{aligned}$$

и

$$(19) \quad \begin{aligned} \Psi(i\varepsilon) = & \pi_u\rho_u \left[y_1(i\varepsilon) \left(\Delta_{1,2}^u\beta_2 - \Delta_{1,1}^u\alpha_2 - \Delta_{2,1}^u\gamma_2 + \Delta_{2,2}^u\delta_2 \right) + \right. \\ & \left. + (1-u)(\Delta_{2,1}^u\gamma_2 - \Delta_{2,2}^u\delta_2) \right]. \end{aligned}$$

В (18) в качестве Q взято

$$Q = -y_1\pi\rho\Delta_{1,1} - (1-u-y_1)\pi\rho\Delta_{1,2} - u\Pi_{1,2}^u.$$

Заметим, что в силу фазовых ограничений все слагаемые (18) неотрицательны и поэтому имеет место $\Phi(i\varepsilon) \geq 0$. Если для первых шести слагаемых это очевидно, то для последнего мы пользуемся тем, что Q ограничено сверху максимумом из $\pi\rho\Delta_{1,1}$, $\pi\rho\Delta_{1,2}$ и $\pi_u\rho_u\Delta_{1,2}^u$, каждое из которых является конечным числом. Таким образом, максимизация $y_1((i+1)\varepsilon)$ при известном значении $y_1(i\varepsilon)$ относительно переменной управления $u_1(i\varepsilon)$, требуемая функционалом (7), обусловлена лишь выражением (19).

Пусть $i = \frac{T}{\varepsilon} - 1$ и значение $y_1(i\varepsilon)$ задано. Определим управление, оптимальное на промежутке $[T-\varepsilon, T]$. Для этого рассмотрим следующие альтернативы:

1. Пусть $\Delta_{2,1}^u\gamma_2 = \Delta_{2,2}^u\delta_2$. Рассмотрим два возможных случая:

а. Если $y_1(i\varepsilon) = 0$, то $u_1(i\varepsilon)$ – любое из множества допустимых управлений $[0, u]$.

б. Пусть $y_1(i\varepsilon) > 0$. Если $\Delta_{1,2}^u\beta_2 - \Delta_{1,1}^u\alpha_2 > 0$, то $u_1(i\varepsilon)$ должен принимать максимальное значение из множества допустимых управлений, т.е. u . Если $\Delta_{1,2}^u\beta_2 - \Delta_{1,1}^u\alpha_2 < 0$, то $u_1(i\varepsilon)$

должен принимать минимальное значение из множества допустимых управлений, т.е. 0. Если $\Delta_{1,2}^u \beta_2 - \Delta_{1,1}^u \alpha_2 = 0$, то оптимальным будет любое управление.

2. Пусть $\Delta_{2,1}^u \gamma_2 < \Delta_{2,2}^u \delta_2$. Рассмотрим два возможных случая:

а. Если $\Delta_{1,2}^u \beta_2 - \Delta_{1,1}^u \alpha_2 - \Delta_{2,1}^u \gamma_2 + \Delta_{2,2}^u \delta_2 \leq 0$, то неравенство $\Psi(i\varepsilon) < 0$ будет выполнено для любого значения $y_1(i\varepsilon)$. Поэтому оптимальным будет управление $u_1(i\varepsilon) = 0$.

б. Пусть $\Delta_{1,2}^u \beta_2 - \Delta_{1,1}^u \alpha_2 - \Delta_{2,1}^u \gamma_2 + \Delta_{2,2}^u \delta_2 > 0$. Неравенство $\Psi(i\varepsilon) > 0$ будет выполнено при $y_1(i\varepsilon) > \Upsilon$, где

$$\Upsilon = \frac{(1-u)(\Delta_{2,2}^u \delta_2 - \Delta_{2,1}^u \gamma_2)}{\Delta_{1,2}^u \beta_2 - \Delta_{1,1}^u \alpha_2 - \Delta_{2,1}^u \gamma_2 + \Delta_{2,2}^u \delta_2}.$$

В таком случае оптимальным будет управление $u_1(i\varepsilon) = u$. Если $y_1(i\varepsilon) \leq \Upsilon$, то тогда оптимальным будет управление $u_1(i\varepsilon) = 0$.

3. Пусть $\Delta_{2,1}^u \gamma_2 > \Delta_{2,2}^u \delta_2$. Рассмотрим два возможных случая:

а. Если $\Delta_{1,2}^u \beta_2 - \Delta_{1,1}^u \alpha_2 - \Delta_{2,1}^u \gamma_2 + \Delta_{2,2}^u \delta_2 \geq 0$, то неравенство $\Psi(i\varepsilon) > 0$ выполнено для любого значения $y_1(i\varepsilon)$. Поэтому оптимальным будет управление $u_1(i\varepsilon) = u$.

б. Пусть $\Delta_{1,2}^u \beta_2 - \Delta_{1,1}^u \alpha_2 - \Delta_{2,1}^u \gamma_2 + \Delta_{2,2}^u \delta_2 < 0$. В таком случае неравенство $\Psi(i\varepsilon) > 0$ выполнено при $y(i\varepsilon) < \Upsilon$. Тогда оптимальным будет управление $u_1(i\varepsilon) = u$. Если $y(i\varepsilon) \geq \Upsilon$, то тогда оптимальным будет управление $u_1(i\varepsilon) = 0$.

Применение принципа оптимальности Беллмана заканчивает доказательство данной теоремы. \square

Теорема 4 не исчерпывает все возможные комбинации параметров модели: в ее доказательстве описаны случаи, которые не вошли в итоговую формулировку. Их мы рассмотрим отдельно. Начнем с ситуации, описанной в случае 1а. Тогда имеет место $y(T) = (1-u)^2 \delta_2 \delta_1 + (1-u)u_{2,2} \delta_2$ и справедливо следующее утверждение.

Утверждение 2. Пусть $\Delta_{2,1}^u \gamma_2 = \Delta_{2,2}^u \delta_2$ и $\delta_1 = 0, \delta_2 = 0$. Пусть для некоторого τ_1 имеет место $y_1(\tau_1) = 0$. Тогда для $\tau > \tau_1$ любое допустимое управление будет оптимальным.

Для случаев 2б и 3б оптимальная стратегия в каждый момент времени определена однозначно, но зависит от текущего состояния системы. Следующие две теоремы показывают, что для этих случаев число возможных переключений не превосходит одного.

Теорема 5. Пусть $\Delta_{2,1}^u \gamma_2 > \Delta_{2,2}^u \delta_2$ и $\Delta_{1,2}^u \beta_2 - \Delta_{1,1}^u \alpha_2 - \Delta_{2,1}^u \gamma_2 + \Delta_{2,2}^u \delta_2 < 0$. Тогда оптимальное управление в момент времени τ зависит от текущего состояния системы $y_1(\tau)$ и устроено следующим образом:

$$(20) \quad u_1(\tau) = \begin{cases} u, & \text{если } y_1(\tau) < \Upsilon, \\ 0 & \text{иначе.} \end{cases}$$

При этом если УС применяет управление (20), то число возможных переключений с $u_1(\tau) = 0$ на $u_1(\tau) = u$ не превосходит одного.

Доказательство. Из-за характера уравнения (5) при соблюдении оптимального управления после попадания системы в зону фазового пространства

$$y_1 < \Upsilon$$

при некотором τ система и далее будет оставаться в этой зоне. Действительно, в силу (9) при $\varepsilon \rightarrow 0$, $u_1(i\varepsilon) = u$ и $y_1(i\varepsilon) < \Upsilon$ будет выполнено $y_1((i+1)\varepsilon) \geq y_1(i\varepsilon)$. Это обусловлено последним слагаемым в (18), которое после умножения на ε и предельного перехода становится равным $y_1(i\varepsilon)$. В таком случае для достаточно малого ε будет иметь место $y_1((i+1)\varepsilon) = y_1(i\varepsilon) + \hat{\Omega}$, где $\hat{\Omega}$ – некоторая небольшая неотрицательная величина. Следовательно, момент переключения оптимального управления, если он будет, произойдёт один раз. \square

Теорема 6. Пусть $\Delta_{2,1}^u \gamma_2 < \Delta_{2,2}^u \delta_2$ и $\Delta_{1,2}^u \beta_2 - \Delta_{1,1}^u \alpha_2 - \Delta_{2,1}^u \gamma_2 + \Delta_{2,2}^u \delta_2 > 0$. Тогда оптимальное управление в момент времени τ зависит от текущего состояния системы $y_1(\tau)$ и устроено следующим образом:

$$(21) \quad u_1(\tau) = \begin{cases} u, & \text{если } y_1(\tau) > \Upsilon, \\ 0 & \text{иначе.} \end{cases}$$

При этом если УС применяет управление (21), то тогда число возможных переключений с $u_1(\tau) = u$ на $u_1(\tau) = 0$ не превосходит одного.

Доказательство. Доказывается аналогично теореме 5. \square

При этом метод нахождения времени переключения управления в теоремах 5 и 6 аналогичен тому, который был описан в разделе 5.

В конце данного раздела отметим, что случай $v = (1, 0)$ разбирается аналогичным образом. Все результаты будут такими же с точностью до замены u_1 на u_2 и y_1 на y_2 .

7. Примеры

Проиллюстрируем полученные результаты на конкретных примерах. Из доказанного выше следует, что оптимальные стратегии управления определяются лишь тем, как нативные агенты воспринимают влияние ботов алгоритмом ранжирования, а точнее той его составляющей, которая отвечает за взаимодействие ботов с нативными агентами и, возможно, текущим состоянием фазовых переменных. Во всех трех нижеследующих примерах будем рассматривать матрицы $P_1^{1,2}$ и $P_2^{1,2}$, определенные следующим образом:

$$(22) \quad P_1^{1,i} = \begin{bmatrix} 0,975 & 0,025 \\ 0,952 & 0,048 \end{bmatrix}, P_2^{1,i} = \begin{bmatrix} 0,066 & 0,934 \\ 0,049 & 0,951 \end{bmatrix},$$

т.е.

$$\alpha_2 = 0,025, \beta_2 = 0,048, \gamma_2 = 0,066, \delta_2 = 0,049.$$

Эти параметры были взяты из таблицы переходов, откалиброванной на эмпирических данных из онлайн-социальной сети ВКонтакте (более подробно см. [12]).

Заметим, что для компонент (22) имеют место неравенства $\beta_2 > \alpha_2$ и $\gamma_2 > \delta_2$, означающие, что сила эффекта конформизма (изменение мнения на противоположное в результате влияния со стороны противоположного мнения – вероятности β_2 и γ_2) выше, чем сила эффекта антиконформизма (изменение мнения на противоположное в результате влияния со стороны когерентного мнения – вероятности α_2 и δ_2) для каждого из мнений Z_1 и Z_2 .

Отсюда можно предположить, что для вектора весов $v = (0, 1)$ (цель УС – максимизировать представленность мнения Z_1 в системе) оптимальной будет стратегия $u_1(\tau) \equiv u$ (т.е. боты должны транслировать мнение Z_1), а для вектора весов $v = (1, 0)$ – $u_1(\tau) \equiv 0$ (боты должны пропагандировать мнение Z_2), по крайней мере в случае, когда алгоритмы ранжирования отключены.

В целом отметим, что ситуация, когда для продвижения в обществе некоторой точки зрения (скажем, Z_1) необходимо распространять аргументы, ей противоречащие (боты должны транслировать мнение Z_2), кажется достаточно абсурдной. Безусловно, соответствующая искусственная настройка компонент таблицы переходов может привести к такому исходу (см. теорему 1, второй случай), однако эмпирические исследования показывают, что конформизм (ассимилятивное влияние) играет более важную роль в формировании мнений людей в социальных группах, чем антиконформизм/диссимилятивное влияние (см. [6, 19]). Отметим также, что подобного рода эффекты возможны, когда азбука мнений ординальна и состоит из нескольких элементов. В таком случае могут существовать более сложные траектории влияния, к примеру, в случае ограниченного доверия (см. [14]).

Теперь проанализируем стратегии управления, которые следуют из доказанных выше теорем. Сначала рассмотрим простейший случай, описанный в разделе 5, т.е. будем считать, что алгоритмы ранжирования отключены.

7.1. Пример 1

Пусть $v = (0, 1)$, т.е. цель УС – максимизировать представленность мнения Z_1 в системе. Для рассматриваемой социальной системы имеют место неравенства $\gamma_2 > \delta_2$ и $\beta_2 - \alpha_2 - \gamma_2 + \delta_2 \geq 0$. Это означает, что выполнены условия случая 3 теоремы 1. Следовательно, оптимальным будет управление $u_1(\tau) \equiv u$, т.е. все боты должны транслировать первое мнение.

7.2. Пример 2

Вместо того чтобы рассматривать противоположный случай $v = (1, 0)$, когда УС старается максимизировать число адептов

мнения Z_2 , инвертируем таблицу переходов (22):

$$P_1^{1,i} = \begin{bmatrix} 0,951 & 0,049 \\ 0,934 & 0,066 \end{bmatrix}, P_2^{1,i} = \begin{bmatrix} 0,048 & 0,952 \\ 0,025 & 0,975 \end{bmatrix}$$

и будем рассматривать вектор весов $v = (0, 1)$. Имеют место неравенства $\gamma_2 > \delta_2$ и $\beta_2 - \alpha_2 - \gamma_2 + \delta_2 < 0$. Это соответствует условиям, описанным в теореме 2. В таком случае оптимальное управление имеет вид

$$u_1(\tau) = \begin{cases} u, & \text{если } y_1(\tau) < \Theta, \\ 0 & \text{иначе.} \end{cases}$$

Для рассматриваемой социальной системы $\Theta = (1 - u)\frac{23}{6}$. Попытаемся понять, возможно ли выполнение неравенства $y_1 \geq \Theta$ в принципе. Пусть оно выполнено. Тогда

$$y_1 \geq (1 - u)\frac{23}{6} = n\frac{23}{6} \geq y_1\frac{23}{6},$$

что возможно, только если $y_1 = 0$ и $\Theta = 0$, откуда следует $u = 1$, что не имеет смысла.

Таким образом, оптимальным управлением будет $u_1(\tau) \equiv u$. Это означает, что для таблицы переходов (22) с вектором весов $v = (1, 0)$ оптимальным будет управление $u_2(\tau) \equiv u$ – для продвижения точки зрения Z_2 в обществе все социальные боты должны транслировать аргументы в ее поддержку.

7.3. Пример 3

Результаты предыдущих двух примеров подтвердили наши гипотезы. Теперь предположим, что алгоритм ранжирования активирован и вектор весов задан как $v = (0, 1)$. Рассмотрим следующий класс алгоритмов ранжирования, параметризуемый двумя коэффициентами:

$$\begin{aligned} \Delta_{1,1}^u &= \Delta_{1,1}^u = x, \\ \Delta_{1,2}^u &= \Delta_{2,1}^u = y. \end{aligned}$$

При $x > y$ алгоритм ранжирования поощряет контакты между агентами с одинаковыми мнениями (в таких случаях говорят, что имеет место гомофилия¹), а при $x < y$ больший приоритет

¹ Считается, что алгоритмы ранжирования в социальных медиа платформах действуют именно таким образом [17].

отдается контактам между представителями разных фракций – имеет место гетерофилия. Рассмотрим два упрощенных случая.

1. Пусть $x \gg y$. Тогда выполнено условие теоремы 6. Следовательно, оптимальной является стратегия

$$u_1(\tau) = \begin{cases} u, & \text{если } y_1(\tau) > \Upsilon, \\ 0 & \text{иначе,} \end{cases}$$

при этом

$$\Upsilon = \frac{(1-u)(49x-66y)}{24x-18y} \approx (1-u)\frac{49}{24}.$$

Для $u = 0,5^1$ получаем $\Upsilon = 49/48$, что означает, что управление $u_1(\tau) \equiv 0$ всегда будет оптимальным. Для $u = 0,75$ имеем $\Upsilon \approx 0,51$, т.е. возможно переключение управления.

2. Пусть $x \ll y$. Тогда выполнено условие теоремы 5. Значит, оптимальной является стратегия

$$u_1(\tau) = \begin{cases} u, & \text{если } y_1(\tau) < \Upsilon, \\ 0 & \text{иначе,} \end{cases}$$

при этом

$$\Upsilon \approx (1-u)\frac{11}{3}.$$

Для $u = 0,5$ получаем $\Upsilon = 11/6$, что означает, что управление $u_1(\tau) \equiv u$ всегда будет оптимальным. Для $u = 0,75$ имеем $\Upsilon = 11/12 \approx 0,92$, т.е. возможно переключение управления.

Сравнивая два рассмотренных выше случая с примером 1, делаем вывод, что работа алгоритма ранжирования может существенно влиять на выбор оптимальной стратегии управления, особенно в условиях, когда ранжирование способствует гомофилии.

8. Заключение

В данной работе была рассмотрена задача управления мнениями агентов, сформулированная в терминах приближения среднего поля для SCARDO-модели, построенного для случая

¹ В работе [20] 2017 года была дана оценка 9%–15% для числа ботов. Более современные оценки из открытых источников составляют порядка 50%.

социальной системы с модулярной сетевой структурой, гетерогенными уровнями активности агентов и персонализацией контента. В рамках модели управления специальная группа агентов (интерпретация – боты в реальных социальных медиа) воздействует на обычных (нативных) агентов, с тем чтобы изменить их мнения в соответствии с интересами управляющего субъекта при конечном горизонте планирования.

Был изучен наиболее простой для анализа случай двухэлементной азбуки мнений, в рамках которой существует только два возможных мнения (интерпретация – выборы между двумя кандидатами/партиями) и одного типа нативных агентов. Для этого случая было построено аналитическое решение задачи управления. Оказалось, что в зависимости от значений параметров модели может быть либо одна оптимальная стратегия управления, либо бесконечно много. При этом в случае, когда оптимальное управление единственно, оно должно быть устроено следующим образом: все социальные боты должны транслировать одно и то же мнение, а переключение распространяемого мнения может произойти не более одного раза лишь для некоторых конфигураций параметров.

С содержательной точки зрения рассматриваемая в данной статье постановка малоперспективна, поскольку для таблиц переходов, полученных эмпирическим путем для двухэлементной азбуки мнений, в типичном случае все сводится к первым двум ситуациям теорем 1 или 4 [12]. Безусловно, учет активности агентов, модулярности сети и персонализации контента делает рассматриваемую постановку более содержательной, однако наиболее перспективным видится рассмотрение азбук мнений, содержащих как минимум три или более элементов. Это позволит формализовать механизмы социального влияния, такие как ограниченное доверие или диссимилиятивное влияние, которые наблюдаются в реальной социальной динамике [13] и получили широкое внимание в литературе [6]. В этом смысле решение задачи управления для двухэлементной азбуки мнений является промежуточным шагом перед изучением азбук мнений размерности $m > 2$.

Литература

1. ГУБАНОВ Д.А., НОВИКОВ Д.А., ЧХАРТИШВИЛИ А.Г. *Социальные сети: модели информационного влияния, управления и противоборства* / Под. ред. чл.-корр. РАН Д.А. Новикова. – М.: Физматлит, 2010. – 228 с.
2. ГУБАНОВ Д.А. *Модели и методы информационного влияния и управления в активных сетевых структурах* : дис. ... докт. техн. наук. – М., 2021.
3. BALIETTI S. et al. *Reducing opinion polarization: Effects of exposure to similar people with differing political views* // Proc. of the National Academy of Sciences. – 2021. – Vol. 118, No. 52. – P. e2112552118. – DOI: <https://doi.org/10.1073/pnas.2112552118>.
4. BINI M. et al. *Graph structure-based heuristics for optimal targeting in social networks* // IEEE Trans. on Control of Network Systems. – 2022. – Vol. 9, No. 3. – P. 1189–1201. – DOI: <https://doi.org/10.1109/TCNS.2022.3163665>.
5. EPSTEIN R. et al. *The “opinion matching effect”(OME): A subtle but powerful new form of influence that is apparently being used on the internet* // PloS One. – 2024. – Vol. 19, No. 9. – P. e0309897. – DOI: <https://doi.org/10.1371/journal.pone.0309897>.
6. FLACHE A. et al. *Models of social influence: Towards the next frontiers* // The Journal of Artificial Societies and Social Simulation. – 2017. – Vol. 20, No. 4. – P. 2. – DOI: <https://doi.org/10.18564/jasss.3521>.
7. FRIEDKIN N.E. *The problem of social control and coordination of complex systems in sociology: A look at the community cleavage problem* // IEEE Control Systems Magazine. – 2015. – Vol. 35, No. 3. – P. 40–51. – DOI: <https://doi.org/10.1109/MCS.2015.2406655>.
8. GEZHA V.N., KOZITSIN I.V. *Binary Opinion Space in the SCARDO Model: Precise Analytical Description* // Advances in Systems Science and Applications. – 2022. – Vol. 22,

- No. 4. – P. 103–115. – DOI: <https://doi.org/10.25728/assa.2022.22.4.1292>.
9. GEZHA V.N., KOZITSIN I.V. *Macroscopic Description of Structured Heterogeneous Online Social Systems with Dynamical Opinions* // 6th Int. Conf. on Control Systems, Mathematical Modeling, Automation and Energy Efficiency (SUMMA–2024). – IEEE, 2024. – P. 133–138. – DOI: <https://doi.org/10.1109/SUMMA64428.2024.10803880>.
 10. HUNTER R.F. et al. *Social network interventions for health behaviours and outcomes: A systematic review and meta-analysis* // PLoS Medicine. – 2019. – Vol. 16, No. 9. – P. e1002890. – DOI: <https://doi.org/10.1371/journal.pmed.1002890>.
 11. KEIJZER M.A., MAS M., FLACHE A. *Polarization on social media: Micro-level evidence and macro-level implications* // JASSS. – 2024. – Vol. 27, No. 1. – P. 7. – DOI: <https://doi.org/10.18564/jasss.5298>.
 12. KOZITSIN I.V. *A general framework to link theory and empirics in opinion formation models* // Scientific Reports. – 2022. – Vol. 12, No. 1. – P. 5543. – DOI: <https://doi.org/10.1038/s41598-022-09468-3>.
 13. 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. – DOI: <https://doi.org/10.1080/0022250X.2021.1956917>.
 14. KOZITSIN I.V. *Optimal control in opinion dynamics models: diversity of influence mechanisms and complex influence hierarchies* // Chaos, Solitons & Fractals. – 2024. – Vol. 181. – P. 114728. – DOI: <https://doi.org/10.1016/j.chaos.2024.114728>.
 15. LEE C., WILKINSON D.J. *A review of stochastic block models and extensions for graph clustering* // Applied Network Science. – 2019. – Vol. 4, No. 1. – P. 1–50. – DOI: <https://doi.org/10.1007/s41109-019-0232-2>.
 16. LI G.J., PORTER M.A. *Bounded-confidence model of opinion dynamics with heterogeneous node-activity levels* // Physical

- Review Research. – 2023. – Vol. 5, No. 2. – P. 023179. – DOI: <https://doi.org/10.1103/PhysRevResearch.5.023179>.
17. PERALTA A.F., KERTESZ J., INIGUEZ G. *Opinion formation on social networks with algorithmic bias: dynamics and bias imbalance* // Journal of Physics: Complexity. – 2021. – Vol. 2, No. 4. – P. 045009. – DOI: <https://doi.org/10.1088/2632-072X/ac340f>.
 18. 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. – DOI: <https://doi.org/10.1016/j.arcontrol.2018.03.005>.
 19. TANG T., SNIJDERS T., FLACHE A. *An Empirical and Simulation Investigation of Bounded Confidence and Negative Influence in Opinion Dynamics Using Stochastic Actor-Oriented Model* // Journal of Artificial Societies and Social Simulation. – 2025. – Vol. 28, No. 1. – P. 1–2. – DOI: <https://doi.org/10.18564/jasss.5566>.
 20. VAROL O. et al. *Online human-bot interactions: Detection, estimation, and characterization* // Proc. of the Int. AAAI Conf. on Web and Social Media. – 2017. – Vol. 11, No. 1. – P. 280–289. – DOI: <https://doi.org/10.1609/icwsm.v11i1.14871>.
 21. ZIMMARO F., CONTUCCI P., KERTESZ J. *Voter-like dynamics with conflicting preferences on modular networks* // Entropy. – 2023. – Vol. 25, No. 6. – P. 838. – DOI: <https://doi.org/10.3390/e25060838>.

CONTROLLING OPINIONS IN THE SCARDO-MODEL OF OPINION DYNAMICS WITH TWO-ELEMENT OPINION ALPHABET AND SINGLE TYPE OF NATIVE AGENTS: ANALYTICAL SOLUTION

Vladislav Gezha, Moscow Institute of Physics and Technology, Moscow, PhD student, (gezha.vn@phystech.edu).

Ivan Kozitsin, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, Moscow Institute of Physics and Technology, Moscow, PhD in mathematics (kozitsin.ivan@mail.ru).

Abstract: This paper investigates a control problem whereby a concerned person attempts to change individuals' opinions by using a group of social bots – stubborn agents that can influence other, native, agents via direct conversations. The control problem is formulated in terms of the mean field approximation for the SCARDO-model, built for the case of a social system with a modular network structure, heterogeneous levels of agent activity and content personalization. We focus on, perhaps, the simplest case of the two-element opinion alphabet, in which there are only two possible opinions (interpretation – elections in a two-party system) and one type of non-bot agents. We derive a precise analytical solution to the control problem and demonstrate that depending on the model parameters, there can be either one optimal control strategy or infinitely many. At the same time, in the case when there is only one optimal control trajectory, it should be organized as follows: all social bots should broadcast the same opinion, and the switch to a different opinion can occur no more than once, only for some parameter configurations. The theoretical results obtained are complemented by insightful examples with data from the online network VKontakte.

Keywords: opinion dynamics models, SCARDO-model, mean-field approximation, optimal control, modular networks, ranking algorithms.

УДК 519.7

ББК 22.18

*Статья представлена к публикации
членом редакционной коллегии В.В. Мазаловым.*

Поступила в редакцию 11.11.2024.

Дата опубликования 31.05.2025.

ПРИБЛИЖЕННЫЙ АНАЛИЗ ДЛИТЕЛЬНОСТИ ПРЕБЫВАНИЯ ТРЕБОВАНИЙ В СЕТИ МАССОВОГО ОБСЛУЖИВАНИЯ С ДЕЛЕНИЕМ И СЛИЯНИЕМ ТРЕБОВАНИЙ

Постнова О. С.¹, Тананко И. Е.², Рогачко Е. С.³

(Саратовский национальный исследовательский
государственный университет имени Н.Г. Чернышевского,
Саратов)

Рассматривается открытая сеть массового обслуживания с делением и слиянием требований. При поступлении в сеть каждое требование разделяется на два фрагмента. Фрагменты требований обслуживаются в одноприборных системах обслуживания с неограниченной очередью, образующих ациклическую структуру рассматриваемой сети. После завершения обслуживания в сети фрагменты требований поступают в выделенную систему сборки требований, где требования собираются из своих фрагментов и покидают сеть. Предлагается приближенный метод вычисления математического ожидания длительности сборки требований и математического ожидания длительности пребывания требований в сети массового обслуживания с делением и слиянием требований при предположениях, что входящий в сеть поток требований является пуассоновским, а длительности обслуживания фрагментов требований являются экспоненциально распределенными случайными величинами. Для оценки точности приближенного метода вычисления соответствующих характеристик сети полученные результаты сравниваются с результатами имитационного моделирования сети. Предложенный метод может применяться при невысокой нагрузке сети массового обслуживания, функционирующей в стационарном режиме. Представленная сеть массового обслуживания может использоваться в качестве моделей современных многопроцессорных вычислительных систем, а также других систем с параллельным и распределённым принципом функционирования.

Ключевые слова: сеть массового обслуживания с делением и слиянием требований, ациклическая сеть массового обслуживания, обслуживание фрагментов требований, приближенный анализ, длительность пребывания требований в сети.

¹ Оксана Сергеевна Постнова, ассистент (oksana.karpenko.2000@mail.ru).

² Игорь Евстафьевич Тананко, к.ф.-м.н., доцент (tanankoie@info.sgu.ru).

³ Екатерина Сергеевна Рогачко, к.ф.-м.н., доцент (rogachkoes@info.sgu.ru).

1. Введение

Одним из способов оперативной обработки больших объемов мало связанной информации является разделение этой информации на части с дальнейшей их одновременной обработкой высокопроизводительными вычислительными системами.

В качестве математических моделей систем параллельной или распределенной обработки информации для решения задач эффективного их использования часто используют системы и сети массового обслуживания с делением и слиянием требований или, как они называются в англоязычной литературе, «fork-join queueing systems». Поступая в такую систему, требование разделяется на фрагменты, которые затем занимают приборы и обслуживаются. Требование считается обслуженным, когда завершится обслуживание последнего из родственных фрагментов требования. Поэтому одним из основных результатов анализа систем fork-join с различной структурой и особенностями функционирования считается нахождение математического ожидания (м.о.) длительности пребывания требований в системе [11, 12].

Первые результаты исследования системы fork-join были представлены для случая разделения требования на два фрагмента с последующим одновременным началом их обработки двумя приборами с различающимися параметрами экспоненциальной функции распределения длительностей обслуживания в работе [7]. Верхние и нижние оценки м.о. длительности пребывания требований в системе с произвольным числом параллельных приборов, пуассоновским входящим потоком и экспоненциальной длительностью обслуживания фрагментов требований получены в работе [15].

Естественным развитием теории и методов анализа систем обслуживания типа fork-join является исследование двух и более параллельно связанных систем массового обслуживания с очередями, образующих сеть массового обслуживания с отдельной обработкой фрагментов требований. Некоторые представления об основных структурах, свойствах и характеристиках таких математических объектов можно получить на основе работ [6, 20].

В [3] получены условие существования стационарного режима и аналитические выражения для нижней и верхней границ м.о. длительности пребывания в fork-join сети из двух параллельных систем обслуживания с неоднородными приборами и общими предположениями о распределении длительностей обслуживания фрагментов.

Для оценки математического ожидания и дисперсии времени реакции fork-join сети обслуживания с различными типами функций распределения длительностей обслуживания фрагментов предлагается подход [9, 10, 21], использующий методы машинного обучения, методы оптимизации и имитационное моделирование. Дальнейшие исследования выявили зависимость между длительностью пребывания фрагментов в системах сети обслуживания и временем реакции этой сети [8].

Работы [18, 19] посвящены изучению fork-join сетей параллельных систем массового обслуживания в условиях высокой загрузки. Например, для сети обслуживания с дискретным временем [19] получен предел при числе систем, стремящемся к бесконечности, непрерывной функции, зависящей от максимально возможной загрузки сети обслуживания. В работе [18] решается задача определения функции распределения времени реакции fork-join сети обслуживания с использованием верхней и нижней границ рабочей нагрузки, определяемой как объем работы, которую должен выполнить прибор системы обслуживания, прежде чем он станет свободным при отсутствии поступления очередного фрагмента.

В работах [14, 17] решается задача оптимального распределения фиксированного количества ресурсов между параллельными системами fork-join сети, в которой обслуживаются фрагменты однородных [14] или неоднородных [17] требований. Предложенные алгоритмы распределения ресурсов минимизируют время реакции рассматриваемой сети обслуживания. В работе [16] решается задача синтеза структуры и алгоритма обслуживания требований fork-join сети массового обслуживания с заданными средним и дисперсией времени реакции.

Рассмотренные fork-join системы и сети массового обслуживания имеют параллельную структуру и могут использоваться в качестве математических моделей многопроцессорных вычислительных систем, оперативной обработки больших объемов информации, поддержки принятия решений с использованием машинного обучения и нейронных сетей. При этом актуальной остается задача анализа и эффективного использования систем передачи информации, особенно сетей с коммутацией пакетов и систем многопутевой маршрутизации. Связь между источником и получателем данных в таких системах осуществляется посредством сети передачи данных с произвольной топологией. В качестве математической модели таких систем может использоваться сеть массового обслуживания произвольной топологии, состоящая из бесконечноприборных систем обслуживания [4].

В данной работе рассматривается открытая сеть массового обслуживания с делением и слиянием требований, включающая системы массового обслуживания, объединенные в ациклическую сеть обслуживания, и выделенную систему сборки фрагментов требований. Эта сеть может быть использована в качестве математической модели, например, распределенной системы обработки и передачи данных, характеризующейся разделением данных на пакеты. Такой способ передачи данных отличается достаточно равномерной загрузкой всех технических устройств на широком диапазоне значений внешней нагрузки на систему. Для рассматриваемой сети массового обслуживания выполняются следующие условия. Входящий в сеть поток требований является пуассоновским. Поступающие требования делятся на два фрагмента. Длительности обслуживания фрагментов требований во всех системах имеют экспоненциальное распределение. Описывается модель системы сборки фрагментов требований. Решается задача приближенного вычисления математических ожиданий длительности сборки требований и длительности пребывания требований в сети. Для оценки точности приближения полученных результатов используется их сравнение с результатами имитационного моделирования рассматриваемой сети.

2. Описание модели

Рассматривается сеть массового обслуживания N^f с делением и слиянием требований. Сеть N^f включает L систем обслуживания S_i типа $M/M/1$ с интенсивностями обслуживания μ_i , $i = 1, \dots, L$, и выделенную систему S_{L+1} , называемую системой сборки.

Из источника S_0 в сеть обслуживания с интенсивностью λ_0 поступает пуассоновский поток требований одного класса. В момент поступления требования в сеть это требование разделяется на два фрагмента, называемых родственными, и оба фрагмента независимо друг от друга поступают в одну из систем сети S_i , $i = 1, \dots, L$, с которыми связан источник.

Фрагменты одного требования независимо друг от друга и независимо от фрагментов других требований переходят между системами сети S_i , $i = 1, \dots, L$, до тех пор, пока не завершится их обслуживание в сети. После этого фрагменты требований поступают в систему сборки S_{L+1} . Родственные фрагменты, ранее принадлежавшие одному требованию, собираются в единое требование, которое покидает систему сборки и сеть N^f . Переходы фрагментов между системами обслуживания сети, источником и системой сборки осуществляются в соответствии с маршрутной матрицей $\Theta = (\theta_{ij})$, $i = 0, 1, \dots, L$, $j = 1, \dots, L + 1$, где θ_{ij} – вероятность того, что после пребывания в системе S_i фрагмент перейдет в S_j .

Сеть массового обслуживания с введенными ранее параметрами систем, в которой обслуживаются фрагменты требований, без системы сборки обозначим через N . Предполагается, что сеть N – ациклическая, т.е. структура сети представляется ориентированным ациклическим графом. Элементы маршрутной матрицы $\bar{\Theta} = (\bar{\theta}_{ij})$, $i, j = 0, 1, \dots, L$, сети N определяются как $\bar{\theta}_{ij} = \theta_{ij}$, если $j \neq 0$, и $\bar{\theta}_{i,0} = \theta_{i,L+1}$. Предполагается, что в сети N число смежных для S_i , $i = 0, 1, \dots, L$, (кроме S_i , для которых $\bar{\theta}_{i,0} = 1$) систем обслуживания S_j , в которые возможен переход фрагментов из S_i , намного больше двух (числа фрагментов

требования), и вероятности $\bar{\theta}_{ij} > 0$, $j = 0, 1, \dots, L$, сравнимы. В этом случае выходящий из сети N поток фрагментов требований можно считать пуассоновским [5] с интенсивностью $2\lambda_0$.

Для сети N необходимым условием существования стационарного режима является

$$(1) \quad \lambda_0 < \frac{1}{2} \min_{i=1, \dots, L} \frac{\mu_i \omega_0}{\omega_i} = a,$$

где $\omega = (\omega_j)$, $j = 0, 1, \dots, L$, – вектор относительных интенсивностей потоков фрагментов требований в сети N – находится как решение уравнения $\omega = \omega \Theta$ с условием нормировки $\sum \omega_j = 1$.

Назначение системы сборки S_{L+1} – сбор требований из родственных фрагментов, поступающих из сети N . Первый из двух поступающих в систему сборки родственных фрагментов будем называть «первым» фрагментом требования. В момент поступления «второго» из родственных фрагментов мгновенно происходит объединение фрагментов в единое требование, которое покидает систему сборки. Таким образом, длительность сборки требования из родственных фрагментов определяется как время пребывания в системе сборки «первого» фрагмента, то есть длительность интервала времени между моментами выхода родственных фрагментов из сети N .

3. Приближенный метод анализа

Математическое ожидание $E\tau_0^f$ длительности пребывания требований в сети N^f определяется как сумма математического ожидания длительности пребывания требований в сети обслуживания N и математического ожидания длительности сборки требований из фрагментов в системе сборки S_{L+1} . Для вычисления последней величины опишем приближенную модель системы сборки требований.

Определим параметры системы сборки в соответствии с символикой Дж. Кендалла [1].

Заметим, что если в систему сборки поступает фрагмент, который не имеет среди уже находящихся фрагментов в этой системе родственного фрагмента, то вновь прибывший фрагмент

занимает свободный прибор. Следовательно, система сборки содержит неограниченное число приборов.

Известно, что выходящий из сети N поток фрагментов требований является пуассоновским [5]. Поэтому в систему сборки также поступает пуассоновский поток фрагментов с интенсивностью λ_0 , так как обслуживаются в системе только «первые» фрагменты.

Результаты имитационного моделирования показывают [2], что длительность сборки требования из родственных фрагментов можно представить экспоненциально распределенной случайной величиной с математическим ожиданием $1/\mu$, а также то, что изменение параметра μ приближенно описывается убывающей линейной функцией интенсивности λ_0 .

Таким образом, систему сборки представим в нашей модели системой массового обслуживания типа $M/M/\infty$.

Определим функцию $\mu = \mu(\lambda_0)$, $\lambda_0 \in (0, a)$. Графиком функции $\mu(\lambda_0)$ является прямая, проходящая через две точки $(0, \nu)$ и $(a, 0)$, где ν – максимальное значение μ , получаемое при $\lambda_0 = \varepsilon$, где $\varepsilon > 0$ – малое число. Действительно, при $\lambda_0 \rightarrow 0$ требование, поступающее в сеть N^f , с вероятностью, стремящейся к 1, заставит сеть пустой. Тогда оба фрагмента требования успевают выйти из сети N , прежде чем в сеть поступит новое требование, и длительность ожидания в системе сборки «первым» фрагментом требования «второго» фрагмента будет минимальной. В этом случае параметр μ , очевидно, принимает максимальное значение ν . Следовательно,

$$(2) \quad \mu = -\frac{\nu}{a}\lambda_0 + \nu, \quad \lambda_0 \in (0, a).$$

Найдем ν . При $\lambda_0 = \varepsilon$ вероятность того, что между двумя поступающими в систему сборки родственными фрагментами требования поступит фрагмент или фрагменты других требований, близка к нулю. Тогда математическое ожидание длительности сборки требований из фрагментов или, иначе, математическое ожидание длительности интервала времени между родственными фрагментами в выходящем из сети N потоке фрагментов совпадает со среднеквадратическим отклонением σ_0 длительности

сти реакции τ_0 сети N (длительности пребывания фрагментов требований в сети N). Тогда

$$(3) \quad \nu = 1/\sigma_0, \quad \sigma_0 = \sqrt{E\tau_0^2 - (E\tau_0)^2},$$

где $E\tau_0$ и $E\tau_0^2$ – первый и второй начальные моменты величины τ_0 .

Из формулы Литтла [1] для сети N следует, что

$$(4) \quad E\tau_0 = \frac{q}{2\lambda_0}, \quad q = \sum_{i=1}^L \frac{\psi_i}{1 - \psi_i},$$

где q – математическое ожидание числа фрагментов требований в сети N , $\psi_i = \lambda_i/\mu_i$ – коэффициент использования системы S_i , $\lambda_i = 2\lambda_0\omega_i/\omega_0$ – интенсивность потока фрагментов требований в систему S_i , $i = 1, \dots, L$.

Чтобы найти второй момент $E\tau_0^2$, введем в рассмотрение τ_i – длительность пребывания в сети N фрагментов требований, поступивших в сеть через систему S_i , $i = 1, \dots, L$, с первым и вторым начальными моментами $E\tau_i$ и $E\tau_i^2$ соответственно, и представим $E\tau_0^2$ в виде:

$$(5) \quad E\tau_0^2 = \sum_{i=1}^L \bar{\theta}_{0i} E\tau_i^2.$$

Известно [13], что величины $E\tau_i$ могут быть найдены из системы уравнений:

$$(6) \quad E\tau_i = \frac{1}{\mu_i(1 - \psi_i)} + \sum_{j=1}^L \bar{\theta}_{ij} E\tau_j, \quad i = 1, \dots, L,$$

а величины $E\tau_i^2$ – из следующей системы уравнений:

$$(7) \quad E\tau_i^2 = \frac{2E\tau_i}{\mu_i(1 - \psi_i)} + \sum_{j=1}^L \bar{\theta}_{ij} E\tau_j^2, \quad i = 1, \dots, L.$$

Введем обозначения векторов и матриц: $\mathbf{E}\boldsymbol{\tau} = (E\tau_i)$, $\mathbf{u} = (1/(\mu_i - \lambda_i))$, $i = 1, \dots, L$, $\bar{\boldsymbol{\Theta}}' = (\bar{\theta}_{ij})$, $i, j = 1, \dots, L$; \mathbf{I} – единичная матрица порядка L , тогда из системы уравнений (6) имеем

$$\mathbf{E}\boldsymbol{\tau} = \mathbf{u}(\mathbf{I} - \bar{\boldsymbol{\Theta}}')^{-1}.$$

Умножив уравнения (7) на λ_i и просуммировав их по i , $i = 1, \dots, L$, получим

$$\sum_{i=1}^L \lambda_i E\tau_i^2 = 2 \sum_{i=1}^L \frac{\lambda_i}{\mu_i(1-\psi_i)} E\tau_i + \sum_{i=1}^L \lambda_i \sum_{j=1}^L \bar{\theta}_{ij} E\tau_j^2,$$

$$\sum_{i=1}^L \lambda_i E\tau_i^2 = 2 \sum_{i=1}^L \frac{\lambda_i}{\mu_i(1-\psi_i)} E\tau_i + \sum_{j=1}^L (\lambda_j - 2\lambda_0 \bar{\theta}_{0j}) E\tau_j^2,$$

или

$$2\lambda_0 \sum_{i=1}^L \bar{\theta}_{0i} E\tau_i^2 = 2 \sum_{i=1}^L \frac{\lambda_i}{\mu_i(1-\psi_i)} E\tau_i,$$

откуда с учетом (5) имеем

$$(8) \quad E\tau_0^2 = \frac{1}{\lambda_0} \sum_{i=1}^L \frac{\psi_i}{1-\psi_i} E\tau_i.$$

Окончательно, подставив найденные по формулам (4) и (8) значения в формулу (3), найдем значение ν , используемое в формуле (2).

Обозначим P_k – стационарная вероятность пребывания системы сборки в состоянии k , $k = 0, 1, 2, \dots$, где k – число фрагментов в системе сборки. Имеем

$$P_k = \frac{(\lambda_0/\mu)^k}{k!} e^{-\lambda_0/\mu}, \quad k = 0, 1, 2, \dots,$$

где μ вычисляется по формуле (2). Математическое ожидание числа фрагментов требований в системе сборки $n = \lambda_0/\mu$.

Определим математическое ожидание $E\tau_0^f$ длительности пребывания требований в сети N^f . Очевидно, что математическое ожидание числа требований в сети N^f равно $(q+n)/2$. С учетом формулы Литтла и найденного математического ожидания $1/\mu$ длительности сборки требований величина $E\tau_0^f$ приближенно определяется выражением

$$(9) \quad E\tau_0^f = \frac{q-n}{2\lambda_0} + 1/\mu.$$

4. Результаты моделирования и точность приближения

Для проверки точности полученных аналитических приближенных результатов была разработана имитационная модель сети N^f [2]. В этой модели реализовано разделение требований на два фрагмента при поступлении требований в сеть, независимая маршрутизация фрагментов в сети N , ожидание «первыми» фрагментами в системе сборки поступления именно родственных «вторых» фрагментов требований.

Рассматривается сеть N^f с $L = 30$ системами обслуживания при различных топологиях и маршрутных матрицах сети N : сеть параллельных систем в три слоя (тип сети « p »): $\bar{\theta}_{0i} = \bar{\theta}_{ij} = \bar{\theta}_{jk} = 0,1$, $\bar{\theta}_{k0} = 1$, $i = 1, \dots, 10$, $j = 11, \dots, 20$, $k = 21, \dots, 30$; сеть с древовидной структурой в три слоя (тип сети « t »): $\bar{\theta}_{0i} = 0,1$, $i = 1, \dots, 10$, $\bar{\theta}_{10} = \bar{\theta}_{20} = 1$, $\bar{\theta}_{ij} = 0,1$, $i = 3, \dots, 10$, $j = 11, \dots, 20$, $\bar{\theta}_{19,0} = \bar{\theta}_{20,0} = 1$, $\bar{\theta}_{jk} = 0,1$, $\bar{\theta}_{k0} = 1$, $j = 11, \dots, 18$, $k = 21, \dots, 30$.

Приводятся результаты вычислительных экспериментов при двух вариантах векторов интенсивностей обслуживания в системах сети N : вектор одинаковых интенсивностей обслуживания (вектор « e »): $\mu_i = 4$, $i = 1, \dots, 30$; вектор различных интенсивностей обслуживания (вектор « d »): $\mu_i = \mu_{i+10} = \mu_{i+20} = 4 + 0,1(i-1)$, $i = 1, \dots, 10$. При таких параметрах сети N согласно условию (1) существования стационарного режима $a = 20$.

Будет использоваться обозначение сети N (тип топологии, тип вектора интенсивностей обслуживания), где «тип топологии» принимает значения из $\{p, t\}$, а «тип вектора интенсивностей обслуживания» принимает значения из $\{e, d\}$.

На рис. 1 представлена интенсивность обслуживания μ в системе сборки при значениях $0 < \lambda_0 < a$, удовлетворяющих условию (1), по результатам имитационного моделирования (ИМ) и результатам аппроксимации (АП) по формуле (2). Результаты имитационного моделирования здесь и далее получены с доверительной вероятностью 0,95 и шириной доверительного интервала 0,01.

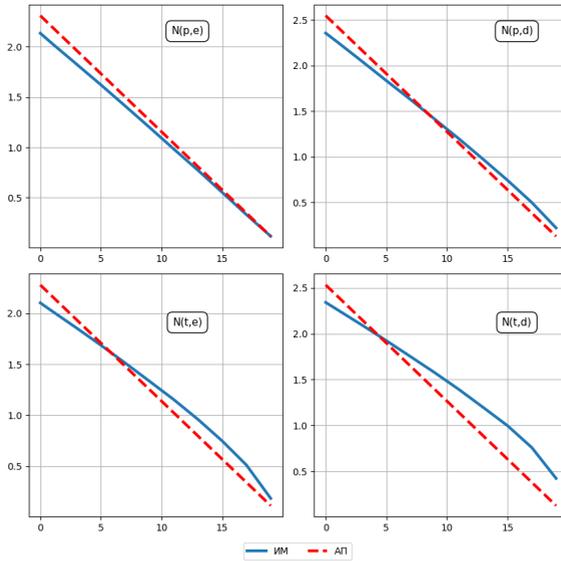


Рис. 1. Интенсивность обслуживания в системе сборки в зависимости от λ_0

Введем величину $\delta = |\tilde{\mu} - \mu|/\tilde{\mu}$, где $\tilde{\mu}$, μ – оценки интенсивности обслуживания в системе сборки, полученные в результате имитационного моделирования и аппроксимации соответственно. Из рис. 1 видно, что для сетей с различными параметрами при $\lambda_0 \rightarrow a$, когда $\mu \rightarrow 0$, результаты имитационного моделирования и аппроксимации могут существенно различаться и погрешность аппроксимации, определяемая величиной δ , может быть большой. Например, при $\lambda_0 \geq 0,55a$ для сетей $N(t, *)$ имеем $\delta > 0,1$. При указанных значениях λ_0 выполняется условие $\lambda_0 \gg \mu$ или, иначе, $\bar{n} \gg 1$. Далее случай $\lambda_0 > 0,55a$ не рассматривается, так как в этом случае наблюдается низкая точность аппроксимации μ .

В таблице 1 представлены оценки \tilde{n} и n математического ожидания числа фрагментов требований в системе сборки, полученные в результате имитационного моделирования и аппроксимации соответственно, и величина $\delta = |\tilde{n} - n|/\tilde{n}$ при λ_0 , изменяющейся от $0,05a$ до $0,55a$ с шагом $0,1a$.

Таблица 1. Математическое ожидание числа фрагментов требований в системе сборки

Тип сети	λ_0	$N(*, e)$			$N(*, d)$		
		\tilde{n}	n	δ	\tilde{n}	n	δ
$N(p, *)$	1,0	0,493	0,456	0,075	0,444	0,413	0,069
	3,0	1,643	1,528	0,070	1,467	1,385	0,056
	5,0	3,087	2,887	0,065	2,727	2,617	0,040
	7,0	4,965	4,663	0,061	4,308	4,227	0,019
	9,0	7,517	7,086	0,057	6,375	6,423	0,007
	11,0	11,200	10,585	0,055	9,199	9,594	0,043
$N(t, *)$	1,0	0,496	0,462	0,067	0,443	0,416	0,061
	3,0	1,617	1,550	0,041	1,432	1,395	0,026
	5,0	2,966	2,928	0,013	2,596	2,634	0,015
	7,0	4,630	4,730	0,022	4,002	4,256	0,063
	9,0	6,758	7,187	0,064	5,715	6,466	0,131
	11,0	9,530	10,736	0,127	7,912	9,660	0,221

Рис. 2 показывает результаты оценки плотности распределения вероятностей длительности обслуживания фрагментов требований в системе сборки. Результаты приведены для сетей рассмотренных типов при $\lambda_0 = 0,55a$. Из результатов имитационного моделирования следует, что длительность сборки требований – экспоненциально распределенная случайная величина с параметром μ . Соответствующие расстояния Колмогорова указываются далее в таблице 2.

По результатам имитационного моделирования число фрагментов требований в системе сборки имеет распределение Пуассона. На рис. 3 показаны результаты для сетей рассмотренных типов при $\lambda_0 = 0,25a$.

В таблице 2 приводятся расстояния Колмогорова для оценки точности аппроксимации распределения вероятностей длительности сборки и распределения вероятностей числа фрагментов в системе сборки. Расстояние Колмогорова для оценки точности аппроксимации распределения вероятностей длительности сборки определяется

$$\Delta_1 = \sup_x |F(x) - \tilde{F}(x)|,$$

где $F(x)$ – функция распределения экспоненциально распределенной случайной величины с найденным параметром μ , $\tilde{F}(x)$ – эмпирическая функция распределения по результатам имитационного моделирования. Расстояние Колмогорова для оценки точности аппроксимации распределения вероятностей числа фрагментов в системе сборки определяется

$$\Delta_2 = \max_{i \geq 0} \left| \sum_{k=0}^i P_k - \sum_{k=0}^i \tilde{P}_k \right|,$$

где P_k и \tilde{P}_k – вероятности состояния k , $k = 0, 1, 2, \dots$, системы сборки, вычисленные аналитически и по результатам имитационного моделирования соответственно.

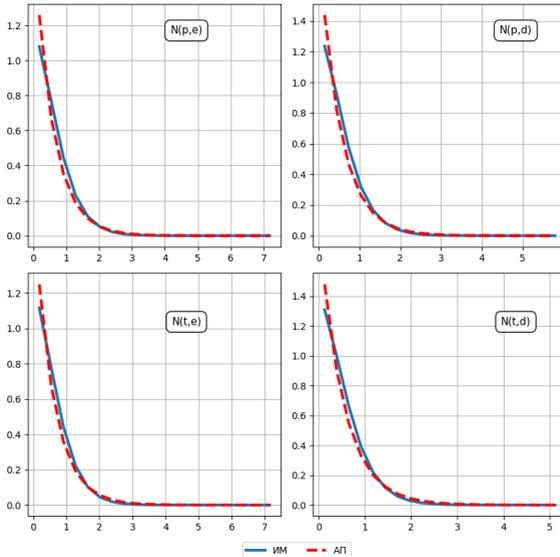


Рис. 2. Плотность распределения вероятностей длительности обслуживания фрагментов в системе сборки при $\lambda_0 = 0,55a$

На основе приведенных результатов можно сделать вывод, что полученная в работе приближенная формула (2) дает достаточную точность аппроксимации при значениях $0 < \lambda_0 < 0,5a$, т.е. для сетей N с невысокой нагрузкой.

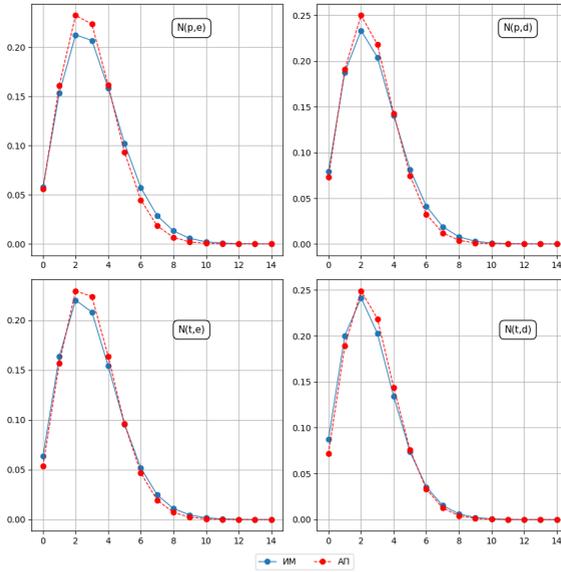


Рис. 3. Распределение вероятностей числа фрагментов требований в системе сборки при $\lambda_0 = 0,25a$

Таблица 2. Расстояния Колмогорова

Тип сети	λ_0	Δ_1		Δ_2	
		$N(*, e)$	$N(*, d)$	$N(*, e)$	$N(*, d)$
$N(p, *)$	1,0	0,082	0,078	0,019	0,016
	3,0	0,078	0,073	0,033	0,025
	5,0	0,074	0,066	0,045	0,030
	7,0	0,071	0,058	0,058	0,031
	9,0	0,070	0,048	0,074	0,053
	11,0	0,068	0,035	0,092	0,098
$N(t, *)$	1,0	0,078	0,075	0,017	0,014
	3,0	0,069	0,063	0,021	0,014
	5,0	0,056	0,049	0,018	0,027
	7,0	0,046	0,034	0,046	0,068
	9,0	0,032	0,048	0,095	0,136
	11,0	0,046	0,071	0,174	0,240

В таблице 3 представлены оценки $E\tilde{\tau}_0^f$ и $E\tau_0^f$ математического ожидания длительности пребывания требований в се-

ти N^f , полученные в результате имитационного моделирования и аппроксимации по формуле (9) соответственно, и величина $\delta = |E\tilde{\tau}_0^f - E\tau_0^f|/E\tilde{\tau}_0^f$ при λ_0 , изменяющейся от 0,05а до 0,55а с шагом 0,1а.

Таблица 3. Математическое ожидание длительности пребывания требований в сети N^f

Тип сети	λ_0	$N(*, e)$			$N(*, d)$		
		$E\tilde{\tau}_0^f$	$E\tau_0^f$	δ	$E\tilde{\tau}_0^f$	$E\tau_0^f$	δ
$N(p, *)$	1,0	1,060	1,017	0,040	0,953	0,916	0,039
	3,0	1,183	1,137	0,039	1,053	1,014	0,036
	5,0	1,338	1,289	0,037	1,173	1,137	0,031
	7,0	1,542	1,487	0,036	1,328	1,294	0,025
	9,0	1,821	1,757	0,035	1,535	1,502	0,021
	11,0	2,225	2,148	0,035	1,814	1,792	0,012
$N(t, *)$	1,0	0,904	0,868	0,040	0,811	0,780	0,038
	3,0	0,989	0,957	0,032	0,879	0,854	0,028
	5,0	1,093	1,067	0,024	0,961	0,945	0,017
	7,0	1,225	1,206	0,015	1,062	1,058	0,004
	9,0	1,396	1,391	0,004	1,189	1,205	0,013
	11,0	1,630	1,648	0,011	1,358	1,404	0,034

Из приведенных результатов видно, что погрешность вычислений по формуле (9) не превосходит 4%. Это свидетельствует о применимости приближенной формулы для оценки математического ожидания длительности пребывания требований в сети N^f при невысокой нагрузке сети.

5. Заключение

Предложен приближенный метод вычисления математического ожидания длительности пребывания требований в сети массового обслуживания с делением и слиянием требований. Особенность рассматриваемой сети состоит в том, что, во-первых, поступающие в сеть требования делятся на два фрагмента. Во-вторых, обслуживание фрагментов требований осуществляется не в параллельных системах обслуживания, а в ациклической сети систем обслуживания. После завершения обслуживания

ния фрагментов их объединение в требования выполняется в системе сборки требований. Описаны приближенная модель системы сборки и способ вычисления ее стационарных характеристик. Темой дополнительных исследований может быть определение соотношения длительности сборки и длительности пребывания требований в сети.

Разработанный метод может быть использован для решения задач анализа и оптимизации систем многопутевой маршрутизации, систем последовательно-параллельной обработки информации, производственных систем, моделируемых сетями обслуживания с делением и слиянием требований. Метод наиболее применим для вычисления характеристик систем с нагрузкой не выше средней, но может быть также использован для выявления потенциально узких мест в системах.

Литература

1. ВИШНЕВСКИЙ В.М. *Теоретические основы проектирования компьютерных сетей*. – М.: Техносфера, 2003. – 512 с.
2. КАРПЕНКО О.С., ТАНАНКО И.Е., РОГАЧКО Е.С. *Исследование имитационной модели открытой сети массового обслуживания с делением и слиянием требований // Системы управления, информационные технологии и математическое моделирование : Материалы V Всероссийской научно-практической конференции с международным участием*. – Омск: ОмГТУ, 2023. – С. 209–214.
3. КЛИМЕНОК В.И. *Характеристики производительности системы массового обслуживания с расщеплением запросов // Информатика*. – 2023. – Т. 20, №3. – С. 50–60.
4. ОСИПОВ О.А., ТАНАНКО И.Е. *Сети массового обслуживания произвольной топологии с делением и слиянием требований: случай бесконечноприборных систем обслуживания // Вестник Тверского государственного университета. Серия: Прикладная математика*. – 2017. – №4. – С. 43–58.
5. ЦИЦИАШВИЛИ Г.Ш., ОСИПОВА М.А. *Стационарные потоки в ациклических сетях массового обслуживания //*

- Дальневосточный математический журнал. – 2016. – №2. – С. 223–228.
6. ENGANTI P., ROSENKRANTZ T., SUN L. et al. *ForkMV: Mean-and-variance estimation of fork-join queuing networks for datacenter applications* // IEEE Int. Conf. on Networking, Architecture and Storage (NAS–2022). – Philadelphia, PA, USA, 2022. – P. 1–8.
 7. FLATTO L., HAHN S. *Two parallel queues created by arrivals with two demands I* // SIAM Journal on Applied Mathematics. – 1984. – Vol. 44, No. 5. – P. 1041–1053.
 8. GORBUNOVA A.V., LEBEDEV A.V. *Correlations of the sojourn times of subtasks in fork-join queueing systems with M/M/1-type subsystems* // Advances in Systems Science and Applications. – 2024. – Vol. 24, No. 2. – P. 1–18.
 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.
 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.
 11. KEMPER B., MANDJES M. *Mean sojourn times in two-queue fork-join systems: bounds and approximations* // OR Spectrum. – 2012. – Vol. 34, No. 3. – P. 723–742.
 12. KO S.S., SERFOZO R. *Response times in M/M/s fork-join networks* // Advances in Applied Probability. – 2004. – Vol. 36, No. 3. – P. 854–871.
 13. LEMOINE A.J. *Networks of queues – a survey of equilibrium analysis* // Management Science. – 1977. – Vol. 24, No. 4. – P. 464–481.
 14. MARIN A., ROSSI S., SOTTANA M. *Dynamic resource allocation in fork-join queues* // ACM Trans. on Modeling

- and Performance Evaluation of Computing Systems. – 2020. – Vol. 5, No. 1. – Article No. 3. – P. 1–28.
15. NELSON R., TANTAWI A.N. *Approximate analysis of fork/join synchronization in parallel queues* // IEEE Trans. on Comp. – 1988. – Vol. 37, No. 6. – P. 739–743.
 16. NGUYEN M., ALESAWI S., LI N. 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.
 17. OZKAN E. *Control of fork-join processing networks with multiple job types and parallel shared resources* // Mathematics of Operations Research. – 2022. – Vol. 47, No. 2. – P. 1310–1334.
 18. RAAIJMAKERS Y., BORST S., BOXMA O. *Fork-join and redundancy systems with heavy-tailed job sizes* // Queueing Systems. – 2023. – Vol. 103, No. 1. – P. 131–159.
 19. SCHOL D., VLASIOU M., ZWART B. *Large fork-join queues with nearly deterministic arrival and service times* // Mathematics of Operations Research. – 2021. – Vol. 47, No. 2. – P. 1335–1364.
 20. THOMASIAN A. *Analysis of fork/join and related queueing systems* // ACM Computing Surveys. – 2014. – Vol. 47, No. 2. – Article No. 17. – P. 1–71.
 21. VISHNEVSKY V.M., KLIMENOK V.I., SOKOLOV A.M. et al. *Investigation of the fork-join system with Markovian arrival process arrivals and phase-type service time distribution using machine learning methods* // Mathematics. – 2024. – Vol. 12, No. 5. – Article No. 659. – P. 1–22.

AN APPROXIMATION FOR THE RESPONSE TIME IN FORK-JOIN QUEUEING NETWORKS

Oksana Postnova, Saratov State University, Saratov, Assistant
(oksana.karpenko.2000@mail.ru).

Igor Tananko, Saratov State University, Saratov, Cand.Sc.,
Associate Professor (tanankoie@info.sgu.ru).

Ekaterina Rogachko, Saratov State University, Saratov, Cand.Sc.,
Associate Professor (rogachkoes@info.sgu.ru).

Abstract: An open fork-join queueing network is considered. An arriving job is split to be serviced into two tasks. The tasks are serviced independently at multiple service-nodes. Each service-node is a queueing system with one server and infinite capacity queue. Service-nodes form a queueing network with an acyclic topology. Two tasks associated with the job synchronize at a join-node before they leave the network. Approximations for the mean time spent by a task in the join-node and the mean response time in the fork-join queueing network are derived by assuming that jobs arrive according to a Poisson process and task service times have an exponential distribution. The accuracy of these approximations is demonstrated by comparing approximate results to simulation results. These approximations can be applied to the analysis of queueing networks with relatively small steady-state workload. The results can be used for the performance analysis of multiprocessor systems and other modern distributed computing systems.

Keywords: fork-join queueing network, acyclic queueing network, task service, approximation, response time.

УДК 519.872

ББК 22.18

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

Поступила в редакцию 14.01.2025.

Дата опубликования 31.05.2025.

ИССЛЕДОВАНИЕ СЕТИ МАССОВОГО ОБСЛУЖИВАНИЯ С НЕОГРАНИЧЕННЫМ ЧИСЛОМ ПРИБОРОВ В УЗЛАХ И ПРЕРЫВАНИЯМИ ОБСЛУЖИВАНИЯ

Салимзянов Р. Р.¹, Моисеев А. Н.², Сущенко С. П.³
(Томский государственный университет, Томск)

Представлено исследование сети массового обслуживания с неограниченным числом обслуживающих приборов в узлах сети и возможностью прерывания обслуживания. Такая модель может быть использована для описания функционирования абонентских сетей доступа, где заявка (подключённый абонент) в процессе обслуживания может перемещаться от одного узла сети к другому или покидать сеть, завершив или не завершив своё обслуживание. Предполагается, что переходы происходят независимо от состояния узлов. Исследование проводится с использованием метода асимптотического анализа при условии высокой интенсивности входящего потока. В рамках проведённого анализа установлено, что при стремлении интенсивности потока к бесконечности совместное стационарное распределение числа заявок в узлах сети приближается к многомерному нормальному распределению. Получены аналитические выражения для параметров этого распределения, таких как математические ожидания и элементы ковариационной матрицы. Проведены численные эксперименты, подтверждающие точность асимптотической аппроксимации. Определена область применимости полученных результатов в зависимости от параметров модели. Кроме того, в работе приведён пример решения задачи оптимизации функционирования исследуемой сети массового обслуживания, иллюстрирующий возможности практического применения разработанного подхода при анализе и управлении телекоммуникационными системами.

Ключевые слова: сеть массового обслуживания, неограниченное число приборов, прерывание обслуживания, асимптотический анализ.

1. Введение

Прогресс технологий и увеличение числа и размеров абонентских сетей связи вызывает необходимость моделирования, исследования и оценки эффективности работы таких сетей.

¹ Радмир Ренатович Салимзянов, аспирант (rsalimzyanov@yahoo.com).

² Александр Николаевич Моисеев, д.ф.-м.н., зав. каф. (moiseev.tsu@gmail.com).

³ Сергей Петрович Сущенко, д.т.н., зав. каф. (ssp.inf.tsu@gmail.com).

Для данных задач активно используют аппарат теории массового обслуживания и имитационное моделирование. С повышением уровня требований к сетям связи нового поколения и изменением сценариев поведения пользователей изменяются и математические модели. Например, в работе [8] рассматривается модель одного узла сети, работающего по технологии интегрированного доступа и транзита [12, 14, 18, 19] в виде системы поллинга [1]. В работах [3, 10, 17] рассматриваются ресурсные системы массового обслуживания, которые позволяют оценить необходимое число радиоресурсов для приема/передачи трафика.

В приведенных работах авторы рассматривают один узел связи, что позволяет оценить время задержки пакетов, число необходимых радио-ресурсов, распределение числа заявок в системе для одной базовой станции. Масштабирование подобных систем для моделирования сетей связи с возможными потерями сигнала и перемещениями пользователя внутри сети является трудоемкой задачей из-за необходимости высокой детализации модели.

В данной работе предложена и исследована модель абонентской сети связи, учитывающая случайный характер перемещения пользователей и возможность прерывания обслуживания вследствие различных факторов.

2. Постановка задачи

Рассмотрим абонентскую сеть, которая состоит из K базовых станций (БС). Абоненты сети перемещаются от одной БС к другой во время обслуживания. Пользователь может выйти за пределы действия БС и не подключиться к другой БС; это может произойти ввиду отсутствия базовой станции в достаточной близости к абоненту для подключения, блокировки сигнала [16] и т.д.

Данную системную модель можно описать в виде системы массового обслуживания (СМО). Рассмотрим простейший случай рассматриваемой модели с двумя базовыми станциями (рис. 1).

На вход системы (рис. 1) поступает простейший поток за-

явок с интенсивностью λ (подключаемые абоненты). Входящая заявка занимает любой свободный прибор (соответствует подключению одного абонента), внутри которого с вероятностью v_1 переходит в состояние 1 (подключается к БС 1) или с вероятностью v_2 – в состояние 2 (подключается к БС 2). Заявка может перейти из состояния 1 в состояние 2 за время Δt с вероятностью $\alpha_{12}\Delta t$ и из состояния 2 в состояние 1 – с вероятностью $\alpha_{21}\Delta t$. Заявка может досрочно покинуть систему (выйти из зоны действия БС) за время Δt с вероятностью $\alpha_{10}\Delta t$ для состояния 1 и с вероятностью $\alpha_{20}\Delta t$ – для состояния 2. По завершении времени обслуживания с произвольным законом распределения $B(x)$ заявка покидает систему независимо от своего текущего состояния (успешное завершение сеанса). Распределение времени пребывания заявки в одном состоянии обозначено как $F_n(x)$, где n – номер состояния.

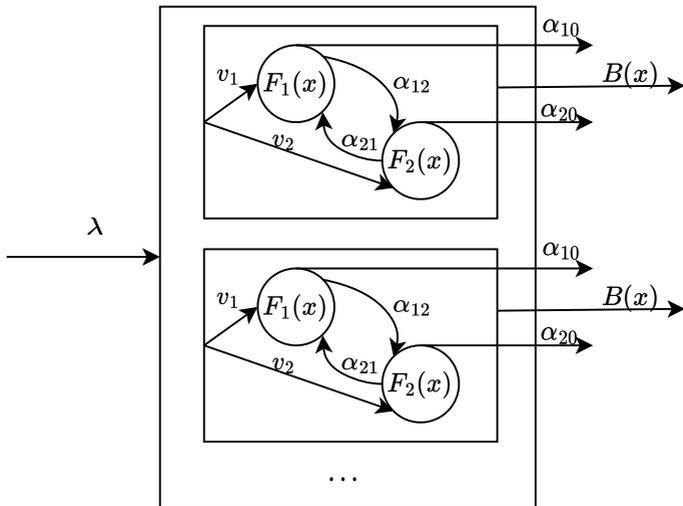


Рис. 1. Базовая модель

Очевидно, что с увеличением числа базовых станций размерность рассматриваемой модели увеличивается и усложняется ее запись. Обозначим число БС в данной сети через K , тогда в модели представление только одного прибора (заявки) будет выглядеть как показано на рис. 2.

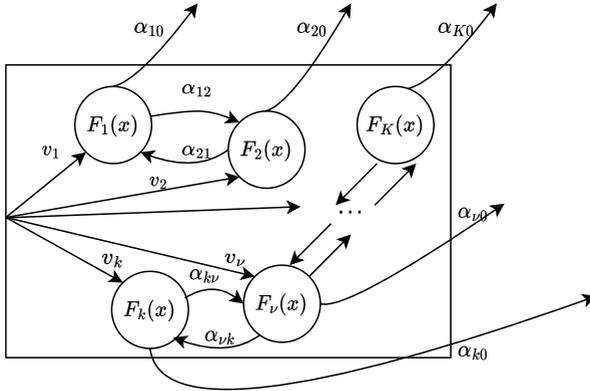


Рис. 2. Один блок обслуживания для СМО с K состояниями

Мы предлагаем представить рассматриваемую СМО в виде сети массового обслуживания (Семо) с неограниченным числом приборов в узлах (рис. 3).

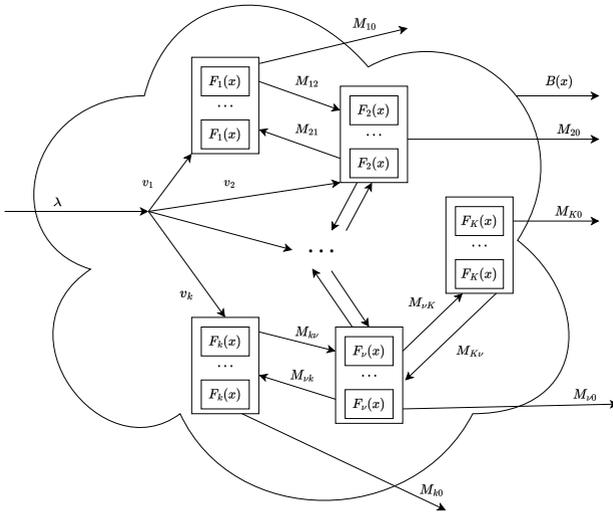


Рис. 3. Схема Семо с прерыванием обслуживания

Состояние обслуживаемых заявок будем моделировать в виде узлов Семо. Таким образом, один узел сети будет соответ-

ствовать одной БС.

Входящая заявка с вероятностью v_n поступает в узел n и пребывает в нем в течение случайного времени с функцией распределения вероятностей $F_n(x)$ без учета условия завершения пребывания в сети по окончании случайного времени с распределением $B(x)$:

$$(1) \quad F_n(x) = 1 - \exp \left\{ -x \sum_{\substack{i=0 \\ i \neq n}}^K \alpha_{ni} \right\}.$$

По окончании пребывания в узле заявка переходит в другой узел или осуществляет досрочный выход из сети согласно матрице маршрутизации \mathbf{M} с элементами

$$(2) \quad M_{ij} = \frac{\alpha_{ij}}{\sum_{k=0}^K \alpha_{ik}} \text{ для } i \neq n, i = 1, \dots, K, j = 1, \dots, K;$$

$$M_{i0} = 1 - \sum_{k=1}^K M_{ik}; \quad M_{ii} = 0.$$

Кроме того, в любой момент заявка может покинуть сеть, успешно завершив обслуживание по истечении общего времени пребывания в сети, которое имеет функцию распределения вероятностей $B(x)$. Для предметной области это событие является успешным завершением обслуживания, однако для рассматриваемой модели сети это выглядит как прерывание обслуживания, так как в момент этого прерывания заявка находилась на обслуживании в каком-то из узлов.

В такой постановке задачи нас интересует совместное распределение вероятностей числа заявок в узлах сети. Для ее решения воспользуемся подходом [6].

3. Метод динамического просеивания и асимптотический анализ

Метод динамического просеивания вместе с методом асимптотического анализа [4, 6, 7] позволяет проводить исследования многофазных СМО и СеМО.

Изложим кратко суть метода динамического просеивания. Если в некоторый момент времени $t_0 \leq T$ система пустая, то распределение вероятностей числа событий $\mathbf{n}(t)$, наступивших в просеянном потоке до момента времени T , совпадает с многомерным распределением вероятностей процесса $\mathbf{i}(t)$ (числа заявок в узлах сети) в момент времени $t = T$, т.е. имеют место равенства [5]

$$(3) \quad P\{\mathbf{i}(T) = \mathbf{m}\} = P\{\mathbf{n}(T) = \mathbf{m}\}$$

для любых неотрицательных значений вектора \mathbf{m} .

В нашем случае искомые вероятности просеивания имеют вид

$$(4) \quad S_n(t) = [1 - B(T - t)]W_n(T - t),$$

где $W_n(T - t)$ – условная вероятность того, что заявка поступившая в сеть в момент времени t , к моменту времени T будет находиться в узле n при условии, что к этому моменту времени ее пребывание в сети не будет прервано [9, 15], множитель $[1 - B(T - t)]$ есть вероятность выполнения этого условия.

Условные вероятности $W_n(T - t)$ могут быть вычислены по известным формулам [6], которые в матричной форме имеют вид

$$(5) \quad \mathbf{W}(t) = \frac{1}{2\pi} \mathbf{v} \int_{-\infty}^{+\infty} e^{-j\alpha t} [\mathbf{I} - \mathbf{F}^*(\alpha)\mathbf{M}]^{-1} [\mathbf{F}^*(\alpha) - \mathbf{I}] \frac{1}{j\alpha} d\alpha.$$

Здесь \mathbf{v} – вектор-строка вероятностей попадания заявки в соответствующий узел при её поступлении в сеть; \mathbf{I} – единичная матрица; $\mathbf{F}^*(\alpha)$ – диагональная матрица, составленная из преобразований Фурье – Стильеса для функций $F_n(x)$ из (1), $j = \sqrt{-1}$.

Теперь найдем стационарное распределение числа заявок в узлах рассматриваемой сети. Решение этой задачи будем искать в асимптотическом условии растущей интенсивности входящего потока. Воспользовавшись результатами [5], применяя полученное выше выражение для вероятностей просеивания $S_n(t)$ (4) и выражение (5) для условных вероятностей $W_n(T - t)$, получим характеристическую функцию $\mathbf{h}(u)$ многомерного стационарного распределения числа заявок в узлах сети в условии высокой интенсивности входящего потока:

$$(6) \quad h(\mathbf{u}) = \exp \left\{ N\lambda j \mathbf{u}^T \mathbf{Y} \mathbf{e} + \frac{1}{2} N j \mathbf{u}^T [\lambda \mathbf{Y} + \kappa \mathbf{V}] j \mathbf{u} \right\},$$

где N – параметр высокой интенсивности входящего потока (сама интенсивность равна $N\lambda$),

$$\mathbf{Y} = \text{diag} \left\{ \int_0^{\infty} [1 - B(\tau)] W_k(\tau) d\tau \right\},$$

$$\mathbf{V} = \int_0^{\infty} [1 - B(\tau)]^2 \mathbf{W}(\tau) \mathbf{W}^T(\tau) d\tau.$$

Значение κ для простейшего входящего потока равно 0. В результате получаем

$$h(\mathbf{u}) = \exp \left\{ N\lambda \mathbf{j} \mathbf{u}^T \mathbf{Y} \mathbf{e} + \frac{1}{2} N\lambda \mathbf{j} \mathbf{u}^T \mathbf{Y} \mathbf{j} \mathbf{u} \right\}.$$

Таким образом, асимптотическое стационарное распределение вероятностей числа заявок в узлах рассматриваемой сети при условии высокой интенсивности входящего потока является многомерным гауссовским с вектором средних значений $N\lambda \mathbf{Y} \mathbf{e}$ и ковариационной матрицей $N\lambda \mathbf{Y}$.

Полученный результат (6), вообще говоря, справедлив и для других типов входящих потоков [6]. Так, в случае рекуррентного входящего потока значение κ вычисляется по формуле

$$\kappa = \lambda^3 (\sigma^2 - a^2),$$

где a – математическое ожидание; σ^2 – дисперсия длин интервалов между последовательными событиями рекуррентного входящего потока, в этом случае $\lambda = \frac{1}{a}$. Выражения для κ для других типов входящих потоков можно найти в [6]. Таким образом, асимптотическое стационарное распределение вероятностей числа заявок в узлах сети в условии высокой интенсивности входящего непуассоновского потока является многомерным гауссовским с вектором средних значений $N\lambda \mathbf{Y} \mathbf{e}$ и матрицей ковариаций $N[\lambda \mathbf{Y} + \kappa \mathbf{V}]$.

4. Численные результаты

Теоретический результат, полученный в данной работе, является аппроксимацией искомого многомерного распределения числа заявок в узлах сети, которая справедлива при достаточно больших значениях интенсивности входящего потока. Возникает

необходимость оценки погрешности данной аппроксимации и области ее применимости. Поскольку аналитическая оценка в данном случае представляется невозможной, предлагается воспользоваться методом имитационного моделирования для получения эмпирического распределения вероятностей числа заявок в сети и сравнения полученной аппроксимации с этим распределением.

Для оценки точности теоретических результатов будем использовать расстояние Колмогорова

$$\Delta = \max_i |F_{\text{sim}}(i) - F_{\text{approx}}(i)|,$$

где $F_{\text{sim}}(i)$ – эмпирическая функция распределения вероятностей, построенная на основе результатов имитационного моделирования; $F_{\text{approx}}(i)$ – функция распределения вероятностей дискретизированной гауссовской аппроксимации, которую будем вычислять по формуле

$$F_{\text{approx}}(i) = \frac{G(i + 0,5) - G(i - 0,5)}{1 - G(-0,5)}$$

для $i = 0, 1, 2, \dots$. Здесь $G(i)$ – функция распределения вероятностей гауссовского распределения с соответствующими параметрами.

Представляем результаты численного эксперимента для СеМО с тремя узлами, параметры которой представлены в таблице 1. Время обслуживания заявок имеет гамма-распределение с параметрами α и β .

В таблице 2 приведены значения расстояния Колмогорова Δ_1, Δ_2 и Δ_3 для распределений вероятностей числа заявок в первом, втором и третьем узлах сети соответственно для разных параметров интенсивности входящего потока $N\lambda$. Из таблицы видно, что при увеличении интенсивности входящего потока, расстояние Колмогорова уменьшается. Если в качестве допустимой погрешности принять $\Delta \leq 0,03$, то предлагаемая аппроксимация может быть применена при $\lambda \geq 20$.

5. Решение задачи оптимизации

Для того чтобы продемонстрировать применение полученного результата, предлагаем рассмотреть задачу оптимизации абонентской сети.

Таблица 1. Параметры СеМО

Вероятность поступления входящей заявки в узел 1, v_1	0,3
Вероятность поступления входящей заявки в узел 2, v_2	0,3
Вероятность поступления входящей заявки в узел 3, v_3	0,4
Интенсивность выхода из состояния 1, α_{10}	0,1
Интенсивность выхода из состояния 2, α_{20}	0,1
Интенсивность выхода из состояния 3, α_{30}	0,1
Интенсивность перехода из состояния 1 в состояние 2, α_{12}	1
Интенсивность перехода из состояния 1 в состояние 3, α_{13}	1
Интенсивность перехода из состояния 2 в состояние 1, α_{21}	0,15
Интенсивность перехода из состояния 2 в состояние 3, α_{23}	1
Интенсивность перехода из состояния 3 в состояние 1, α_{31}	12
Интенсивность перехода из состояния 3 в состояние 2, α_{32}	0,5
Параметр формы гамма-распределения времени обслуживания, α	0,5
Параметр масштаба гамма-распределения времени обслуживания, β	0,1

Таблица 2. Расстояние Колмогорова для разных значений интенсивности входящего потока $N\lambda$

$N\lambda$	1	10	20	50	100
Δ_1	0,0637	0,0590	0,0271	0,0086	0,0030
Δ_2	0,0706	0,0448	0,0230	0,0087	0,0033
Δ_3	0,0616	0,0594	0,0267	0,0097	0,0033

Переходы пользователя от одного узла сети к другому являются случайными, и на перемещение пользователя повлиять невозможно, но можно повлиять на то, как долго пользователь будет находиться в пределах диапазона работы одной БС. Это можно сделать путем изменения радиуса распространения сигнала базовой станции, который зависит от излучаемой мощности антенны [11]. При увеличении мощности антенн БС увеличивается зона покрытия, что позволяет пользователю дольше находиться в пределах радиуса действия БС и не переключаться к другим БС.

Пусть в исследуемой системе все оборудование в узлах сети является однотипным и работает с одинаковой эффективностью.

Тогда можем рассмотреть следующий показатель эффективности сети:

$$(7) \quad \Theta(a) = C_1 \sum_{k=1}^K E\{i_k(a)\} - C_2 \cdot K \cdot a.$$

Здесь a – некоторый коэффициент эффективности работы (мощности) одной БС, который прямо пропорционален стоимости покупки и содержания ее оборудования; $i_k(a)$ – число абонентов, подключенных к k -й БС; C_1 – доход, получаемый за одного подключенного абонента в единицу времени; C_2 – расходы на содержание одной БС сети в единицу времени; K – число БС.

Будем считать, что радиус действия БС увеличивается линейно от мощности ее оборудования a . Тогда матрицу маршрутизации M можно представить в виде

$$M := \frac{1}{a} \cdot M.$$

Значения элементов этой матрицы влияют на число заявок в узлах сети (абонентов, подключенных к базовым станциям) $i_k(a)$.

Ставится задача поиска наибольшего значения показателя (7) при разных значениях мощности оборудования a . Для сети с параметрами, представленными в таблице 1 и значениями стоимостных коэффициентов $C_1 = 4$, $C_2 = 100$ решение найдено численно при $N\lambda = 100$. На рис. 4 показана зависимость показателя $\Theta(a)$ от параметра a . Максимум достигается при $a_{\text{opt}} = 1,3$, при этом показатель эффективности сети составляет $\Theta(a_{\text{opt}}) = 877,8$.

6. Заключение

В работе предложена и исследована сеть массового обслуживания с прерыванием обслуживания для моделирования абонентской сети связи миграцией абонентов. Получена аппроксимация многомерного распределения вероятностей числа заявок в сети и проведены численные эксперименты, которые показывают адекватность полученных теоретических результатов. Установлена область применимости полученной аппроксимации. Приведен пример решения оптимизационной задачи, в котром найдено значение параметра, характеризующего мощность установленного

оборудования, обеспечивающее максимальное значение критерия эффективности сети.

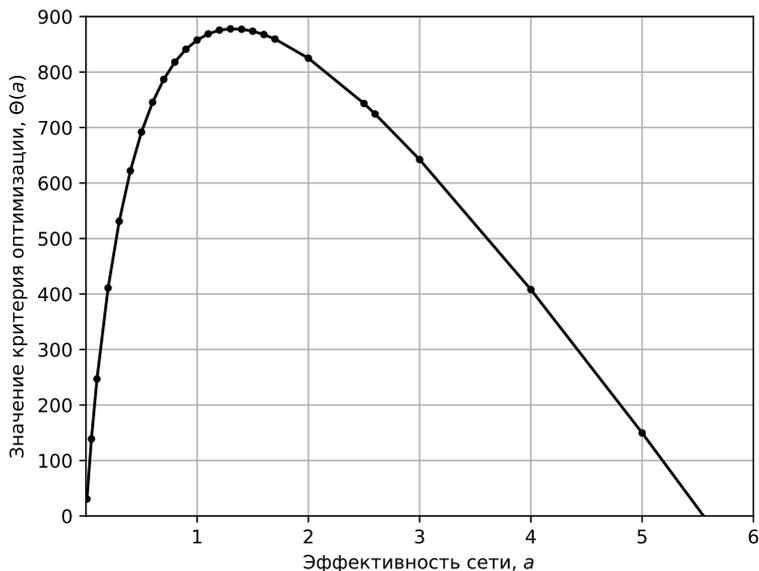


Рис. 4. Зависимость критерия оптимальности от параметра эффективности сети

Литература

1. ВИШНЕВСКИЙ В.М., СЕМЕНОВА О.В. *Обзор моделей систем поллинга и их применение в телекоммуникационных сетях* // Проблемы информатики. — 2020. — Т. 48, №3. — С. 30–59.
2. ГАЙДАМАКА Ю.В., ЗАРИПОВА Э.Р., САМУЙЛОВ К.Е. *Модели обслуживания вызовов в сети сотовой подвижной связи* — М.: Изд-во РУДН, 2008. — 72 с.
3. ГОРБУНОВА А.В., НАУМОВ А.В., ГАЙДАМАКА Ю.В. и др. *Ресурсные системы массового обслуживания с произвольным обслуживанием* // Информатика и её применения. — 2019. — Т. 13, №1. — С. 99–107.

4. ЛИСОВСКАЯ Е.Ю., МОИСЕЕВА С.П. *Асимптотический анализ немарковской бесконечнолинейной системы обслуживания требований случайного объема с входящим рекуррентным потоком* // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. – 2017. – №39. – С. 30–38.
5. МОИСЕЕВ А.Н. *Исследование математических моделей систем и сетей массового обслуживания с высокоинтенсивными непурассоновскими входящими потоками*: дис. ... д-ра физ.-мат. наук. – 2016. – 333 с.
6. МОИСЕЕВ А.Н., НАЗАРОВ А.А. *Бесконечнолинейные системы и сети массового обслуживания*. – Томск: Изд-во НТЛ, 2015. – 240 с.
7. МОИСЕЕВА С.П., НАЗАРОВ А.А. *Метод асимптотического анализа в теории массового обслуживания*. – Томск: Изд-во НТЛ, 2006. – 109 с.
8. НИКОЛАЕВ Д.И., ГАЙДАМАКА Ю.В. *Модель узла IAB в виде поллинговой системы с двумя очередями* // Информационные технологии и математическое моделирование (ИТММ–2023): Материалы XXII Междунар. конф. имени А.Ф. Терпугова, Томск, 4–9 декабря 2023 г. – Томск: Изд-во Томского государственного ун-та, 2023 – Ч. 1. – 2023. – С. 193–198.
9. САЛИМЗЯНОВ Р.Р., МОИСЕЕВ А.Н. *Моделирование абонентской сети, реализованной на основе IAB технологии в виде системы с двумя состояниями и произвольным временем обслуживания* // Сб. трудов научной конф. «Информационно-телекоммуникационные технологии и математическое моделирование высокотехнологичных систем», Москва, 8–12 апреля 2024 г. – Москва, 2024 – С. 77–80.
10. САЛИМЗЯНОВ Р.Р., ЛИСОВСКАЯ Е.Ю. *Модели обслуживания трафика на базовой станции в виде систем массового обслуживания* // Информационные технологии и математическое моделирование (ИТММ–2021): Материалы XX

- Международ. конф. имени А.Ф. Терпугова, Томск, 1–5 декабря 2021 г. — Томск: Изд-во Томского государственного ун-та, 2022 — С. 240–245.
11. CEMIL C., ENDER A. *Energy-Efficient Base Station Deployment in Heterogeneous Networks* // IEEE Wireless Communications Letters. — 2014. — Vol. 3(12). — P. 593–596.
 12. HUANG C., WANG X. *A Bayesian Approach to the Design of Backhauling Topology for 5G IAB Networks* // IEEE Trans. on Mobile Computing. — April 2023. — Vol. 22, No. 4. — P. 1867–1879. — DOI: 10.1109/TMC.2021.3118958.
 13. LEE W. *Mobile cellular telecommunications: analog and digital systems*. — N.Y.: McGraw-Hill, 1995. — 664 p.
 14. RANJAN S., JHA P., KARANDIKAR A. et al. *A Flexible IAB Architecture for Beyond 5G Network* // IEEE Communications Standards Magazine. — September 2023. — Vol. 7, No. 3. — P. 64–71. — DOI: 10.1109/MCOMSTD.0009.2200018.
 15. SALIMZYANOV R., MOISEEV A. *Infinite-Server Queueing System with Two States of Service and Abandonments: Information Technologies and Mathematical Modelling. Queueing Theory and Applications*. — Springer Nature Switzerland: Cham. — P. 134–145.
 16. SOPIN E., MOLTCHANOV D., MASLOV A. et al. *User Persistence in 5G/6G mmWave/Sub-THz Systems With Blockage: Does It Pay Off?* // IEEE Trans. on Wireless Communications. — October 2024. — Vol. 23, No. 10. — P. 14864–14878. — DOI: 10.1109/TWC.2024.3419918.
 17. SHI B., ZHENG F.-C., SHE C. et al. *Risk-Resistant Resource Allocation for eMBB and URLLC Coexistence Under M/G/1 Queueing Model* // IEEE Trans. on Vehicular Technology. — June 2022. — Vol. 71, No. 6. — P. 6279–6290. — DOI: 10.1109/TVT.2022.3159247.
 18. TEZERGIL B., ONUR E. *Wireless Backhaul in 5G and Beyond: Issues, Challenges and Opportunities* // IEEE Communications Surveys & Tutorials. — 2022. — P. 1-1. — DOI: 10.1109/COMST.2022.3203578.

19. YIN H., ROY S., CAO L. *Routing and Resource Allocation for IAB Multi-Hop Network in 5G Advanced* // IEEE Trans. on Communications. – October 2022. – Vol. 70, No. 10. – P. 6704–6717. – DOI: 10.1109/TCOMM.2022.3200673.

INFINITE-SERVER QUEUEING NETWORK WITH ABANDONMENTS

Radmir Salimzyanov, Tomsk State University, Tomsk, Graduate Student (rsalimzyanov@yahoo.com).

Alexander Moiseev, Tomsk State University, Tomsk, Doctor of Science, Head of Department (moiseev.tsu@gmail.com).

Sergey Sushchenko, Tomsk State University, Tomsk, Doctor of Science, Head of Department (ssp.inf.tsu@gmail.com).

Abstract: The paper presents a study of a queueing network with an unlimited number of servers in the nodes and service abandonments. Using such a model, a subscriber access network can be described. In the considered network, a connected subscriber can move from one node of the network to another during the service process or leave the network, having completed or not completed its service. It is assumed that such transitions occur independently of the current state of the nodes. The study is carried out using the method of asymptotic analysis under the condition of high input flow intensity. It is found that in the specified asymptotic regime, the joint stationary probability distribution of the number of subscribers in the network nodes converges to a multi-dimensional Gaussian distribution. Explicit expressions for the parameters of this distribution, including the mean vector and the covariance matrix, are obtained. Numerical experiments are performed to evaluate the accuracy of the approximation, and the domain of applicability of the results is established depending on the model parameters. In addition, an example of solving an optimization problem for the studied queueing network is provided, demonstrating the potential practical application of the proposed model and analytical methods for the analysis and management of telecommunication systems.

Keywords: queueing network, unlimited number of servers, service abandonments, asymptotic analysis.

УДК 519.2

ББК 22.18

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

Поступила в редакцию 03.04.2025.

Дата опубликования 31.05.2025.

МОДЕЛИ ЭЛЕКТРОННОГО ДОКУМЕНТА В РАМКАХ ПАРАДИГМЫ «ИНДУСТРИЯ 4.0»

Тали Д. И.¹, Финько О. А.²

(Краснодарское высшее военное училище, Краснодар)

Цифровая трансформация общества и производства постепенно вступает в стадию Индустрии 4.0, когда обслуживающая инфраструктура, в том числе и автоматизированные информационные системы юридически значимого электронного документооборота, будут способны адаптироваться под текущие условия обстановки. Однако современный этап развития таких систем характеризуется несовершенством терминологического аппарата и отождествлением объекта воздействия (электронного документа) с его бумажным прототипом, что приводит к ограничению их функциональных возможностей. Таким образом, существующая парадигма восприятия электронного документа не соответствует перспективному уровню развития автоматизированных информационных систем юридически значимого электронного документооборота в рамках концепции Индустрии 4.0. Цель работы – построение моделей электронного документа и описание метода обеспечения его целостности, позволяющего гарантировать достижение требуемого уровня защищенности автоматизированных информационных систем юридически значимого электронного документооборота, функционирующих в соответствии с положениями концепции Индустрии 4.0. Используемые методы: применение криптографических хэши-функций к компонентам (контенту и метаданным) электронного документа, результаты которых вычисляются по типу бинарного дерева Меркла в сочетании с технологией цепной записи данных. Результаты исследования: введено понятие «интегративный электронный документ», предложена новая структура электронного документа, что позволило сократить длительность его верификации и, как следствие, снизить риск реализации деструктивных воздействий на автоматизированную информационную систему.

Ключевые слова: Индустрия 4.0, контент, метаданные, управление электронными документами, целостность электронного документа.

1. Введение

Цифровые технологии с каждым годом совершенствуются и интегрируются во все сферы жизни, вызывая трансформацию общества и глобальной экономики, что становится завершаю-

¹ Дмитрий Иосифович Тали, к.т.н. (dimatali@mail.ru).

² Олег Анатольевич Финько, д.т.н., профессор (ofinko@yandex.ru).

щим этапом третьей промышленной революции. Подобные тенденции являются катализатором зарождения очередного этапа развития науки и техники, характеризующегося использованием новых подходов, позволяющих обеспечить эффективность государственного управления [25].

Цифровая трансформация связана с полномасштабным применением различных видов информационных систем и информационных ресурсов, предполагающих массовое использование электронных документов (ЭлД) и автоматизированных информационных систем юридически значимого электронного документооборота (АИС ЮЗ ЭД). Применение подобных технологий является одним из наиболее результативных механизмов комплексного решения задач повышения эффективности управления информацией и документацией [14].

Однако существующая нормативная и организационная база для формирования полноценного ЮЗ ЭД, основанного на принципах «Индустрии 4.0» (распределенность, виртуализация, интеллектуализация и т.д.), в настоящее время не соответствует требованиям, предъявляемым современным этапом технологического развития. Данное обстоятельство приводит к ряду трудностей, выражающихся в дублировании электронных и бумажных технологий, препятствующих переходу к безбумажному цифровому взаимодействию [13].

Подтверждением выше сказанному является неопределённость толкования Регулятором основных определений, используемых в данной предметной области. Так, под термином «электронный документ» понимается документированная информация, созданная, полученная и сохраняемая организацией или частным лицом в качестве доказательства и актива для подтверждения правовых обязательств или деловой транзакции [8]. При этом определение понятия «документ» аналогично указанному с той лишь разницей, что акцентируется внимание на фиксации информации на материальном носителе [5]. Однако состав ЭлД, в отличие от бумажного, определяется как совокупность двух компонентов: контента и метаданных, описывающих контекст, контент и структуру документов, а также управление ими в течение времени. Вместе с тем защита ЭлД, обеспечиваемая

подсистемой защиты информации АИС ЮЗ ЭД, сводится к защите метаданных посредством разграничения прав доступа и выполнению соответствующих правил, исключая криптографические средства защиты в виде электронной подписи (ЭП), которыми обеспечивается защищенность контента [8].

Исходя из этого при сложившейся структуре ЭД его целостность имеет функциональную зависимость от двух компонент (контент и метаданные), имеющих между собой только логическую связь, что в условиях преднамеренных или непреднамеренных деструктивных воздействий уполномоченных пользователей (инсайдеров) не позволит обеспечить целостность ЭД при полной или частичной утрате одного из них (рис. 1) [4, 23].

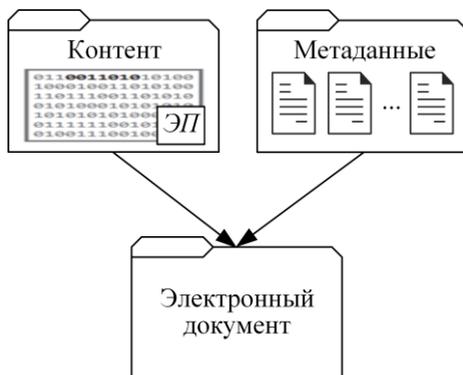


Рис. 1. Существующая схема синтеза ЭД

Архаичность принятого в настоящее время представления о структуре ЭД приводит к нарушению ряда требований, предъявляемых к ним в период жизненного цикла (рис. 2), выражающихся в:

- 1) невозможности обеспечения условий, исключающих несанкционированную модификацию ЭД;
- 2) получении доступа к содержанию ЭД для документов или их частей, не подлежащих автоматическому раскрытию государственным органам или третьим сторонам;
- 3) невозможности обеспечения доступности ЭД в течение всего срока хранения [13].

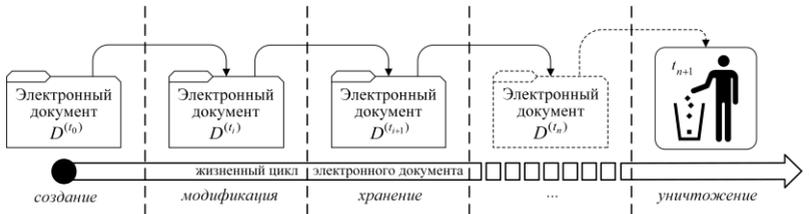


Рис. 2. Схема жизненного цикла ЭД

Эти обстоятельства приводят к возникновению затруднений в проектировании документных систем ввиду снижения уровня их защищенности при подобной организации структуры ЭД. Кроме того, повсеместное внедрение технологий, основывающихся на принципах Индустрии 4.0, заставляет совершенствовать и модернизировать такие системы до уровня систем обслуживания информации, характеризующихся централизацией обслуживаемых процессов и масштабируемостью обрабатываемой информации. В этих условиях основной задачей становится обеспечение целостности не столько ЭД, сколько самой системы (инфраструктуры), обслуживающей циркулирующую в ней информацию.

Таким образом, возникает необходимость модификации не только самой структуры ЭД, но и подходов к обеспечению его целостности, в рамках создания перспективных образцов АИС ЮЗ ЭД, функционирующих на основе принципов новой концепции.

2. Обзор существующих методов и средств контроля целостности данных

Выполненный анализ существующих методов и средств контроля целостности данных позволил установить, что в их основе лежит применение криптографической хэш-функции (ХФ) [6]. Применение подобного механизма позволит реализовать криптографическую связь компонент ЭД и, как следствие, обеспечить формирование его единой криптографической структуры.

Наибольший интерес для решения такой задачи представляют метод «однократной записи» [27], способ аутентификации сообщений на основе НМАС (hash-based message authentication) [26], а также техническое решение [18].

Недостатком метода [27] является отсутствие возможности установления дополнительных параметров ввода ключевых данных различных пользователей и порядка их применения (секретных для самих пользователей или для различных групп пользователей, состав которых самим пользователям не известен), позволяющих обеспечить соответствующий уровень защищенности записей данных в файле.

В способе [26] полученный код аутентичности позволяет убедиться в том, что данные не изменялись с тех пор, как они были созданы, переданы или сохранены доверенным источником. Для такого рода проверки необходимо, чтобы две доверяющие друг другу стороны заранее договорились об использовании секретного ключа, который известен только им, что и является ограничением. Тем самым гарантируется аутентичность источника и сообщения.

Техническое решение [18] позволяет обеспечить необходимый уровень защиты записей данных в файле на основе задания соответствующих параметров метода «однократной записи»: степени вложенности в блок данных, количество используемых внешних ключей. При этом для контроля целостности данных могут быть использованы как бесключевые, так и ключевые ХФ/ЭП.

Недостатком данного технического решения является относительно высокая сложность операции контроля целостности записей данных, так как для ее реализации требуется использование всего множества ключей на всех этапах преобразования данных, что представляет существенное ограничение и техническую сложность для практического использования.

При этом, несмотря на несомненные достоинства вышеперечисленных решений, у них имеется общий недостаток, выражающийся в отсутствии технической возможности локализации записей данных с признаками нарушения целостности при реализации деструктивных воздействий уполномоченными пользователями АИС ЮЗ ЭД.

Наиболее подходящим решением является способ [19], позволяющий повысить уровень защищенности метаданных Элд, обрабатываемых АИС ЮЗ ЭД, с возможностью контроля их целостности. Основным преимуществом этого решения является реализация функции обнаружения и локализации несанкционированно модифицированных записей метаданных Элд, в случае нарушения их целостности уполномоченными пользователями (инсайдерами). Однако его недостатком является отсутствие функциональной возможности обеспечения «интегративной целостности Элд» (свойство целостного системного объекта, возникающее в результате взаимодействия его компонент и не присущее ни одной из них в отдельности [3, 12, 17, 23]), достигаемой за счет криптографической связности контента и метаданных, входящих в его состав, что в условиях преднамеренных или непреднамеренных деструктивных воздействий уполномоченных пользователей (инсайдеров) не обеспечивает целостности Элд в течение всего жизненного цикла при полной или частичной утрате одной из его компонент.

Перечисленные способы [18, 19, 26, 27] основаны на общих принципах построения цепной записи данных, однако в отличие от известных вариантов реализации технологии «блокчейн» [28–31] носят универсальный характер, что позволяет использовать их в том числе и в АИС ЮЗ ЭД с низкой интенсивностью проводимых транзакций ввиду отсутствия высоких вычислительных мощностей и малым количеством рабочих мест, входящих в общую сеть [22].

3. Постановка задачи

Воспользуемся теоретико-множественным подходом, получившим широкое распространение при описании различных технических систем [10, 16].

Пусть задана модель подсистемы ЗИ АИС ЮЗ ЭД: $Z = \{T, X, Y, D, R, C\}$, где

T – множество моментов времени в которые наблюдается подсистема ЗИ АИС ЮЗ ЭД, $t_i \in T$, $i = 0, 1, \dots, n$; t_0 – начало эксплуатации АИС ЮЗ ЭД, t_n – конец эксплуатации АИС ЮЗ ЭД.

X – множество входных инициирующих событий (воздействий) (ИнС) на подсистему ЗИ АИС ЮЗ ЭД (например: ошибки персонала АИС ЮЗ ЭД/действия злоумышленников, отказы технических компонентов АИС ЮЗ ЭД, нарушения технологии обработки ЭЛД, истечение сроков действия ключей ЭП и т.д.).

Y – множество функциональных состояний подсистемы ЗИ АИС ЮЗ ЭД в результате влияния ИнС (например: функционирование в штатном режиме; функционирование в режиме несанкционированного доступа к ресурсам АИС ЮЗ ЭД внутренним/внешним нарушителем; функционирование в режиме воздействия вредоносного программного обеспечения (ВПО), сбой в работе АИС ЮЗ ЭД и т.д.).

D – множество ЭЛД, изменяющихся под влиянием ИнС на подсистему ЗИ АИС ЮЗ ЭД.

R – множество выходных качественных состояний, обрабатываемых ЭЛД, причем

$$(1) \quad R = \left\{ R_{(+)}^{(x_j)}, R_{(-)}^{(x_j)} \right\},$$

где $R_{(+)}^{(x_j)}$ – состояние ЭЛД, при котором его целостность под влиянием ИнС $x_j \in X$ обеспечена, $R_{(-)}^{(x_j)}$ – состояние ЭЛД, при котором под влиянием ИнС $x_j \in X$ он утрачивает целостность или обеспечение указанного свойства не доступно в текущий момент функционирования АИС ЮЗ ЭД.

S – множество сценариев функционирования подсистемы ЗИ АИС ЮЗ ЭД под влиянием ИнС $x_j \in X$.

Исходя из определения [9] признаком «целостности системы» является реализация ею своего целевого предназначения. Целевым предназначением рассматриваемых систем является обеспечение целостности ЭЛД. Из чего следует, что нахождение ЭЛД в состоянии $R_{(+)}^{(x_j)}$ характеризует состояние целостности АИС ЮЗ ЭД, а переход ЭЛД в состояние $R_{(-)}^{(x_j)}$ – состояние нарушения целостности АИС ЮЗ ЭД.

Таким образом, справедливо утверждать, что в первом случае такое развитие событий приведет к нормальному сценарию

(штатному режиму) функционирования АИС ЮЗ ЭД, а во втором – к его нарушению. Однако ЭлД $D^{(t_i)}$ в состоянии момента времени t_i представляет из себя совокупность контента $K^{(t_i)}$ и подмножества метаданных $M^{(t_i)}$ [8], что может быть записано как

$$(2) \quad D^{(t_i)} = \{K^{(t_i)}, M^{(t_i)}\},$$

причем $M^{(t_i)} = \{m_{i1}, m_{i2}, m_{i3}, \dots, m_{in}\}$, где $m_{i1}, m_{i2}, m_{i3}, \dots, m_{in}$ – реквизиты ЭлД (автор, дата создания, гриф и т.д.).

Таким образом, всякое состояние R ЭлД $D^{(t_i)}$ характеризуется в каждый момент времени $t_i \in T$ состоянием контента $K^{(t_i)}$ и состоянием метаданных $M^{(t_i)}$, что может быть записано как $K^{(t_i)} = \{k_{(+)}^{(x_j)}, k_{(-)}^{(x_j)}\}$ и $M^{(t_i)} = \{m_{(+)}^{(x_j)}, m_{(-)}^{(x_j)}\}$, изменяющихся под влиянием ИнС $x_j \in X$ на подсистему ЗИ АИС ЮЗ ЭД, где $k_{(+)}^{(x_j)}$ – состояние ЭлД $D^{(t_i)}$, при котором в ходе реализации ИнС x_j целостность его контента не нарушена («1»), и $k_{(-)}^{(x_j)}$ – состояние ЭлД $D^{(t_i)}$, при котором целостность его контента нарушена («0»), $m_{(+)}^{(x_j)}$ – состояние ЭлД $D^{(t_i)}$, при котором в ходе реализации ИнС x_j целостность его метаданных не нарушена («1»), и $m_{(-)}^{(x_j)}$ – состояние ЭлД $D^{(t_i)}$, при котором целостность его метаданных нарушена («0»).

Так как ЭлД $D^{(t_i)}$ имеет функциональную зависимость от двух компонент – контента $K^{(t_i)}$ и метаданных $M^{(t_i)}$, нахождение подсистемы ЗИ АИС ЮЗ ЭД в состоянии («0») (нарушение функционирования) увеличивается до трех вариантов: (0, 0), (0, 1), (1, 0) (на рис. 3 отмечены пунктиром).

В таком случае требуется минимизировать количество состояний АИС ЮЗ ЭД с признаками нарушения целостности:

$$(3) \quad Y(R(X)) \rightarrow \min |C| \quad \Big| \quad t_i \in T, \quad i = \overline{0, n},$$

где $|C|$ – количество (мощность множества) сценариев нарушения функционирования подсистемы ЗИ АИС ЮЗ ЭД.

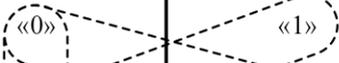
Компоненты Элд $D^{(t_i)}$	Состояния функционирования АИС ЮЗ ЭД	
контент $K^{(t_i)}$		
метаданные $M^{(t_i)}$		

Рис. 3. Варианты сценариев нарушения функционирования АИС ЮЗ ЭД

Достичь результата возможно посредством разработки метода обеспечения целостности Элд за счет связности его компонент. Выполнение критерия качества функционирования подсистемы ЗИ АИС ЮЗ ЭД определяется неравенством

$$(4) \quad |C_{\text{разр.}}| < |C_{\text{сущ.}}|,$$

где $|C_{\text{разр.}}|$ и $|C_{\text{сущ.}}|$ – количество сценариев нарушения функционирования подсистемы ЗИ АИС ЮЗ ЭД посредством нарушения целостности компонент $K^{(t_i)}$ и $M^{(t_i)}$ при использовании разработанной структуры Элд $D^{(t_i)}$ и при использовании существующей структуры Элд $D^{(t_i)}$ соответственно.

При этом в целях контроля целостности Элд, обрабатываемых АИС ЮЗ ЭД, разрабатываемый метод не должен уступать существующим по таким параметрам, как время вычисления $t_{\text{выч.}}$ (наиболее сложной операции ХФ/ЭП), а также время локализации $t_{\text{лок.}}$ записей данных (контента и/или метаданных Элд) с признаками нарушения целостности.

В таком случае в качестве преимущества разработанного метода требуется выполнение следующих условий:

$$(5) \quad t_{\text{выч.разр.}} \leq t_{\text{выч.прототипа}};$$

$$(6) \quad t_{\text{лок.разр.}} \leq t_{\text{лок.прототипа}},$$

где $t_{\text{выч.прототипа}}$, $t_{\text{лок.прототипа}}$, – время, затрачиваемое на вычисление наиболее сложной операции ХФ/ЭП и локализацию записей данных (контента и/или метаданных Элд) с признаками нару-

шения целостности при использовании известных методов контроля целостности данных.

4. Обеспечение целостности электронных документов на основе бинарного дерева Меркла

Предлагаемое решение основывается на таких принципах системного анализа, как принцип целостности и принцип интегративности. Под первым предполагается рассмотрение исследуемого объекта в единстве его взаимодействующих частей, а под вторым – изучение свойств электронного документа как целостного системного объекта, возникающих в результате взаимодействия его компонент и не присутствующих ни одной из компонент в отдельности [1, 2, 10, 23].

В целях реализации этих принципов разработан соответствующий метод, содержащий совокупность правил формирования и верификации состояний Элд и реализующие эти правила формальные модели.

4.1. КОНЦЕПТУАЛЬНАЯ МОДЕЛЬ АИС ЮЗ ЭД, УЧИТЫВАЮЩАЯ ПРИНЦИПЫ ЦЕЛОСТНОСТИ И ИНТЕГРАТИВНОСТИ ЭЛД

Анализ публикаций в области теории систем и системного анализа [3, 11, 12, 17] позволяет сделать вывод о том, что АИС ЮЗ ЭД может рассматриваться как среда существования Элд, а сам Элд как системный объект, состоящий из двух компонент: контента и метаданных (2). Причем их функциональная зависимость очевидна: без создания контента невозможно формирование метаданных, в свою очередь без метаданных невозможно управление Элд, в то время как в среде АИС ЮЗ ЭД эти элементы в настоящее время имеют разную степень защиты.

Концептуальная модель АИС ЮЗ ЭД, соответствующая принципам целостности и интегративности, поясняется с помощью рис. 4. При этом предлагается достичь интегративности Элд за счет использования стандартных криптографических протоколов.

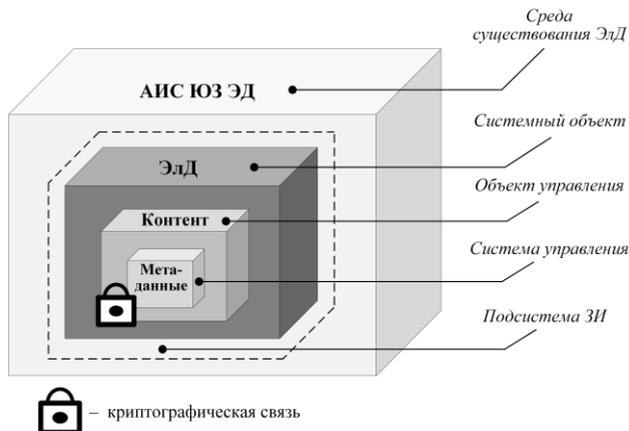


Рис. 4. Пояснение концептуальной модели АИС ЮЗ ЭД, учитывающей принципы целостности и интегративности ЭД

4.2. ФОРМАЛЬНЫЕ МОДЕЛИ ОБЕСПЕЧЕНИЯ ЦЕЛОСТНОСТИ ЭД НА ОСНОВЕ БИНАРНОГО ДЕРЕВА МЕРКЛА

Применим криптографические ХФ к компонентам ЭД (контенту и метаданным), результаты которых вычисляются по типу бинарного дерева Меркла в сочетании с технологией цепной записи данных [27–31]. Подобное решение позволит повысить уровень защищенности ЭД, обрабатываемых АИС ЮЗ ЭД, за счет обнаружения и локализации контента и записей метаданных, подвергшихся несанкционированной модификации со стороны уполномоченных пользователей (инсайдеров).

Метод обеспечения целостности ЭД на основе бинарного дерева Меркла можно представить следующим образом. Модель формирования ЭД (с модифицированной структурой): множество ключей V_U разбивается на два подмножества: $V_U \in \{V_U^{(1)}, V_U^{(2)}\}$ и $V_U^* \in \{V_U^{*(1)}, V_U^{*(2)}\}$, при этом каждая часть подмножеств содержит ключи $V_U^{(q)} \in \{v_1^{(q)}, v_2^{(q)}, \dots, v_\zeta^{(q)}\}$ и $V_U^{*(q)} \in \{v_1^{*(q)}, v_2^{*(q)}, \dots, v_\zeta^{*(q)}\}$, где $q=1, 2$ – принадлежность ключей соответствующей части подмножеств, для всех $U=1, \dots, \zeta$ – номер ключа в соответствующей части подмножеств.

В рассматриваемом варианте реализации данного метода ключи $V_U^{(1)}$ являются ключами исполнителя (автора) электронного документа, ключи $V_U^{(2)}$ – внутренними системными ключами.

В момент формирования t_0 ЭЛД в памяти системы обработки данных образуется информационный блок $D^{(t_0)} = \{K^{(t_0)}, M^{(t_0)}\}$, представляющий собой ЭЛД, создаваемый подмножеством контента $K^{(t_0)}$ и подмножеством записей статических метаданных $M^{(t_0)} = \{m_{01}, m_{02}, m_{03}, \dots, m_{0n}\}$, над которыми одновременно производятся операции криптографического преобразования:

$$\begin{aligned} H_0 &= f^{(v_x^{(1)})}(K^{(t_0)}) \text{ на ключах } v_x^{(1)} \in V_U^{(1)}, \\ h_{01} &= f^{(v_x^{(2)})}(m_{01}) \text{ на ключах } v_x^{(2)} \in V_U^{(2)}, \\ h_{02} &= f^{(v_x^{(2)})}(m_{02}) \text{ на ключах } v_x^{(2)} \in V_U^{(2)}, \\ h_{03} &= f^{(v_x^{(2)})}(m_{03}) \text{ на ключах } v_x^{(2)} \in V_U^{(2)}, \\ &\dots, \\ h_{0n} &= f^{(v_x^{(2)})}(m_{0n}) \text{ на ключах } v_x^{(2)} \in V_U^{(2)}. \end{aligned}$$

Над вычисленными значениями сигнатур $h_{01}, h_{02}, h_{03}, \dots, h_{0n}$ записей статических метаданных производятся двухместные операции конкатенации: $h_{01}||h_{02}, h_{03}||h_{0n}$, над результатами которых выполняются операции криптографического преобразования на ключах $v_x^{(2)} \in V_U^{(2)}$:

$$\begin{aligned} h_{0102} &= f^{(v_x^{(2)})}(h_{01} | h_{02}), \\ h_{030n} &= f^{(v_x^{(2)})}(h_{03} | h_{0n}). \end{aligned}$$

После чего над вновь вычисленными сигнатурами последовательно выполняются операции конкатенации и криптографического преобразования на ключах $v_x^{(2)} \in V_U^{(2)}$:

$$h_0^\Sigma = f^{(v_x^{(2)})}(h_{0102} || h_{030n}),$$

результатом которой является формирование корневой сигнатуры h_0^Σ записей статических метаданных (рис. 5).

Затем операции конкатенации и криптографического пре-

образования на ключах $v_x^{(2)} \in V_U^{(2)}$ производятся над значением сигнатуры H_0 контента и корневой сигнатурой h_0^Σ записей статических метаданных, в целях получения общей корневой сигнатуры ЭлД, созданного в момент времени t_0 :

$$S_0 = f^{(v_x^{(2)})}(H_0 | h_0^\Sigma)$$

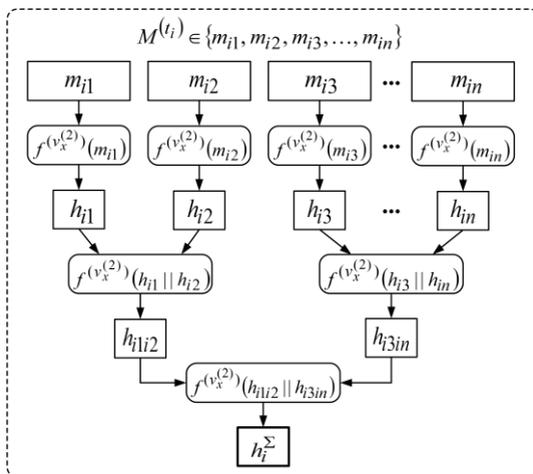


Рис. 5. Диаграмма функциональной модели формирования корневой сигнатуры записей метаданных ЭлД

После произведенных операций ЭлД вводится в систему, причем во время его жизненного цикла t_1, t_2, \dots, t_n при выполнении записи и редактирования повторно производятся вышеописанные операции.

При этом общая корневая сигнатура S_1 ЭлД, вычисленная в момент времени t_1 , конкатенирует с общей корневой сигнатурой S_0 ЭлД, вычисленной в момент времени t_0 , с одновременным криптографическим преобразованием на ключах $v_x^{(2)} \in V_U^{(2)}$, тем самым образуя сигнатуру жизненного цикла S_{01} ЭлД:

$$S_{01} = f^{(v_x^{(2)})}(S_0 | S_1).$$

По мере изменения ЭлД операции повторяются в аналогичном порядке в соответствии с моментами времени t_3, t_4, \dots, t_n его

жизненного цикла, что формирует его модифицированную структуру (далее ЭлД с модифицированной структурой будем называть «интегративным»).

Введем понятие «интегративный электронный документ» (ИЭлД) – информационный объект, состоящий из взаимодействующих составных компонентов (контента и метаданных), возникающий в результате реализации механизма их криптографической связности, и характеризующийся целостностью [7, п.5.9]. Иллюстрированный пример ИЭлД представлен на рис. 6.

Контент $K^{(t_i)}$ и вычисленные значения его сигнатур H_i , записи метаданных $m_{i1}, m_{i2}, m_{i3}, \dots, m_{in}$, их сигнатуры $h_{i1}, h_{i2}, h_{i3}, \dots, h_{in}$, общие сигнатуры h_{i1i2}, h_{i3in} записей метаданных, корневые сигнатуры h_i^Σ записей метаданных, общие корневые сигнатуры S_i , а также корневые сигнатуры S_{in} его жизненного цикла сохраняются в памяти системы обработки данных и используются для контроля целостности ИЭлД в моменты времени его жизненного цикла $t_0, t_1, t_2, \dots, t_n$.

Модель верификации ИЭлД: контроль целостности осуществляется на основе извлечения из памяти системы обработки данных контента $K^{*(t_i)}$ и вычисленных значений его сигнатур H_i^* , записей метаданных $m_{i1}^*, m_{i2}^*, m_{i3}^*, \dots, m_{in}^*$ и их сигнатур $h_{i1}^*, h_{i2}^*, h_{i3}^*, \dots, h_{in}^*$, общих сигнатур h_{i1i2}^*, h_{i3in}^* записей метаданных, корневых сигнатур $h_i^{\Sigma*}$ записей метаданных, общих корневых сигнатур S_i^* , а также корневых сигнатур S_{in}^* его жизненного цикла, прошедших процедуру хранения и подлежащих контролю целостности.

После чего выполняются повторные операции криптографического преобразования над извлеченным контентом $K^{*(t_i)}$ и записями метаданных $m_{i1}^*, m_{i2}^*, m_{i3}^*, \dots, m_{in}^*$. В результате чего вновь вычисленные сигнатуры

$$H_i^{**}; h_{i1}^{**}, h_{i2}^{**}, h_{i3}^{**}, \dots, h_{in}^{**}; h_{i1i2}^{**}, h_{i3in}^{**}; h_i^{\Sigma**}; S_i^{**}; S_{in}^{**}$$

попарно сравниваются с ранее извлеченными:

$$H_i^*; h_{i1}^*, h_{i2}^*, h_{i3}^*, \dots, h_{in}^*; h_{i1i2}^*, h_{i3in}^*; h_i^{\Sigma*}; S_i^*; S_{in}^*.$$

Заключение об отсутствии нарушения целостности делается при выполнении равенств: $S_{in}^{**} = S_{in}^*$; $S_i^{**} = S_i^*$; $H_i^{**} = H_i^*$; $h_i^{\Sigma**} = h_i^{\Sigma*}$; $h_{i1i2}^{**} = h_{i1i2}^*$, $h_{i3in}^{**} = h_{i3in}^*$; $h_{i1}^{**} = h_{i1}^*$, $h_{i2}^{**} = h_{i2}^*$, $h_{i3}^{**} = h_{i3}^*$, ..., $h_{in}^{**} = h_{in}^*$, в противном случае делается заключение о нарушении целостности для соответствующих номеров сигнатур (рис. 7).

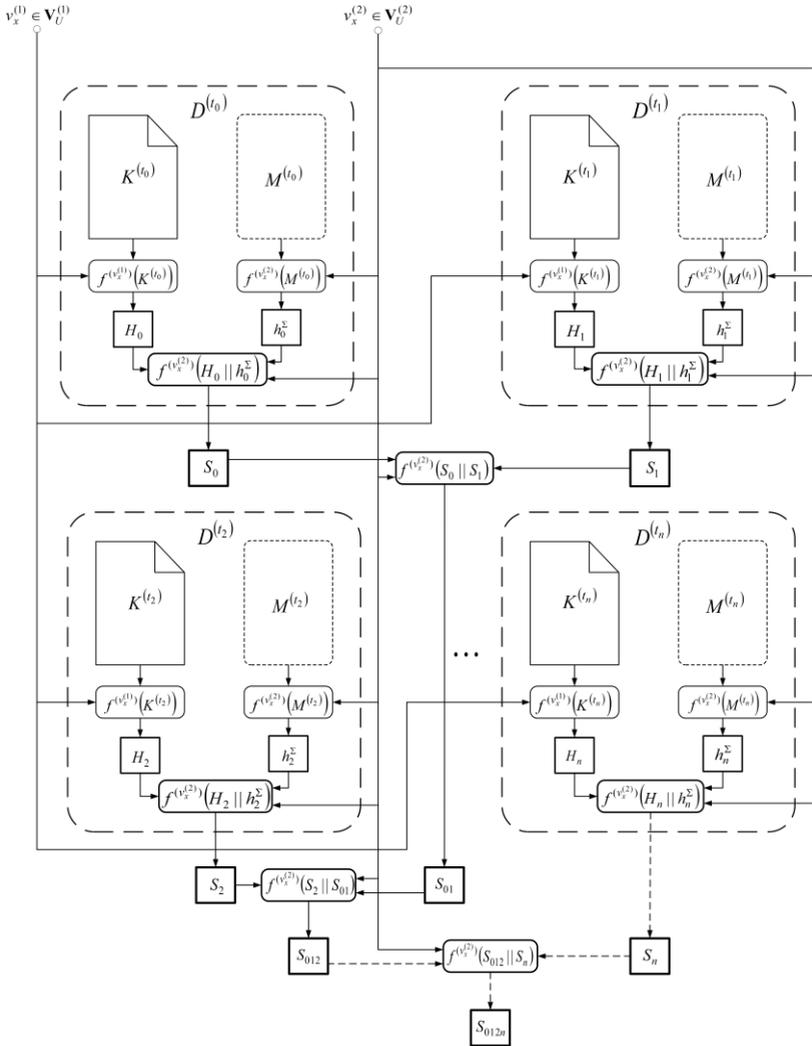


Рис. 6. Диаграмма функциональной модели формирования ИЭЛД

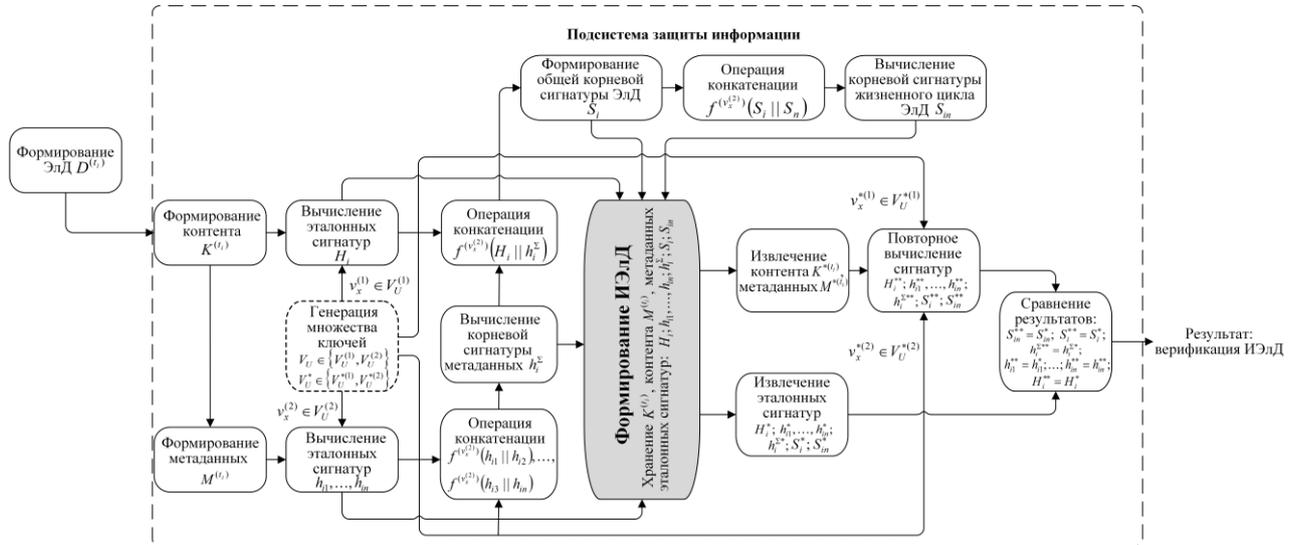


Рис. 7. Функциональная модель формирования и верификации ИЭлд

В качестве криптографического преобразования $f^{(v_x^{(1)})}(K^{(t_i)})$, $f^{(v_x^{(2)})}(m_{ij})$ используется либо ключевая ХФ, при этом $V_U = V_U^*$ (равные множества), либо ЭП, при этом $V_U \neq V_U^*$ (неравные множества).

4.3. ПРИКЛАДНЫЕ СРЕДСТВА РЕАЛИЗАЦИИ МЕТОДА ОБЕСПЕЧЕНИЯ ЦЕЛОСТНОСТИ ЭЛД НА ОСНОВЕ БИНАРНОГО ДЕРЕВА МЕРКЛА

Процесс обеспечения целостности ЭлД на основе бинарного дерева Меркла включает в себя процедуры формирования ИЭлД и верификации его состояний. В целях практического использования разработанного метода сформированы соответствующие правила, реализующие данные процедуры, представленные в виде схем и пошагового описания, выполняемых в них операций.

Так, правило формирования ИЭлД можно сформулировать следующим образом (рис. 8):

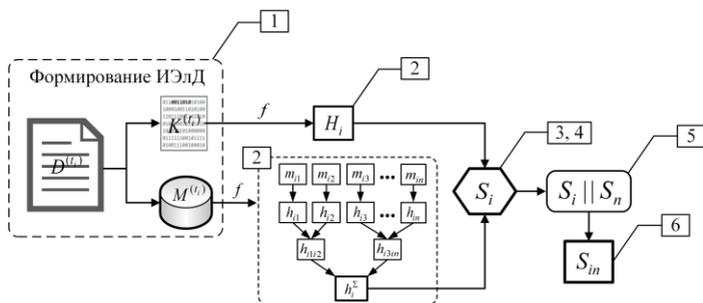


Рис. 8. Схема формирования ИЭлД

Шаг 1. Формирование $D^{(t_i)}$ посредством создания контента $K^{(t_i)}$ и соответствующих ему записей метаданных $M^{(t_i)}$.

Шаг 2. Вычисление сигнатуры H_i контента $K^{(t_i)}$ и корневой сигнатуры записей метаданных h_i^Σ .

Шаг 3. Вычисление общей корневой сигнатуры S_i (рис. 8 – гексагон).

Шаг 4. Повтор произведенных операций для каждого последующего момента времени.

Шаг 5. Выполнение операции конкатенации общей корневой сигнатуры S_i с полученной в каждый последующий момент времени.

Шаг 6. Вычисление сигнатуры S_{in} его жизненного цикла.

В таком случае правило верификации ИЭЛД представляет собой следующую последовательность выполнения операций (рис. 9).

Шаг 1. Формирование $D^{(t_i)}$, заключающееся в повторном вычислении сигнатур $H_i^{**}; h_{i1}^{**}, h_{i2}^{**}, h_{i3}^{**}, \dots, h_{in}^{**}; h_{i1i2}^{**}, h_{i3in}^{**}; h_i^{\Sigma**}; S_i^{**}; S_{in}^{**}$.

Шаг 2. Извлечение эталонных сигнатур (прошедших процедуру хранения) $H_i^*; h_{i1}^*, h_{i2}^*, h_{i3}^*, \dots, h_{in}^*; h_{i1i2}^*, h_{i3in}^*; h_i^{\Sigma*}; S_i^*; S_{in}^*$.

Шаг 3. Сравнение результатов: $S_{in}^{**} = S_{in}^*; S_i^{**} = S_i^*; H_i^{**} = H_i^*; h_i^{\Sigma**} = h_i^{\Sigma*}; h_{i1i2}^{**} = h_{i1i2}^*, h_{i3in}^{**} = h_{i3in}^*; h_{i1}^{**} = h_{i1}^*, h_{i2}^{**} = h_{i2}^*, h_{i3}^{**} = h_{i3}^*, \dots, h_{in}^{**} = h_{in}^*$.

Шаг 4. Заключение об отсутствии (нарушении) целостности (результат: верификация ИЭЛД).

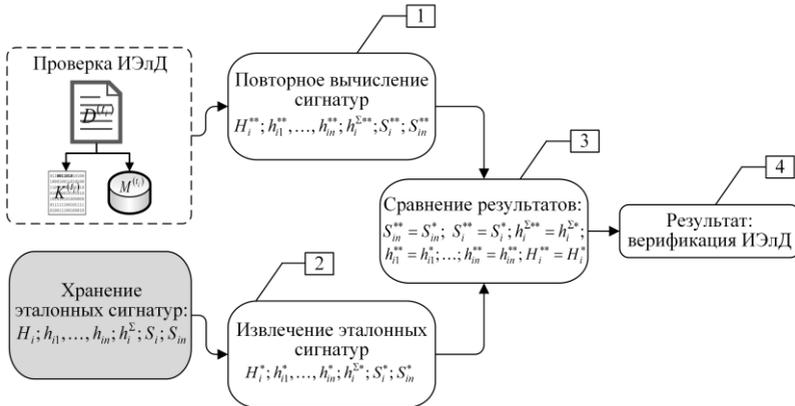


Рис. 9. Схема верификации ИЭЛД

Таким образом, указанные правила позволяют создать алгоритмы и программное обеспечение, реализующее метод

обеспечения целостности электронных документов на основе бинарного дерева Меркла.

5. Результаты исследования разработанного метода

Исходя из постановки задачи, а также учитывая существующую структуру ЭлД (2), процесс функционирования АИС ЮЗ ЭД может быть описан следующей моделью: в любой произвольный момент времени t_i , в процессе обработки ЭлД, подсистема ЗИ АИС ЮЗ ЭД может находиться в одном из четырех состояний по отношению к обрабатываемому ЭлД $D^{(t_i)}$ при возникновении ИнС $x_j \in X$ в АИС ЮЗ ЭД [21]:

$$(7) \quad R_{(+)}^{(x_j)}(t_i) = D^{(t_i)}(k_{(+)}^{(x_j)}, m_{(+)}^{(x_j)}) \rightarrow 1,$$

где $R_{(+)}^{(x_j)}(t_i)$ – состояние ЭлД $D^{(t_i)}$, при котором в ходе реализации ИнС x_j целостность его компонентов не нарушена («1»); символ « \rightarrow » далее по тексту читать как «отсюда следует».

$$(8) \quad R_{(-)}^{(x_j)}(t_i) = D^{(t_i)}(k_{(+)}^{(x_j)}, m_{(-)}^{(x_j)}) \rightarrow 0,$$

где $R_{(-)}^{(x_j)}(t_i)$ – состояние ЭлД $D^{(t_i)}$, при котором в ходе реализации ИнС x_j целостность его метаданных нарушена («0»).

$$(9) \quad R_{(-)}^{(x_j)}(t_i) = D^{(t_i)}(k_{(-)}^{(x_j)}, m_{(+)}^{(x_j)}) \rightarrow 0,$$

где $R_{(-)}^{(x_j)}(t_i)$ – состояние ЭлД $D^{(t_i)}$, при котором в ходе реализации ИнС x_j целостность его контента нарушена («0»);

$$(10) \quad R_{(-)}^{(x_j)}(t_i) = D^{(t_i)}(k_{(-)}^{(x_j)}, m_{(-)}^{(x_j)}) \rightarrow 0,$$

где $R_{(-)}^{(x_j)}(t_i)$ – состояние ЭлД $D^{(t_i)}$, при котором в ходе реализации ИнС x_j целостность его компонентов нарушена («0»).

Переход ЭлД в состояние «0» рассматривается как событие, нарушающее состояние его целостности и, как следствие, приводящее к нарушению функционирования АИС ЮЗ ЭД. При этом насчитывается до трех таких сценариев с учетом существующей структуры ЭлД (8), (9), (10) (рис. 10б, в, г).

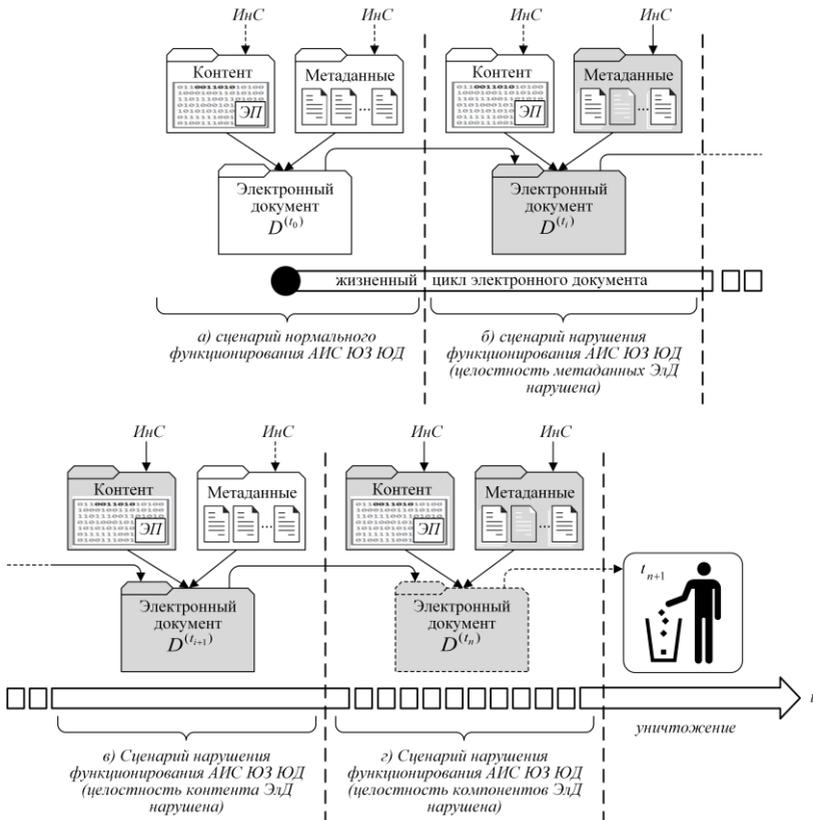


Рис. 10. Сценарии перехода АИС ЮЗ ЭД в различные режимы функционирования в процессе жизненного цикла ЭД

В качестве практического примера можно привести ситуацию, при которой подсистема ЗИ АИС ЮЗ ЭД успешно нейтрализовала ИнС $x_j \in X$, вызванные воздействием ВПО/действиями злоумышленников, чем обеспечила целостность компонентов обрабатываемого ЭД и, как следствие, штатный режим функционирования АИС ЮЗ ЭД (рис. 10а).

В противном случае те же ИнС $x_j \in X$ вызывают нарушение целостности метаданных при защищенном контенте (например, посредством уничтожения базы данных, в которой осуществляется их хранение, что приведет к невозможности нахождения,

обрабатываемого Элд) и, как следствие, самого Элд, что вызовет переход АИС ЮЗ ЭД в режим нарушения функционирования (рис. 10б).

Другим вариантом реализации ИнС $x_j \in X$ станет нарушение целостности контента при защищенных метаданных (например, компрометация ключей подписи контента и внесение в него нелегитимных изменений злоумышленником без внесения изменений в метаданные, что приведет к несанкционированному изменению содержательной части Элд) и, как следствие, самого Элд, что также вызовет переход АИС ЮЗ ЭД в режим нарушения функционирования (рис. 10в).

В качестве последнего варианта реализации ИнС $x_j \in X$ рассмотрим ситуацию, при которой злоумышленником является администратор АИС ЮЗ ЭД, имеющий доступ к ключам подписи контента Элд, а также расширенные права на внесение изменений в его метаданные. При таком сценарии доступ к обрабатываемому Элд может быть утрачен ввиду нарушения его целостности (контента и метаданных), что приведет к переходу АИС ЮЗ ЭД в режим нарушения функционирования (рис. 10г).

Разработанный метод позволяет сократить количество сценариев нарушения функционирования АИС ЮЗ ЭД за счет формирования ИЭлд, иллюстрированный пример приведен на рис. 11.

Таким образом, предлагаемый метод учитывает требования (3) по минимизации количества сценариев $|C|$ нарушения функционирования АИС ЮЗ ЭД, а также позволяет достичь необходимого критерия качества функционирования подсистемы ЗИ АИС ЮЗ ЭД (4). Кроме того, предполагается, что новая структура Элд позволит сократить временные затраты на его верификацию (обнаружение и локализацию компонентов Элд с признаками нарушения целостности).

В целях получения численной оценки результатов исследования произведем сравнительный анализ разработанного технического решения [20] с прототипом, в качестве которого используется способ [19].

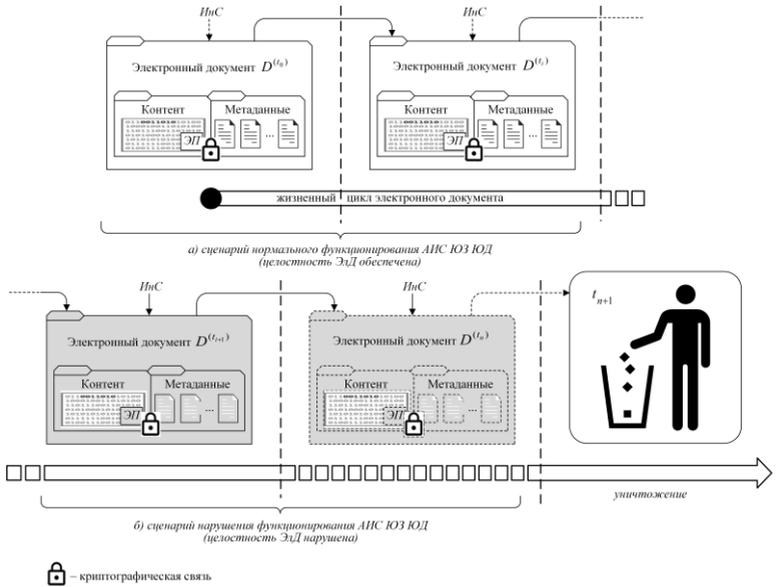


Рис. 11. Сценарии перехода АИС ЮЗ ЭД в различные режимы функционирования в процессе жизненного цикла ИЭД

Для чего произведем расчет количества операций по вычислению ХФ/ЭП в зависимости от количества совершаемых преобразований в течение жизненного цикла ЭД.

При использовании известного технического решения [19]:

$$(11) t_{\text{выч. прототипа}} = (k - 1) (n + 1).$$

При использовании разработанного технического решения [20]:

$$(12) t_{\text{выч. разр.}} = \sum_{i=0}^{\log_2 n} \frac{n}{2^i} + k + 1,$$

где n – количество реквизитов (метаданных), содержащихся в ЭД; k – количество моментов времени, в которые производятся преобразования ЭД.

Построим графики соответствующих зависимостей для ЭД, содержащего 64 реквизита, при условии 10 преобразований в период его жизненного цикла (рис. 12).



Рис. 12. Результат оценки количества операций вычисления ХФ/ЭП в зависимости от количества совершаемых преобразований в течение жизненного цикла ЭЛД

Рассчитаем среднее количество операций по вычислению ХФ/ЭП при использовании известного и разработанного решения:

$$(13) H_{\text{ср.}} = \frac{H_2 + H_3 + \dots + H_{10}}{n(k-1)},$$

где n – количество реквизитов (метаданных), содержащихся в ЭЛД; k – количество моментов времени, в которые производятся преобразования ЭЛД; H_i – операции по вычислению ХФ/ЭП.

Вычисления производятся при условии 10 преобразований ЭЛД в период его жизненного цикла, при этом $k=2$ – начало преобразований.

Воспользовавшись формулой (13), получим:

$$H_{\text{ср.прототипа}} = 167,5;$$

$$H_{\text{ср.разр.}} = 73.$$

Полученный результат позволяет сделать вывод о том, что разработанное техническое решение [20] позволяет сократить количество операций по вычислению ХФ/ЭП в 2,3 раза в срав-

нении с известным [19], что в процентном соотношении соответствует сокращению $t_{\text{выч.}}$ на 56,4 % [24].

Таким образом, выполнение условия (5) оценки разработанного метода достигнуто.

Далее произведем расчет количества сравнений компонентов ЭЛД при выполнении процедуры верификации в зависимости от момента времени, в котором было обнаружено первое несоответствие.

При использовании известного технического решения [19]:

$$(14) t_{\text{лок.прототипа}} = (k - i) n + (k - i - 1) (n - 1).$$

При использовании разработанного технического решения [20]:

$$(15) t_{\text{лок.разр.}} = 2 + \sum_{i=0}^{\log_2 n} 2^i + (k - i),$$

где n – количество реквизитов (метаданных), содержащихся в ЭЛД; k – количество моментов времени, в которые производятся преобразования ЭЛД; i – момент времени, в который было обнаружено первое несоответствие.

Построим графики соответствующих зависимостей для ЭЛД, содержащего 64 реквизита, при условии обнаружения первого несоответствия на 5-м моменте времени его жизненного цикла (рис. 13).

На основе полученных результатов вычислим среднее количество сравнений ХФ/ЭП с эталонными значениями при использовании известного и разработанного решения:

$$(16) L_{\text{ср.}} = \frac{L_1 + L_2 + \dots + L_n}{n (i - 1)},$$

где n – количество реквизитов (метаданных), содержащихся в ЭЛД; i – количество моментов времени, в которые производятся операции сравнения ХФ/ЭП с эталонными значениями; L_i – операции сравнения ХФ/ЭП с эталонными значениями.

При использовании формулы (16) получим

$$L_{\text{ср.прототипа}} = 288,5;$$

$$L_{\text{ср.разр.}} = 75.$$

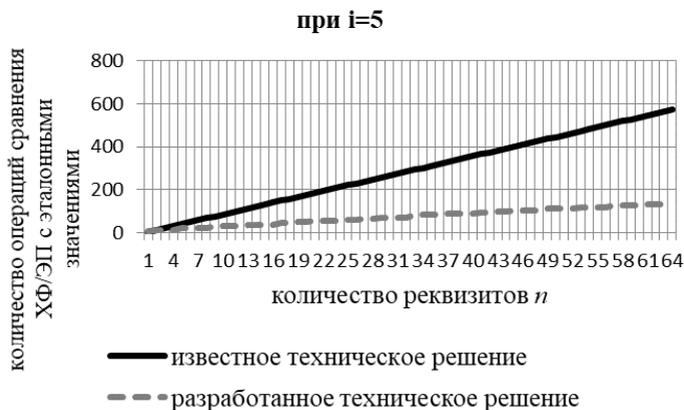


Рис. 13. Результат оценки количества операций сравнения ХФ/ЭП в зависимости от момента времени, в котором было обнаружено первое несоответствие

Разработанное техническое решение [20] позволяет сократить количество операций сравнения ХФ/ЭП с эталонными значениями в 3,84 раза в сравнении с известным [19], что в процентном соотношении соответствует сокращению $t_{\text{лок}}$ в среднем на 74,5 % [24].

Исходя из того, что ущерб, наносимый АИС ЮЗ ЭД, увеличивается пропорционально времени реализации угрозы [15], анализ полученных результатов (13), (16) позволяет сделать вывод о том, что сокращение количества криптографических операций приводит к снижению длительности верификации Элд, что позволяет уменьшить вероятность реализации деструктивных воздействий на подсистему ЗИ АИС ЮЗ ЭД и, как следствие, повысить уровень защищенности АИС ЮЗ ЭД.

В целях подтверждения данного утверждения произведем анализ данных, полученных в результате использования (14), (15) (таблица 1).

Для представления зависимости вероятности реализации деструктивных воздействий от времени выполнения процедуры контроля целостности Элд используем функцию вида

$$(17) f(x) = \frac{1}{1+e^{-x}} - 0,5.$$

Таблица 1. Результаты расчета количества операций сравнения ХФ/ЭП с эталонными значениями в соответствии с моментом времени жизненного цикла ЭД

<i>i</i>	1	2	3	4	5	6
<i>t</i> _{лок.прототипа}	1080	953	826	699	572	445
<i>t</i> _{лок.разр.}	138	137	136	135	134	133
<i>i</i>	7		8		9	
<i>t</i> _{лок.прототипа}	318		191		64	
<i>t</i> _{лок.разр.}	132		131		130	

Исходя из особенностей моделируемого процесса введём допущения:

- 1) $P(0) = 0$;
- 2) $\lim_{t \rightarrow \infty} P(t) = 1$,

где P – вероятность успешной реализации деструктивных воздействий.

В таком случае формула (17) примет вид

$$(18) P(T_{\text{контр}}) = \frac{2}{1+e^{-T_{\text{контр}}}} - 1.$$

При этом общее время выполнения процедуры контроля целостности ЭД зависит от количества операций сравнения ХФ/ЭП с эталонными значениями, что можно записать как

$$(19) T_{\text{контр.}} = t_{\text{опер.}} \times t_{\text{лок.}},$$

где $T_{\text{контр.}}$ – общее время выполнения процедуры контроля целостности ЭД; $t_{\text{опер.}}$ – время выполнения операции, причем $t_{\text{опер.}} = 10^{-3}$ с.

Используя таблицу 1 и формулы (18), (19), построим график соответствующих зависимостей для известного и разработанного решений (рис. 14).

С помощью полученных результатов определим вероятность успешной реализации деструктивных воздействий при

использовании известного и разработанного решений в процессе жизненного цикла ЭД (таблица 2).

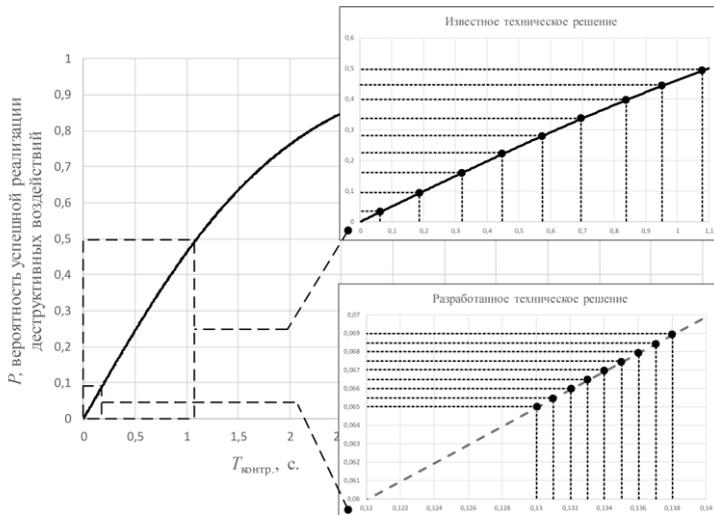


Рис. 14. Зависимость вероятности успешной реализации деструктивных воздействий от времени выполнения процедуры контроля целостности ЭД

Таблица 2. Вероятность успешной реализации деструктивных воздействий в соответствии с моментом времени жизненного цикла ЭД

i	1	2	3	4	5	6
$P_{\text{прототипа}}$	0,5	0,45	0,4	0,34	0,28	0,23
$P_{\text{разр.}}$	0,069	0,0685	0,068	0,0675	0,067	0,0665
ΔP_i	0,431	0,3815	0,332	0,2725	0,213	0,1635
i	7		8		9	
$P_{\text{прототипа}}$	0,17		0,09		0,04	
$P_{\text{разр.}}$	0,066		0,0655		0,065	
ΔP_i	0,104		0,0245		-0,025	

Значения разности $P_{\text{прототипа}} - P_{\text{разр.}} = \Delta P_i$ характеризуют снижение вероятности успешной реализации деструктивных воздействий на подсистему ЗИ АИС ЮЗ ЭД при использовании

предложенного решения, что свидетельствует о повышении уровня защищенности АИС ЮЗ ЭД. Однако при несанкционированном изменении ЭД в 9-й (в примере) момент времени результат уступает известному за счет особенностей реализации криптографических операций (структура бинарного дерева Меркла). Поэтому данный механизм целесообразно применять для ЭД долгосрочного хранения, в то время как известное решение подходит для контроля целостности ЭД с коротким сроком жизни.

Таким образом, поставленная задача на разработку метода обеспечения целостности ЭД на основе бинарного дерева Меркла, удовлетворяющего установленным критериям (4), (5), (6) полностью решена.

Оценка объема памяти W , необходимого для хранения данных формируемых на протяжении всего жизненного цикла ЭД, при условии размера ХФ 64 бита:

– при использовании известного технического решения [19]:

$$(20) W_{\text{прототипа}} = 64(k-1)(n+1);$$

– при использовании разработанного технического решения [20]:

$$(21) W_{\text{разр.}} = \left(\sum_{i=0}^{\log_2 n} \frac{n}{2^i} + 2 \right) 64k + 64(k-1).$$

Зависимости для ЭД, содержащего 64 реквизита, в процессе его жизненного цикла $k = 10$ представлены на рис. 15.

Средний объем памяти, требуемый для хранения данных преобразуемых на протяжении всего жизненного цикла ЭД, при использовании известного и разработанного решения:

$$(22) W_{\text{ср.}} = \frac{W_2 + W_3 + \dots + W_{10}}{n(k-1)},$$

где n – количество реквизитов (метаданных), содержащихся в ЭД; k – количество моментов времени, в которые производятся преобразования ЭД; W – объем памяти.

Вычисления производятся при условии выполнения десяти преобразований ЭД в период его жизненного цикла, при этом $k = 2$ – начало преобразований.

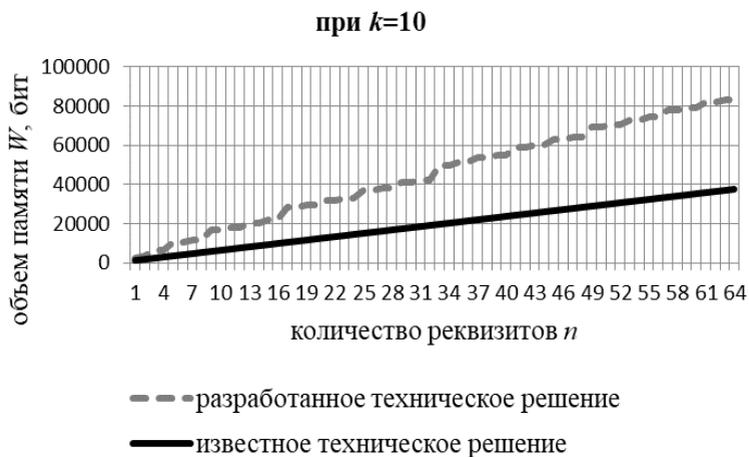


Рис. 15. Результат оценки необходимого объема памяти для хранения данных

С помощью (22) получим:

$$W_{\text{ср.прототипа}} = 10720 \text{ (бит);}$$

$$W_{\text{ср.разр.}} = 27181 \text{ (бит).}$$

Таким образом, представленные результаты достигаются за счет увеличения объема памяти хранимых данных в среднем в 2,53 раза [24] в сравнении с известным решением [19]. Предполагается, что увеличение объема памяти станет незначительным для АИС ЮЗ ЭД ввиду меньшего количества Элд длительного хранения в сравнении с количеством Элд с коротким сроком хранения.

6. Выводы

Введено понятие «интегративный Элд», характеризующийся целостностью, достигаемой за счет связности его компонентов. Представленный подход в формировании Элд учитывает особенности структуры, динамику и непрерывность его изменений в процессе жизненного цикла. Это соответствует принципам концепции «Индустрия 4.0», заключающимся в реализации функции аудита в режиме реального времени как за счет цен-

трализации обслуживаемых процессов (на примере АИС ЮЗ ЭД с низкой интенсивностью проводимых транзакций), так и за счет их виртуализации (технологии цифровых двойников) (на примере АИС ЮЗ ЭД крупных компаний, имеющих распределенную архитектуру).

Полученные результаты позволяют повысить уровень защищенности АИС ЮЗ ЭД посредством сокращения количества криптографических операций, производимых над ЭД, а также длительности выполнения процедуры его верификации (обнаружения и локализации компонентов с нарушенной целостностью).

Литература

1. ВАСИЛЬЕВ Н.В., ТИТОВ Г.С., РАКОВ И.В. *Технологии построения систем защищенного электронного документооборота* // Техника средств связи. – 2022. – №2(158). – С. 53–61. – DOI 10.24412/2782-2141-2022-2-53-61.
2. ВЕЛИГУРА А.В. *Моделирование управления большими стохастическими динамическими системами* // Социально-экономические и технические системы: исследование, проектирование, оптимизация. – 2021. – №2(88). – С. 32–39.
3. ВОЛКОВА В.Н., ЛОГИНОВА А.В., ЛЕОНОВА А.Е. и др. *Закономерности теории систем: состояние исследований и применения* // В сб.: «Системный анализ в проектировании и управлении. Сборник научных трудов XXVI Международной научно-практической конференции. В 3-х частях». – СПб., 2023. – С. 65–74.
4. ГВОЗДЕВ В.Е., ВАСИЛЬЕВ В.И., ГУЗАИРОВ М.Б. и др. *Поддержка управления функциональной безопасностью аппаратно-программных комплексов на основе системных архетипов* [Электронный ресурс] // Моделирование, оптимизация и информационные технологии. – 2022. – №10(2). – URL: <https://moitvivr.ru/ru/journal/pdf?id=1181> (дата обращения: 10.08.2024). – DOI: 10.26102/2310-6018/2022.37.2.025.
5. *ГОСТ Р ИСО 15489-1-2007. Система стандартов по информации, библиотечному и издательскому делу. Управле-*

- ние документами. – Введ. 01.01.2007. – М.: Стандартинформ, 2007. – 34 с.
6. *ГОСТ Р 34.11-2012. Информационная технология. Криптографическая защита информации. Функция хэширования.* – М.: Изд-во стандартов, 2012. – 16 с.
 7. *ГОСТ Р 43.0.10-2017. Информационное обеспечение техники и операторской деятельности. Информационные объекты, объектно-ориентированное проектирование в создании технической информации.* – М.: Стандартинформ, 2018. – 23 с.
 8. *ГОСТ Р ИСО 15489-1-2019 Система стандартов по информации, библиотечному и издательскому делу. Управление документами. Ч. 1. Понятия и принципы.* – М.: Стандартинформ, 2019. – 23 с.
 9. *ГОСТ Р 59341-2021. Системная инженерия. Защита информации в процессе управления информацией системы.* – М.: Стандартинформ, 2021. – 58 с.
 10. ДИЧЕНКО С.А. Модель контроля целостности многомерных массивов данных // Автоматическое управление и компьютерные науки. – 2021. – Т. 55, № 8. – С. 1188–1193.
 11. КАЛИНИН В.И., ЮСУПОВ Р.М., СОКОЛОВ Б.В. *Междисциплинарное взаимодействие и развитие теории систем, кибернетики и информатики* // В сб.: «Системный анализ в проектировании и управлении. Сборник научных трудов XXVI Международной научно-практической конференции. В 3-х частях». – СПб., 2023. – С. 7–13.
 12. КАЛИНИН В.Н., СОКОЛОВ Б.В. *Теория системных исследований как основа фундаментальной общенаучной подготовки молодых ученых* // В сб.: «XIV Всероссийское совещание по проблемам управления. Сборник научных трудов». – М., 2024. – С. 3735–3740.
 13. *Концепция развития электронного документооборота в хозяйственной деятельности. Решение президиума Правительственной комиссии по цифровому развитию, использованию информационных технологий для улучшения качества жизни и условий ведения предпринимательской деятельности от 25 декабря 2020 г. № 34.* // Официальный

- интернет-портал правовой информации [Электронный ресурс]. – 2022. – URL: [http:// https://www.nalog.gov.ru/html/sites/www.new.nalog.ru/docs/edo/edo_concept.pdf](http://https://www.nalog.gov.ru/html/sites/www.new.nalog.ru/docs/edo/edo_concept.pdf) (дата обращения: 10.08.2024).
14. ЛАРИН М.В. СУРОВЦЕВА Н.Г., ТЕРЕНТЬЕВА Е.В. и др. *Управление документами в цифровой экономике: организация, регламентация, реализация.* – М.: РГГУ, 2021. – 242 с.
 15. ЛИПАЕВ В.В. *Надежность и функциональная безопасность комплексов программ реального времени.* – М., 2013. – 176 с.
 16. МАХОВ Д.С. *Повышение устойчивости управления параметрами функционирования пространственно-распределенных радиотехнических систем робототехнических комплексов на основе нечетких множеств // Вопросы оборонной техники. Серия 16: Технические средства противодействия терроризму.* – 2020. – №5–6(143–144). – С. 36–44.
 17. НОВИКОВ Д.А. *Принцип декомпозиции в задачах управления организационно-техническими системами // В сб.: «Математическая теория управления и ее приложения (МТУиП-2020). Материалы конференции».* – СПб.: Государственный научный центр Российской Федерации АО «Концерн «ЦНИИ «Электроприбор», 2020. – С. 256–259.
 18. *Патент РФ 2637486, 04.12.2017.*
 19. *Патент РФ 2726930, 16.07.2020.*
 20. *Патент РФ 2812304, 29.01.2024.*
 21. РЯБИНИН И.А., СТРУКОВ А.В. *Решение одной задачи оценки надежности структурно-сложной системы разными логико-вероятностными методами // Труды Международной научной школы «Моделирование и анализ безопасности и риска в сложных системах» (МАБР-2019).* – 2019. – С. 159–172.
 22. ТАЛИ Д.И., ФИНЬКО О.А. *Криптографический рекурсивный контроль целостности метаданных электронных документов. Часть 1. Математическая модель // Вопросы кибербезопасности.* – 2020. – №5(39). – С. 2–18. – DOI: 10.21681/2311-3456-2020-05-2-18.

23. ТАЛИ Д.И. *Принцип целостности и интегративности в формировании электронного документа* // Правовая информатика. – 2022. – №3. – С. 72–83. – DOI: 10.21681/1994-1404-2022-3-72-83.
24. ТАЛИ Д.И., ЗАХАРОВ И.Л., СЕВРЮГИН А.А. и др. *Программная модель оценивания способа обеспечения интегративной целостности электронных документов, обрабатываемых автоматизированными информационными системами электронного документооборота* // Программа для ЭВМ № 2023613837. – 2023. – Бюл. №2.
25. ШВАБ К. *Четвертая промышленная революция*. – М.: Эксмо, 2021. – 208 с.
26. BELLARE M. *New Proofs for NMAC and HMAC: Security without Collision-Resistance* // CRYPTO. ePrint Archive, Report 2006/043. – 2006. – P. 1-16.
27. HARADA A., NISHIGAKI M., SOGA M. et al. *A Write-Once Data Management System* // ICITA. Shizuoka University, Johoku, Hamamatsu. –2002. – P. 432–8011.
28. KIAYIAS A., RUSSELL A., BERNARDO D. et al. *Ouroboros: A Provably Secure Proof-of-Stake Blockchain Protocol* // Lecture Notes in Computer Science. – 2017. – Vol. 10401 LNCS. – Springer, Cham. – P. 357–388. – DOI: 10.1007/978-3-319-63688-7_12.
29. KORPELA K., HALLIKAS J., DAHLBERG T. *Digital Supply Chain Transformation toward Blockchain Integration* // Proc. of the 50th Hawaii Int. Conf. on System Sciences. – 2017. – P. 4182–4191. – DOI: 10.24251/HICSS.2017.506.
30. MADHWAL Y., NURLYBAY D., YANOVICH Y. *Blockchain Extension for PostgreSQL Data Storage* // Proc. of the 3rd Blockchain and Internet of Things Conf. (BIOTC-2021), July 8–10, 2021, Ho Chi Minh City, Vietnam. – ACM, New York, NY, USA. – 6 p. – DOI: 10.1145/3475992.3476002.
31. TZIPORA H., BENHAMOUDA F., ANGELO DE CARO et al. *Initial Public Offering (IPO) on Permissioned Blockchain Using Secure Multiparty Computation* // Proc. of the IEEE Int. Conf. on Blockchain (Blockchain). IEEE, 2019. – P. 91–98. – DOI: 10.1109/Blockchain.2019.00021.

ELECTRONIC DOCUMENT MODEL WITHIN THE INDUSTRY 4.0 PARADIGM

Dmitry Tali, Krasnodar Higher Military School, Krasnodar, Cand.Sc., Phd student (dimatali@mail.ru).

Oleg Finko, Krasnodar Higher Military School, Krasnodar, Doctor of Science, professor (ofinko@yandex.ru).

Abstract: The digital transformation of society and production is gradually entering the stage of Industry 4.0, when the servicing infrastructure, including automated information systems of legally significant electronic document management, will be able to independently understand their environment and adapt to the current conditions of the situation. However, the current stage of development of such systems is characterized by the imperfection of the terminology and the identification of the object of influence (electronic document) with its paper prototype, which leads to a limitation of their functionality. Thus, the existing paradigm of perception of an electronic document does not correspond to the promising level of development of automated information systems of legally significant electronic document management within the framework of the concept of Industry 4.0. The purpose of the work: the formation of models of an electronic document and a description of the method of ensuring its integrity, which allows to guarantee the achievement of the required level of security of automated information systems of legally significant electronic document management, functioning in accordance with the provisions of the concept of Industry 4.0. The methods used: application of cryptographic hash functions to the components (content and metadata) of an electronic document, the results of which are calculated using the binary Merkle tree type in combination with the chain data recording technology. Results of the study: the concept of "integrative electronic document" was introduced, a new structure of the electronic document was proposed, which made it possible to reduce the duration of its verification, and as a result, reduce the risk of destructive impacts on the automated information system.

Keywords: Industry 4.0, content, metadata, electronic document management, electronic document integrity.

УДК 519.7
ББК 32.811

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

*Поступила в редакцию 17.09.2024.
Опубликована 31.05.2025.*

ЭВОЛЮЦИОННАЯ МЕТОДОЛОГИЯ МНОГОАЛЬТЕРНАТИВНОГО УПРАВЛЕНИЯ СЛОЖНЫМИ СИСТЕМАМИ

Подвальный С. Л.¹, Васильев Е. М.²
(Воронежский государственный технический
университет, Воронеж)

Излагаются результаты исследований, посвящённых свойству многоальтернативности строения и функционирования сложных биологических систем. Целью работы являются: обобщение этих результатов в виде эволюционной методологии многоальтернативного управления большими искусственными системами, выявление места этой методологии в ряду известных кибернетических принципов управления и определение её возможности при практической реализации. Отмечено, что актуальность упомянутого обобщения определяется назревшей необходимостью систематизации различных направлений в рамках самой кибернетики как науки. Раскрыто кибернетическое основание методологии многоальтернативного управления, заключающееся в воспроизведении эволюционных механизмов живых систем, выходящих за рамки механизма естественного отбора и традиционно рассматриваемых как свойство многообразия этих систем. На основе анализа содержания указанных механизмов сформулированы достаточно простые принципы предлагаемой методологии: принцип разнообразия и разделения функций, принцип многоуровневости структуры и принцип модульности. Показан конструктивный характер этих принципов, непосредственно указывающий на методы и способы их практической реализации. Представлена обобщённая модель системы многоальтернативного управления, данописание её функциональных узлов. Показана системообразующая роль методологии многоальтернативного управления по отношению к ряду базовых кибернетических принципов управления сложными объектами, заключающаяся в выявлении у этих принципов общего эволюционного содержания. Рассмотрены конкретные примеры применения методологии для синтеза систем управления различными объектами, и продемонстрирована её высокая результативность.

Ключевые слова: эволюционная кибернетика, кибернетические принципы эволюции, методология многоальтернативного управления.

¹ Семён Леонидович Подвальный, д.т.н, профессор (spodalny@yandex.ru).

² Евгений Михайлович Васильев, д.т.н, с.н.с. (vgtu-aits@yandex.ru).

1. Введение

Кибернетика как наука о наиболее общих принципах управления и обработки информации в животном, машине и обществе характеризуется значительным разнообразием направлений своего развития в технике, экономике, биологии и социологии [29, 30]. Это разнообразие предметных областей кибернетики сопровождается соответствующим расширением перечня её базовых принципов:

- принцип обратной связи и концепция чёрного ящика [10];
- принцип критической сложности Дж. фон Неймана [28];
- принцип дискретности информации [21];
- принцип магистрального развития [1, 2, 27];
- принцип необходимого разнообразия [55];
- принципы иерархической организации и рациональной централизации [29] и т.д. [30, 31, 43, 44, 59].

Проводя анализ истории прогресса кибернетики, Д.А. Новиков отмечает: «Но в начале 70-х годов [её] развитие затормозилось, целостный поток разветвился на множество частных и, наконец, «потерялся в деталях»: число научных направлений росло, каждое из них продолжало развиваться, а общих закономерностей почти не выявлялось и не систематизировалось» [30, с. 46].

Кроме того, в практическом плане применительно к задачам управления большими системами перечисленные базовые принципы во многих случаях требуют своего конструктивного раскрытия, указывающего на методы, способы и приёмы их реализации при построении сложных систем.

Приведём некоторые примеры.

Непосредственное использование принципа обратной связи на основе классической концепции чёрного ящика (управление по выходу) для управления сложными системами не позволяет обеспечить приемлемое качество управления такими системами [18, 19, 20, 34, 51, 57]. Практическое применение этого принципа в сложных системах выразилось в виде многоконтурных систем подчиненного регулирования [6, 60]; систем модального управления [4, 7, 25] и синергетических систем [9, 18, 19, 20, 52,

53], в которых предусмотрен частичный или полный отказ от концепции управления по выходу.

Принцип критической сложности в большинстве случаев трактуется лишь как концептуальная парадигма, реализация которой не выходит за рамки моделей клеточных автоматов [5, 61, 64].

Принцип необходимого разнообразия практически воплотился в системах с переменной структурой и скользящими режимами [15, 22, 23, 26, 48, 49, 50], идеи которых первоначально возникли без явной методической связи с этим принципом.

Таким образом, с точки зрения собственно кибернетики (Кибернетики с большой буквы [29, с. 9]) упомянутый выше перечень её принципов необходимо предполагает:

во-первых, существование некоторой единой концепции, объединяющей эти принципы в рамках кибернетики;

во-вторых, наличие вытекающей из этой концепции общей методологии управления, которая бы не только включала в себя указанные принципы, но и содержала указания на конкретные методы, способы и приёмы их реализации, т.е. была бы конструктивной.

Целью предлагаемой работы является методологическое обобщение указанных кибернетических принципов на основе эволюционной концепции многоальтернативного управления [33,39], которая, объединяя их в содержательном плане, одновременно будет обладать конструктивным характером, т.е. будет определять конкретные пути реализации этих принципов независимо от физической природы управляемых систем.

Для изложения заявленной методологии многоальтернативного управления воспользуемся схемой, предложенной в работе [33]:

- основания методологии;
- конструктивные принципы;
- внутренняя структура, определяющая объект и предмет методологии, её модели и методы;
- оценка практической результативности;
- системообразующая (объединяющая) функция методологии многоальтернативности как кибернетической концепции.

2. Кибернетические основания методологии многоальтернативного управления сложными системами

Исходным, эмпирическим установленным феноменом, инициировавшим формирование концепции многоальтернативности, является то обстоятельство, что эволюционная стабильность биологических систем с ростом их сложности возрастает [18, с. 12], т.е. в живых открытых системах, в отличие от антропогенных систем, не существует проклятия размерности.

Раскроем содержание понятия сложности биологических систем и роль этой сложности в их эволюционной стабильности.

Наиболее очевидной характеристикой сложности живых систем является их многообразие.

Многообразие биосистем на уровне элементарной эволюционной ячейки – популяции одного вида, заключается в многообразии её генофонда, содержащего, в том числе, нейтральные в текущих условиях гены, которые «активируются» при изменении этих условий и обеспечивают сохранение вида в целом [17]. Ярким проявлением эволюционной роли такого молекулярно-генетического многообразия является понятие критической численности популяции – аналогии понятия критической сложности системы по Дж. фон Нейману [28]: при достижении числа особей в популяции значения, меньшего чем некоторое критическое значение, последняя обречена на вырождение. Причина этого заключается в том, что интенсивность потока генетических мутаций в малой популяции становится ниже интенсивности потока изменений внешней среды, и популяция в целом на уровне своего генетического фонда уже не в состоянии приспособиться к этим изменениям. Теоретический анализ описываемого эволюционного процесса в популяции проведён в работе [37] на основе дискретной марковской модели накопления генофонда. На рис. 1 и рис. 2 показаны результаты этого анализа для популяции с начальной численностью особей $n_{\text{нач}} = 40$ – выше критического значения (рис. 1) и $n_{\text{нач}} = 10$ – ниже критического значения, рис. 2. Интенсивности потока генетических изменений в одной особи $\mu = 1$ и потока событий внешней

среды $\lambda = 25$ таковы, что суммарный поток накопление генофонда $n_{\text{нач}} \cdot \mu$ в первом случае превышает значение λ ($40 > 25$), а во втором случае – меньше его ($10 < 25$).

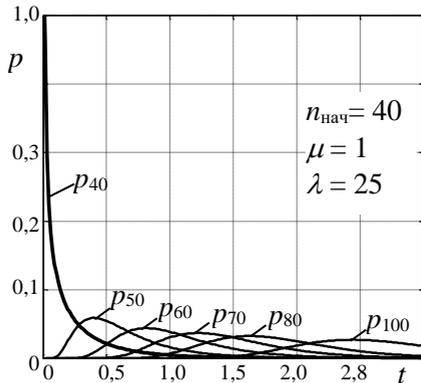


Рис. 1. Рост генофонда популяции при её начальной численности $n_{\text{нач}} = 40$

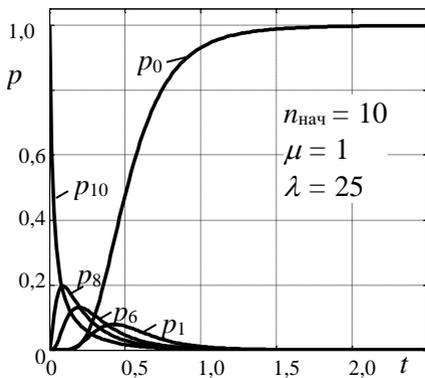


Рис. 2. Вырождение генофонда популяции при её начальной численности $n_{\text{нач}} = 10$

Рис. 1 и рис. 2 содержат графики изменения вероятностей p_i накопления популяцией генофонда объёмом i . Сопоставление этих рисунков показывает, что в популяции с количеством особей выше критического значения с течением времени t про-

исходит неограниченный рост объёма генофонда, при обратном соотношении популяция постепенно приходит в поглощающее состояние с нулевым объёмом генофонда, т.е. вырождается.

Эволюционная роль разнообразия в рамках одного вида проявляется также в форме разделения функций между органами, предназначенными для функционирования в разных условиях окружающей среды: светочувствительные клетки для дневного зрения и клетки для сумеречного зрения у приматов; ионофильтрующие клетки для пресной воды и для морской воды в жаберном эпителии мигрирующих рыб; жёсткие крылья для передвижения в грунте или древесине и мягкие крылья для полёта у жуков. Такая специализация функций значительно расширяет приспособительные возможности вида для обитания в разной среде.

На уровне крупных экосистем эволюционная роль многообразия проявляется, прежде всего, в разнообразии видов живых организмов, обладающих разными приспособительными механизмами и обеспечивающих сохранение живой массы экосистемы в разных условиях внешней среды. Возникающая при этом межвидовая конкуренция приводит к освоению новых экологических ниш отдельными видами, т.е. приводит к расширению ареала экосистемы. Э.Д Коуп в работе [58] указывает, что при возникновении экологического кризиса отмирание ранее господствующих видов сопровождается быстрым ростом числа ранее угнетённых видов, оказавшихся лучше приспособленными к новым условиям. Современным примером такого расширения является экосистема Аральского моря, при усыхании которого возросла его солёность, что привело к практическому исчезновению пресноводных донных организмов с одновременным компенсирующим взрывообразным ростом численности солоноустойчивых видов [3].

Помимо молекулярно-генетического многообразия и многообразия видов, в больших экологических системах появляется дополнительный способ обеспечения их эволюционной стабильности: многоуровневость пищевой цепочки, в которой происходит регулирующее взаимодействие между уровнями

(по схемам хищник – жертва или симбиоза) и, в конечном счёте, их сохранение в замкнутом круговороте вещества [13].

Аналитическая модель в виде системы дифференциальных уравнений для изменения биомассы каждого уровня замкнутой экосистемы, подтверждающая стабилизирующие свойства многоуровневости, приведена, например, в работах [37, 39]. Воспользуемся полученными в этих работах результатами исследования переходных процессов биомассы x_i^j в экосистемах с разным количеством трофических уровней $j=2$ и $j=3$, с числом видов $i=1, \dots, 3$ на каждом уровне (уровень 3 высший), рис. 3 и рис. 4.

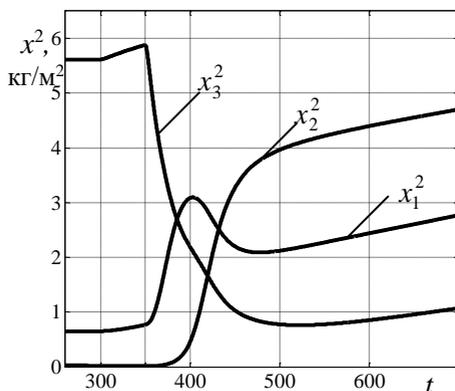


Рис. 3. Изменение биомассы x двухуровневой биосистемы в результате возмущения, вызвавшего угнетение организмов третьего вида второго уровня x_3^2

Возмущение (изменение условий внешней среды) поступает на второй уровень системы в момент времени $t=350$ и в двухуровневой системе приводит к шестикратному уменьшению биомассы третьего вида второго уровня x_3^2 . При этом наблюдается многократный рост биомассы x_2^2 второго вида, который стал занимать доминирующее положение в своём уровне, а также увеличение биомассы x_1^2 первого вида с колебательным характером переходного процесса.

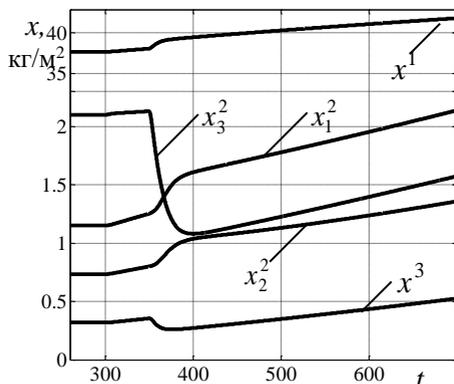


Рис. 4. Изменение биомассы x трёхуровневой биосистемы в результате возмущения, вызвавшего угнетение организмов третьего вида второго уровня x_3^2

При переходе к трёхуровневой системе (см. рис. 4) пере- численные выше кризисные изменения существенно уменьши- лись по амплитуде и стали монотонными, т.е. введение допол- нительного уровня иерархии в систему улучшило её стабилизирующие свойства за счёт «перераспределения» воз- мущения по всем уровням системы – биомассы x^1 и x^3 также претерпели соответствующие изменения.

Анализируя механизмы регулирования во многоуровневых системах живых организмов, У. Эшби [55, 56] отмечает, что общая стабильность состояния сложной живой системы (гомео- стазис) при сравнительно небольших отклонениях от этого состояния обеспечивается процессам регулирования в рамках некоторого одного уровня системы. При исчерпании ресурсов регулирования на текущем уровне системы управление переда- ётся на вышестоящий по своим ресурсам уровень, механизмы регулирования на котором могут носить качественно иной характер. Как пример им приводится пятиуровневая система регулирования содержания глюкозы в крови человека [56, с. 99].

Отметим ещё одно свойство живых систем, «материально» обеспечивающее упомянутые выше характеристики разнообра- зия и многоуровневости, – свойство блочности (модульности):

всё многообразие живого мира уже на генетическом уровне обусловлено дискретными комбинациями небольшого числа базовых блоков – нуклеотидов, образующих цепочки аминокислот и различных генов [47]. Именно свойство дискретности генетических структур, по мнению Р. Докинза [14], определяет их эволюционную устойчивость – «бессмертность гена».

Отмечая эволюционную роль дискретности живых структур, Э.М. Галимов в [12, с. 217] указывает, что «наиболее экономный способ производства низкоэнтропийного продукта состоит в комбинировании уже имеющихся низкоэнтропийных структур, эволюция в нашем понимании должна происходить не только и даже не столько путем малых изменений, сколько скачками, обусловленными новыми сочетаниями старых структур».

Перечисленные свойства многообразия, многоуровневости и модульности, определяющие собою сложность экосистем, позволяют выдвинуть тезис о том, что многоальтернативность является эволюционной стратегией живых систем, обеспечивающей им эволюционную стабильность [37]. В совокупности эти свойства можно объединить в единую эволюционную концепцию многоальтернативности строения и функционирования биологических систем, которая, в соответствии с основной идеей кибернетики, может быть положена в основу методологии многоальтернативного управления сложными системами любой природы.

3. Конструктивные принципы методологии многоальтернативного управления

Для того чтобы концепция многоальтернативного строения и функционирования сложных систем приобрела практический, конструктивный характер, необходимо сформулировать её принципы (механизмы), которые бы непосредственно указывали на методы, способы и приёмы реализации этой концепции в искусственных системах.

Изложенные выше кибернетические основы методологии многоальтернативного управления позволяют к таким принципам отнести [36, 37, 39]:

– принцип разнообразия и разделения функций в системе, заключающийся в том, что сложная система разделяется на подсистемы по функциональному признаку с высоким уровнем специализации каждой из них, включая качественно разные принципы построения. При изменении условий функционирования системы происходит автоматическая передача управления от одной подсистемы к другой, в результате чего достигается адаптация системы к воздействиям внешней среды. Узкая специализация каждой подсистемы создаёт предпосылки для упрощения реализации каждой из них с одновременным повышением качества выполнения своих функций;

– принцип многоуровневости структуры сложной системы даёт возможность выстроить «глубокоэшелонированное» в энергетическом смысле управление, при котором переход к более высокому уровню происходит по мере исчерпания возможностей управления низкого уровня. Такая иерархия системы управления позволяет обеспечить в ней аналог биологического свойства гомеостаза;

– принцип модульности (блочности) сложной системы является необходимой предпосылкой для физической реализации перечисленных выше принципов разнообразия и иерархии в системе. Модульность системы даёт возможность создать многообразие вариантов функционирования системы, которое обеспечивает ей полное или частичное функционирование (живучесть) при открытом взаимодействии с внешней средой в широком диапазоне изменений её параметров.

4. Объект и предмет методологии, её модели и методы

Объектом рассматриваемой методологии является собственно Кибернетика как наука об управлении, в частности, её раздел – эволюционная кибернетика [45, 46]. Отмечая актуальность выделения этого раздела кибернетики, В.Г. Редько пишет [46, с. 17]: «В процессе биологической эволюции возникли чрезвычайно сложные и вместе с тем удивительно эффективно функционирующие живые организмы. Эффективность, гармо-

ничность и согласованность работы компонент живых существ обеспечивается биологическими управляющими системами», – и, далее, о его предмете: «Какие архитектуры и принципы функционирования биологических систем управления обеспечивают способность животных приспосабливаться, адаптироваться к постоянно меняющимся условиям во внешней среде?».

К предмету рассматриваемой методологии относятся модели и методы эволюционной кибернетики, позволяющие осуществить практический синтез систем управления искусственными сложными системами на основе изложенных выше эволюционных принципов многоальтернативности.

Исходя из содержания этих принципов (раздел 3) обобщённая модель системы многоальтернативного управления должна включать в себя следующие функциональные блоки:

- функциональные блоки управления, соответствующие различным функциям (режимам) системы или различным алгоритмам управления в рамках одной функции;

- несколько уровней передачи управления, отличающихся физическими принципами работы или энергетическими возможностями;

- блоки оценки текущего состояния системы и выбора, в зависимости от этого состояния, альтернативного варианта управления;

- контуры параметрического управления в рамках одного алгоритма (контуры малых отклонений) и контур переключения альтернативных вариантов алгоритмов.

Общий вид модели системы многоальтернативного управления в виде функциональной схемы представлен на рис. 5.

Аналитическое представление показанных на рис. 5 процедур:

$$(1) \begin{cases} \dot{x} = B(x, t) + N(x, t) \cdot u + F(x, t) \cdot f; \\ y = A(x, t) + K(x, t) \cdot u + H(x, t) \cdot f, \end{cases}$$

$$(2) \quad x_j \rightarrow u_{ij}; \quad i = \overline{1, k}; \quad j = \overline{1, N_i},$$

$$(3) \quad b_i \rightarrow u_{ij}; \quad i = \overline{1, k}; \quad j = \overline{1, N_i},$$

$$(4) \quad \varepsilon = g - y; \quad u_{ij} = u_{ij}(\varepsilon).$$

где (1) – объект управления с нестационарными матрицами входов u , f , состояния x и выходов y ; (2) – процедуры выбора альтернативы управления u_{ij} по состоянию x_j объекта; (3) – процедуры программного назначения альтернативы управления u_{ij} в зависимости от требуемого режима b_i работы объекта; (4) – процедуры параметрического регулирования по отклонению ε . Практическое содержание этих процедур будет раскрыто ниже (раздел 5).

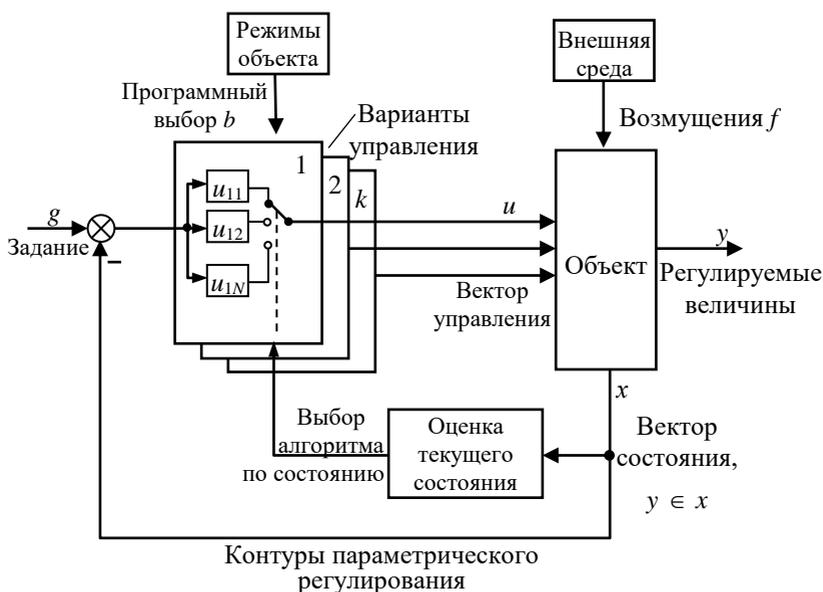


Рис. 5. Обобщённая модель системы многоальтернативного управления с двумя уровнями регулирования

Высокая степень общности принципов многоальтернативности, реализуемых в моделях вида (1)–(4), оставляет широкие возможности для использования в этих системах всего объёма методов, накопленных теорией автоматического управления, сохраняя лишь характерные для таких систем признаки разделения функций, многоуровневости и модульности.

5. Оценка практической результативности методологии многоальтернативного управления

Рассмотрим примеры целенаправленного практического применения методологии многоальтернативности для построения систем управления техническими объектами с различной степенью неопределённости происходящих в них процессов.

5.1. СИСТЕМА УПРАВЛЕНИЯ ПРОЦЕССОМ ТЕПЛОВЛАЖНОСТНОЙ ОБРАБОТКИ БЕТОННЫХ ИЗДЕЛИЙ

Процесс ускоренного затвердевания бетонных изделий происходит в паровлажностных камерах, температура $T(t)$ в которых должна изменяться по требуемой функции $T_{\text{тр}}(t)$, соответствующей режимам нагрева, выдерживания и остывания изделий, рис. 6 [8]. Изменение $T(t)$ по закону $T_{\text{тр}}(t)$ обеспечивается релейным регулятором подачи пара в единственном канале полной мощности $W_{\text{очн}}(t)$. Недостатками такого способа управления являются значительные пульсации фактической температуры $T(t)$ и большое число переключений дорогостоящего парового клапана полной мощности, приводящее к необходимости его частой замены, рис. 6 [8].

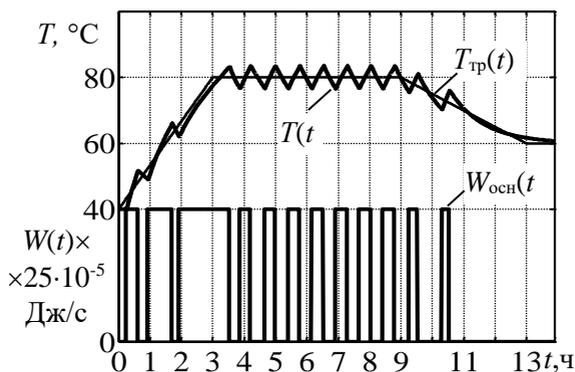


Рис. 6. Процесс регулирования температуры в одноуровневой системе управления

Анализ физического содержания режимов работы камеры указывает на то, что большая мощность при подаче пара необ-

ходима только в режиме нагрева камеры или при больших отклонениях фактической температуры $T(t)$ от требуемой $T_{тр}(t)$; в остальных режимах для управления температурой достаточна подача пара (тепла) с малой дополнительной мощностью $W_{доп}(t)$. Кроме того, для постоянной компенсации потерь тепла во внешнюю среду нужна мощность $W_{пост}$ подачи тепла, не изменяющаяся во времени.

Таким образом, анализ технологического процесса указывает на целесообразность разделения функций управления по трём каналам системы с иерархически различным уровнем мощности подачи пара: $W_{осн}(t)$, $W_{доп}(t)$, $W_{пост}$.

Результат такого разделения функций показан на рис. 7.

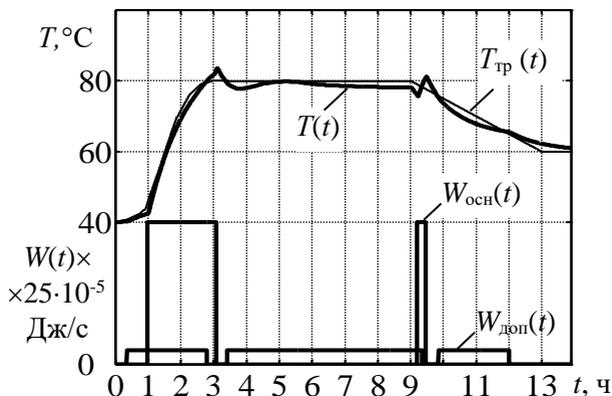


Рис. 7. Процесс регулирования температуры в трехуровневой системе управления

Сопоставление рис. 6 и рис. 7 показывает, что использование в рассматриваемой системе принципа разделения функций и иерархичности уровней управления позволяет существенно снизить среднее отклонение фактической температуры $T(t)$ от требуемой $T_{тр}(t)$ и в то же время уменьшить число переключений мощного клапана с 24 до 4 раз и общее число переключений всех трёх клапанов — в два раза по сравнению с одноканальной системой. Полученные преимущества улучшают прочностные характеристики обрабатываемых бетонных изде-

лий с одновременным уменьшением эксплуатационных расходов на замену паровых клапанов.

5.2. СИСТЕМА МОДАЛЬНОГО УПРАВЛЕНИЯ НЕСТАЦИОНАРНЫМ КОЛЕБАТЕЛЬНЫМ ОБЪЕКТОМ

Задача подавления собственных колебаний в объектах вида

$$(5) \quad \begin{aligned} \dot{x}(t) &= Bx(t) + Nu(t); \\ y(t) &= Ax(t) \end{aligned}$$

с характеристическим полиномом

$$(6) \quad C(s) = s^n + c_{n-1}s^{n-1} + \dots + c_1s + c_0$$

традиционно может быть решена построением замкнутой системы с модальным регулятором R :

$$(7) \quad \begin{aligned} \dot{x}(t) &= Bx(t) + N(g(t) - Rx(t)); \\ y(t) &= Ax(t), \end{aligned}$$

и желаемым характеристическим полиномом

$$(8) \quad D(s) = s^n + d_{n-1}s^{n-1} + \dots + d_1s + d_0$$

или

$$(9) \quad D(s) = s^n + (c_{n-1} + r_n)s^{n-1} + \dots + (c_1 + r_2)s + c_0 + r_1,$$

корни $\lambda_1, \lambda_2, \dots, \lambda_n$ которого путём выбора соответствующих коэффициентов регулятора $R = [r_1 \ r_2 \ \dots \ r_n]$ могут быть размещены требуемым образом на вещественной полуоси $(-\infty, 0]$ комплексной плоскости корней [25].

В том случае, если характеристическая матрица B нестационарная, т.е. имеет вид $B(t)$, на модальный регулятор может быть возложена задача обеспечения робастности системы управления к изменению коэффициентов $c_{n-1}, c_{n-2}, \dots, c_1, c_0$.

Из (9) следует, что такая робастность обеспечивается при выполнении n неравенств:

$$(10) \quad r_{i+1} \gg |c_i|, \quad i = \overline{0, n-1}.$$

Поскольку уравнение (9) имеет единственное решение относительно неизвестных коэффициентов r_1, r_2, \dots, r_n , то это решение в общем случае не удовлетворяет условию робастности (10), т.е. задачи (9) и (10) несовместны.

Принцип разделения функций указывает в этом случае на необходимость отдельного решения указанных задач.

Такое разделение можно обеспечить, если часть корней $\lambda_1^*, \lambda_2^*, \dots, \lambda_k^*$, $k < n$, определяющих требуемые динамические свойства системы, разместить на доминирующие позиции комплексной плоскости корней, а оставшиеся $n - k$ корней расположить на этой плоскости так, чтобы выполнялось условие (10) и одновременно корни $\lambda_1^*, \lambda_2^*, \dots, \lambda_k^*$ сохранили своё доминирующее положение, т.е.:

$$(11) \operatorname{Re}(\lambda_j) \ll \min[\operatorname{Re}(\lambda_i)]; \quad i = \overline{1, k}; \quad j = \overline{k+1, n}.$$

Выразим коэффициенты d_i полинома $D(s)$ через его корни:

$$(12) d_i(r_1, r_2, \dots, r_n) = d_i(\lambda_1, \lambda_2, \dots, \lambda_n); \quad i = \overline{0, n-1}.$$

В результате будет сформирована общая система ограничений:

$$(13) \begin{cases} d_i(r_1, r_2, \dots, r_n) = d_i(\lambda_1^*, \lambda_2^*, \dots, \lambda_k^*, \lambda_{k+1}, \dots, \lambda_n); & i = \overline{0, n-1}; \\ r_{\ell+1} \gg |c_\ell|; & \ell = \overline{0, n-1}; \\ \operatorname{Re}(\lambda_j) \ll \min[\operatorname{Re}(\lambda_i)]; & j = \overline{k+1, n}, \end{cases}$$

в которой неизвестны $2n - k$ величин: $r_1, r_2, \dots, r_n, \lambda_{k+1}, \lambda_{k+2}, \dots, \lambda_n$, и имеется n ограничений-равенств и $2n - k$ ограничений-неравенств. Таким образом, система ограничений (13) совместна и имеет $n - k$ свободных переменных $\lambda_{k+1}, \dots, \lambda_n$, т.е. искомое решение не единственно.

При практическом решении поставленной задачи степень доминирования (неравенства) в ограничениях определяются эмпирически, обычно 5–10-кратная. Кроме того, неединственность решения целесообразно использовать для совместного решения задачи (13) с задачей нахождения экстремума некоторой функции, например:

$$(14) J = \frac{\sum_{i=1}^n (r_i - \Omega)^2}{\Omega^2} \rightarrow \min ,$$

в которой $\Omega = \sqrt[n]{r_1 \cdot r_2 \cdot \dots \cdot r_n}$ – среднегеометрическое значение коэффициентов модального регулятора, а критерий J оптимизации обеспечивает минимальный разброс этих коэффициентов относительно среднегеометрического значения.

В качестве примера рассмотрим существенно колебательный объект [41]:

$$(15) W_{об}(s) = \frac{0,2(s + 0,2 + 400j)(s + 0,2 - 400j)}{(s + 0,5 + pj)(s + 0,5 - pj)(s + 100)},$$

собственная частота p колебаний которого изменяется в диапазоне $p = 6 \dots 60 \text{ с}^{-1}$.

Синтез модального регулятора с разделением функций по процедурам (13), (14) с заданными значениями доминирующих полюсов системы $\lambda_1^* = -20 \text{ с}^{-1}$, $\lambda_2^* = -100 \text{ с}^{-1}$ привёл к значению третьего полюса $\lambda_3 = -587 \text{ с}^{-1}$ и коэффициентам регулятора $R = [1133975; 71939,75; 606]$.

Подавление колебаний в системе и её грубость при изменении параметра p подтверждается рис. 8 и таблицей 1.

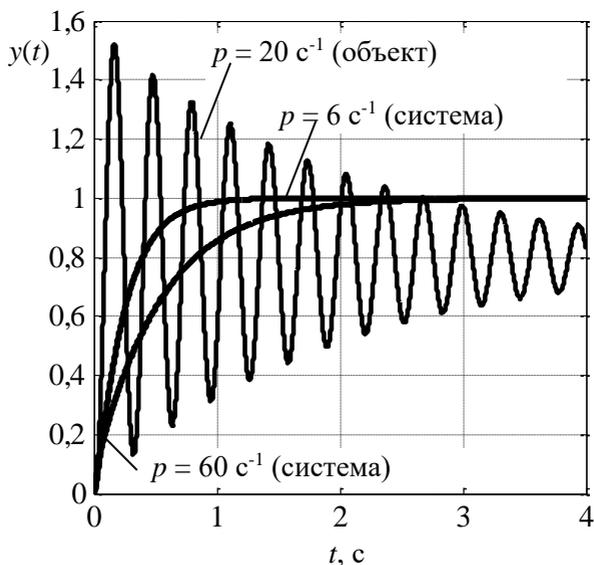


Рис. 8. Иллюстрация подавления колебаний в объекте с одновременным обеспечением робастности системы

Таблица 1. Чувствительность полюсов системы к изменению собственной частоты p колебаний объекта

p, c^{-1}	λ_1, c^{-1}	λ_2, c^{-1}	λ_3, c^{-1}
60	-26	-100	-581
20	-20	-100	-587
6	-19	-100	-588

Содержание таблицы 1 показывает, что при десятикратном изменении собственной частоты p колебаний объекта полюса системы с модальным регулятором с разделением функций остаются вещественными, т.е. обеспечивают монотонный характер переходных процессов, при этом значение доминирующего корня λ_1 изменяется всего в 1,3 раза.

5.3. СИСТЕМА С СИГНАЛЬНОЙ АДАПТАЦИЕЙ

Типовые структуры систем с сигнальной адаптацией, использующие эталонную модель и обратную связь с большим коэффициентом передачи, отличаются высоким быстродействием и простой реализацией, но обеспечивают малую глубину адаптации [16].

На рис. 9 представлен пример структуры такой системы [40], в которой $W(s), B, N, A$ – передаточная функция и матрицы (1) объекта управления; $\hat{W}(s), \hat{B}, \hat{N}, \hat{A}$ – передаточная функция и соответствующие матрицы эталонной модели; k – коэффициент передачи обратной связи.

Уравнения движения представленной системы

$$(16) \begin{cases} \dot{x} = (B - NkA)x + N(k\hat{A}\hat{x} + u); \\ \dot{\hat{x}} = \hat{B}\hat{x} + \hat{N}u \end{cases}$$

можно переписать в виде:

$$(17) \begin{bmatrix} \dot{x} \\ \dot{\hat{x}} \end{bmatrix} = \begin{bmatrix} B - NkA & Nk\hat{A} \\ 0 & \hat{B} \end{bmatrix} \cdot \begin{bmatrix} x(t) \\ \hat{x}(t) \end{bmatrix} + \begin{bmatrix} N \\ \hat{N} \end{bmatrix} \cdot u$$

и получить характеристический полином системы:

$$(18) \left| \begin{array}{c|c} sE - B + NkA & -NkA \\ \hline 0 & sE - \hat{B} \end{array} \right| = |sE - B + NkA| \cdot |sE - \hat{B}|,$$

где $E = \text{diag}(1 \dots 1)$.

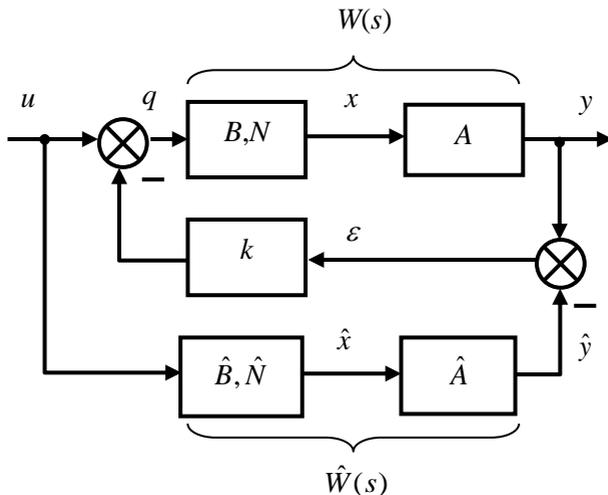


Рис. 9. Типовая структурная схема системы с сигнальной адаптацией и эталонной моделью

Выражение (18) показывает, что характеристический полином замкнутой системы разбивается на два независимых сомножителя, соответствующие объекту с обратной связью и эталонной модели. Поскольку полином эталонной модели заведомо гурвицев, то коэффициент k обратной связи влияет только на устойчивость и свойства замкнутого контура с объектом.

Анализ этого факта, проведённый в [40], показал, что в общем случае стремление увеличить глубину адаптации путём увеличения коэффициента k ограничено ростом колебательности и потерей устойчивости системы, однако в объектах с динамическим порядком не выше второго указанное ограничение отсутствует. Этот вывод указывает на целесообразность физического представления объекта в виде блочной структуры из элементов низкого порядка и построения многоуровневой схе-

мы сигнальной адаптации в виде последовательного иерархически связанного соединения одноуровневых систем, рис. 10.

На рис. 10 показан вариант структуры двухкаскадной системы, в которой $W_1(s)$, $W_2(s)$ обозначают передаточные функции каскадов объекта; $\hat{W}_1(s)$, $\hat{W}_2(s)$ – передаточные функции соответствующих каскадов эталонной модели.

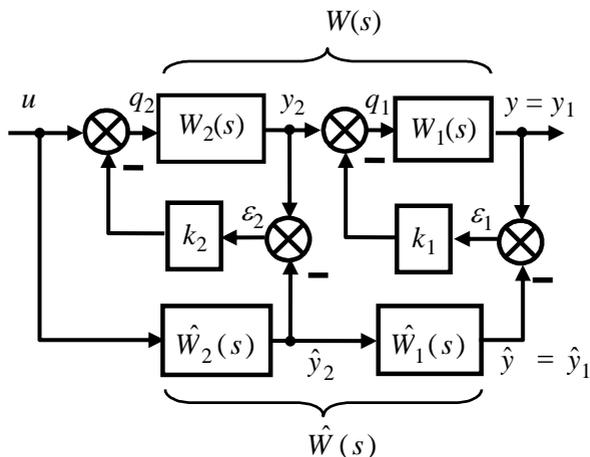


Рис. 10. Двухуровневая система с сигнальной адаптацией

Общая передаточная функция $\tilde{W}(s, k_1, k_2)$ двухуровневой системы имеет вид

$$(19) \quad \tilde{W}(s, k_1, k_2) = \frac{y(s)}{u(s)} = \frac{W(s)(k_2 \hat{W}_2(s) + 1)}{(1 + k_1 W_1(s))(1 + k_2 W_2(s))} + \frac{W_1(s) k_1 \hat{W}(s)}{1 + k_1 W_1(s)},$$

из которого следует соотношение

$$(20) \quad \lim_{\substack{k_1 \rightarrow \infty \\ k_2 \rightarrow \infty}} \tilde{W}(s, k_1, k_2) = \hat{W}(s),$$

т.е. с ростом k_1 и k_2 динамические свойства системы приближаются к свойствам эталонной модели.

Результативность двухуровневой системы иллюстрируется рис. 11, на котором показаны переходные процессы в одноуров-

невой и двухуровневой системах с сигнальной адаптацией при пятикратном изменении параметров объекта. Одноуровневая система $y_{1,1}(t)$ не подавила собственные колебания объекта. При этом в двухуровневой системе обеспечена высокая степень совпадения переходного процесса $y_{1,2}(t)$ с эталонной моделью $\hat{y}(t)$.

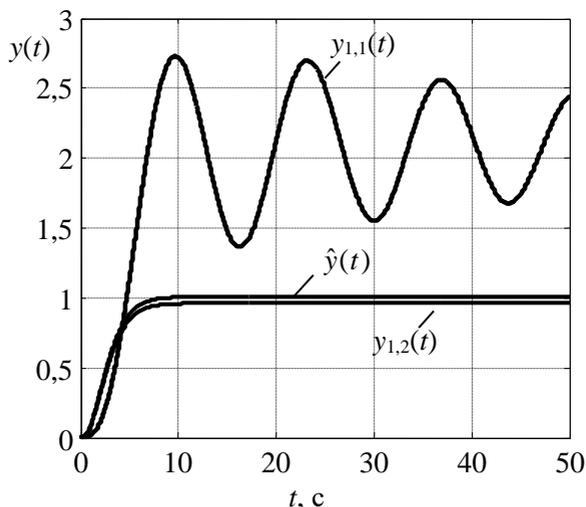


Рис. 11. Сопоставление адапционных свойств одноуровневой $y_{1,1}(t)$ и двухуровневой систем сигнальной адаптации $y_{1,2}(t)$ относительно эталонной модели $\hat{y}(t)$

Полученные результаты, подтверждающие результативность рассматриваемой методологии для синтеза систем с сигнальной адаптацией, справедливы и в том случае, если объект физически невозможно разделить на части с низким динамическим порядком, но количество доминирующих корней характеристического полинома каждой части не превышает двух [42].

5.4. СИСТЕМА УПРАВЛЕНИЯ БЕСПИЛОТНЫМ ЛЕТАТЕЛЬНЫМ АППАРАТОМ

Рассмотрим беспилотный летательный аппарат (БПЛА) вертолётного типа, рис. 12, описываемый системой дифферен-

циальных уравнений [63]:

$$(21) \dot{\omega}_x = \frac{(J_y - J_z) \cdot \omega_y \omega_z + Q_x}{J_x};$$

$$(22) \dot{\omega}_y = \frac{(J_z - J_x) \cdot \omega_z \omega_x + Q_y}{J_y};$$

$$(23) \dot{\omega}_z = \frac{Q_z}{J_z};$$

$$(24) \dot{\phi} = \omega_x + \operatorname{tg}(\theta) \cdot (\omega_y \cdot \sin(\varphi) + \omega_z \cdot \cos(\varphi));$$

$$(25) \dot{\theta} = \omega_y \cdot \cos(\varphi) - \omega_z \cdot \sin(\varphi);$$

$$(26) \dot{\psi} = \frac{\omega_y \cdot \sin(\varphi) + \omega_z \cdot \cos(\varphi)}{\cos(\theta)};$$

$$(27) \dot{V}_x = \frac{m \cdot g \cdot \sin(\theta) - m(V_z \cdot \omega_y - V_y \cdot \omega_z) - k_x \cdot (V_x - V_{W_x}) \cdot |V_x - V_{W_x}|}{m};$$

$$(28) \dot{V}_y = \frac{-m \cdot g \cdot \sin(\varphi) \cdot \cos(\theta) - m(V_x \cdot \omega_z - V_z \cdot \omega_x) - k_y \cdot (V_y - V_{W_y}) \cdot |V_y - V_{W_y}|}{m};$$

$$(29) \dot{V}_z = \frac{U_z - m \cdot g \cdot \cos(\theta) \cdot \cos(\varphi) - m(V_y \cdot \omega_x - V_x \cdot \omega_y) - k_z \cdot (V_z - V_{W_z}) \cdot |V_z - V_{W_z}|}{m};$$

$$(30) \begin{bmatrix} \dot{X} \\ \dot{Y} \\ \dot{Z} \end{bmatrix} = C^T \cdot \begin{bmatrix} V_x \\ V_y \\ V_z \end{bmatrix}.$$

Принятые обозначения: НСК – неподвижная (земная) система координат XYZ ; ПСК – подвижная (связанная с аппаратом) система координат xuz ; $\Omega_1, \dots, \Omega_4$ и u_1, \dots, u_4 – частоты вращения и силы тяги винтов; $U_z = u_1 + u_1 + u_3 + u_4$; $\omega_x, \omega_y, \omega_z$ и Q_x, Q_y, Q_z – угловые скорости и вращающие моменты БПЛА относительно осей координат x, y, z ПСК; J_x, J_y, J_z – моменты инерции корпуса БПЛА относительно осей x, y, z ; V_x, V_y, V_z – составляющие скорости БПЛА в ПСК; k_x, k_y, k_z и $V_{W_x}, V_{W_y}, V_{W_z}$ – коэффициенты сопротивления воздуха и скорости ветра по осям x, y, z ; m – масса аппарата; g – ускорение свободного падения; C – матрица перехода между подвижной и неподвижной системами координат; θ, ψ, φ – тангаж, курс и крен аппарата, см. рис. 12;

$$(31) C = \begin{bmatrix} c\theta \cdot c\psi & c\theta \cdot s\psi & -s\theta \\ s\varphi \cdot s\theta \cdot c\psi - & s\varphi \cdot s\theta \cdot s\psi + & s\varphi \cdot c\theta \\ -c\varphi \cdot s\psi & +c\varphi \cdot c\psi & \\ c\varphi \cdot s\theta \cdot c\psi + & c\varphi \cdot s\theta \cdot s\psi - & c\varphi \cdot c\theta \\ +s\varphi \cdot s\psi & -s\varphi \cdot c\psi & \end{bmatrix},$$

$c\theta = \cos(\theta)$, $s\theta = \sin(\theta)$, $c\varphi = \cos(\varphi)$, $s\varphi = \sin(\varphi)$, $c\psi = \cos(\psi)$, $s\psi = \sin(\psi)$.

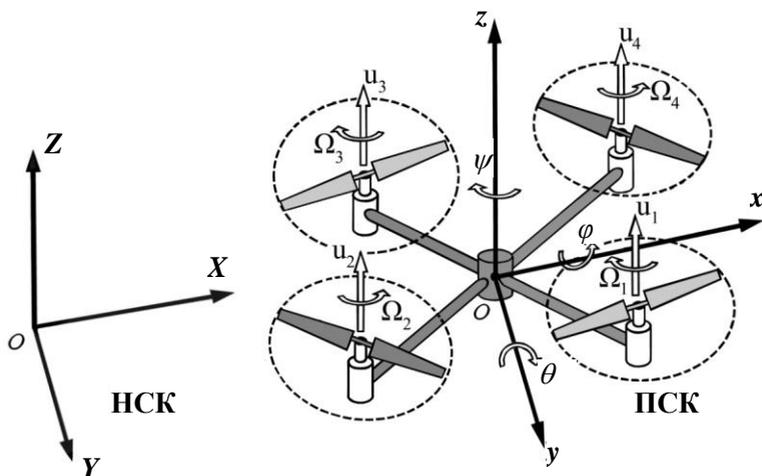


Рис. 12. Компоновочная схема и системы координат БПЛА

Основная трудность управления, в частности, стабилизации такого аппарата в условиях ветровых возмущений F_w , заключается, как это следует из (21)–(31), в наличии существенных перекрёстных связей между координатами его состояния. Синтез автономных каналов регулирования, компенсирующих эти взаимосвязи, существенно усложняет систему управления; синтез системы без учёта этих взаимосвязей не позволяет обеспечить высокие показатели качества регулирования.

Для преодоления этой трудности в работах А.А. Колесникова [18, 19, 20] предложен метод аналитического конструи-

рования агрегированных регуляторов, который предусматривает разделение переменных состояния объекта на группы, образующие физически связанную последовательность величин. В результате общий многомерный регулятор разбивается на части с малой динамической размерностью, т.е. регулятор становится агрегированным и его синтез можно выполнить последовательно – по частям. При этом учитываются все внутренние взаимосвязи многомерного объекта. Такой покаскадный синтез значительно проще синтеза полного регулятора, не требует линеаризации модели, обеспечивает получение высоких показателей качества системы и в целом соответствует принципу разделения функций и многоуровневости методологии многоальтернативного управления.

Для данного объекта указанное разделение координат осуществляется следующим образом [63].

Поскольку задачей управления является формирование требуемой траектории $L_{тр}(t) = [X_{тр}(t) \ Y_{тр}(t) \ Z_{тр}(t)]^T$ перемещения $L(t)$ БПЛА в пространстве, то в соответствии с (30) для построения такой траектории необходимо сформировать вектор скоростей $V_{тр}(t) = [V_{тр,x}(t) \ V_{тр,r}(t) \ V_{тр,z}(t)]^T$, текущие значения которого можно вычислять исходя из условия монотонной сходимости макропеременной $L_{тр}(t) - L(t) = \Psi_L(t)$ к нулю с заданной постоянной времени T_L :

$$(32) \quad T_L \cdot \dot{\Psi}_L(t) + \Psi_L(t) = 0,$$

где $L(t)$ – координаты центра масс: $L(t) = [X(t) \ Y(t) \ Z(t)]^T$.

Таким образом, в качестве первой группы управляющих переменных следует использовать вектор фактических скоростей $V(t) = [V_x(t) \ V_r(t) \ V_z(t)]^T$, рис. 13.

Далее, из уравнений (24)–(29) можно увидеть, что для создания вектора $V(t)$ линейных скоростей и курса ψ следует использовать очередную (вторую) группу переменных: вектор угловых скоростей $\Omega(t) = [\omega_x(t) \ \omega_y(t) \ \omega_z(t)]^T$ и силу тяги U_z , требуемые значения которых будет вычисляться аналогично (32) из условий (33):

$$(33) \quad \begin{aligned} T_V \cdot \dot{\Psi}_V(t) + \Psi_V(t) &= 0; \\ T_\Psi \cdot \dot{\Psi}_\Psi(t) + \Psi_\Psi(t) &= 0, \end{aligned}$$

в которых $\Psi_V(t) = V_{тр}(t) - V(t)$; $\Psi_\Psi(t) = \psi_{тр}(t) - \psi(t)$, рис. 13.

Третья группа управляющих переменных – вектор вращающихся моментов $Q(t)$ – определяется уравнениями (21)–(23), требуемое значение $Q_{тр}(t)$ которого будет вычисляться с использованием макропеременной $\Psi_\Omega = \Omega_{тр} - \Omega$ по условию

$$(34) \quad T_\Omega \cdot \dot{\Psi}_\Omega(t) + \Psi_\Omega(t) = 0.$$

Выделение указанных управляющих переменных $V(t)$, $\Omega(t)$, $Q(t)$, вычисляемых иерархически последовательно, позволяет разбить регулятор на три части, которые синтезируются по условиям (32)–(34) независимо друг от друга, рис. 13.

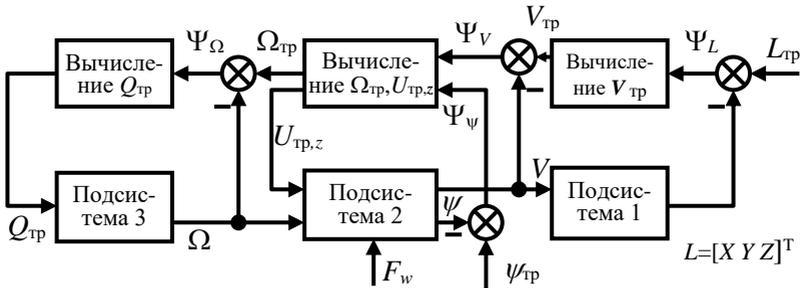


Рис. 13. Структурная схема объекта с выделенными подсистемами и соответствующим агрегированным регулятором с разделением функций управления по группам переменным

Результаты моделирования процесса стабилизации заданного курса $\psi(t)$ БПЛА при его движении со скоростью 3 м/с при возникновении порыва бокового ветра со скоростью 20 м/с на интервале $t = [5;8]$ с представлены на рис. 14.

Рис. 14 подтверждает способность системы автоматически поддерживать заданное направление полёта путём энергичного управления угловым положением аппарата, т.е. управлением вектором его тяги.

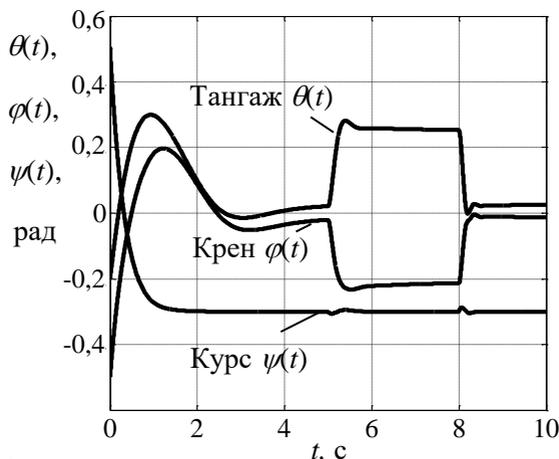


Рис. 14. Стабилизация курса БИЛА изменением вектора его тяги при боковом порыве ветра на интервале $t = [5;8]$ с

Важно отметить, что использование принципа иерархии управления и разделения функций при построении агрегированного регулятора не является его тривиальной декомпозицией на независимые линеаризованные части меньшей размерности, так как в нём сохраняется многосвязность и нелинейность управления многомерным объектом высокого порядка, в то время как каждая часть регулятора представляет собою динамическое звено первого порядка.

5.5. СИСТЕМА ПРИНЯТИЯ РЕШЕНИЙ В КОМБИНАТОРНЫХ ЗАДАЧАХ ОПТИМИЗАЦИИ

Рассмотрим пример применения методологии многоальтернативности для построения системы принятия оптимальных решений как частного случая системы управления.

В качестве объекта выберем класс комбинаторных задач, отличающийся следующими особенностями:

- наличие большого количества локальных экстремумов;
- расположение глобального экстремума в пространстве поиска в общем случае изолировано от расположения локальных

оптимальных решений, т.е. область глобального экстремума нельзя обнаружить по шлейфу локальных, рис. 15 [35].

Указанное обстоятельство затрудняет использование для решения таких задач генетических алгоритмов, исключая, по существу, использование свойства преэмергентности этих алгоритмов. Например, если при решении тестовых задач квадратичного программирования размерностью $n = 12$ классический генетический алгоритм обеспечивал нахождение глобального экстремума в 80% поисков, то при увеличении размерности до $n = 20$ такой результат достигался лишь в 2–5% поисков [35].

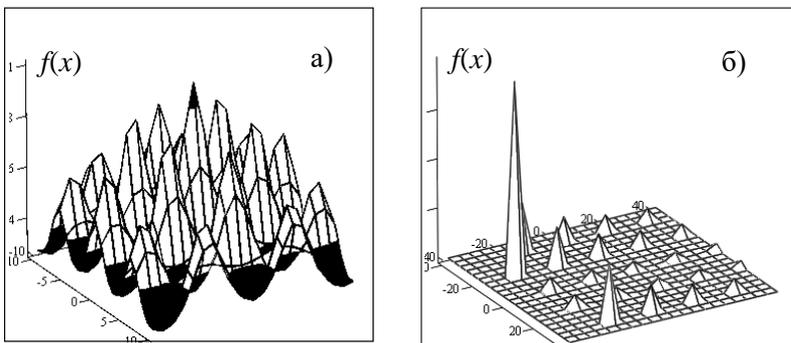


Рис. 15. Иллюстрация расположения глобального экстремума функции $f(x)$: а) в окружении шлейфа локальных экстремумов; б) изолированное

Методология многоальтернативности указывает на целесообразность использования в данной задаче принципа блочности, который применительно к генетическим процессам был сформулирован М. Эйгеном [54] в виде гипотезы о том, то в добиологический период рост макромолекул происходил не путём скрещивания родительских особей, а на основе матричной репликации: накопительного присоединения новых блоков в результате самоинструирования – избирательной предпочтительности такого присоединения, определяемой уже сформированной частью макромолекулы.

Практические процедуры генетического алгоритма с матричной репликацией содержат следующие этапы:

- формирование множества вариантов решений – наборов хромосом таким образом, что во все хромосомы входит одна и та же матрица: в дальнейшем неизменяемые по содержанию и расположению участки – гены. Пример такой хромосомы с матрицей репликации показан на рис. 16. Создаётся несколько вариантов таких исходных множеств (популяций), отличающихся своими матрицами. Например, если использовать матрицу, состоящую из одного элемента h^1 , то размещая её в одной популяции на первое место h_1^1 , во второй популяции – на второе место h_2^1 и т.д., мы получим 10 популяций, одна из которых уже содержит один элемент из варианта решения, соответствующему глобальному экстремуму X ;

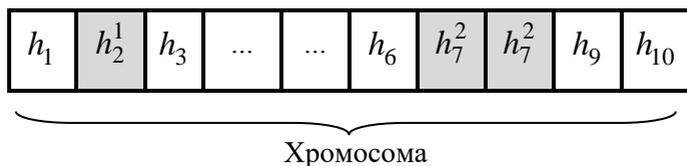


Рис. 16. Вариант хромосомы с генами h_i , $i = 1, \dots, 10$, содержащей трёхместную матрицу репликации $\{h_2^1, h_7^2, h_8^3\}$

- реализация процесса эволюции каждой популяции с помощью генетического алгоритма любой модификации. При этом, в соответствии с гипотезой М. Эйгена, предполагается, что та популяция, которая содержит неизменяемую матрицу глобального решения X , за счёт самоинструирования обеспечивает нахождение этого решения с вероятностью, большей, чем при «безматричной» эволюции. Отметим, что процедуры естественного отбора осуществляются только в пределах каждой популяции, т.е. эволюция каждой матрицы происходит независимо от других и исключение какой-либо матрицы из эволюционного процесса не допускается;

- анализ решений, полученных из каждой матрицы, и выбор лучших (или группы конкурентоспособных) из них.

В качестве демонстрации результативности применения принципа блочности в виде поиска с матрицей репликации приведём решение задачи определения кратчайшего маршрута сверления печатной платы с 99 отверстиями (задача коммивояжера), рис. 17а [35].

Высокая размерность этой задачи практически исключила возможность нахождения глобального экстремума: пятисоткратное повторение поиска традиционным генетическим алгоритмом с 1000 циклов эволюции в каждом привело к получению ряда близких маршрутов длиной $L = 261 \dots 264$ мм, как это оказалось, далёких от кратчайшего пути.

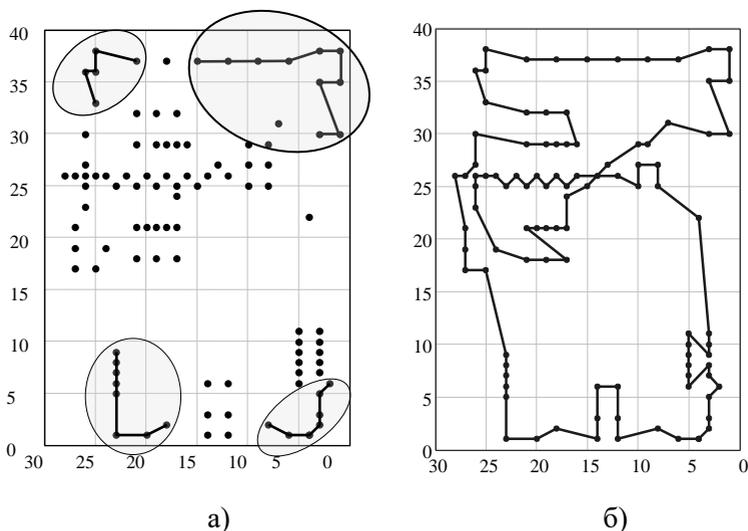


Рис. 17. Решение задачи коммивояжера генетическим алгоритмом с матричной репликацией

Вместе с тем анализ полученных в результате такого традиционного поиска решений показал, что в значительной части решений присутствуют одинаковые элементы маршрутов (выделены на рис. 17а). Эти элементы были использованы в качестве четырёх матриц репликации и введены в алгоритм поиска в виде неизменяемой части. Уже 50-кратного повторения поиска

оказалось достаточным для получения значительно лучшего варианта маршрута с длиной $L = 232$ мм, рис. 17б.

Разнообразие представленных выше задач, решаемых на основе методологии многоальтернативного управления, подтверждает высокую степень общности её принципов, заложенных в эволюционных механизмах природы.

6. Системообразующая функция методологии многоальтернативности как кибернетической концепции

Рассмотрим взаимосвязь методологии многоальтернативного управления с другими кибернетическими принципами.

Концепция чёрного ящика, заключающаяся в сознательном абстрагировании от внутреннего строения объектов управления и происходящих в них процессов, постепенно, с ростом сложности последних, уступает место управлению по вектору состояния, адекватная оценка которого возможна только при раскрытии внутренней структуры и процессов, протекающих в этих объектах. Принципы многоальтернативного управления в своей основе в полной мере отвечают этой тенденции, так как раскрывают содержание внутренних эволюционных механизмов сложных живых систем, обеспечивающих им высокие адаптационные свойства.

Принцип критической сложности, сформулированный в работах Дж. фон Неймана и указывающий на возможность самоорганизации только в тех искусственных системах, уровень сложности которых выше некоторого критического значения, непосредственно согласуется с многоальтернативным принципом многообразия и разделения функций, обеспечивающим сложной системе, как это было показано выше, устойчивую работоспособность и живучесть в широких пределах изменяющихся условий функционирования.

Принцип дискретности информации и комбинаторный подход к определению её количества [14, 21] полностью совпадает с принципом блочности структур сложных систем, порождаю-

щим многообразие их конфигураций и приспособительных функций.

Принцип магистрального развития предусматривает существование у сложных многоагентных систем такой стратегии своего развития, которая делает невыгодным ни одному агенту отклонение от этой стратегии, несмотря на то, что интересы отдельно взятых агентов не совпадают [62]. Реализация этого принципа в технических системах осуществляется с помощью многоальтернативного принципа иерархичности строения и функционирования сложных систем, в которых выполнение общей (магистральной) функции системы возможно только при согласованном, синергетическом функционировании её подсистем, подчинённом этой общей функции. Примечательно, что принцип иерархического строения хорошо согласуется результатами работы [11], в которой доказывается правило рациональной организации: с ростом сложности системы и условий её функционирования количество уровней иерархии в ней необходимо увеличивать.

Принцип необходимого разнообразия [55], утверждающий, что для противодействия многообразным условиям изменяющейся внешней среды открытая система должна обладать соответствующим многообразием своих состояний, т.е. управлений, и по названию и по своему содержанию совпадает с сущностью методологии многоальтернативного управления, изложенной выше.

Таким образом, определяя место методологии многоальтернативного управления в кибернетике, можно констатировать, что она выполняет функцию обобщения основных кибернетических принципов, методологически объединяя в себе их содержание на основе общих механизмов эволюционных процессов живой природы.

7. Заключение

Феномен высокой адаптационной стабильности сложных биологических систем указывает на существование в них неко-

торых эволюционных механизмов, выходящих за рамки естественного отбора.

В качестве таких механизмов предложено выделить ряд достаточно общих принципов – принципы многоуровневости, многообразия и разделения функций и принцип модульности, составляющих в своей совокупности общую эволюционную концепцию многоальтернативного строения и функционирования биологических систем.

Математическое моделирование процессов эволюции живых систем с учётом механизмов многоальтернативности подтвердило эволюционные свойства последних и открыло возможность для направленного воспроизведения этих свойств в сложных антропогенных системах. Такое воспроизведение необходимо предполагает обобщение соответствующих процедур синтеза в эволюционную методологию многоальтернативного управления.

Раскрытые в работе состав и кибернетическое содержание этой методологии позволили выявить её системообразующую (обобщающую) роль по отношению к ряду базовых кибернетических принципов построения сложных систем, заключающуюся в выявлении у этих принципов единого – эволюционного смысла.

Представленные примеры практического применения методологии многоальтернативного управления не только подтверждают её высокую результативность, но и раскрывают конструктивный характер её принципов, непосредственно указывающих на способы их реализации при построении сложных искусственных систем.

Литература

1. АБАКУМОВ А.И. *Модели Неймана – Гейла*. – Владивосток: ДВГУ, 2004. – 44 с.
2. АБАКУМОВ А.И., ПИДЮРА Т.А. *Магистральные характеристики матричных моделей экономической динамики*. – Владивосток: Вестник ТГЭУ. – 2010. – № 2. – С. 96–103.

3. АНДРЕЕВА С.И., АНДРЕЕВ Н.И. *Эволюционные преобразования двустворчатых моллюсков Аральского моря в условиях экологического кризиса.* – Омск: Изд-во Омского государственного педагогического университета, 2003. – 382 с.
4. АНИСИМОВ А.А., ТАРАРЫКИН С.В. *Особенности синтеза параметрически грубых систем модального управления с наблюдателями состояния // Известия РАН. Теория и системы управления.* – 2012.– №5.– С. 3–14.
5. АРИСТОВ А.О. *Теория квазиклеточных сетей.* – М: МИ-СиС, 2014. – 188 с.
6. БАШАРИН А.В., НОВИКОВ В.А., СОКОЛОВСКИЙ Г.Г. *Управление электроприводами.* – Л.: Энергоиздат, 1982. – 392 с.
7. ВАСИЛЬЕВ Е.М., ГУСЕВ К.Ю. *Модальное управление нестационарными системами // Вестник Воронежского государственного технического университета.* – 2008. – Т. 4, №8. – С. 46–54.
8. ВАСИЛЬЕВ Е.М., ТАРАТЫНОВ О.Ю. *Алгоритмы управления тепловлажностной обработкой бетонных изделий // Вестник Воронежского государственного технического университета.* – 2012. – Т. 8, №2. – С. 13–16.
9. ВЕСЕЛОВ Г.Е., ПОПОВ А.Н., КУЗЬМЕНКО А.А. *Синергетический подход в теории управления: истоки, развитие, результаты // Тр. X Всероссийской науч. конф. «Системный синтез и прикладная синергетика».* – Ростов–на–Дону, Таганрог: Изд-во ЮФУ, 2021. – С. 12–18.
10. ВИНЕР Н. *Кибернетика, или управление и связь в животном и машине.* – М.: Наука, 1983. – 344 с.
11. ВОРОНИН А.А., ГУБКО М.В., МИШИН С.П. и др. *Математические модели организаций.* – М.: Ленанд, 2008. – 360 с.
12. ГАЛИМОВ Э.М. *Феномен жизни: между равновесием и нелинейностью. Происхождение и принципы эволюции.* – М.: Едиториал УРСС, 2006. – 256 с.
13. ГРИНИН Л.Е., МАРКОВ А.В., КОРОТАЕВ А.В. *Макроэволюция в живой природе и обществе.* – М.: Либроком, 2009. – 248 с.
14. ДОКИНЗ Р. *Эгоистичный ген.* – М.: АСТ, 2013. – 512 с.

15. ЕМЕЛЬЯНОВ С.В. *Системы автоматического регулирования с переменной структурой*. – М.: Наука, 1967. – 336 с.
16. ЖМУДЬ В.А. *Адаптивные системы автоматического управления с единственным основным контуром* // Автоматика и программная инженерия. – 2014. – №2(8). – С. 106–122.
17. КИМУРА М. *Молекулярная эволюция: теория нейтральности*. – М.: Мир, 1985. – 394 с.
18. КОЛЕСНИКОВ А.А. *Синергетическая теория управления*. – М.: Энергоатомиздат, 1994. – 334 с.
19. КОЛЕСНИКОВ А.А. *Синергетическая теория управления: концепция, методы, тенденции развития* // Известия Южного федерального университета. Технические науки. – 2001. – Т. 23, №5. – С. 7–27.
20. КОЛЕСНИКОВ А.А. *Синергетические методы управления сложными системами: теория системного синтеза*. – М.: КомКнига, 2006. – 240 с.
21. КОЛМОГОРОВ А.Н. *Три подхода к определению понятия «количество информации»* // Проблемы передачи информации. – 1965. – Т. 1, вып. 1. – С. 3–11.
22. КРАСНОВА С.А., УТКИН В.А., УТКИН А.В. *Блочный синтез систем управления роботами-манипуляторами в условиях неопределённости*. – М.: Ленанд, 2014. – 208 с.
23. КРАСНОВА С.А. *Каскадный синтез наблюдателей состояния для нелинейных систем при наличии возмущений* // Автоматика и телемеханика. – 2003. – №1. – С. 3–26.
24. КРАСОВСКИЙ А.А. *Проблемы физической теории управления* // Автоматика и телемеханика. – 1990. – №11. – С. 3–28.
25. КУЗОВКОВ Н.Т. *Модальное управление и наблюдающие устройства*. – М.: Машиностроение, 1976. – 184 с.
26. ЛЁТОВ А.М. *Условно устойчивые регулируемые системы (об одном классе оптимальных регулируемых систем)* // Автоматика и телемеханика. – 1957. – №7. – С. 601–604.
27. МАКАРОВ В.Л., РУБИНОВ А.М. *Математическая теория экономической динамики и равновесия*. – М.: Наука, 1973. – 336 с.
28. НЕЙМАН ДЖ. *Теория самовоспроизводящихся автоматов*. – М.: URSS, 2010. – 384 с.

29. НОВИКОВ Д.А. *Кибернетика: Навигатор. История кибернетики, современное состояние, перспективы развития.* – М.: Ленанд, 2016. – 160 с.
30. НОВИКОВ Д.А. *Методология управления.* – М.: Либроком, 2011. – 128 с.
31. НОВИКОВ Д.А. *Теория управления образовательными системами.* – М.: Народное образование, 2009. – 452 с.
32. НОВИКОВ С.П. *Вторая половина XX века и ее итог: кризис физико-математического сообщества в России и на Западе* // Вестник ДВО РАН. – 2006. – Вып. 4. – С. 3–22.
33. НОВИКОВ А.М., НОВИКОВ Д.А. *Методология.* – М.: Синтег, 2007. – 668 с.
34. *Новые концепции общей теории управления: Сборник научных трудов* / Под ред. А.А. Красовского. – М.-Таганрог: ТРТУ, 1995. – 183 с.
35. ПОДВАЛЬНЫЙ С.Л., ВАСИЛЬЕВ Е.М. *Матричная репликация в NP-полных задачах комбинаторной оптимизации* // Вестник Воронежского государственного технического университета. – 2022. – Т. 18, №4. – С. 7–14.
36. ПОДВАЛЬНЫЙ С.Л. *Многоальтернативные системы: обзор и классификация* // Системы управления и информационные технологии. – 2012. – №2. – С. 4–13
37. ПОДВАЛЬНЫЙ С.Л., ВАСИЛЬЕВ Е.М. *Многоальтернативное управление открытыми системами: концепция, состояние и перспективы* // Управление большими системами. Вып. 48. – М.: ИПУ РАН, 2014. – С. 6–58.
38. ПОДВАЛЬНЫЙ С.Л., ВАСИЛЬЕВ Е.М. *Многоальтернативность: эволюционная стратегия биологических систем* // Управление большими системами. – 2019. Вып. 77. – С. 125–170.
39. ПОДВАЛЬНЫЙ С.Л., ВАСИЛЬЕВ Е.М. *Системы многоальтернативного управления: эволюционный подход.* – Старый Оскол: Изд-во ТНТ, 2023. – 376 с.
40. ПОДВАЛЬНЫЙ С.Л., ВАСИЛЬЕВ Е.М. *Многоуровневая сигнальная адаптация в нестационарных системах управления* // Вестник Воронежского государственного технического университета. – 2022. – Т. 18, №5. – С. 38–47.

41. ПОДВАЛЬНЫЙ С.Л., ЗАХВАТОВ В.И., ВАСИЛЬЕВ Е.М. *Принцип разделения функций в задачах модального управления* // Вестник Воронежского государственного технического университета. – 2022. – Т. 18, №1 – С. 7–16.
42. ПОДВАЛЬНЫЙ С.Л., ВАСИЛЬЕВ Е.М. *Синтез систем высокого порядка с сигнальной адаптацией* // Вестник Воронежского государственного технического университета. – 2024. – Т. 20, №3. – С. 7–14.
43. ПРАНГИШВИЛИ И.В. *Системный подход и общесистемные закономерности*. – М.: СИНТЕГ, 2000. – 528 с.
44. ПРАНГИШВИЛИ И.В. *Энтропийные и другие системные закономерности*. – М.: Наука, 2003. – 428 с.
45. РЕДЬКО В.Г. *Эволюционная кибернетика*. – М.: Наука, 2003. – 155 с.
46. РЕДЬКО В.Г. *Эволюция, нейронные сети, интеллект. Модели и концепции эволюционной кибернетики*. – М.: Либроком, 2013. – 224 с.
47. ТИМОФЕЕВ-РЕСОВСКИЙ Н.В., ВОРОНЦОВ Н.Н., ЯБЛОКОВ А.В. *Краткий очерк теории эволюции*. – М.: Наука, 1977. – 303 с.
48. УТКИН В.И. *Метод разделения движений в задачах наблюдения* // Автоматика и телемеханика. – 1990. – №3. – С. 27–37.
49. УТКИН В.И. *Скольльзящие режимы и их применение в системах с переменной структурой*. – М.: Наука, 1974. – 272 с.
50. УТКИН В.И. *Условно устойчивая система с переменной структурой в работе А.М. Лётова* // Автоматика и телемеханика. – 2011. – №11. – С. 140–142.
51. ФИЛИМОНОВ Н.Б. *Методологический кризис «всепобеждающей математизации» современной теории управления* // Мехатроника, автоматизация, управление. – 2016. – Т. 17, №5. – С. 291–300.
52. ХАКЕН Г. *Синергетика*. – М.: Мир, 1980. – 404 с.
53. ХАКЕН Г. *Информация и самоорганизация: макроскопический подход к сложным системам*. – М.: УРСС, 2014. – 320 с.
54. ЭЙГЕН М. *Самоорганизация материи и эволюция биологических макромолекул*. – М.: Мир, 1973. – 224 с.

55. ЭШБИ У.Р. *Введение в кибернетику*. – М.: КомКнига, 2005. – 432 с.
56. ЭШБИ У.Р. *Конструкция мозга. Происхождение адаптивного поведения*. – М.: Изд. иностр. лит., 1962. – 397 с.
57. ABBOTT D. *The reasonable ineffectiveness of mathematics* // Procs. of the IEEE. – 2013. – Vol. 101, No. 10. – P. 2147–2153.
58. COPE E.D. *The primary factors of organic evolution*. – Sydney: Wentworth Press, 2016. – 588 p.
59. HEYLIGHEN F., JOSLYN C. *Cybernetics and Second-Order Cybernetics* // Encyclopedia of Physical Science & Technology. – New York: Academic Press, 2001. – P. 155–170.
60. KESSLER C. *Ein Beitrag zur theorie mehrschleifiger regelungen* // At-Automatisierungstechnik. – 1960. – Vol. 8, No. 1–12. – P. 261–266.
61. KROC J., JIMÉNEZ-MORALES F., GUISTADO J.L. et al. *Building efficient computational cellular automata models of complex systems: background, applications, results, software, and pathologies* // Advances in Complex Systems. – 2019. – Vol. 22, No. 5. – 1950013. – 38 p.
62. *Mechanism design and management: mathematical methods for smart organizations* / Ed. by Prof. D. Novikov. – New York: Nova Science Publishers, 2013. – 163 p.
63. PODVALNY S.L., VASILJEV E.M. *Analytical synthesis of aggregated regulators for unmanned aerial vehicles* // Journal of Mathematical Sciences. – 2019. – Vol. 239, No. 2. – P. 135–145.
64. ZHANG J., XIONG X., WANG Y. et al. *Simulation model for cascading failure in complex network: a cellular automata approach* // Proc. of the 2nd World Symposium on Software Engineering. – New York: Publisher Association for Computing Machinery, 2020. – P. 274–277.

EVOLUTIONARY METHODOLOGY FOR MULTI-ALTERNATIVE CONTROL OF COMPLEX SYSTEMS

Semen Podvalny, Voronezh State Technical University, Voronezh, Doctor of Science, professor (spodvalny@yandex.ru).

Eugeny Vasiljev, Voronezh State Technical University, Voronezh,
Doctor of Science, senior researcher (vgtu-aits@yandex.ru).

Abstract: The article presents the results of studies devoted to the property of multi-alternative structure and functioning of complex biological systems. The purpose of the work is: generalization of these results in the form of an evolutionary methodology of multi-alternative control of large artificial systems, identification of the place of this methodology among known cybernetic principles of control, and determination of its capabilities in practical implementation. It is noted that the relevance of the above-mentioned generalization is determined by the urgent need to systematize various areas within the framework of cybernetics itself as a science. The cybernetic basis of the methodology of multi-alternative control is revealed, consisting in the reproduction of evolutionary mechanisms of living systems that go beyond the mechanism of natural selection and are traditionally considered as a property of the diversity of these systems. Based on the analysis of the content of these mechanisms, fairly simple principles of the proposed methodology are formulated: the principle of diversity and separation of functions, the principle of multi-level structure and the principle of modularity. The constructive nature of these principles is shown, directly indicating the methods and ways of their practical implementation. A generalized model of a multi-alternative control system is presented, and a description of its functional units is given. The system-forming role of the multi-alternative control methodology in relation to a number of basic cybernetic principles of control of false objects is shown, consisting in identifying the general evolutionary content of these principles. Specific examples of the application of the methodology for the synthesis of control systems for various objects are considered, and its high efficiency is demonstrated.

Keywords: evolutionary cybernetics, cybernetic principles of evolution, methodology of multi-alternative control.

УДК 007:575.85

ББК 22.18 28.02

*Статья представлена к публикации
членом редакционной коллегии А.К. Погодаевым.*

Поступила в редакцию 28.02.2025.

Опубликована 31.05.2025.

ЭФФЕКТИВНЫЙ МЕТОД ЧИСЛЕННОГО РЕШЕНИЯ ОБЪЕМНЫХ ИНТЕГРАЛЬНЫХ УРАВНЕНИЙ ФРЕДГОЛЬМА ДЛЯ ЗАДАЧ МОДЕЛИРОВАНИЯ РАСПРОСТРАНЕНИЯ АКУСТИЧЕСКИХ ВОЛН

Юрченков И. А.¹

(МИРЭА – Российский технологический университет,
Москва)

Целью исследования является разработка численной схемы с использованием итерационных методов решения систем уравнений для решения объемных задач акустики с неоднородным индексом рефракции. Приводится постановка задачи распространения акустических волн в виде объемного интегрального уравнения Фредгольма второго рода. Для дискретизации задачи с целью последующего численного решения использована структурированная объемная прямоугольная сетка. С использованием дискретизации постановка задачи сводится к дискретизированному оператору в виде системы уравнений с большим количеством неизвестных и матрицей оператора высокой размерности. Учитывая особенности интегральных ядер уравнения Гельмгольца в интегральной форме, приводятся численные методы решения систем уравнений с использованием модификаций матрично-векторного умножения трёхплечевых матриц на вектор на основе быстрого дискретного преобразования Фурье. Продемонстрированы численные результаты работы комплекса программ моделирования распространения реализаций модели плоской волны в объемной среде с неоднородным индексом рефракции. Особое внимание уделяется возможности быстрого решения задач математической физики на структурированной сетке большой размерности, что позволит рассматривать особенности решения на сложных неоднородных границах, а также упростить аппроксимацию решения. В завершение будут сделаны выводы о качестве получаемых решений на различных примерах неоднородностей рассматриваемой объемной области.

Ключевые слова: объемные интегральные уравнения, задача акустики, уравнение Фредгольма, итерационные методы, быстрое преобразование Фурье, уравнение Гельмгольца.

1. Введение

В настоящем исследовании рассматриваются задачи распространения и рассеяния акустических волн, которые формулируются в виде объемного интегрального уравнения Фредгольма второго рода [3, с. 19]. Это уравнение получается напрямую из

¹ Иван Александрович Юрченков, аспирант, *ст. преп.* (yurchenkov@mirea.ru).

дифференциального уравнения Гельмгольца, описывающего поведение волновых процессов в выбранной среде. При численном решении такого класса задач возникает необходимость в дискретизации пространства, что ведет к большим системам линейных алгебраических уравнений (СЛАУ).

В представленном исследовании обсуждается специфика итерационных алгоритмов решения крупных систем линейных алгебраических уравнений (СЛАУ), возникающих при численном моделировании волновых процессов методом коллокации на равномерной сетке. Ключевые критерии эффективности (по памяти и времени вычисления) формализуются соотношениями

$$(1) \quad M \sim M_A, \quad T \sim L \cdot T_A,$$

где M_A – количество элементов массива значений матрицы системы, которое необходимо хранить в памяти вычислительного устройства для достижения решения; L – количество итераций для получения решения с заданной точностью; T_A – число арифметических операций для умножения матрицы системы уравнений на вектор.

Скорость алгоритма оценивается самой требовательной операцией для вычисления, коей является умножение матрицы на вектор для получения очередной итерации приближения искомого решения. Параметр T_A является основным критерием для оценки скорости алгоритма, и для каждого итерационного метода его стараются минимизировать, как и значение числа итераций L без потери качества решения конечной задачи.

При переходе к численному решению используется метод коллокации на равномерной сетке [2, с. 32], который позволяет аппроксимировать искомые функции в дискретных узлах. В результате такой дискретизации получается система линейных алгебраических уравнений (СЛАУ), чья матрица обладает почти трёхдиагональной структурой, т.е. элементы зависят главным образом от разности индексов строк и столбцов.

Объемные интегральные уравнения Фредгольма второго рода возникают при моделировании задач рассеяния акустических волн на неоднородных структурах. Такие уравнения выводятся из дифференциального уравнения Гельмгольца, которое описывает гармонические волновые процессы [12, с. 34]. Неоднородность задач понимается в выражении рефракции среды

в виде функции координат области решения задачи. Для задачи распространения акустических волн в форме интегрального уравнения также доказаны теоремы существования и единственности решения [6, с. 1261].

Для аппроксимации интегральных уравнений применяются методы коллокации на равномерных или адаптивных сетках. В случае сложных границ или неоднородностей используется неравномерная сетка, которая затем преобразуется во вспомогательную равномерную через интерполяцию. Это позволяет сохранить структуру матрицы системы линейных алгебраических уравнений (СЛАУ) близкой к блочно-тёплицевой, что критично для эффективного умножения матрицы на вектор [8, с. 879]. Метод коллокации на равномерной сетке обеспечивает высокую точность за счет использования B -сплайнов или кубических сплайнов, которые минимизируют погрешность интерполяции [5]. Например, в работах [8, с. 879, 9] показано, что переход к равномерной сетке с числом узлов, сопоставимым с исходной сеткой, сохраняет вычислительную эффективность.

Для эффективного решения систем линейных алгебраических уравнений (СЛАУ), которые характеризуются большим количеством неизвестных, часто используются итерационные методы, позволяющие существенно сократить вычислительные затраты и упростить процедуру нахождения точного или приближённого решения.

Среди итерационных методов заметную роль играет обобщенный метод минимальных невязок GMRES [14], отличающийся высокой эффективностью при решении широкого класса задач, а также стабилизированный метод бисопряженных градиентов (BiCGSTAB) [13], предназначенный для работы с несимметричными системами, которые нередко встречаются при решении интегральных уравнений. Дополнительный интерес вызывают модификации итерационных методов градиентного спуска, о которых подробно изложено в работе [7]. Здесь авторы исследуют различные варианты алгоритмов, ориентированных на решение систем уравнений высокой размерности, и демонстрируют, что двухшаговая модификация итерационного метода градиентного спуска, основанная на принципе минимизации не-

вязки, отличается не только хорошей доказываемой сходимостью, но и практической эффективностью. Благодаря этим преимуществам данные методы могут применяться в задачах из самых разных областей, где требуется решить крупные системы уравнений с высокими требованиями к точности и быстродействию вычислений.

Полная исчерпывающая методологическая база исследования постановки задачи была предложена ранее в работе [6]. Для данного исследования предлагается модификация численного метода решения соответствующих уравнений на случай блочно-тёплицевой матрицы оператора, а также сравнительные численные результаты сходимости метода.

2. Постановка задачи

2.1. ИНТЕГРАЛЬНОЕ УРАВНЕНИЕ

Будем решать интегральное уравнение, описывающее распространение и рассеяние акустических волн на прозрачных структурах, характеризующихся индексом рефракции $\eta(x)$ [12, с. 45]:

$$(2) \quad u(x) - k_0^2 \int_Q G(R) \hat{n}(x) u(y) dy = u^0(x);$$
$$\hat{n}(x) = \eta(x) - 1; \quad G(R) = \frac{\exp(ikR)}{4\pi R};$$
$$u^0(x) = - \int_Q G(R) f(y) dy; \quad x \in Q,$$

где $G(R)$ – ядро интегрального уравнения, зависящее от разности координат $R=|x - y|$; Q – объемная область задачи, для которой ищется распространение скалярного поля $u(x)$, $u(x)$ – неизвестная функция скалярного поля в области Q , k_0 – волновое число внешнего источника излучения, $u_0(x)$ – внешнее моделируемое излучение источника $f(x)$, заданное интегралом.

Данное интегральное уравнение Фредгольма второго рода будем решать численно, применяя итерационные методы решения систем уравнений. Для дискретизации задачи будем использовать метод кусочно-постоянных аппроксимаций и коллокаций [2, с. 32], используя прямоугольную структурированную сетку Π ,

сведя уравнение к дискретизированному виду [3, с. 21], неизвестное решение u которого будет определено в дискретных точках, принадлежащих Q [4, с. 54].

В прямоугольной декартовой системе координат определим параллелепипед Π , внутри которого находится область Q . Ребра параллелепипеда параллельны осям координат, а длины ребер равны $N_l \times h_l$, $l = 1, 2, 3$, где h_l – шаги сетки по декартовым координатам. Тогда параллелепипед Π можно представить как объединение ячеек (элементарных параллелепипедов) $\Pi(p)$, $p = (p_1, p_2, p_3)$, $p_l = 0, 1, \dots, N_l - 1$. Определим область \hat{Q} как объединение N_Q ячеек, центры которых лежат внутри области Q . Узловые точки, в которых определяются значения функций, будем задавать в центрах ячеек и обозначать как $x(p)$, а значения функций в этих точках как $f(p)$.

Таким образом получим дискретный вид для задачи (2) в форме системы уравнений [11]:

$$(3) \quad u(p) - \sum_{y(q) \in Q} A(p - q) \hat{\eta}(q) u(q) = u^0(p), x(p) \in Q,$$

$$A(p - q) = k^2 \int_{\Pi(q)} \frac{\exp(ik|x(p) - y|)}{4 \cdot \pi \cdot |x(p) - y|} dy, p \neq q, A(0) = 0.$$

Поскольку точки коллокации находятся в центрах прямоугольных разбиений, то для дискретизации применяется оценка сверху по точности аппроксимации в $O(h^2)$, $h = (h_1^2 + h_2^2 + h_3^2)^{1/2}$.

Основные затраты в (3) связаны с вычислением сумм:

$$(4) \quad W(p) = \sum_{q_1=0}^{N_1-1} \sum_{q_2=0}^{N_2-1} \sum_{q_3=0}^{N_3-1} A(p_1 - q_1, p_2 - q_2, p_3 - q_3) u(q_1, q_2, q_3).$$

Ввиду особенности ядра интегрального уравнения (2) матрица системы (3) является трехуровневой блочно-тёплицевой. На основе определения значений матрицы A в точках $\hat{p} = (\hat{p}_1, \hat{p}_2, \hat{p}_3)$, $\hat{p}_l = N_l, \dots, 2N_l$, определяя таким образом прямоугольную дискретизацию Π_2 за пределами дискретизации Π [10, с. 173], а также доопределив в этих точках искомое приближение решения $u(\hat{p}_1, \hat{p}_2, \hat{p}_3)$ нулями в этих точках, получим возможность использовать быстрое умножение уже блочно-циркулянтной матрицы на вектор, ускорив вычисления (4) с помощью вычисления прямого и обратного быстрого преобразований Фурье для функций трех аргументов:

$$(5) \quad W(p) = \sum_{q_1=0}^{2N_1-1} \sum_{q_2=0}^{2N_2-1} \sum_{q_3=0}^{2N_3-1} A(p_1 - q_1, p_2 - q_2, p_3 - q_3) u(q_1, q_2, q_3),$$

$$(6) \quad W^F(k_1, k_2, k_3) = A^F(k_1, k_2, k_3) u^F(k_1, k_2, k_3), k \in \Pi_2.$$

Количество операций T_A и необходимая память M_A для реализации алгоритма в таком случае оцениваются как [10, с. 174]

$$(7) \quad T_A \sim 10N \text{ LOG}(N), \quad M_A \sim 10N, \quad N = N_1 N_2 N_3.$$

Представленная оценка асимптотически меньше полиномиальной сложности, в результате чего ожидается выигрыш в скорости решения.

Задачу (3) в дискретном виде будем решать с помощью модификации двухшагового итерационного метода градиентного спуска [7] в случае блочно-тёплицевой матрицы оператора.

2.2. ИТЕРАЦИОННЫЙ МЕТОД

Задача (3) сводится к поиску решения системы уравнений и состоит в решении операторного уравнения, записанного в символическом виде:

$$(8) \quad (H)u = b, \quad (H)u = u - A(\hat{\eta}u),$$

где (H) – заранее известный (при этом не исключается комплексный характер) матричный оператор; b – вектор правой части (также в общем случае комплексный); u – вектор, подлежащий определению.

В качестве итерационного метода решения системы (3) будем использовать модификацию двухшагового метода градиентного спуска [7] с использованием умножений (5)–(6) в (8). Итерационный метод ставит своей задачей поиск приближения $u_\tau(p) \approx u(x(p))$ неизвестной искомой функции. Итерации модифицированного итерационного метода градиентного спуска определяются следующим образом:

$$r_0 = (H)u_0 - b;$$

$$u_1 = u_0 - \|(H^*)r_0\|^2 / \|(H)(H^*)r_0\|^2 \cdot (H^*)r_0;$$

$$(9) \quad r_k = (H)u_k - b; \quad \Delta r_k = r_k - r_{k-1};$$

$$t_k \|\Delta r_k\|^2 + h_k \text{Re}(\Delta r_k, (H)(H^*)r_k) = \text{Re}(r_k, \Delta r_k),$$

$$t_k \text{Re}(\Delta r_k, (H)(H^*)r_k) + h_k \|(H)(H^*)r_0\|^2 = \text{Re}(r_k, (H)(H^*)r_k),$$

$$u_{k+1} = u_k - t_k(u_k - u_{k-1}) - h_k(H^*)r_k, \quad k = 1, 2, \dots, \tau,$$

где u_0 – начальное приближение неизвестного вектора, а u_k – приближения неизвестного вектора на итерациях.

Единственным ограничением на итерационный метод (9) является существование ограниченного обратного оператора к H . Доказательство сходимости итераций (9) представлены в работе [7].

В итерациях модифицированного итерационного метода градиентного спуска [7] также представлено действие сопряженного оператора (H^*) на вектор u . Учитывая, что по определению для матричного оператора

$$H^* = \overline{H}^T,$$

где \overline{H} – это матрица с комплексно-сопряженными элементами к элементам матрицы H , то действие оператора (8) может быть переписано в виде

$$(10) (H^*)u = u - \overline{\eta} (\overline{A}^T u),$$

$$(H^*)u = u(p) - \overline{\eta}(p) \sum_{y(q) \in Q} \overline{A}(q-p)u(q), p \in Q,$$

где $\overline{\eta}$ – также вектор рефракции с комплексно-сопряженными элементами к вектору $\hat{\eta}$.

Критерием остановки итераций (9) выберем метрику δ_k относительной ошибкой приближаемого вектора на шаге k итераций:

$$(11) \delta_k = \|u_k - u_{k-1}\| / \|b\| < \varepsilon,$$

где ε – заданная точность итераций, которая чаще всего устанавливается равной 10^{-d+1} , где d – количество значащих цифр компьютерного представления чисел с плавающей точкой.

В результате имеем метод, который эффективно справляется с задачей решения операторных уравнений [6]. В качестве апробации результатов используем в сравнении стабилизированный итерационный метод бисопряженных градиентов (BiCGStab) [13] для которого также будем использовать быстрые матрично-векторные умножения (5)–(6) и операторную форму матрицы системы (3).

3. Результаты

3.1. ПОСТАНОВКА ЭКСПЕРИМЕНТОВ

Для вычисления поставленной задачи объемного интегрального уравнения итерационными методами был разработан ком-

плекс программ с использованием библиотеки numpy. Все проделанные вычислительные эксперименты, представленные далее, были проведены на вычислительной машине с центральным процессором Intel Core i5-10400 под управлением ОС Windows 10. Среда выполнения программы – Jupyter Notebook.

Для численных результатов предложено два эксперимента. Первый эксперимент показывает работу итерационных методов решения для простой задачи распространения волн без неоднородности рефракционной характеристики среды. Второй эксперимент добавляет сферическую неоднородность с уменьшением длины плоской волны моделируемого источника излучения.

Для простоты изложения условимся сокращать названия итерационных методов, применяемых в экспериментах. Для стабилизированного метода бисопряженных градиентов (BiCGStab) [1] примем сокращение в виде BG, а для модификации двухшагового метода градиентного спуска (TwoSGD) примем сокращение TS.

Для сравниваемых итерационных методов в процессе итераций будем измерять их метрики качества сходимости от количества операторных умножений m на итерациях k , т.е. $k \times m$. Первой метрикой является норма невязки итераций $r_k = (H)u_k - b$, второй метрикой станет значение критерия останова δ_k . В завершение для методов сравним значения последней невязки системы r_τ и замерим время работы алгоритма.

Главным отличием реализаций методов является возможность минимизации числа операторных умножений (5)–(6), (8), (10) на одной итерации приближения как самой ресурсоемкой операции. Для метода BG количество операторных умножений равно двум [13], в то время как для метода TS оно равно трём [7]. В этом смысле установим разное максимальное количество итераций методов. Для TS установим количество итераций $\tau_1 = 50$, в то время как для BG установим $\tau_2 = 75$. Целевое значение для остановки метода для δ_k установим, как $\varepsilon = 10^{-7}$.

Прямоугольная область решения Q характеризуется линейными размерами $L = 3$ по каждой из декартовых осей $x = (x_1, x_2, x_3)$, а также точкой центра кубической области $c = (0, 0, 0)$ в начале координат. Дискретизация объемной обла-

сти решения выбрана как $N_j = 100$ элементов по каждому из измерений $j = 1, 2, 3$, т.е. итоговое количество неизвестных равно $N_j = 10^6$, что в действительности отвечает системе с большим количеством неизвестных. Реализация численных экспериментов с таким количеством неизвестных является задачей моделирования с достаточно большими требованиями к вычислениям.

Внешнее излучение $f(x)$ из (2) моделируется в виде плоской волны комплексных гармонических колебаний:

$$(12) f_0(x) = \exp(ik_0(x, \mathbf{d})), \quad x \in Q,$$

с заданным значением волнового числа k_0 и вектором направления распространения \mathbf{d} . После вычисления $f_0(x)$ в точках $x(p)$ их значения подставляются вместо $f(y)$ в (2) для получения вектора свободных членов $u^0(p)$ системы уравнений (3). Далее данные значения будут фигурировать как вектор b свободных значений СЛАУ в итерациях (9) и итерациях метода бисопряженных градиентов.

3.1. РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ

Первый численный эксперимент проведен в объемной области Q с однородным комплексным индексом рефракции $\eta(x) = 2 + 1,5i$. Для задачи было выбрано волновое число $k_0 = 5$, вектор направления плоской волны $\mathbf{d} = (-1/\sqrt{3}, -1/\sqrt{3}, -1/\sqrt{3})$ из области положительных значений сетки в отрицательную, амплитуда колебаний волны равна $E_0 = 1$.

Для первого эксперимента отобразим график сходимости методов на рис. 1, где по оси абсцисс отложено количество операторных умножений метода $k \cdot m$, а по оси ординат – метрика δ_k в логарифмическом масштабе для детальной визуализации сходимости.

В первом эксперименте на графике сходимости рис. 1 по критерию останова видим, что метод ВГ сходится быстрее до целевого значения останова ε , чем метод ТС. В то время как метод ТС не дошел до целевого значения сходимости ε , метод ВГ дошел до него с запасом по количеству итераций. Конечная норма невязки интегрального уравнения в первом эксперименте для метода ТС приблизительно составляет $\|r_\tau\| = \|(H)u_\tau^{TS} - b\| = 0,22 \times 10^{-4}$,

в то время как для метода BG $\|r_i\| = \|(H)u_i^{BG} - b\| = 0,26 \times 10^{-5}$. Вторым методом по конечной норме невязки системы оказался приоритетным по сравнению с первым.

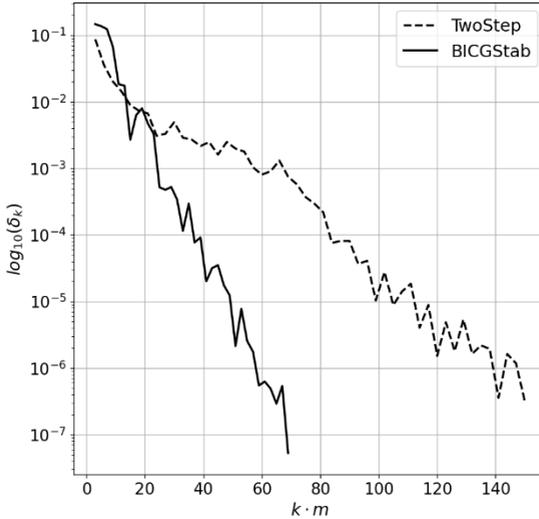


Рис. 1 Сходимость итерационных методов, оцененная по динамике δ_k относительно числа итераций k в первом эксперименте

Рядом покажем трёхмерную визуализацию полученного численного решения для дискретизации скалярного поля методом TS (рис. 2). Цветом на рисунке показано вычисленное значение амплитуды вещественных значений скалярного поля в точках дискретизации области П. На визуализации затухающего скалярного поля видим хорошую ожидаемую картину решения, соответствующую обоим методам.

Второй численный эксперимент проведен на неоднородности в форме окружности единичного радиуса. Волновое число для внешнего излучения $k_0 = 20$, вектор направления плоской волны $\mathbf{d} = (-1/\sqrt{3}, -1/\sqrt{3}, -1/\sqrt{3})$ из области положительных значений сетки в отрицательную, амплитуда колебаний волны равна $E_0 = 1$.

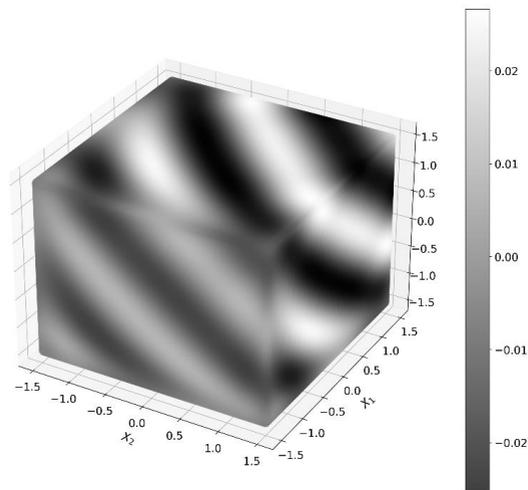


Рис. 2 Приближение скалярного поля $u(x(p))$ в объемной дискретизации Π , найденное с помощью модификации двухшагового метода градиентного спуска в первом эксперименте

Внутри окружности единичного радиуса значение коэффициента рефракции равно $\eta_2(x) = 4 + 1,5i$, вне окружности на остальной объемной области данное значение равно $\eta_1(x) = 2 + 0i$. Таким образом индекс рефракции для задачи составлен в виде кусочно-постоянной функции координат:

$$\eta(x) = \begin{cases} \eta_1(x) = 2 + 0i, & x_1^2 + x_2^2 + x_3^2 > 1, \\ \eta_2(x) = 4 + 1.5i, & x_1^2 + x_2^2 + x_3^2 \leq 1. \end{cases}$$

Для второго эксперимента отобразим график сходимости методов на рис. 3, где по оси ординат отложим уже норму невязки r_k в логарифмическом масштабе для детальной визуализации сходимости.

Во втором эксперименте по графику сходимости на рис. 3 видим, что результаты работы метода TS являются более удовлетворительными ввиду того, что сходимость происходит гладко и без скачков, а конечная норма невязки системы минимальна на интервале. Метод BG не достиг сходимости в процессе итераций, что видно на графике. Характер сходимости после достижения

минимального значения невязки резко меняется, что может быть связано с реализацией быстрых умножений матрицы на вектор.

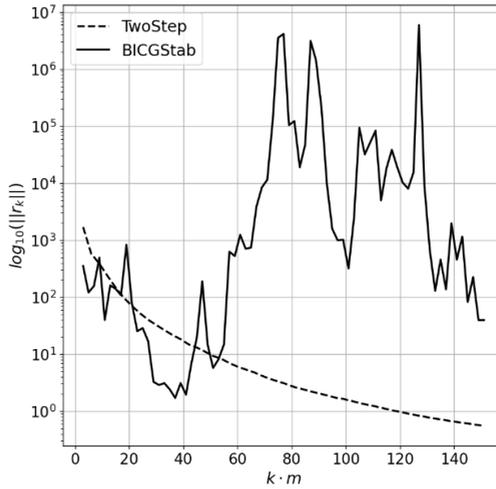


Рис. 3 Сходимость итерационных методов, оцененная по динамике $\|r_k\|$ относительно числа итераций k во втором эксперименте

Конечная норма невязки интегрального уравнения для метода TS приблизительно составляет $\|r_\tau\| = \|(H)u_\tau^{TS} - b\| = 0,7333$, в то время как для стабилизированного метода BG составляет $\|r_\tau\| = \|(H)u_\tau^{BG} - b\| = 6,2869$, что показывает более качественное приближение решения первого рассматриваемого метода относительно второго.

Далее на рис. 4, 5 покажем сравнение полученного численного решения для дискретизации скалярного поля методом TS (рис. 4) и BG (рис. 5) в виде двумерных диаграмм уровня на среде значения дискретизации $x_3 = 0$. Цветом на рисунке показано вычисленное значение амплитуды вещественных значений u_τ .

На визуализации двумерного среза решения на рис. 4 отметим более качественную картину решения для скалярного поля, полученного методом TS. Решение, полученное методом BG на рис. 5 не отвечает дискретизированной системе по конечному

значению нормы невязки, что также подтверждается визуализацией.

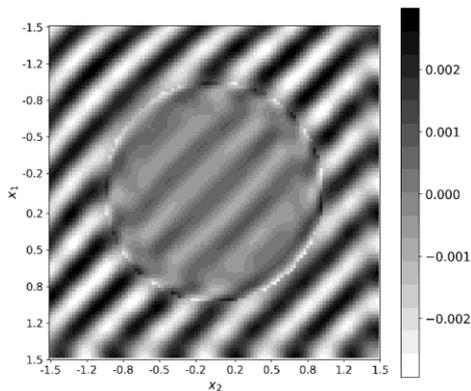


Рис. 4 Визуализация двумерного среза решения в плоскости x_1Ox_2 для вещественных значений приближения u_τ , полученных методом TS

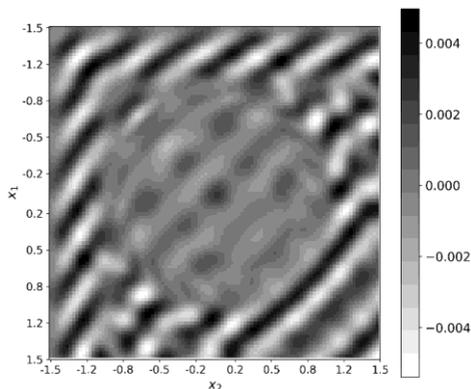


Рис. 5 Визуализация двумерного среза решения в плоскости x_1Ox_2 для вещественных значений приближения u_τ , полученных методом BG

В таблице 1 представлено время работы методов и конечная норма невязки по завершению итераций на экспериментах.

Таблица 1. Сводная таблица экспериментов.

Методы	TwoSGD		BiCGStab	
	Время работы, с	Невязка	Время работы, с	Невязка
Эксперимент № 1	172,9	$0,22 \times 10^{-4}$	86,68	$0,26 \times 10^{-5}$
2	171,9	0,6333	164,5	6,2869

Из таблицы видим, что время работы алгоритма BiCGStab меньше, чем для алгоритма TwoSGD, однако отметим более стабильную сходимость второго метода в условиях повышающихся значений волнового числа k_0 .

В результате реализации метода получаем возможность моделировать процессы, связанные с распространением или рассеянием скалярного поля в области неоднородности с реально высокой дискретизацией ($N_Q > 10^6$). На это влияет настройка параметра дискретизации N_i , $i = 1, 2, 3$, количество разбиений объемной области решения задачи вдоль каждой из осей декартовой системы координат.

4. Обсуждение

Условием применимости данного вычислительного метода заключается в требовании наличия ядра интегрального оператора, зависящего от разности координат, ввиду чего основные вычисления сопряжены с вычислениями коэффициентов оператора в виде блочно-теплицевой матрицы. В представленном численном эксперименте наглядно продемонстрирован эффективный подход к решению такой постановки задач, а также качественные картины решения задач распространения акустической волны в среде с затуханием и рассеяния волны на границе рефракции прозрачных объемных сред.

Примечательным является то, что требований к структуре функции рефракции, за исключением физических ограничений моделируемых материалов, не ставится, что позволяет производить математическое моделирование представленных выше задач на областях высокой сложности. Основным ограничением будет являться размер области, на которой будет задана функция.

Предложенный для решения задачи модифицированный итерационный метод градиентного спуска показал себя как эффективный метод решения операторного уравнения с плотной матрицей оператора по количеству арифметических операций, времени вычисления на однопроцессорном выполнении, а также эффективности использования памяти системы.

Для реализации алгоритма требуется хранить и использовать для вычисления не квадратные матрицы огромной размерности, а только компоненты системы, необходимые для построения трёхлинейного оператора до циркулянтного. Этот аспект позволяет рассматривать решение задач на сверхбольших сетках дискретизации с количеством неизвестных более миллиона на одном вычислительном узле в задачах моделирования с точечными неоднородностями.

По результатам численного эксперимента показано, что разработанный метод устойчиво сходится при изменении размерности задачи, а также стабильно сходится за разумное число умножений матрицы на вектор. Полученная численная схема позволяет серьезно ускорить вычисления в общих постановках решения СЛАУ или операторных уравнений, а с применением специальной дискретизации также позволит решать большие по числу неизвестных прикладные задачи моделирования.

5. Заключение

Представленные в работе результаты и обсуждение подтверждают высокую эффективность разработанного подхода к моделированию волновых процессов в средах со сложной рефракционной структурой. Возможность описывать широкий класс включений прямоугольными параллелепипедами в сочетании с реализацией плотных матриц большого размера позволяет решать многомерные акустические задачи при различных физических ограничениях на свойства материалов.

Ключевым моментом при этом является использование интегрального оператора с ядром, зависящим от разности координат, что упрощает приведение к задаче с трёхлинейной матрицей и обеспечивает значительный выигрыш в вычислительной производительности.

Литература

1. БУКАНСУС С., МАНДЭ С., ТАИР Б. и др. *Построение обобщенных итерационных методов, используемых для решения интегрального уравнения Фредгольма* // Вычислительные методы и программирование. – 2022. – Т. 23, №4. – С. 350–364.
2. ДЕМИДОВИЧ Б.П. *Численные методы анализа. Приближение функций, дифференциальные и интегральные уравнения: учебное пособие.* – СПб.: Лань, 2022. – 400 с.
3. КУДРЯШОВА Н.Ю. *Граничные интегральные уравнения: учебное пособие.* – Пенза: ПГУ, 2018. – 72 с.
4. НУРУТДИНОВА И.Н., ПОЖАРСКИЙ Д.А. *Интегральное исчисление. Дифференциальные уравнения. Ряды: учебное пособие.* – Ростов-на-Дону: Донской ГТУ, 2021. – 96 с.
5. ТАИР Б., СЕГНИ С., ГИББИ Х. и др. *Два численных метода решения линейного интегро-дифференциального уравнения Фредгольма со слабо сингулярным ядром* // Вычислительные методы и программирование. – 2022. – Т. 23, №2. – С. 117–136.
6. САМОХИН А.Б., САМОХИНА А.С., ЮРЧЕНКОВ И.А. *Интегральное уравнение Фредгольма для задач акустического рассеяния на трёхмерных прозрачных структурах* // Дифференциальные уравнения. – 2023. – Т. 59, №9. – С. 1260–1265.
7. САМОХИН А.Б., САМОХИНА А.С., ШЕСТОПАЛОВ Ю.В. и др. *Итерационные методы градиентного спуска для решения линейных уравнений* // Журнал вычислительной математики и математической физики. – 2019. – Т. 59, №8. – С. 1331–1339.
8. САМОХИН А.Б. *Численный метод решения объемных интегральных уравнений на неравномерной сетке* // Журнал вычислительной математики и математической физики. – 2021. – Т. 61, №5. – С. 878–884.
9. САМОХИН А.Б., ЮРЧЕНКОВ И.А. *Численные методы решения стационарных объемных интегральных уравнений Фредгольма второго рода для задач распространения и рассеяния электромагнитных волн* // Материалы Всероссийской открытой научной конференции «Современные проблемы дистанционного зондирования, радиолокации, распростране-

ния и дифракции волн: Всероссийские открытые Армандовские чтения», Муром, 25–27 июня 2024 года. – Муром: Владимирский государственный университет им. Александра Григорьевича и Николая Григорьевича Столетовых, 2024. – С. 103–112.

10. САМОХИН А.Б. *Объемные сингулярные интегральные уравнения электродинамики: монография.* – М.: Техносфера, 2021. – 218 с.
11. САМОХИН А.Б. *Методы и эффективные алгоритмы решения многомерных интегральных уравнений* // Russian Technological Journal. – 2022. – Т.10, №6. – С. 70–77.
12. COLTON D., KRESS R. *Inverse acoustic and electromagnetic scattering theory.* – Springer-Verlag, Berlin, 1992. – 328 p.
13. SCHOUTROP C.E.M., TEN THIJJE BOONKKAMP J.H.M., VAN DIJK J. *Reliability Investigation of BiCGStab and IDR Solvers for the Advection-Diffusion-Reaction Equation.* // Communications in Computational Physics. – 2022. – Vol. 32, No. 1. – P. 156–188.
14. ZOU Q. *GMRES algorithms over 35 years* // Applied Mathematics and Computation. – 2023. – Vol. 445. – P. 127869.

AN EFFICIENT METHOD FOR NUMERICAL SOLUTION OF VOLUME INTEGRAL FREDHOLM EQUATIONS FOR ACOUSTIC WAVE PROPAGATION MODELING PROBLEMS

Ivan Yurchenkov, MIREA – Russian Technological University, Moscow, Graduate Student, Senior Lecturer (yurchenkov@mirea.ru).

Abstract: The purpose of the research is to develop a numerical scheme using iterative methods for solving systems of equations for solving bulk acoustic problems with inhomogeneous refraction index. The paper presents a formulation of the acoustic wave propagation problem in the form of a volume integral Fredholm equation of the second kind. A structured volume rectangular mesh is used to discretize the problem for the purpose of subsequent numerical solution. Using discretization, the problem formulation is reduced to a discretized operator in the form of a system of equations with a large number of variables and an operator matrix of high dimensionality. Taking into account the peculiarities of the integral kernels of the Helmholtz equation in integral form, numerical methods for solving the systems of equations using modifications of the matrix-vector multiplication of Toeplitz matrices by a vector based on the

fast discrete Fourier transform are given. Numerical results of a set of programs for modeling propagation realizations of a plane wave model in a volumetric medium with inhomogeneous refraction index are demonstrated. Special attention in this paper is paid to the possibility of fast solution of mathematical physics problems on a structured grid of high dimensionality, which will allow us to consider the features of the solution on complex inhomogeneous boundaries, as well as to simplify the approximation of the solution. Finally, conclusions will be drawn about the quality of the obtained solutions on different examples of inhomogeneities of the considered volume domain.

Keywords: volume integral equations, acoustic problem, Fredholm equation, iterative methods, fast Fourier transform, Helmholtz equation.

УДК 519.642+51-73

ББК 22.19

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

Поступила в редакцию 17.02.2025.

Опубликована 31.05.2025.

ОПТИМИЗАЦИЯ РАСПРЕДЕЛЕНИЯ ВОЗДУШНЫХ СУДОВ ПО АВИАЛИНИЯМ ПРИ НЕЧЁТКИХ ИСХОДНЫХ ДАННЫХ

Романенко В. А.¹

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

Решен вариант задачи оптимальной расстановки воздушных судов (ВС) заданных типов на заданной сети авиалиний. Распределение ВС по авиалиниям является одной из узловых проблем, определяющей эффективность деятельности авиакомпании. Рассмотренная задача состоит в определении для каждой авиалинии недельного числа рейсов ВС, обеспечивающего максимальный экономический эффект от перевозки пассажиров. Новизна постановки задачи заключается в том, что уровни спроса на пассажирские перевозки приняты не полностью определёнными экспертно заданными нечёткими величинами, что соответствует этапу предварительного проектирования расписания. Оптимизация расстановки ВС по авиалиниям сформулирована как целочисленная задача математического программирования с нечётким критерием и чёткими ограничениями. Благодаря использованию приёма дефазификации нечёткая задача сведена к обычной задаче математического программирования, решаемой имеющимися эффективными методами на базе доступного программного обеспечения. С использованием программного пакета IBM ILOG OPL получено решение модельных примеров задачи в нечёткой и «чёткой» постановках. Выполненное сравнение выявило существенные различия наиболее значимых результатов решения оптимизационной задачи в нечёткой и «чёткой» постановках, что свидетельствует о целесообразности учёта нечёткой неопределённости исходных данных.

Ключевые слова: оптимизация, математическое программирование, распределительная задача, нечёткое число, воздушные суда, авиалинии.

1. Введение

Рассматривается вариант одной из типичных распределительных задач математического программирования, часто встречающейся в практике планирования работы авиатранспортных предприятий. Задача состоит в оптимальной расстановке на заданной сети авиалиний воздушных судов (ВС) определённых типов таким образом, чтобы обеспечить экстремум

¹ Владимир Алексеевич Романенко, к.т.н., доцент (vla_rot@mail.ru).

некоторого показателя, имеющего, как правило, экономический смысл. Расстановка ВС по авиалиниям является одной из узловых проблем, в значительной степени определяющей эффективность эксплуатации авиакомпанией своего авиапарка. Особенностью описанной ниже оптимизационной модели является учёт неопределённого характера некоторых величин, входящих в состав исходных данных, посредством задания этих величин в нечёткой форме.

Рациональное распределение по авиалиниям ВС, имеющих в распоряжении авиаперевозчика, возможно при наличии информации относительно уровней пассажиропотоков на рассматриваемых авиалиниях. На этапе решения распределительной задачи спрос на перевозки может быть известен лишь предположительно. Для формирования оценок уровня спроса на тех авиалиниях, по которым перевозки авиакомпанией уже выполнялись, может быть использована накопленная статистика. В этом случае прогнозы спроса будут описываться в терминах теории вероятностей. Для открываемых же вновь авиалиний или при недостаточном объеме статистики источником информации, вероятно, станут мнения экспертов. Разброс мнений, неизбежно присущий экспертам, их представления об изменчивости оцениваемого параметра традиционно находят отражение в выборе нечёткой формы представления параметра, который рассматривается как нечёткая величина с экспертно заданной функцией принадлежности. Чтобы обеспечить возможность учёта неопределённости в уровнях спроса при решении оптимизационной задачи, будем рассматривать все неопределённые величины как нечёткие, предварительно приводя случайные величины к нечёткой форме представления. Использование нечётких величин вместо статистических распределений позволит упростить оптимизационную модель, сохранив возможность учёта влияния на результат неопределённости в предполагаемых уровнях пассажиропотоков. Таким образом, задача оптимизации распределения ВС по авиалиниям будет решаться при нечётких данных о пассажиропотоках.

Оптимизация распределения ВС по авиалиниям была одной из первых задач математического программирования, успешно применённых на практике еще в середине 1950-х годов. Впер-

вые её решение в детерминированной постановке было представлено в работе [15]. Практически в то же время теми же авторами началась разработка методов решения рассматриваемой задачи с учётом стохастичности пассажиропотоков [3, 16]. Ранние этапы исследований в рассматриваемой области отражены в работах отечественных авторов [1, 7].

К настоящему времени разработаны методики решения задачи в весьма сложных и детальных постановках, базирующиеся на двух основных методических подходах. Первый подход, основанный на предложенной и развитой в работах [9, 24] пространственной модели, сводит рассматриваемую задачу к задаче целочисленного программирования при возможности учёта движения многочисленного парка ВС на разветвлённой сети авиалиний со многими промежуточными посадками. Ограничения накладываются на уровень спроса и провозные ёмкости. Могут учитываться также временные ограничения на параметры графика оборота ВС, необходимость возвращения ВС в базовый аэропорт и некоторые другие факторы. Второй подход, отличающийся большей детализацией и, соответственно, приводящий к большей размерности задачи, основан на предложенной в работе [18] пространственно-временной модели сети авиалиний. Введение бинарных переменных позволило учесть ряд важных ограничений, таких как, например, возможность «стыковки» авиарейсов, что необходимо для авиакомпаний со значительной долей трансферных перевозок. В рамках обоих подходов для эффективного решения задачи либо минимизации расходов, либо максимизации прибыли разработаны весьма изощрённые многоэтапные алгоритмы.

Рядом авторов справедливо отмечается [1, 10, 26], что задача распределения ВС по авиалиниям является лишь частью общей проблемы принятия решений в управлении авиакомпанией и должна решаться совместно с другими оптимизационными задачами. С ростом возможностей вычислительной техники наблюдается тенденция к расширению комплекса оптимизационных задач, решаемых совместно. К настоящему времени разработаны интегрированные детерминированные оптимизационные модели распределения ВС по авиалиниям совместно со следующими задачами: формирование расписания полётов

[23, 24, 27–29], формирование авиалиний [10, 11], планирование технического обслуживания ВС [12, 13], составление расписания и назначение экипажей на рейсы [10–12, 17, 26].

Эволюция исследований стохастической задачи идет по пути как расширения комплекса совместно решаемых оптимизационных задач, так и увеличения числа учитываемых случайных факторов. Если в более ранних работах [16, 22] оптимизировалось распределение ВС по авиалиниям при случайном спросе, то, например, в [20] был учтен также стохастический характер цен на авиатопливо, а в [25] распределение ВС рассмотрено параллельно с формированием авиалиний, составлением расписания полётов и планированием технического обслуживания ВС при случайном спросе и отклонениях от расписания. Комплекс задач в одной из наиболее сложных постановок представлен в статье [19], где одновременно решаются задачи распределения ВС, формирования авиалиний и составления расписания с учётом необходимости технического обслуживания ВС и наличия код-шеринговых соглашений между перевозчиками. Случайными считаются спрос, время наземной стоянки ВС и отклонения от расписания. обстоятельные обзоры работ по рассматриваемой тематике представлены в [19, 30]. Вызывает некоторое недоумение тот факт, что из пяти десятков наиболее значимых работ, упомянутых в этих обзорах, ни в одной из них неопределённость не описывается в терминах нечёткости. Настоящая работа призвана отчасти восполнить пробел в исследованиях задачи распределения авиапарка при частичной нечёткости исходных данных.

2. Описание неопределённости исходных данных

Напомним используемые ниже понятия и приемы нечёткой арифметики. Под нечётким множеством \tilde{A} на универсальном множестве U понимается совокупность кортежей вида $\langle \mu_{\tilde{A}}(u), u \rangle$, где $\mu_{\tilde{A}}(u)$ – степень принадлежности элемента $u \in U$ нечёткому множеству \tilde{A} , которая задается как действительное число из интервала $[0, 1]$. Функция, позволяющая вычислить степень принадлежности универсальному множеству

произвольного его элемента, называется функцией принадлежности. Под нечёткой величиной понимается нечёткое множество, заданное на множестве действительных чисел [5].

Ограничимся использованием нечётких величин, относящихся к типу нормальных с треугольными профилями функций принадлежности, или «треугольных нечётких чисел» (ТНЧ). ТНЧ характеризует неопределённость типа «приблизительно равно» и является одним из наиболее часто используемых, интуитивно понятных, простых и удобных для практических вычислений типов нечётких величин. ТНЧ \tilde{A} может быть представлено в виде кортежа $\tilde{A} = \langle a^L, a^M, a^R \rangle$, включающего координаты опорных точек функции принадлежности $\mu_{\tilde{A}}(u)$:

$$(1) \quad \mu_{\tilde{A}}(u) = \begin{cases} \frac{u - a^L}{a^M - a^L}, & a^L \leq u \leq a^M, \\ \frac{a^R - u}{a^R - a^M}, & a^M \leq u \leq a^R, \\ 0, & u \leq a^L, a^M \leq u, \end{cases}$$

где a^M – наиболее возможное значение u (мода); a^L, a^R – соответственно наименьшее и наибольшее из возможных значений u (соответственно левая и правая границы ТНЧ), $a^L \leq a^M \leq a^R$.

Разности $\Delta a^L = a^M - a^L$ и $\Delta a^R = a^R - a^M$ называются соответственно левым и правым коэффициентами нечёткости, а их отношения к модальному значению a^M – левым k^{a^L} и правым k^{a^R} относительными коэффициентами нечёткости:

$$k^{a^L} = \frac{\Delta a^L}{a^M}, \quad k^{a^R} = \frac{\Delta a^R}{a^M}.$$

ТНЧ \tilde{A} принято называть симметричным (СТНЧ) в случае равенства величин его левых и правых коэффициентов:

$$(2) \quad \Delta a^L = \Delta a^R = \Delta a, \quad k^{a^L} = k^{a^R} = k^a.$$

Основываясь на принципе обобщения Заде [5] определим необходимые ниже операции с ТНЧ. Пусть даны ТНЧ \tilde{A} и «обычное» (не нечёткое) число b . Результатами их нечёткого

сложения (+), перемножения (\times), взятия максимума (max) и минимума (min) будут ТНЧ, определяемые соответственно как

$$\begin{aligned} \tilde{A} + (\times) b &= \langle a^L + (\times) b, a^M + (\times) b, a^R + (\times) b \rangle, \\ \max(\min)(\tilde{A}, b) &= \langle \max(\min)(a^L, b), \max(\min)(a^M, b), \\ (3) \quad \max(\min)(a^R, b) \rangle. \end{aligned}$$

Для дефаззификации ТНЧ, т.е. приведения к чёткой форме, используем далее метод центроида [5], в соответствии с которым чёткое число \bar{a} – результат дефаззификации ТНЧ $\tilde{A} = \langle a^L, a^M, a^R \rangle$ – определяется как

$$(4) \quad \bar{a} = \text{def}(\tilde{A}) = \frac{a^L + a^M + a^R}{3}.$$

Очевидно, что если \tilde{A} – СТНЧ, то $\bar{a} = \text{def}(\tilde{A}) = a^M$.

Примем, что неопределённость экспертного прогноза учитывается путем задания оценок уровней спроса в форме СТНЧ. Если же благодаря наличию статистических данных формируется вероятностный прогноз, подчиняющийся нормальному распределению, то СТНЧ может быть использовано для его аппроксимации. Пусть в результате обработки накопленных за предыдущие периоды статистических данных получен частотный прогноз спроса, подчиняющийся нормальному распределению. Необходимо адаптировать его для решения задачи оптимизации распределения авиапарка, трансформировав в СТНЧ.

Рассмотрим задачу формирования функции принадлежности СТНЧ на базе нормального распределения случайной величины. Предположим, что имеется нормальное частотное распределение [2] для переменной u с известными параметрами математического ожидания m , среднеквадратического отклонения σ и функцией плотности

$$(5) \quad f(u) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{u-m}{\sigma}\right)^2}.$$

Требуется определить параметры a^M и k^a СТНЧ \tilde{A} с функцией принадлежности $\mu_{\tilde{A}}(u)$ вида (1), (2), аппроксимирующего

распределение (5). В силу симметричности функции плотности $f(u)$ и функции принадлежности $\mu_{\tilde{A}}(u)$ параметр a^M СТНЧ \tilde{A} , очевидно, совпадает с параметром m нормального частотного распределения:

$$a^M \equiv m.$$

Проведём нормализацию функции плотности нормального распределения, понимая в данном случае под этим процедуру получения на основе $f(u)$ такой функции $f'(u)$, максимальное значение которой, достигаемое в точке $u = m$, равно единице. Учитывая, что для $u = m$ плотность $f(u)$ принимает значение

$$f(u = m) = \frac{1}{\sigma\sqrt{2\pi}},$$

получаем следующую нормализованную функцию

$$(6) \quad f'(u) = \frac{f(u)}{f(u = m)} = \sigma\sqrt{2\pi} f(u),$$

для которой $f'(u = m) = 1$.

В качестве наилучшего приближения нормального распределения (6) нечётким числом (1) будем рассматривать СТНЧ с $a^M \equiv m$ и такими координатами a^L и a^R , которые обеспечивают равенство двух площадей над осью абсцисс:

$$(7) \quad S_f = S_{\mu},$$

где S_f , S_{μ} – площади, ограничиваемые функцией распределения (6) и функцией принадлежности (1) соответственно.

Площадь S_f для нормального распределения:

$$(8) \quad S_f = \int_{-\infty}^{\infty} f'(u) du = \sigma\sqrt{2\pi} \int_{-\infty}^{\infty} f(u) du = \sigma\sqrt{2\pi}.$$

Величина S_{μ} определяется как площадь равнобедренного треугольника единичной высоты $\mu_{\tilde{A}}(u = m) = 1$, построенного на основании длиной $(a^R - a^L) = 2 \cdot \Delta a = 2 \cdot k^a \cdot m$:

$$(9) \quad S_{\mu} = \frac{1}{2} \cdot \mu_{\tilde{A}}(u = m) \cdot (a^R - a^L) = k^a \cdot m.$$

Приравнивая в соответствии с условием (7) выражения (8) и (9), получаем искомый относительный коэффициент нечёткости аппроксимирующего СТНЧ:

$$(10) k^a = \frac{\sigma\sqrt{2\pi}}{m}.$$

Следует отметить, что равенство площадей (7) является не единственным условием, используемым при подборе параметров нечёткого приближения нормального распределения. Например, в [8] для определения численно-аналитическим методом наилучшего нечёткого интервала, аппроксимирующего нормальное распределение, наряду с (7), использовано условие минимума интеграла квадратичных отклонений значений плотности нормального распределения и функции принадлежности нечёткого интервала. В случае с СТНЧ \tilde{A} указанное условие в общем виде запишется как:

$$(11) \int_{-\infty}^{\infty} (f'(u, m, \sigma) - \mu_{\tilde{A}}(u, a^M = m, k^a))^2 du \xrightarrow{k^a} \min.$$

Однако процедура поиска величины k^a , доставляющей минимум целевой функции (11), существенно более сложна, чем расчёт по формуле (10). При этом, как показала серия численных решений задачи (11) в широком диапазоне исходных значений m и σ , различия в величинах k^a , найденных двумя описанными выше подходами не превышают 5–6%.

3. Постановка задачи

Ограничимся одной из простых постановок задачи распределения ВС, позволяющей оценить целесообразность учёта нечёткости исходных данных и её влияние на результаты.

Пусть авиакомпания, располагающая ВС J типов, планирует выполнять беспосадочные рейсы из своего базового аэропорта в I других аэропортов. Задано количество ВС каждого (j -го, $j = 1, \dots, J$) типа n_j . Для ВС каждого типа известны дальность d_j и пассажироместимость b_j .

В соответствии с практикой предположим, что пассажиропотоки в прямом направлении – из базового аэропорта в i -й аэропорт – в общем случае отличаются от пассажиропотоков в обратном направлении. Различаются также затраты на обслуживание одного пассажира в аэропортах на рейсах в прямом

и обратном направлениях. Совпадающими примем средние тарифы на перевозку пассажира «туда» и средние тарифы на перевозку «обратно». Обозначим h направление перевозки и будем присваивать h одно из двух значений: «1» – в случае перевозки из базового аэропорта, «2» – в случае перевозки в базовый аэропорт. Поскольку цикличность повторяемости дней выполнения рейсов в расписании, как правило, соответствует недельному интервалу [6], то при необходимости будем приводить значения параметров к указанному временному интервалу. Примем заданными следующие параметры для каждой (i -й, $i = 1, \dots, D$) авиалинии: расстояние между базовым и i -м аэропортами (протяжённость авиалинии) l_i , средний тариф на перевозку пассажира в одном направлении s_i , прогнозы недельного спроса \tilde{q}_i^h и затраты c_i^h на аэропортовое обслуживание одного пассажира в прямом и обратном направлениях, $h = 1, 2$. Учтем, что сокращение фактической недельной частоты выполнения рейсов ниже минимально допустимого уровня может привести к неприемлемому с точки зрения потенциальных пассажиров увеличению промежутков времени между рейсами, что послужит причиной труднопрогнозируемого падения спроса на перевозки на данной авиалинии. Поэтому набор перечисленных выше параметров авиалинии дополним заданным минимально допустимым недельным числом рейсов k_i .

Будем считать, что авиакомпания, действуя в конкурентной среде, не является единственным перевозчиком на каждой авиалинии, поэтому пассажиры, которым не находится мест на рейсах рассматриваемой авиакомпании, имеют возможность уйти к конкурентам. Таким образом, неудовлетворенный в течение рассматриваемой недели спрос не переходит на следующие периоды и считается упущенным.

Предположим известными следующие величины, зависящие как от типа ВС j , так и от авиалинии i : затраты C_{ij} на выполнение парного («туда и обратно») рейса ВС j -го типа на авиалинии i ; максимально возможное недельное количество K_{ij} рейсов одного ВС j -го типа на авиалинии i . Затраты C_{ij} не включают расходы на обслуживание пассажиров в аэропортах. Максимальное количество рейсов K_{ij} должно быть определено зара-

нее исходя из протяжённости авиалинии, рейсовой скорости ВС, продолжительности наземной стоянки ВС в аэропортах, затрат времени на техническое обслуживание ВС и т.д.

Введём принадлежащие множеству целых неотрицательных чисел \mathbf{Z}_+ переменные x_{ij} , $i = 1, \dots, I$, $j = 1, \dots, J$, представляющие собой численности парных рейсов ВС j -го типа на авиалинии i . Назовём «плановым» недельное число пассажиров i -й авиалинии в направлении h , на перевозку которых авиакомпания будет вправе рассчитывать при некоторых x_{ij} , $i \in \{1, \dots, I\}$, $j = 1, \dots, J$. Ясно, что плановое недельное число пассажиров не сможет превысить как прогнозируемого недельного спроса \tilde{q}_i^h , так и суммарного числа мест P_i на всех рейсах в одном направлении i -й авиалинии:

$$(12) \quad P_i = \sum_{j=1}^J x_{ij} b_j .$$

Поскольку плановое недельное число пассажиров будет зависеть от ТНЧ \tilde{q}_i^h , то оно также будет ТНЧ. Обозначим это ТНЧ \tilde{p}_i^h и будем, исходя из приведённых выше соображений, определять её по правилам нечётких вычислений (3) как

$$(13) \quad \tilde{p}_i^h = \min(P_i, \tilde{q}_i^h) .$$

В качестве максимизируемой целевой функции логично выбрать операционную прибыль от выполнения перевозок, которая представляет собой разность между доходами от продаж перевозок и расходами на выполнение рейсов с учётом затрат на обслуживание пассажиров и определяется с использованием введенных выше величин и правил (3) как ТНЧ \tilde{E} :

$$(14) \quad \tilde{E} = \sum_{i=1}^I \sum_{h=1}^2 \tilde{p}_i^h s_i - \left(\sum_{i=1}^I \sum_{j=1}^J C_{ij} x_{ij} + \sum_{i=1}^I \sum_{h=1}^2 \tilde{p}_i^h c_i^h \right) \xrightarrow{\{x_{ij}\}} \max .$$

Используются следующие ограничения.

На минимальное недельное число рейсов на авиалинии, которое не может быть ниже заданного минимально допустимого недельного числа:

$$(15) \quad \sum_{j=1}^J x_{ij} \geq k_i, \quad \forall i \in \{1, \dots, I\} .$$

На максимальное количество используемых ВС j -го типа, которое не может быть меньше имеющегося в составе авиапарка перевозчика числа ВС того же типа:

$$(16) \sum_{i=1}^I \frac{x_{ij}}{K_{ij}} \leq n_j, \quad \forall j \in \{1, \dots, J\}.$$

Запрет назначения на авиалинию ВС, дальность которого меньше протяжённости авиалинии:

$$(17) x_{ij} = 0, \quad \forall i \in \{1, \dots, I\}, \quad \forall j \in \{1, \dots, J\} : l_i > d_j.$$

Оптимизационная задача с чёткими ограничениями (15)–(17) и нечёткой целевой функцией (14) в силу нечёткости последней является задачей с бесконечным числом целевых функций [4]. Поэтому формулировку (14)–(17) следует рассматривать как условную. Чтобы придать постановке задачи определённую, используем в качестве целевой функции значение \bar{E} , полученное в результате дефаззификации (4) ТНЧ \tilde{E} :

$$(18) \bar{E} = \text{def}(\tilde{E}) \rightarrow \max.$$

Таким образом, задача оптимального распределения ВС по авиалиниям сводится к определению множества $\{x_{ij} \in \mathbf{Z}_+, i = 1, \dots, I, j = 1, \dots, J\}$, обеспечивающего максимум целевой функции \bar{E} (18) и удовлетворяющего ограничениям (15)–(17) при заданных $n_j, d_j, b_j, l_i, s_i, k_i, \tilde{q}_i^h, c_i^h, C_{ij}, K_{ij}$ ($i = 1, \dots, I, j = 1, \dots, J, h = 1, 2$). Решение задачи, относящейся к классу задач целочисленного математического программирования, достигается известными «чёткими» методами математического программирования, в том числе при небольшой размерности задачи, используемыми программной надстройкой «Поиск решения» табличного процессора Microsoft Excel. Решение рассмотренных ниже примеров получено с помощью программного пакета IBM ILOG OPL.

4. Модельные примеры

Источником исходных данных для модельных примеров послужили результаты транспортной работы авиакомпаний, выполняющих перевозки из одного из крупных региональных

аэропортов РФ. Рассматривается распределение по 20 авиалиниям ($I = 20$) ВС пяти типов ($J = 5$), характеристики и численность которых представлены в таблице 1.

Исходные данные по авиалиниям сведены в таблицу 2. Величины, имеющие экономический смысл, выражены в условных денежных единицах (усл. ед.).

Таблица 1. Характеристики авиапарка

j	1	2	3	4	5
b_j , пас.	210	160	100	70	50
d_j , км	5100	5100	2950	1370	2500
n_j	2	5	5	2	2

Таблица 2. Характеристики авиалиний

i	l_i , км.	s_i , усл. ед.	$q_i^{(1)M}$, пас.	$q_i^{(2)M}$, пас.	k_i , рейсов, по вариантам			
					I	II	III	IV
1	870	6,5	14760	16400	110	104	88	54
2	1360	9,4	6312	6801	66	62	52	32
3	1350	9,3	6039	5929	31	28	25	16
4	3200	20,1	1859	1828	10	10	8	5
5	1630	11,0	1185	1261	10	10	8	5
6	2100	13,7	1118	1117	14	13	11	7
7	2350	15,2	1021	902	6	6	5	3
8	1150	8,2	938	902	8	8	6	4
9	1800	12,0	881	770	10	10	8	5
10	765	5,9	740	826	12	11	10	6
11	1600	10,8	552	481	5	5	4	3
12	1200	8,5	460	410	3	3	2	2
13	1915	12,6	397	399	2	2	2	1
14	2250	14,6	408	244	2	2	2	1
15	2005	13,2	310	331	5	5	4	3
16	2200	14,3	284	308	2	2	2	1
17	1770	11,8	257	296	2	2	2	1
18	1040	7,5	198	278	3	3	2	2
19	510	4,4	180	176	2	2	2	1
20	280	3,0	146	105	2	2	2	1
Всего			38042	39762	305	290	245	153

Для всех авиалиний затраты на обслуживание пассажира в аэропортах приняты одинаковыми: $c_i^h = 0,6$ усл. ед., $i = 1, \dots, I$. Недельный спрос на перевозки по авиалиниям задан модальными значениями $q_i^{(1)M}$ и $q_i^{(2)M}$ СТНЧ $\tilde{q}_i^{(1)}$ и $\tilde{q}_i^{(2)}$ соответственно и относительными коэффициентами нечёткости, которые для всех СТНЧ приняты одинаковыми: $k^{\tilde{q}_i^{(1)}} = k^{\tilde{q}_i^{(2)}} = 0,15$, $i = 1, \dots, I$.

Решение получено для четырёх приведённых в таблице 2 вариантов ограничений на минимальное недельное число рейсов. Вариант I, соответствующий наиболее жестким требованиям к частотам выполнения рейсов, практически совпадает с фактическим расписанием. Варианты II, III и IV допускают сокращение суммарного недельного числа рейсов на 5%, 20% и 50% соответственно.

Модельные затраты на выполнение парных рейсов C_{ij} и максимально возможные недельные численности рейсов K_{ij} одного ВС приведены в таблице 3. Прочерки соответствуют случаям недостаточной дальности ВС.

Для оценки влияния нечёткости на результаты оптимизации задача оптимального распределения ВС по авиалиниям была решена также в «чёткой» постановке с целевой функцией E , вычисляемой по формулам (12)–(14), где ТНЧ заменены обычными «чёткими» числами. В составе исходных данных вместо нечётких величин спроса \tilde{q}_i^h использованы их чёткие аналоги q_i^h , равные дефаззифицированным значениям СТНЧ \tilde{q}_i^h , совпадающим с их модальными значениями. Далее для задач в нечёткой и «чёткой» постановках использованы сокращения «НЗ» и «ЧЗ», соответственно.

Результаты для варианта ограничений I, представленные более подробно, приводятся в таблицах 4 и 5. Несмотря на то, что подсчитанные для каждой (i -й) авиалинии суммарные недельные численности рейсов x_i ВС всех типов в ЧЗ и НЗ практически совпадают, распределения числа рейсов ВС различных типов по авиалиниям в этих двух задачах заметно различаются.

В таблице 5 приводятся значения операционной прибыли, как по отдельным авиалиниям, так и суммарные, и параметры

плановых пассажиропотоков, как по каждому направлению, так и суммарные.

Таблица 3. Характеристики авиалиний для ВС различных типов

i	C_{ij} , усл. ед., по типам ВС (j)					K_{ij} , рейсов, по типам ВС (j)				
	1	2	3	4	5	1	2	3	4	5
1	1890	1440	1260	650	960	21	35	28	28	35
2	2800	2140	1880	980	1430	21	28	28	21	28
3	2780	2120	1860	970	1420	21	28	28	21	28
4	6230	4760	–	–	–	14	14	–	–	–
5	3300	2520	2220	–	1690	14	21	21	–	28
6	4180	3190	2800	–	2150	14	21	21	–	21
7	4650	3550	3120	–	2390	14	14	14	–	21
8	2420	1850	1620	840	1230	21	28	28	21	35
9	3620	2770	2430	–	1860	14	21	21	–	21
10	1700	1300	1130	580	860	21	35	35	28	42
11	3250	2480	2180	–	1660	14	21	21	–	28
12	2510	1920	1680	870	1280	21	28	28	21	28
13	3840	2930	2570	–	1970	14	21	21	–	21
14	4460	3410	2990	–	2290	14	21	14	–	21
15	4000	3060	2690	–	2050	14	21	21	–	21
16	4370	3330	2930	–	2240	14	21	21	–	21
17	3570	2720	2390	–	1830	14	21	21	–	21
18	2210	1680	1480	760	1120	21	28	28	28	35
19	1210	930	810	400	610	28	42	42	42	49
20	790	600	530	240	390	28	49	49	56	63

Несмотря на то, что общая сумма прибыли будет положительной, часть авиалиний могут стать убыточными. ЧЗ по сравнению с НЗ даёт более «оптимистичный» прогноз по прибыли. Как следует из результатов НЗ, лишь в редких случаях на некоторых направлениях ожидается удовлетворение всего имеющегося спроса вплоть до максимально возможных уровней. Это предполагается на тех авиалиниях, где наблюдается значительный дисбаланс в пассажиропотоках «туда» и «обратно». Как правило, выделяемые провозные ёмкости позволяют перевезти близкое к модальному число пассажиров. В целом дефазифицированные плановые пассажиропотоки в НЗ несколько ниже, чем в ЧЗ.

Таблица 4. Оптимальное количество рейсов

i	НЗ						ЧЗ					
	x_{ij} , рейсов, по типам ВС (j)					x_i , рейсов	x_{ij} , рейсов, по типам ВС (j)					x_i , рейсов
	1	2	3	4	5		1	2	3	4	5	
1	0	78	0	32	0	110	1	77	0	32	0	110
2	0	0	50	16	0	66	5	9	28	14	10	66
3	0	32	0	0	0	32	13	20	0	0	0	33
4	0	10	0	0	0	10	5	5	0	0	0	10
5	0	7	0	0	3	10	0	3	7	0	0	10
6	2	1	0	0	11	14	0	2	4	0	8	14
7	3	0	3	0	0	6	1	5	0	0	0	6
8	2	0	5	0	1	8	0	4	1	2	1	8
9	0	0	9	0	1	10	1	2	0	0	7	10
10	0	0	3	0	9	12	0	2	0	0	10	12
11	0	3	0	0	2	5	0	3	0	0	2	5
12	0	2	1	0	0	3	1	1	0	0	1	3
13	2	0	0	0	0	2	1	1	0	0	0	2
14	0	2	0	0	0	2	0	2	0	0	0	2
15	0	0	1	0	4	5	0	0	1	0	4	5
16	0	2	0	0	0	2	0	2	0	0	0	2
17	0	1	1	0	0	2	0	2	0	0	0	2
18	0	1	0	1	1	3	0	0	1	2	0	3
19	0	1	0	1	0	2	0	0	1	1	0	2
20	0	0	0	2	0	2	0	0	0	0	2	2
Σ	9	140	73	52	32	306	28	140	43	51	45	307

Рис. 1 иллюстрирует механизм трансформации СТНЧ «прогнозируемый недельный спрос» в ТНЧ «плановый недельный пассажиропоток» под влиянием ограничений на провозные способности «назначенных» ВС. Рассмотрены характерные примеры результатов НЗ для трех направлений. Судя по графикам функций принадлежности, суммарное число P_5 мест на ВС в аэропорт 5 является недостаточным для удовлетворения спроса, если он будет близок к максимальному прогнозируемому уровню. Суммарное число P_{10} мест на ВС из аэропорта 10 не «покрывает» не только максимального возможного уровня прогнозируемого спроса, но и модального его уровня. На направле-

нии в аэропорт 14 число P_{14} выделяемых мест заметно ниже любого возможного уровня спроса.

Таблица 5. Результаты оптимизации

i	НЗ							ЧЗ		
	\bar{E}_i , усл. ед.	$P_i^{(1)M}$, пас.	$P_i^{(2)M}$, пас.	$k^{P_i^{(1)L}}$	$k^{P_i^{(1)R}}$	$k^{P_i^{(2)L}}$	$k^{P_i^{(2)R}}$	E_i , усл. ед.	$P_i^{(1)}$, пас.	$P_i^{(2)}$, пас.
1	34415	14720	14720	0.15	0.00	0.05	0.00	40309	14760	14770
2	-5328	6120	6120	0.12	0.00	0.06	0.00	967	6312	6770
3	20873	5120	5120	0.00	0.00	0.02	0.00	24522	5930	5929
4	14361	1600	1600	0.01	0.00	0.03	0.00	16746	1850	1828
5	1712	1185	1261	0.15	0.07	0.15	0.01	1424	1180	1180
6	-7289	1118	1117	0.15	0.01	0.15	0.01	-5547	1118	1117
7	2599	930	902	0.07	0.00	0.15	0.03	5088	1010	902
8	-663	938	902	0.15	0.03	0.15	0.08	1973	930	902
9	-5188	881	770	0.15	0.08	0.15	0.15	-3391	880	770
10	-3578	740	750	0.15	0.01	0.06	0.00	-2993	740	820
11	-445	552	481	0.15	0.05	0.15	0.15	-259	552	481
12	821	420	410	0.07	0.00	0.15	0.02	838	420	410
13	1568	397	399	0.15	0.06	0.15	0.05	2105	370	370
14	1076	320	244	0.00	0.00	0.15	0.15	1076	320	244
15	-3595	300	300	0.12	0.00	0.06	0.00	-3364	300	300
16	1246	284	308	0.15	0.13	0.15	0.04	1433	284	308
17	502	257	260	0.15	0.01	0.03	0.00	739	257	296
18	-390	198	278	0.15	0.15	0.15	0.01	9	198	240
19	18	180	176	0.15	0.15	0.15	0.15	74	170	170
20	78	140	105	0.11	0.00	0.15	0.15	98	100	100
Σ	52793	36400	36223	-	-	-	-	81847	37681	37907

В двух из трех рассмотренных выше случаях СТНЧ $\tilde{q}_i^{(h)}$ ($\langle h, i \rangle \in \{(1, 5), \langle 2, 10 \rangle\}$) преобразуются в ТНЧ $\tilde{p}_i^{(h)}$, не являющиеся симметричными, с коэффициентами нечёткости более низкими, чем у исходных СТНЧ $\tilde{q}_i^{(h)}$. В случае же с направлением в аэропорт 14 исходное СТНЧ $\tilde{q}_{14}^{(1)}$ вообще редуцируется в «чёткое» число.

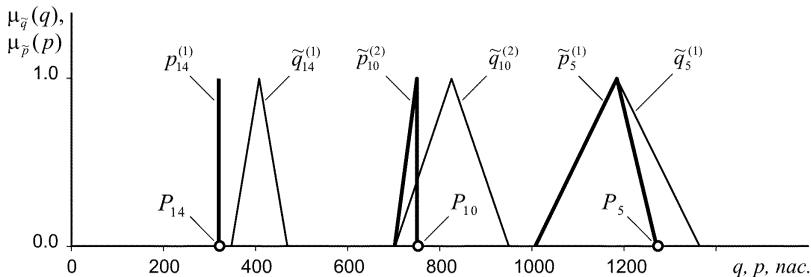


Рис. 1. Примеры функций принадлежности нечётких параметров пассажиропотоков

Уместно отметить, что введение нечёткости в постановку задачи приводит к некоторому сокращению среднего уровня плановых пассажиропотоков, что выше уже упоминалось. Поясним этот момент на примере сравнения чётких оценок пассажиропотоков для двух постановок задачи. В рассмотренном для направления в аэропорт 5 примере при «чёткой» постановке задачи плановое недельное чёткое число пассажиров $p_5^{(1)}$ совпадает с прогнозируемым спросом $q_5^{(1)}$ (как было сказано выше, в ЧЗ принимается $q_5^{(1)} = \text{def}(\tilde{q}_5^{(1)}) = q_5^{(1)M}$). Однако в случае нечёткой постановки дефаззифицированное плановое недельное число пассажиров $\bar{p}_5^{(1)} = \text{def}(\tilde{p}_5^{(1)})$, очевидно, будет меньше «чёткого» аналога $p_5^{(1)}$. Следовательно, введение нечёткости должно приводить к различиям в результатах решения задачи в «чёткой» и нечёткой постановках.

На практике результаты из представленных выше примеров означают, что на направлении в аэропорт 5 перевозчику следует ориентироваться на пассажиропоток, который будет близок к прогнозируемому, но не достигнет пиковых прогнозов. На направлении из аэропорта 10 авиакомпания может с высокой степенью определённости рассчитывать на перевозку P_{10} пассажиров либо немного меньшего их числа при высоких коэффициентах занятости мест на рейсах. Наконец, на рейсах в аэропорт 14 все места определённо будут заняты пассажирами, число которых составит P_{14} . Учёт авиакомпанией такого рода ре-

зультатов в своей деятельности позволит ей скорректировать свою коммерческую политику, стратегию и тактику в отношении использования собственных ресурсов.

Материалы таблицы 6, рис. 2 и 3 дают возможность сравнить результаты НЗ и ЧЗ для всех четырёх вариантов ограничений. В таблицу 6 сведены недельные величины операционной прибыли E , суммарные численности рейсов x^Σ и пассажиров p^Σ , а также средние коэффициенты занятости пассажирских мест K_3 , подсчитанные как отношение суммарного планового недельного числа пассажиров p^Σ к общему недельному числу мест на запланированных рейсах всех ВС. Для НЗ приведены дефаззифицированные значения нечётких величин, при этом символика, обозначающая дефаззификацию, опущена. Приводятся также относительные отклонения δ результатов ЧЗ относительно НЗ.

Таблица 6. Результаты оптимизации для различных вариантов ограничений

Вариант	E , тыс. усл. ед.		δ_E , %	x^Σ , рейсов		δ_x , %	p^Σ , пас.		δ_p , %	K_3		δ_K , %
	НЗ	ЧЗ		НЗ	ЧЗ		НЗ	ЧЗ		НЗ	ЧЗ	
	I	54,4	81,8	50,4	306	307	0,3	70,7	75,6	6,9	0,960	0,984
II	70,6	96,6	36,8	290	291	0,3	72,1	76,2	5,7	0,951	0,975	2,5
III	108,5	121,0	11,5	245	259	5,7	68,8	73,7	7,1	0,975	0,995	2,1
IV	121,7	129,8	6,7	220	236	7,3	66,0	69,3	5,0	0,986	0,997	1,1

Столбчатые диаграммы, позволяющие сравнить распределение рейсов по типам ВС для обеих задач и различных вариантов ограничений, отображены на рис. 2.

Оптимальные значения целевой функции (18) НЗ для вариантов ограничений I, ..., IV в виде точек с координатами $\bar{E}_I, \dots, \bar{E}_{IV}$ и графики функций принадлежности соответствующих ТНЧ $\tilde{E}_I, \dots, \tilde{E}_{IV}$ представлены на рис. 3. Для тех же четырёх вариантов ограничений показаны оптимальные значения целевой функции ЧЗ E_I, \dots, E_{IV} .

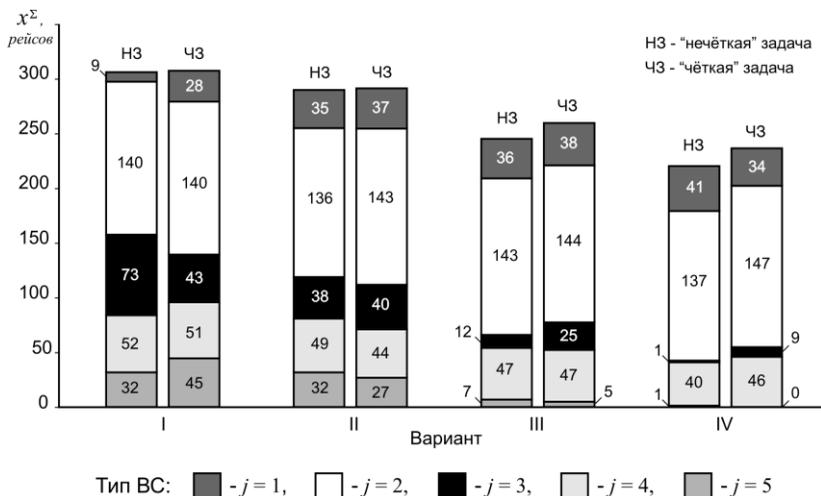


Рис. 2. Оптимальное недельное число рейсов ВС

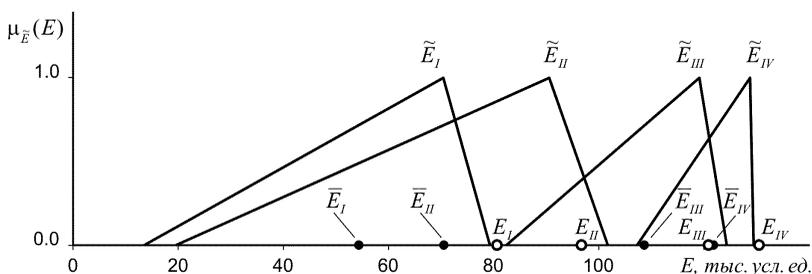


Рис. 3. Оптимальные значения целевой функции

Анализ представленных результатов позволяет сделать следующие выводы. По всем четырём вариантам суммарные и средние значения параметров, за исключением операционной прибыли, полученные при решении НЗ и ЧЗ, близки. Нечёткое решение по сравнению с «чётким» предполагает выполнение несколько меньшего числа рейсов при меньшем плановом числе пассажиров, что объясняется описанной выше особенностью нечёткой модели. С увеличением порядкового номера варианта ограничений, т.е. со снижением жёсткости требований по минимальным обязательным частотам выполнения рейсов на авиали-

ниях, суммарные численности рейсов и пассажиров вполне логично сокращаются. Оптимизация распределения ВС по авиалиниям позволяет весьма эффективно использовать авиапарк, о чем говорит очень высокая занятость мест на ВС, имеющая тенденцию к росту со снижением минимальных обязательных частот. Эта закономерность объясняется тем, что уменьшение минимального обязательного числа рейсов позволяет снизить занятость ВС выполнением рейсов и более эффективно перераспределить ВС различных типов.

При более детальном анализе результатов различия между «чёткой» и нечёткой моделями становятся более заметными. Как следует из рис. 2, недельные численности рейсов при близости суммарных значений значительно различаются для отдельных типов ВС. Так, например, относительные отклонения численности рейсов ВС небольшой пассажировместимости (не более 100 мест) для ЧЗ относительно НЗ составляют от 7 до 31%. По отдельным типам ВС эти расхождения могут быть еще более значительными. При этом с сокращением потребной (и фактической) частоты рейсов снижается интенсивность использования небольших ВС и, соответственно, возрастает нагрузка на большие ВС. Описанная тенденция вполне объяснима: если нет необходимости выполнять рейсы часто, то затраты при тех же доходах можно сократить, если выбирать самолёты большей пассажировместимости, позволяющие перевезти всех желающих меньшим числом рейсов.

Как следует из материалов таблицы 6 и рис. 3, различия нечёткого и «чёткого» решений наиболее ярко проявляются, если судить по полученным значениям целевой функции. Отклонения операционной прибыли для ЧЗ относительно НЗ составляют от 7 до 50% в зависимости от варианта исходных данных. С ужесточением требований к минимальной недельной частоте выполнения рейсов – в «реальных» задачах такие требования могут быть вполне жёсткими – отклонения возрастают.

Таким образом, результаты нечёткой и «чёткой» оптимизации распределения ВС на авиалиниях, схожие по менее значимым параметрам, заметно разнятся в наиболее существенных характеристиках: ожидаемой величине прибыли от выполнения транспортной операции и распределении ВС по авиалиниям.

При представленном наборе исходных данных на базе современной персональной вычислительной техники с использованием пакета IBM ILOG OPL затраты машинного времени на получение оптимальных решений задачи в нечёткой постановке составили около 3–4 минут, в чёткой постановке – порядка 1 минуты.

Уместно остановиться ещё на одном моменте, связанном с целесообразностью решения в нечёткой постановке задачи оптимального распределения ВС. Из представленного выше описания следует, что решение ЧЗ и НЗ умеренной размерности достигается при сравнительно малых затратах машинного времени. Однако с повышением размерности задач следует ожидать значительно больших временных затрат на их решение, особенно ощутимых в случае НЗ. Проверим возможность отказа от использования более сложной нечёткой модели с игнорированием фактической неопределённости спроса. С этой целью оптимальные значения x_{ij} , $i = 1, \dots, I$, $j = 1, \dots, J$, полученные при решении ЧЗ, используем для расчёта недельных показателей деятельности авиакомпании в условиях нечёткого спроса в соответствии с нечёткой моделью. Такой подход позволит оценить последствия неучёта возможных нарушений прогноза пассажирского спроса при поиске оптимального распределения ВС. Расчёты, выполненные для рассмотренных выше вариантов исходных данных, показали, что сокращение дефазифицированной прибыли в случае «подстановки» решения ЧЗ в нечёткую модель достигает 12% по сравнению с использованием соответствующего решения НЗ. Ещё большее падение прибыли фиксируется при повышении степени неопределённости прогноза. Так, при относительных коэффициентах нечёткости спроса, равных 0,3 и 0,45, использование решения ЧЗ становится причиной снижения дефазифицированной прибыли по сравнению с решением НЗ на 14% и 25% соответственно. При этом в отличающемся низкой рентабельностью случае жестких требований к частотам выполнения рейсов различия по прибыли не столь велики, однако использование решения ЧЗ при неопределённом спросе чревато угрозой возникновения убытков. В отношении других показателей перевозочного процесса, таких как уровень пассажиропотока, суммарная численность рейсов и коэффици-

ент занятости пассажирских мест, можно сказать, что использование решения ЧЗ вместо решения НЗ приводит к их незначительному росту. Можно сказать, что это означает для авиакомпании необходимость несколько больших затрат ресурсов на осуществление перевозок при меньшей прибыли. Таким образом, решение задачи оптимального распределения ВС в нечёткой постановке при заданном в нечёткой форме прогнозе пассажирского спроса следует считать весьма желательным; замена нечёткого оптимального решения чётким нецелесообразна.

5. Заключение

В статье решена задача оптимального распределения ВС известных типов и численности на заданной сети авиалиний. Новизна постановки задачи состоит в том, что уровни спроса на пассажирские перевозки приняты не полностью определёнными экспертно заданными нечёткими величинами, что соответствует этапу предварительного проектирования сезонного расписания. Значимость задачи определяется тем, что результаты её решения напрямую влияют на экономическую эффективность перевозочной деятельности авиакомпании. Для случая, когда одна часть исходных данных задается экспертами в нечёткой форме, а другая часть определяется в результате обработки имеющейся статистики как частотные распределения, предложена методика, позволяющая аппроксимировать частотные распределения нечёткими числами, сводя таким образом все исходные данные к единой нечёткой форме. Следует отметить, что и в случае комбинирования в составе исходных данных нечётких с обычными числами постановка и подход к решению оптимизационной задачи не изменится.

Задача оптимизации расстановки ВС по авиалиниям сформулирована как целочисленная задача математического программирования с нечётким критерием и чёткими ограничениями. С целью придания постановке определённости использован приём дефазификации, позволивший нечёткую задачу свести к обычной задаче математического программирования, решаемой имеющимися действенными методами на базе доступного программного обеспечения. Решение модельных примеров по-

лучено с использованием программного обеспечения IBM ILOG OPL. Затраты машинного времени на решение задачи в нечёткой постановке при принятых исходных данных составили не более нескольких минут, что свидетельствует о практической реализуемости решения в условиях действующих предприятий гражданской авиации.

Выполненное сравнение позволило сделать вывод о существенных различиях наиболее значимых результатов решения оптимизационной задачи в нечёткой и «чёткой» постановках, что свидетельствует о желательности учёта неопределённости исходных данных, заданной в нечёткой форме. Задача решена на примере перевозок пассажиров, однако оптимизационная модель может быть без существенных переработок применена для случаев любых других видов корреспонденций. Целесообразным следует считать использование нечётких моделей и в случае более детальных и комплексных постановок задачи распределения ВС и смежных задач управления на воздушном транспорте.

Литература

1. АНДРОНОВ А.М., ХИЖНЯК А.Н. *Математические методы планирования и управления производственно-хозяйственной деятельностью предприятий гражданской авиации*. – М.: Транспорт, 1977. – 215 с.
2. ВАДЗИНСКИЙ Р.Н. *Справочник по вероятностным распределениям*. – СПб.: Наука, 2001. – 294 с.
3. ДАНЦИГ Д.Б. *Линейное программирование, его применения и обобщения*. – М.: Прогресс, 1966. – 600 с.
4. ОРЛОВСКИЙ С.А. *Проблемы принятия решений при нечёткой исходной информации*. – М.: Наука, 1981. – 208 с.
5. ПЕГАТ А. *Нечёткое моделирование и управление*. – М.: БИНОМ. Лаборатория знаний, 2013. – 798 с.
6. *Приказ Министерства транспорта РФ от 12 декабря 2011 г. N 310 «Об утверждении Порядка формирования, утверждения и опубликования расписания регулярных воздушных перевозок пассажиров и (или) грузов, выполняемых перевозчиками, имеющими соответствующие лицензии»*. –

- URL: <https://mintrans.gov.ru/documents/2/2207> (дата обращения: 10.12.2024).
7. САКАЧ Р.В., ПИНАЕВ Е.Г., ГЛАДЫШЕВСКАЯ Г.Н. и др. *Моделирование в планировании гражданской авиации*. – М.: Транспорт, 1983. – 173 с.
 8. СТОРОЖЕВ С.В. *Алгоритм двухпараметрической аппроксимации нормального частотного распределения нечётким интервалом* // Вестник Донецкого национального университета. Серия А: Природные науки. – 2014. – №2. – С. 78–80.
 9. AVARA J. *Applying integer linear programming to the fleet assignment problem* // Interfaces. – 1989. – Vol. 19, No. 4. – P. 20–28. – DOI: 10.1287/inte.19.4.20.
 10. BARNHART C., SMITH B. *Quantitative Problem Solving Methods in the Airline Industry* // International Series in Operations Research and Management Science. – Springer Science & Business Media, 2012. – 462 p. – DOI: 10.1007/978-1-4614-1608-1.
 11. BAZARGAN M. *Airline operations and scheduling*. – Routledge, 2010. – 302 p.
 12. CLARKE L.W., HANE C.A., JOHNSON E.L. et al. *Maintenance and crew considerations in fleet assignment* // Transportation Science. – 1996. – Vol. 30, No. 3. – P. 249–260.
 13. CLARKE M., SMITH B. *Impact of operations research on the evolution of the airline industry* // Journal of Aircraft. – 2004. – Vol. 41, No. 1. – P. 62–72. – DOI: 10.2514/1.900.
 14. DESAULNIERS G., DESROSIERS J., DUMAS Y. et al. *Daily aircraft routing and scheduling* // Management Science. – 1997. – Vol. 43, No. 6. – P.841–855. – DOI: 10.1287/mnsc.43.6.841.
 15. FERGUSON A.R., DANTZIG G.B. *The problem of routing aircraft – a mathematical solution* // Aeronaut. Engng. Rev. – 1955. – Vol. 14, No. 4. – P. 51–55.
 16. FERGUSON A.R., DANTZIG G.B. *The allocation of aircraft to routes – an example of linear programming under uncertain demand* // Manag. Sci. – 1956. – Vol. 3, No. 1. – P. 45–73. – DOI: 10.1515/9781400884179-029.
 17. GARG A., AGARWAL Y., SRIVASTAVA R.K. et al. *Integrated commercial and operations planning model for schedule de-*

- sign, aircraft rotation and crew scheduling in airlines // Networks*. – 2024. – Vol. 83, Iss. 4. – P. 653–672. – DOI: 10.1002/net.22211.
18. HANE C.A., BARNHART C., JOHNSON E.L. et al. *The fleet assignment problem: Solving a large-scale integer program // Mathematical Programming*. – 1995. – Vol. 70. – P. 211–232.
 19. KIZILOGLU K., SAKALLI U.S. *Integrating Flight Scheduling, Fleet Assignment, and Aircraft Routing Problems with Codesharing Agreements under Stochastic Environment // Aerospace*. – 2023. – Vol. 10, No. 12. – DOI: 10.3390/aerospace 10121031.
 20. NAUMANN M., SUHL L., FRIEDEMANN M. *A stochastic programming model for integrated planning of re-fleeting and financial hedging under fuel price and demand uncertainty // Procedia – Soc. Behav. Sci.* – 2012. – Vol. 54. – P. 47–55. – DOI: 10.1016/j.sbspro.2012.09.724.
 21. NUGROHO A., SUHARTO A.M. *Airline Fleet Assignment and Schedule Planning // Jurnal Manajemen Transportasi & Logistik*. – 2014. – Vol. 1, No. 1. – P. 31–42. – DOI: 10.54324/j.mtl.v1i1.5.
 22. PILLA V.L., ROSENBERGER J.M., CHEN V.C. et al. *A statistical computer experiments approach to airline fleet assignment // IIE Trans.* – 2008. – Vol. 40, No. 5. – P. 524–537. – DOI: 10.1080/07408170701759734.
 23. REXING B., BARNHART C., KNIKER T.S. et al. *Airline fleet assignment with time windows // Transportation Science*. – 2000. – Vol. 34, No. 1. – P. 1–20. – DOI: 10.1287/trsc.34.1.1.12277.
 24. RUSHMEIER R.A., KONTOGIORGIS S.A. *Advances in the optimization of airline fleet assignment // Transportation Science*. – 1997. – Vol. 31, No. 2. – P. 159–169.
 25. SAFAK O., CAVUS O., AKTURK M.S. *Multi-stage airline scheduling problem with stochastic passenger demand and non-cruise times // Transp. Res. Part B Methodol.* – 2018. – Vol. 114. – P. 39–67. – DOI: 10.1016/j.trb.2018.05.012.
 26. SHERALI H.D., BISH E.K., ZHU X. *Airline fleet assignment concepts, models, and algorithms // European Journal of Opera-*

- tional Research. – 2006. – Vol. 172, No. 1. – P. 1–30. – DOI: 10.1016/j.ejor.2005.01.056.
27. XU Y., ADLER N., WANDELT S. et al. *Competitive integrated airline schedule design and fleet assignment* // European Journal of Operational Research. – 2024. – Vol. 314, No. 1. – P. 32–50. – DOI: 10.1016/j.ejor.2023.09.029.
 28. YAN C., BARNHART C., VAZE V. *Choice-Based Airline Schedule Design and Fleet Assignment: A Decomposition Approach* // Transportation Science. – 2022. – Vol. 56, No. 6. – P. 1410–1431. – DOI: 10.1287/trsc.2022.1141.
 29. YAN S., TSENG C.-H. *A passenger demand model for airline flight scheduling and fleet routing* // Computers and Operations Research. – 2002. – Vol. 29, No. 11. – P. 1559–1581. – DOI: 10.1016/s0305-0548(01)00046-6.
 30. ZHOU L., CHOU C.-A., CHAOVALITWONGSE W.A. et al. *Airline planning and scheduling: Models and solution methodologies* // Front. Eng. Manag. – 2020. – Vol. 7, No. 1. – P. 1–26. – DOI: 10.1007/s42524-020-0093-5.

OPTIMIZATION OF AIRCRAFT ASSIGNMENT TO AIRLINES WITH FUZZY INITIAL DATA

Vladimir Romanenko, Samara National Research University, Samara, Cand.Sci., associate professor (vla_rom@mail.ru).

Abstract: A variant of the problem of optimal assignment of aircraft of specified types and numbers on a given airline network has been solved. The distribution of aircraft by airline is one of the key problems determining the efficiency of an air transport company. The considered task is to determine for each airline the weekly number of aircraft flights, which ensures the maximum economic effect of passenger transportation. The novelty of the problem statement lies in the fact that the levels of demand for passenger transportation are not fully defined expertly set fuzzy values, which corresponds to the stage of preliminary schedule design. Optimization of aircraft assignment by airlines is formulated as an integer mathematical programming problem with a fuzzy criterion and clear constraints. Thanks to the use of the defuzzification technique, the fuzzy problem is reduced to an ordinary mathematical programming problem solved by the available effective methods based on accessible software. Using the IBM ILOG OPL software package, the solution of model examples of the problem in fuzzy and "crisp" formulations was obtained. The comparison revealed significant differences between the most significant results of solving the

optimization problem in fuzzy and "crisp" formulations, which indicates the expediency of taking into account the fuzzy uncertainty of the initial data.

Keywords: optimization, mathematical programming, distribution problem, fuzzy number, aircraft, airlines.

УДК 519.85 + 510.644.4 + 656.7.022.1

ББК 22.185.432 + 22.126 + 39.580.3

*Статья представлена к публикации
членом редакционной коллегии В.Г. Лебедевым.*

Поступила в редакцию 23.12.2024.

Опубликована 31.05.2025.

БАЙЕСОВСКАЯ МОДЕЛЬ ОБЪЕДИНЕНИЯ ПРИЗНАКОВЫХ ПРЕДСТАВЛЕНИЙ В ЗАДАЧЕ РЕИДЕНТИФИКАЦИИ ЛИЧНОСТИ¹

Русаков К. Д.²

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

Рассматривается задача реидентификации человека по изображениям, полученным из систем видеонаблюдения, с использованием признаков различной природы. В отличие от традиционных подходов, опирающихся на один биометрический фактор, предложена универсальная методика объединения признаков, полученных из разных источников информации. В основе метода лежит вероятностная модель, реализующая объединение признаков в пространстве решений по правилу байесовского вывода. Каждый признак рассматривается как наблюдение, аппроксимируемое нормальным распределением с параметрами, оцениваемыми по обучающим данным. Принятие решения осуществляется путём максимизации апостериорной вероятности принадлежности объекта к одной из известных личностей. В качестве средств извлечения признаков используются модели на основе архитектур с механизмами внимания, обеспечивающие устойчивость к искажениям. Проведено сравнение с методом линейного объединения признаков. Результаты экспериментов на открытом наборе данных демонстрируют повышение точности идентификации, устойчивость к частичному отсутствию данных и возможность количественной оценки степени достоверности принимаемого решения, что актуально для применения в системах обеспечения безопасности.

Ключевые слова: реидентификация личности, объединение признаков, байесовская модель, компьютерное зрение, трансформерные архитектуры, видеонаблюдение.

1. Введение

Реидентификация человека (person re-identification, ReID) — задача поиска и распознавания одного и того же человека на различных снимках или в разных камерах наблюдения [8]. Актуальность ReID обусловлена активным развитием систем видеонаблюдения и вопросами безопасности: автоматическое отсле-

¹ Автор признателен д.т.н., профессору РАН, г.н.с. Р.В. Мецержякову за ценное обсуждение содержания статьи.

² Константин Дмитриевич Русаков, н.с. (rusakov@ipu.ru).

живание перемещений подозреваемых, поиск пропавших людей и т.п. [19].

Традиционно ReID опирается либо на биометрию лица, либо на общие признаки фигуры (силуэт, одежда) человека по отдельности. Однако только на одном источнике (лицо или тело) достичь высокой надёжности распознавания сложно из-за вариаций позы, освещения, разрешения и перекрытий. Совмещение нескольких биометрических признаков позволяет повысить точность за счёт взаимодополняющей информации различных модальностей [12, 13].

В частности, силуэт человека считается почти таким же уникальным, как и лицо, и уже используется на практике. Так, новейшие отечественные системы умного видеонаблюдения отслеживают траекторию по силуэтам и осуществляют поиск по фигуре в реальном времени [7]. Вместе с тем возникает вопрос эффективной интеграции разнородных признаков.

Наиболее простой подход объединения признаков — линейная комбинация, например, конкатенация векторов лица и тела или усреднение оценок совпадения [13]. Однако жёсткое фиксированное слияние имеет ряд ограничений. Во-первых, оно не учитывает надёжность каждой модальности: «шумные» данные (например, размытое лицо или скрытая одежда) могут снижать итоговую точность, тогда как оптимально было бы придать им меньший вес. Во-вторых, линейное слияние не отражает вероятностную природу биометрических измерений и не даёт информации о степени уверенности. В результате система может ошибаться в условиях искажённого входа, не имея механизма выражения неуверенности в принятом решении.

Цель данной работы — разработать байесовскую схему интеграции признаков лица и силуэта для задачи реидентификации, которая учитывает статистическую неопределённость данных и обеспечивает более надёжное распознавание по сравнению с традиционным детерминированным объединением. В данной работе впервые предлагается использовать вероятностную модель на основе теоремы Байеса для объединения многомодаль-

ных представлений лица и тела.

Научная новизна заключается в том, что байесовский подход позволяет адаптивно взвешивать вклад каждой модальности через апостериорную вероятность, повышая устойчивость системы к шуму, изменению условий съёмки и отсутствию одного из биометрических факторов.

В работе представлены:

а) формальная модель и алгоритм байесовской интеграции признаков;

б) её программная реализация с использованием современных архитектур сверточных нейронных сетей (ArcFace и трансформер Vision Transformer);

в) экспериментальное сравнение с традиционным линейным подходом на открытом датасете.

Ниже рассматриваются связанные исследования, описывается метод, экспериментальные результаты, обсуждаются преимущества и ограничения, и даются выводы.

2. Связанные работы

2.1. РЕИДЕНТИФИКАЦИЯ ПО ЛИЦУ И ТЕЛУ

Задача распознавания лица по изображениям длительное время развивалась независимо от задачи ReID по силуэту, формируя собственные подходы и методы. С появлением глубоких нейросетей были достигнуты выдающиеся результаты: так, архитектура FaceNet с триплетным обучением [17] достигла точности 99,6% на эталонном датасете LFW.

В дальнейшем были предложены специальные функции потерь, направленные на повышение дискриминативности векторов лица. Одной из наиболее эффективных стала функция ArcFace [10], которая вводит аддитивную угловую поправку в классификационное обучение, обеспечивая более плотные кластеры векторов для каждого человека. Сегодня ArcFace де-факто является стандартом в задачах распознавания лиц; его эффективность подтверждена как зарубежными, так и отечественными исследованиями [4].

Тем не менее даже лучшие модели распознавания лиц могут давать сбои в условиях неконтролируемой съёмки: при низком разрешении, поворотах головы, частичном закрытии лица (маски, шарфы) [6]. В частности, отечественные работы отмечают снижение качества на «нетипичных» лицах, которые существенно отличаются от доминирующих в обучающих данных (например, лица детей или представителей редких рас) [6], и предлагают методы предварительной фильтрации «сложных» случаев для повышения надёжности [5, 6].

Одновременно со стороны ReID по фигуре также наблюдается значительный прогресс. Классические подходы опирались на признаки внешнего вида: цветовые гистограммы, HOG-описатели, локальные инвариантные признаки и др. [8]. С развитием глубокого обучения акцент сместился на обучаемые векторы изображений полноростового человека [19].

Архитектуры на основе сверточных нейронных сетей (например, ResNet-50) обучаются либо классифицировать идентичности в тренировочном наборе с использованием функции Softmax, либо непосредственно оптимизировать векторы с помощью метрического обучения (триплет-лосс, контрастивный лосс и т.п.) [11]. Для повышения устойчивости к вариациям позы используются стратегии деления тела на части и механизмы внимания.

Так, Harmonious Attention Network (HAN) объединяет пространственное и каналовое внимание для выделения наиболее информативных областей изображения человека [16]. Другие работы задействуют скелетную позу как условие для объединения признаков или применяют графовые нейронные сети (Graph Convolutional Networks, GCN) для моделирования взаимосвязей между частями тела и визуальными атрибутами [14].

В целом современные модели person re-identification демонстрируют высокое качество на стандартных бенчмарках (Market-1501, MSMT17 и др.), особенно при использовании нескольких масштабов и агрегации локальных признаков частей тела [2]. Тем не менее, большинство существующих решений опираются ис-

ключительно на визуальную модальность — внешний вид человека.

2.2. ИНТЕГРАЦИЯ МНОГОМОДАЛЬНЫХ БИОМЕТРИЧЕСКИХ ПРИЗНАКОВ

Комбинация признаков лица и полного силуэта представляет собой мультибиометрический подход к идентификации. Ещё в ранних работах отмечалось, что объединение нескольких биометрических идентификаторов (например, лица, радужной оболочки, голоса) повышает общую точность системы за счёт разнообразности используемых источников [9].

В задаче ReID предпринимались различные подходы к multi-modal fusion на разных уровнях. Простейший из них — объединение на уровне признаков: например, конкатенация векторов лица и тела с последующей классификацией или вычислением расстояний [13]. Такой метод использован в работе [18], где отдельные нейросети извлекают признаки лица (FaceNet) и тела (OSNet), после чего векторы объединяются и сравниваются с эталонными комбинациями; авторы демонстрируют прирост порядка 5% в метрике Rank-1 по сравнению с использованием только тела.

Более сложные схемы реализуют слияние на уровне оценок (score-level fusion), например, посредством усреднения или обучаемого взвешивания сходства между лицом и фигурой [13]. В системах высокого уровня применяются также байесовские сети доверия, где выходы детектора лица и дополнительная информация (пол, возраст, рост) интегрируются для повышения надёжности идентификации в робототехнических приложениях [12].

Важным направлением является применение механизмов перекрёстного внимания между модальностями, позволяющих модели самостоятельно выявлять, каким признакам следует уделять больше внимания. В недавней работе [20], объединяющей визуальные и языковые описания, демонстрируется улучшение ReID за счёт одновременного использования лица и тела: на основе изображения генерируется текстовое описание, и модули внимания сопоставляют текст с соответствующими областями (лицо, тело) на изображении, обеспечивая более полную

идентификацию.

Графовые подходы, в частности основанные на GCN, представляют мультимодальные признаки в виде графа, где узлы соответствуют векторам разных модальностей одного человека, а рёбра отражают их принадлежность к одной личности [14]. Обучение такого графа позволяет эффективно распространять информацию между модальностями.

В большинстве рассмотренных подходов (за исключением байесовских) интеграция осуществляется либо через обучаемые веса (что по сути имитирует байесовский вывод, но неявно), либо через механизмы внимания. Однако явная вероятностная интеграция по правилу Байеса до сих пор не применялась для слияния векторов лица и тела. Между тем, в смежных областях байесовские методы доказали свою эффективность при принятии решений на основе нескольких источников информации [9]. В данной работе мы восполняем этот пробел, предложив явную байесовскую модель для объединения векторов в задаче ReID.

Таким образом, предложенный подход концептуально близок к мультибиометрическим системам, осуществляющим объединение признаков на уровне принятия решения. Однако, в отличие от существующих решений, он основан не на эвристических или обучаемых весах, а на строгом байесовском выводе.

Разработанная схема может быть интерпретирована как наивный байесовский классификатор идентичности, в котором признаками выступают векторы лица и силуэта. Ранее наивный Байес использовался в биометрических системах для принятия решений по отпечаткам пальцев и изображениям лиц [9], но не в контексте глубоких нейросетевых векторов.

Следует также отметить, что в работе применяются современные высокоуровневые подходы — ArcFace и Vision Transformer, обеспечивающие компактное и устойчивое представление каждой модальности. Это принципиально отличает предложенный метод от более ранних работ, основанных на низкоуровневых визуальных признаках [8].

3. Байесовская схема интеграции признаков

3.1. БАЙЕСОВСКАЯ МОДЕЛЬ ИНТЕГРАЦИИ

Задача распознавания моделируется как статистическая классификацию личности y по двум наблюдениям: вектору признаков лица f и вектору признаков силуэта s . Пусть y – дискретная случайная переменная, принимающая значения из множества идентичностей Y (известных системе лиц). Тогда по теореме Байеса апостериорная вероятность того, что наблюдения (f, s) принадлежат человеку y , определяется как

$$(1) \quad P(y | f, s) \propto P(f | y) \cdot P(s | y) \cdot P(y),$$

где $P(f | y)$ – правдоподобие наблюдаемого признака лица при условии личности y , $P(s | y)$ – правдоподобие признака силуэта, а $P(y)$ – априорная вероятность личности y . В нашем случае априори все кандидаты равновероятны (непредвзятая идентификация), поэтому далее полагаем $P(y) = const$. Таким образом, решение о том, что на изображении представлен человек y^* , принимается по правилу максимума апостериорной вероятности (MAP)

$$(2) \quad y^* = \arg \max_{y \in Y} \left\{ \ln P(f | y) + \ln P(s | y) \right\},$$

где переход к логарифмическим правдоподобиям заменяет произведение плотностей на сумму логарифмов и предотвращает переполнение при работе с малыми величинами вероятностей, а также обеспечивает численную устойчивость вычислений. Выражение (2) показывает, что решение складывается из двух логарифмических правдоподобий: от лица и от силуэта. Если, скажем, лицо плохо распознано (низкое $P(f | y)$), но силуэт совпал уверенно (высокое $P(s | y)$), то личность всё равно может быть правильно определена за счёт второй части. И наоборот, при противоречивых признаках будет выбран тот кандидат y , для которого произведение плотностей (или сумма логарифмов) максимизируется, т.е. оба признака достаточно соответствуют этому кандидату лучше, чем любому другому. Такая схема эквивалентна наивному байесовскому классификатору, учитывающему независимые «признаковые условия» f и s для каждого класса y [9].

В нашем случае независимость признаков лица и тела – упрощающее предположение, оправданное тем, что алгоритмы извлечения векторов обучены разделять эти признаки в различные пространства.

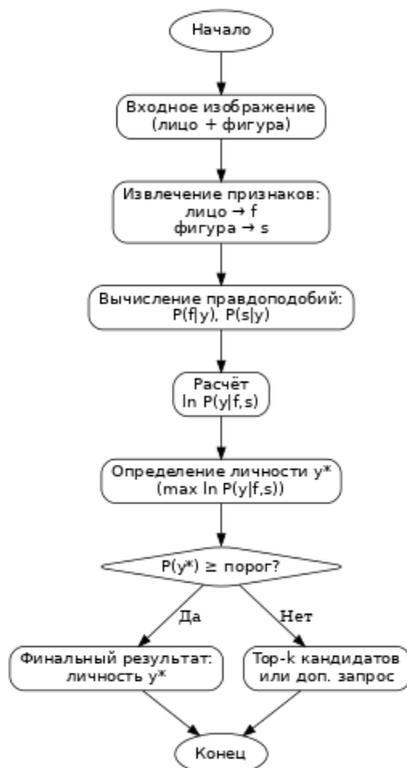


Рис. 1. Блок-схема предлагаемого алгоритма байесовской реидентификации

Для вычисления $P(f | y)$ и $P(s | y)$ необходимо построить вероятностную модель распределения векторов каждой модальности для каждой личности. В данном исследовании выбирается простейшая модель – многомерное нормальное распределение (гауссиан) в пространстве векторов. Предполагается, что для каждой личности y признаки лица f распределены нормаль-

но с некоторым средним μ_y^f и ковариационной матрицей Σ_y^f , то же верно для признаков силуэта с параметрами μ_y^s , Σ_y^s . Эти параметры можно оценить по обучающей выборке векторов (т.е. по нескольким изображениям каждого человека из базы). Тогда плотность имеет вид

$$(3) P(f | y) = \frac{1}{(2\pi)^{d/2} |\Sigma_y^f|^{1/2}} \exp\left(-\frac{1}{2} (f - \mu_y^f)^\top \Sigma_y^f^{-1} (f - \mu_y^f)\right),$$

и аналогично для $P(s | y)$. Здесь d — размерность пространства вектора (для ArcFace $d = 128$). На практике для устойчивости и уменьшения числа параметров возможно использование диагональной, или изотропной (скалярной $\sigma_y^2 I$) ковариации, тогда формулы упрощаются до зависимости от евклидова или махалановисова расстояния $|f - \mu_y^f|$.

В эксперименте была использована $\Sigma_y^f = \sigma^2 I$ (одинаковая фиксированная ковариационная матрица для всех y , оценённая по совокупности данных), чтобы избежать проблемы ограниченной статистики на каждого человека. Это допущение превращает выражение (2) в простой критерий по сумме квадратов расстояний: максимизируется $-|f - \mu_y^f|^2 - |s - \mu_y^s|^2$.

Таким образом, байесовский подход в данном упрощении сводится к распознаванию по ближайшему центроиду в объединённом пространстве, что эквивалентно взвешенному (по дисперсии) сложению признаков. Полноценная же байесовская модель (3) позволяет задать разные ковариации, учитывая, например, что разброс признаков лица для данной личности меньше, чем разброс признаков фигуры (что логично, так как одежда меняется).

На рис. 1 представлена блок-схема предлагаемого алгоритма. Сначала из входного изображения с помощью двух нейросетевых моделей извлекаются векторы: f — 128-мерный вектор лица и s — d_s -мерный вектор тела (в эксперименте $d_s = 256$).

Затем эти признаки поступают в модуль байесовского классификатора, где вычисляются правдоподобия (3) для каждого из N зарегистрированных в системе пользователей (личностей). Перемножая плотности (или суммируя лог-правдоподобия) со-

гласно формуле (1), модуль выдаёт апостериорные вероятности $P(y | f, s)$ по всем классам $y = 1, \dots, N$.

Финальное решение — идентификатор y^* — выбирается по максимальной апостериорной вероятности. Кроме того, вероятности могут быть использованы для оценки уверенности: если максимум $P(y^* | f, s)$ низок, система может воздержаться от однозначного ответа (выдать топ- k кандидатов либо запросить дополнительный образ). Ниже приведён алгоритм.

Алгоритм 1 Байесовская идентификация

1: **Вход:** изображение с лицом и фигурой человека

Извлечение признаков:

2: $f \leftarrow \text{Embedder}_{\text{face}}(\text{изображение лица})$

3: $s \leftarrow \text{Embedder}_{\text{body}}(\text{изображение тела})$

Байесовский классификатор:

4: **for all** $y \in Y$ **do**

5: $L_y^f \leftarrow P(f | y)$ ▷ по формуле (3)

6: $L_y^s \leftarrow P(s | y)$

7: $\text{posterior}_y \leftarrow L_y^f \times L_y^s$ ▷ так как $P(y) = \text{const}$

8: **end for**

Идентификация:

9: $y^* \leftarrow \arg \max_y \text{posterior}_y$

10: вывести y^* как опознанную личность

11: при необходимости вывести posterior_{y^*} как степень уверенности

3.2. АРХИТЕКТУРА КОДИРОВЩИКОВ

В предлагаемом подходе применяются две нейросетевые модели (кодировщика), извлекающие векторные признаки из изображений: одна — для лица, другая — для силуэта человека. Обе модели построены на базе Vision Transformer (ViT) и используют функцию потерь ArcFace [10], доказавшую свою эффективность для задач биометрической идентификации.

Для каждого изображения соответствующая модель извлекает вектор:

$$(4) \quad f := \text{Embedder}_{\text{face}}(I_{\text{лицо}}), \quad s := \text{Embedder}_{\text{body}}(I_{\text{тело}}),$$

где $f \in \mathbb{R}^{128}$ — вектор лица, $s \in \mathbb{R}^{256}$ — вектор силуэта.

Архитектура ViT представлена на рис. 2 [22]. Она разбивает изображение на непересекающиеся патчи, кодирует их в линейном пространстве и передаёт в стек self-attention-блоков трансформера. На выходе берётся вектор [CLS] как обобщённое представление всего изображения. Такая структура эффективна как для лиц, так и для тел, что делает её универсальной основой для обеих модальностей.

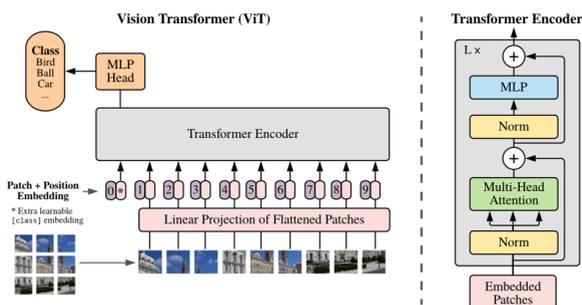


Рис. 2. Архитектура Vision Transformer для извлечения признаков изображения. Входное изображение делится на патчи фиксированного размера, которые подаются в трансформерный энкодер. На выходе используется классификационный токен ([CLS]) как вектор изображения. MLP-голова отбрасывается

Функция потерь ArcFace [10] используется для повышения дискриминативности представлений, вводя аддитивную угловую поправку в классификационную границу. Это позволяет векторам формироваться в плотные и разделённые кластеры. Альтернативно в задачах re-identification может применяться Triplet Loss [11] — метрическая функция, минимизирующая расстояние между положительными парами и увеличивающая расстояние до

отрицательных. Однако в наших экспериментах ArcFace показал лучшую сходимость и стабильность обучения.

Перед вычислением плотностей векторы нормализуются по L_2 -норме:

$$(5) \quad \hat{f} := \frac{f}{\|f\|_2}, \quad \hat{s} := \frac{s}{\|s\|_2}.$$

В отличие от классического подхода, центрирование признаков относительно среднего по обучающей выборке не производится, поскольку параметры многомерного нормального распределения (среднее и ковариационная матрица) для каждого класса оцениваются напрямую по обучающим данным.

Для объединения информации из разных модальностей — лица и силуэта — применяется байесовская схема. Итоговая логарифмическая апостериорная вероятность принадлежности к классу вычисляется как сумма логарифмов модальности-зависимых плотностей:

$$(6) \quad \ln P(y|f, s) \propto \ln P(f|y) + \ln P(s|y).$$

При этом весовые коэффициенты не используются, т.е. вклад каждой модальности считается равным. Эксперименты показали, что даже при равномерной агрегации признаки обеих модальностей вносят значимый вклад в классификацию, а предварительная L_2 -нормализация обеспечивает стабильную работу подхода вне зависимости от размерности векторов.

4. Эксперимент

4.1. ПРОТОКОЛ ЭКСПЕРИМЕНТА

В данном разделе приводится полное описание проведения эксперимента по оценке предложенной байесовской схемы интеграции признаков лица и тела на валидационном сплите датасета SUNK03 [15]. Для каждой личности в каждой модальности было случайным образом отобрано по два изображения в качестве опорного набора, остальные изображения той же личности вошли в проверочный набор. Классы включались в эксперимент только при наличии не менее двух кадров и в лице, и в силуэте, что позволило синхронизировать разбиение.

Изображения приводились к размеру 112×112 и нормировались по каналам с параметрами ($\mu = 0,5, \sigma = 0,5$). Векторы извлекались двумя моделями MobileFaceNet [21]: для лица — размерностью 128, для силуэта — 256. Все векторы нормировались по L_2 -норме. На опорном наборе для каждого класса рассчитывались средний вектор и полная ковариационная матрица с регуляризацией $10^{-3}I$.

Байесовский метод принимал решение по формуле

$$\ln P(y | f, s) = \ln \mathcal{N}(f; \mu_y^f, \Sigma_y^f) + \ln \mathcal{N}(s; \mu_y^s, \Sigma_y^s),$$

а в качестве признанного класса выбирал y с максимальным значением лог-апостериорного шанса. Для сравнения рассматривались также следующие подходы:

- метод объединения векторов с конкатенацией и ближайшим соседом по косинусному сходству;
- метод только по лицу;
- метод только по силуэту.

4.2. РЕЗУЛЬТАТЫ МЕТРИК

В таблице 1 приведены ключевые метрики классификации, полученные на проверочном наборе данных. Рассматриваются четыре подхода: предложенный байесовский алгоритм, метод объединения признаков (конкатенация векторов лица и силуэта), а также базовые одномодальные методы — по лицам и по силуэтам соответственно. Для каждого метода рассчитаны значения точности (precision), полноты (recall), F_1 -меры и площади под ROC-кривой (AUC).

Таблица 1. Ключевые метрики на проверочном наборе

Метод	Точность	Полнота	F_1 -мера	AUC
Байесовский метод	0,9565	0,9554	0,9431	0,9993
Метод объединения	0,9240	0,9198	0,9052	0,9980
Метод по лицам	0,6222	0,6763	0,6028	0,9454
Метод по силуэтам	0,8579	0,8582	0,8329	0,9862

Из таблицы 1 следует, что байесовский метод демонстрирует наивысшие значения по всем метрикам. Особенно заметно его преимущество по F_1 -мере и AUC: при $F_1 = 0,9431$

и $AUC=0,9993$ он превосходит метод объединения, а также значительно опережает одномодальные варианты, особенно в части работы по лицевым признакам. На рис. 3 представлены ROC-кривые (micro-AUC) для всех подходов; байесовский метод достигает 0,9993, что выше, чем у метода объединения (0,9980) и одномодальных решений (лицо — 0,9454; силуэт — 0,9862). Кривые Precision–Recall на рис. 4 показывают $AP = 0,951$ для байесовского метода против 0,932 у объединения. На рис. 5 приведена t-SNE-визуализация векторов байесовского метода для десяти наиболее частых классов: кластеры оказались более плотными и разделёнными по сравнению с конкатенированными векторами.

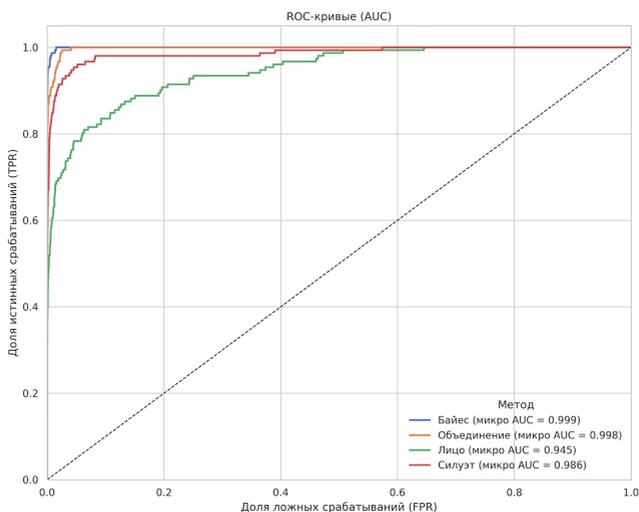


Рис. 3. ROC-кривые (micro-AUC) для всех методов

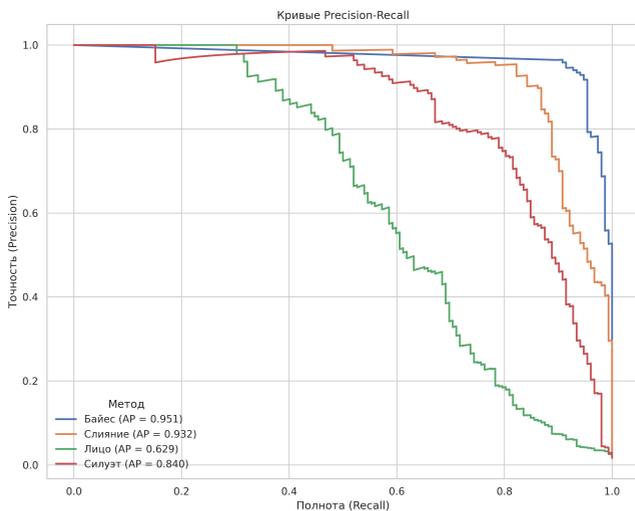


Рис. 4. Кривые «точность — полнота» (*micro-AP*) для всех методов

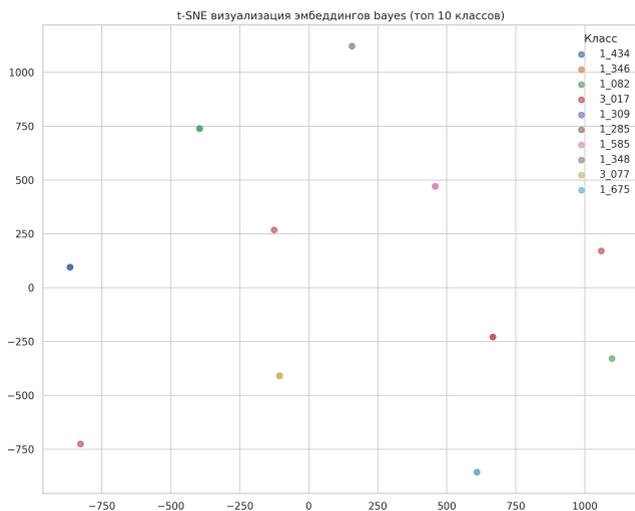


Рис. 5. *t-SNE*-визуализация векторов байесовского метода для топ-10 классов

5. Выводы и перспективы

В статье предложена универсальная байесовская схема объединения признаков лица и силуэта для задачи реидентификации личности в системах видеонаблюдения. В основе метода лежит вычисление апостериорных вероятностей по правилу Байеса, где каждый вектор моделируется нормальным распределением с параметрами, оценёнными по опорным данным. Такой подход позволяет адаптивно учитывать надёжность каждой модальности: при повреждённых или отсутствующих фрагментах изображения система автоматически снижает вклад шумного признака и фокусируется на более информативном.

Экспериментальная проверка на датасете SUNK03 показала, что предложенная схема превосходит как классическую конкатенацию векторов, так и одномодальные методы. Байесовская модель продемонстрировала прирост точности идентификации более чем на 3% по сравнению с линейным объединением и обеспечила micro-AUC 0,9993 и micro-AP 0,951. Кроме того, оцениваемая апостериорная вероятность позволяет в реальном времени формировать порог уверенности и, при необходимости, запрашивать дополнительный образ или выдавать несколько наиболее вероятных кандидатов.

Одним из важных преимуществ предложенного байесовского подхода является его низкая вычислительная сложность и хорошая масштабируемость. В отличие от сложных методов с обучаемыми коэффициентами или методов на основе глубоких сетей с перекрестным вниманием, предложенная модель требует лишь расчета евклидовых расстояний до центроидов классов в низкоразмерных пространствах признаков, что реализуется крайне эффективно даже для больших баз идентифицируемых личностей. При увеличении количества личностей (классов) вычислительная нагрузка растёт линейно, обеспечивая лёгкое масштабирование решения. Кроме того, подход позволяет легко добавлять новые модальности и адаптировать систему под изменения условий съёмки без необходимости переобучения сложных глубоких

архитектур.

Одним из ограничений предложенного подхода является использование предположения о нормальности распределения признаков в каждой модальности, что может не всегда идеально соответствовать реальным данным, особенно в случае наличия выбросов или нетипичных наблюдений. Для преодоления этого ограничения рекомендуется в дальнейшем рассмотреть применение более гибких статистических моделей, таких как смеси гауссовых распределений или непараметрические методы оценки плотностей. Также можно обсудить устойчивость модели к шумам и искажениям в данных и возможности её улучшения через использование робастных статистических методов или предварительной фильтрации входных данных.

Дальнейшие направления работы связаны с расширением модели несколькими биометрическими модальностями и внедрением более гибких ковариационных структур, позволяющих учитывать попарную зависимость признаков. Также перспективно исследовать онлайн-обучение параметров распределений для поддержки динамических баз персоналий и адаптацию порога уверенности под изменяющиеся условия съёмки.

Литература

1. АМИРГАЛИЕВА З., САДЫКОВА А., КЕНШИМОВ К. *Разработка модифицированного алгоритма Виолы – Джонса для распознавания лиц* // Известия НАН РК. Серия физико-математическая. – 2022. – №4. – С. 258–267.
2. ИГНАТЬЕВА С.А., БОГУШ Р.П. *Алгоритм реидентификации людей по изображениям систем видеонаблюдения с использованием нейросетевого составного дескриптора* // Системный анализ и прикладная информатика. – 2024. – №1 – С. 54–65.
3. КОСУЛИН К.Э., КАРПОВ А.А. *Методы аудиовизуального распознавания людей в масках* // Научно-технический вестник информационных технологий, механики и оптики. – 2022. – Т. 22, № 3. – С. 415–432.

4. МУБИНОВА Э.С. *Эволюция технологий распознавания лиц и их влияние на общественное восприятие безопасности* // Вестник науки. – 2024. – №9(78). – С. 168–171.
5. САВЧЕНКО А.В. *Метод максимально правдоподобных рас-согласований в задаче распознавания изображений на основе глубоких нейронных сетей* // Компьютерная оптика. – 2017. – №41(5). – С. 819–828.
6. СОКОЛОВА А.Д., САВЧЕНКО А.В., НИКОЛЕНКО С.И. *Поиск редких данных в задаче распознавания лиц на изображениях* // Компьютерная оптика. – 2022. – №46(2). – С. 318–327.
7. ТЕЛЕНЬКОВ А. *Система распознавания силуэтов применяется МВД в пяти регионах РФ* // ТАСС (ЦИПР). – 24.06.2021.
8. ХАЧУМОВ М.В., НГУЕН Т.З. *Распознавание лиц по фотографиям на основе инвариантных моментов* // Современные проблемы науки и образования. – 2015. – №2 – С. 10.
9. САВЧУК В.П. *Байесовские методы статистического оценивания: надежность технических объектов.* – М.: Наука, Гл. ред. физ.-мат. лит., 1989.
10. DENG J., GUO J., XUE N. et al. *ArcFace: Additive Angular Margin Loss for Deep Face Recognition* // Proc. CVPR. – 2019. – P. 4690–4699.
11. HERMANS A., BEYER L., LEIBE B. *In Defense of the Triplet Loss for Person Re-Identification* // arXiv preprint. – 2017. – arXiv:1703.07737.
12. IRFAN B., GARCIA ORTIZ M., LYUBOVA N. et al. *Multi-modal Open World User Identification* // ACM Trans. Human-Robot Interaction. – 2022. – Vol. 11(1). – P. 1–50.
13. KOO J.H., CHO S.W., BAEK N.R. et al. *CNN-based multimodal human recognition in surveillance environments* // Sensors. – 2018. – Vol. 18(9). – P. 3040.
14. LIAO Z., DI D., HAO J. et al. *MMM-GCN: Multi-Level Multi-Modal Graph Convolution Network for Video-Based Person Identification* // LNCS. – 2023. – Vol. 13834. – P. 3–15.

15. LI W., ZHAO R., XIAO T. et al. *DeepReID: Deep Filter Pairing Neural Network for Person Re-identification* // Proc. CVPR. – 2014. – P. 152–159.
16. LI W., ZHU X., GONG S. *Harmonious Attention Network for Person Re-Identification* // Proc. CVPR. – 2018. – P. 2285–2294.
17. SCHROFF F., KALENICHENKO D., PHILBIN J. *FaceNet: A unified embedding for face recognition and clustering* // Proc. CVPR. – 2015. – P. 815–823.
18. SHOUKRY N., ABD EL GHANY M.A., SALEM M.A.-M. *Multi-Modal Long-Term Person Re-Identification Using Physical Soft Bio-Metrics and Body Figure* // Applied Sciences. – 2022. – Vol. 12(6). – P. 2835.
19. VEZZANI R., BALTIERI D., CUCCHIARA R. *People re-identification in surveillance and forensics: A survey* // ACM Computing Surveys. – 2013. – Vol. 46(2). – P. 1–37.
20. ZHANG M., LIU R., ABE N. *Face Helps Person Re-Identification: Multi-modality Person Re-ID Based on Vision-Language Models* // Proc. IJCB. – 2024.
21. CHEN S., LIU Y., GAO X. et al. *MobileFaceNets: Efficient CNNs for Accurate Real-Time Face Verification on Mobile Devices* // arXiv preprint. – 2018. – arXiv:1804.07573.
22. DOSOVITSKIY A. et al. *An image is worth 16x16 words: Transformers for image recognition at scale* // arXiv preprint. – 2020. – arXiv:2010.11929.

BAYESIAN MODEL FOR FEATURE REPRESENTATION FUSION IN PERSON RE-IDENTIFICATION

Konstantin Rusakov, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, research fellow (rusakov@ipu.ru).

Abstract: This paper addresses the problem of person re-identification in surveillance systems based on the analysis and integration of heterogeneous descriptive features. Unlike traditional approaches relying on a single biometric modality, a general-purpose probabilistic method is proposed to combine features derived from various sources. The method is based on Bayesian inference and models each feature as a random variable following a multivariate normal distribution. Parameters of the distributions are estimated from the training data. Decision making is performed by maximizing the posterior probability of identity given the available evidence. To extract features, transformer-based architectures with attention mechanisms are used, ensuring robustness to visual noise and viewpoint variation. The proposed model is compared with classical approaches based on linear combination of feature vectors. Experiments conducted on an open re-identification dataset demonstrate that the Bayesian scheme improves recognition accuracy and remains effective in cases of partial information loss. Moreover, the method provides a quantitative confidence score associated with each decision, making it particularly suitable for deployment in safety-critical environments such as automated video surveillance and access control systems.

Keywords: person re-identification, feature fusion, Bayesian model, computer vision, transformer architectures, video surveillance.

УДК 004.93

ББК 32.81

*Статья представлена к публикации
членом редакционной коллегии Н.А. Коргиным.*

Поступила в редакцию 22.04.2025.

Дата опубликования 31.05.2025.

ИССЛЕДОВАНИЕ АУКЦИОННОЙ СХЕМЫ РАСПРЕДЕЛЕНИЯ ЗАКАЗА ВНУТРИ АССОЦИАЦИИ ПРОИЗВОДИТЕЛЕЙ

Гасанов И. И.¹
(ФИЦ ИУ РАН, Москва)

Продолжается описание проекта, представленного в статье "Проект системы распределения заказа внутри ассоциации производителей посредством множественных аукционов" // Управление большими системами. – 2024. – Вып. 112. Исследуется схема сетевого взаимодействия внутри ассоциации предприятий, занимающейся сборочным производством фиксированного набора изделий. Рассматривается производство, представленное в виде направленного ациклического графа, вершинам которого сопоставлены производимые участниками ассоциации комплектующие, причём для сборки деталей более низкого ранга используются детали, имеющие более высокий ранг. Взаимодействие базируется на системе множественных локальных аукционов, посредством которых производится распределение производства очередного заказа между предприятиями – членами ассоциации. Если в предыдущей работе дан краткий обзор проекта и очерчены проблемы, связанные с его реализацией, то в настоящей статье делается детальный разбор рассматриваемой схемы, предлагаются правила, которых рекомендуется придерживаться участникам схемы при выборе решений. Исследуется проблема рационального выбора участниками цен производимых изделий. В заключении указываются риски, связанные с реализацией проекта, и намечается план исследования эффективности схемы в режиме компьютерной имитации.

Ключевые слова: аукцион, сетевое взаимодействие, планирование производства.

1. Введение

В статье рассматривается ассоциация предприятий, которая осуществляет сборочное производство некоторого набора изделий. Модель ассоциации и общее описание аукционной схемы даны в статье [3]. Ниже для удобства чтения даётся лишь краткое описание изучаемой конструкции, повторяются некоторые основные обозначения и определения.

Участники производят как сами изделия, так и комплектующие для их монтажа. Производство может быть пред-

¹ Игорь Искендерович Гасанов, с.н.с. (gasanov48@yandex.ru).

ставлено как ациклический граф, вершинам которого соответствуют производимые продукты. Первому уровню графа сопоставлены конечные продукты. Изделия каждого уровня монтируются из комплектующих, находящихся на более высоких уровнях графа. Данный граф будем называть технологическим.

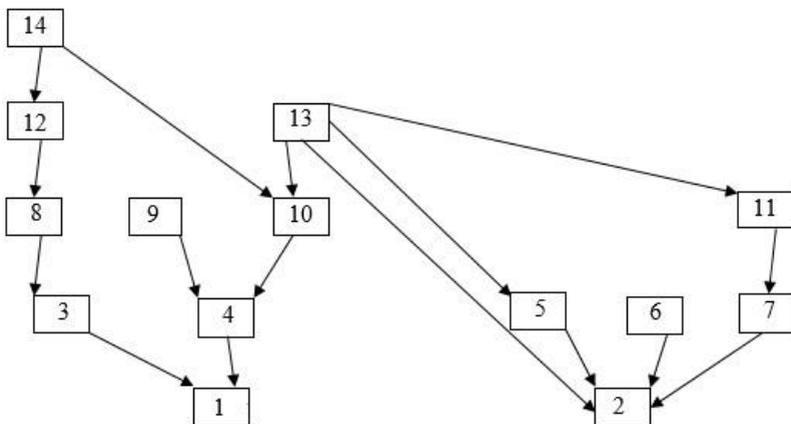


Рис. 1. Иллюстративный пример технологического графа

Условимся обозначать векторы прописными буквами с верхней крышкой, а их компоненты – строчными буквами с индексами, т.е. по шаблону $\bar{Q} = (q_1, q_2, \dots, q_N)$.

Множество участников ассоциации обозначим $\hat{L} = \{l\}_{l=0}^L$, множество изделий, производимых участниками ассоциации, обозначим $\hat{M} = \{m\}_{m=1}^M$.

Помимо производителей в ассоциации есть выделенный участник – Центр, который принимает заказы на производство изделий из набора $M^1 \subset \hat{M}$ и координирует их сборку. Присвоим ему номер 0.

Обозначим \bar{W} вектор продукции, в котором компоненты w_m соответствуют количеству (объёму) изделий $m \in \hat{M}$. Запись \bar{w}_m будет означать вектор \bar{W} с единственной ненулевой компо-

нентой t , равной w_m . Вектор \bar{w}_m , у которого компонента $w_m = 1$, будем обозначать \bar{e}_m .

Обозначим через $G(m) \subset \hat{M}$ множество тех деталей $\mu \in \hat{M}$, из которых собирается изделие m , а через $\bar{G}(\bar{w}_m)$ – вектор объёмов комплектующих $G(m)$, требующихся для производства изделия m в объёме w_m . В $G(m)$ входят только детали финальной сборки изделия m : если деталь $\mu \in G(m)$ сама является сборной, то объёмы комплектующих $G(\mu)$, необходимые для производства вектора $\bar{G}(\bar{\mu}_m)$ в компоненты вектора $\bar{G}(\bar{w}_m)$ не входят.

Назовём проекцией вектора $\bar{W} = (w_1, w_2, \dots, w_M)$ и обозначим $\bar{G}(\bar{W})$ сумму $\sum_{m \in M} G(\bar{w}_m)$.

Будем называть комплектующие из множества $G(m)$ предшественниками m , а изделие m – их потомком. На рис. 1 вершинам графа соответствуют изделия $m \in \hat{M}$, а дуги ведут от предшественников к потомкам.

Обозначим через $\bar{H}(\bar{W})$ и назовём полной проекцией вектора \bar{W} вектор деталей $m \in \hat{M}$ всех уровней, необходимых для сборки \bar{W} . Вектор $\bar{H}(\bar{W})$ – это сумма последовательных проекций $\bar{G}(\bar{W}) + \bar{G}(\bar{G}(\bar{W})) + \bar{G}(\bar{G}(\bar{G}(\bar{W}))) + \dots$

Назовём рангом изделия $m \in \hat{M}$ и обозначим $R(m)$ максимальное число дуг на технологическом графе от вершины m до его корня плюс 1.

2. Подробное описание схемы аукционов

Обзор аукционной схемы был сделан в работе [3]. Её эффективность следует оценивать по тому, насколько надёжно она генерирует план исполнения заказов и как сильно себестоимость продукции, произведённой по такому плану, отличается от оптимума исходной задачи (1)–(3) [3], поиск которого сложно реализовать на практике.

Рассматриваемая схема – это способ декомпозировать исходную задачу, в которой векторы производства всех участни-

ков и матрица поставок определяются Центром. Аукционная схема предоставляет участникам возможность самостоятельного выбора решений. Следуя традиции исследований в области теории игр [2, 4, 5], полагаем, что цель каждого участника – это максимизация прибыли от участия в выполнении заказа. При этом, формируя аукционные предложения, он вынужден учитывать несколько противоположных факторов. Участнику выгодно увеличение объёмов своего участия в заказе и стоимости производимых в рамках заказа изделий. Но его предложения по поставкам не должны входить в противоречие с его ресурсными возможностями по их выполнению, а цена его продукции должна быть конкурентной относительно цен других участников. Он должен учитывать, что от его предложений зависит, будет ли вообще сформирован допустимый план, без которого его прибыль окажется нулевой. Пример 2 из работы [3] демонстрирует, что существование допустимого плана производства в исходной задаче не гарантирует его создания посредством аукционной схемы.

Проблема в том, что в рамках схемы производитель вынужден делать выбор в условиях очень существенной, многоступенчатой неопределённости. Формируя свои предложения для текущего аукциона, он не знает, какими будут решения других участников на этом и последующих аукционах. А значит, не знает ни того, каковы будут задания по сформированному по результатам всех аукционов плану, ни того, как эти задания зависят от его собственных предложений. С точки зрения теории игр рассматриваемая схема представляет собой сложную многошаговую игру многих лиц с неполной информацией [7, 8, 10]. Шаги, на которых участник принимает решение, это: первоначальный выбор желаемого вектора производства; затем по мере завершения аукционов разных рангов выбор скорректированного объёма производства для каждого из планируемых им к производству изделий, а также выбор объёмов предложений по поставкам этих изделий на последующих аукционах; наконец, распределение заданий поставщикам комплектующих после того как к нему поступает окончательный заказ на производство изделий. Обобщённо данную игру можно описать следующим образом.

Пусть $1, \dots, \Lambda$ – последовательность шагов, на которых участники принимают решения. Обозначим $U_l = (u_{1,l}, u_{2,l}, \dots, u_{\Lambda,l})$ стратегию участника l . (Не обязательно, что участник принимает решения на всех шагах.) Совокупность стратегий всех участников обозначим $U = (U_1, U_2, \dots, U_L)$. Отрезок стратегии участника l до шага r обозначим $\tilde{U}_l^r = (u_{1,l}, u_{2,l}, \dots, u_{r,l})$, а совокупность таких отрезков всех участников – через \tilde{U}^r . Организация аукционной схемы такова, что управление участника l на шаге r – это функция от отрезка стратегий \tilde{U}^{r-1} : $u_{r,l} = \Phi_l(\tilde{U}^{r-1}, \eta_l)$, где η_l – параметры, состояния внутренних ресурсов участника.

Цель каждого участника – максимизация своей прибыли. Её величина зависит от решений всех участников на всех шагах схемы, т.е. прибыль $I_l = I_l(\bar{U})$. В то же время при выборе решения на шаге r участник l не знает выбор других участников на шагах больше $r - 1$. Для поиска оптимального решения в вероятностных категориях участнику потребовалась бы гипотеза о распределении вероятностей выборов у остальных игроков, и это было бы вероятностное распределение очень большой размерности. Если искать решение с позиций гарантированного результата, то участнику при любой его стратегии гарантирован только 0 – ничто не мешает остальным участникам проигнорировать заказ и тогда он не будет исполнен. Если искать для участников какое-то равновесное решение, то получим задачу едва ли менее сложную, чем задача (1)–(3) в работе [3], т.е. то, от чего мы пытались уйти посредством аукционной схемы. Сложность задачи не даёт шансов на поиск оптимальных стратегий. Тем не менее представляется, что аукционный выбор участников можно разумным образом регламентировать, предложив для этого некоторый набор правил. Далее в процессе описания схемы формулируются такие правила.

Ниже по тексту используются величины с множественными индексами по образцу $X_{\alpha,\beta}^{n,\gamma,\delta}$. Условимся, что индекс α указывает на поставляемую или изготавливаемую комплектующую; индекс β – на изделие, для которого предназначена эта комплектующая; индекс n – на этап процедуры принятия решений; индекс γ – на поставщика изделия; индекс δ – на его получателя. Переменные в рамках модели полагаем непрерывными.

Пусть Центру поступил заказ на производство изделия $m \in M^1$ в объёме w_m^* . Обозначим его \bar{W}^* . Формирование плана производства разбивается на несколько этапов.

Этап 1. Центр рассчитывает и объявляет участникам вектор всех необходимых для исполнения заказа деталей $\bar{H}(\bar{W}^*) = (h_1(\bar{W}^*), \dots, h_M(\bar{W}^*))$.

Этап 2. Участники выбирают векторы изделий $\bar{W}^{2,l} = (w_1^{2,l}, w_2^{2,l}, \dots, w_M^{2,l})$, которые они хотели бы произвести в рамках данного заказа. Для этого участник l выбирает точку на подмножестве пространства \mathbb{R}^M , заданном ресурсными ограничениями

$$(1) \quad \forall l: \bar{\Psi}^l(\bar{W}^{2,l}) \leq \bar{S}^l,$$

и требованиями заказа, т.е. неравенствами

$$(2) \quad w_n^{2,l} \leq h_n(\bar{W}^*) \quad \forall n \in \hat{M}.$$

Здесь $\bar{\Psi}^l(\bar{W}) : \mathbb{R}^M \rightarrow \mathbb{R}^p$ – вектор-функция, отражающая расход ресурсов каждого вида p участником l при производстве вектора продукции \bar{W} , \bar{S}^l – текущие объёмы ресурсов участника.

Исходим из того, что каждый производитель в первую очередь заинтересован в том, чтобы заказ был выполнен. Понятно, что недоминируемый вектор $\bar{W}^{2,l}$ предоставляет большие возможности при формировании плана производства, чем доминируемый им. Поэтому формулируем следующее требование.

Правило 1. Каждый участник $l \in \hat{L}$ выбирает вектор $\bar{W}^{2,l}$ среди недоминируемых векторов, удовлетворяющих неравенствам 1, 2.

Выбрав вектор $\bar{W}^{2,l}$ участник l объявляет аукционы A_k^l на поставки комплектующих, нужных для производства этого вектора продукции, т.е. объёмы $g_k(\bar{W}^{2,l})$, $k \in \hat{M}$.

Этап 3. Это этап собственно аукционов. Они проводятся последовательно от изделий более высоких рангов к более низким. Аукцион A_k^l представляет собой сбор участником l предложений по объёмам и ценам поставок от производителей изделия k . Присвоим такому аукциону ранг комплектующей k , т.е. $R(k)$. Все предложения от поставщиков на аукционе A_k^l образуют пул комплектующих Π_k^l . Объём пула будем обозначать $v(\Pi_k^l)$.

Используя объёмы пулов $v(\Pi_k^l)$, предприятие l определяет объёмы поставок своей продукции, которые он планирует предложить на последующих аукционах. Обозначим эти объёмы как $y_n^{3,j,l}$. Рассмотрим процедуру выбора значений $y_n^{3,j,l}$ подробнее.

Предприятию l следует скорректировать объёмы производства $w_n^{2,l}$ и определить новые объёмы $w_n^{3,l}$, достижимые с пулами комплектующих, собранных на объявленных им аукционах. Пусть $\alpha^l = \{n \in \hat{M} \mid w_n^{2,l} > 0, g_k(\bar{w}_n^{2,l}) > 0\}$, иначе говоря, это множество изделий, которые l запланировал для производства на этапе 2 и для которых требуются комплектующие. Изделия $n \in \alpha^l$ могут иметь разный ранг. По мере проведения аукционов по изделиям $k \in G(n)$, $n \in \alpha^l$, участник l выделяет объёмы комплектующих из пулов Π_k^l для обеспечения производства изделий n . Остаток от пула $v(\Pi_k^l)$ к периоду проведения аукционов ранга r обозначим как $v_r(\Pi_k^l)$.

При проведении аукционов ранга $R(n)$, $n \in \alpha^l$, участник l может рассчитать максимальный объём изделия n , который он может изготовить, используя пулы $v_{R(n)}(\Pi_k^l)$, $k \in G(n)$. Это величина

$$\chi(n) = \max(w_n \mid \bar{g}_k(w_n) \leq v_r(\Pi_k^l) \forall k \in g(n)).$$

Участник l имеет возможность запланировать любой объём $w_n^{3,l}$ производства изделия n , не превышающий значение $\min(\chi(n); w_n^{2,l})$, но он должен учитывать, что комплектующие из

множества $G(n)$ могут использоваться в других изделиях из множества α^l с рангами $R(n)$ или ниже. При этом показатель $\chi(n')$ при $R(n') < R(n)$ определится только к началу аукционов ранга $R(n')$. Характер неопределённости таков, что среди разных вариантов распределения пулов комплектующих между изделиями вряд ли можно указать оптимальный. Предлагается следующая процедура.

Пусть $\alpha^l(r) \subset \alpha^l$ множество изделий, имеющих ранг r , $\bar{W}(r)$ – некоторый вектор объёмов таких изделий. Следующие неравенства определяют допустимую область O_r^l для векторов $\bar{W}(r)$, которые предприятие l может изготовить, используя пулы комплектующих $v_r(\Pi_k^l)$ и не превышая объёмы $\bar{W}^{2,j}$:

$$w_n \leq w_n^{2,j}, \forall n \in \hat{M},$$

$$g_k(\bar{W}(r)) \leq v_r(\Pi_k^l).$$

Здесь, как и выше, $g_k(\bar{W}(r))$ обозначает объём комплектующих k , используемых при производстве вектора продукции $\bar{W}(r)$.

Правило 2. Участник l выбирает в качестве целевого вектора производства $\bar{W}(r)$ один из недоминируемых векторов выпуклого многогранника O_r^l .

Выбрав такой вектор производства, который обозначим $\bar{W}_r^{j,0}$, участник l формирует свои аукционные предложения $y_n^{3,j}$, $n \in \alpha^l$, согласно следующему правилу.

Правило 3. Участник l , определяя свои предложения $y_n^{3,j}$ по аукционам ранга r , действует следующим образом:

А) если $w_{nr}^{j,0} = h_n(\bar{W}^*) = w_n^{2,j}$, то $y_n^{3,j} = g_n(\bar{W}^{2,j}) \quad \forall j \in \hat{L}$,

т.е. если объём $w_{nr}^{j,0}$ достаточен для удовлетворения всей потребности в изделии n по заказу \bar{W}^* , то l предлагает каждому производителю j тот объём комплектующих n , который j объявил на этапе 2 (в силу неравенств, определяющих множество O_r^l , $w_{nr}^{j,0}$ не может быть больше $h_n(\bar{W}^*)$);

Б) если $w_{nr}^{j0} < h_n(\bar{W}^*)$, то l выбирает значения $y_n^{3,l,j}$, решая следующую задачу:

$$\sum_{j \in L} y_n^{3,l,j} \rightarrow \max$$

при ограничениях

$$\sum_{j \in L} y_n^{3,l,j} \leq w_{nr}^{j0},$$

$$y_n^{3,l,j} \leq g_n(\bar{W}^{2,j}) \forall j \in \hat{L}.$$

Обоснование. Предложения $y_n^{3,l,j}$ носят характер обязательств: после того, как на этапе 4 план заказа будет сформирован, участник l обязан осуществить любую плановую поставку $y_n^{4,l,j} \leq y_n^{3,l,j}$. Но на этапе 3 объёмы заказов $y_n^{4,l,j}$ участнику l ещё не известны. Однако процедура формирования плана такова, что $\sum_{l \in L} y_n^{4,l,j}$ не может превысить $\min[w_{nr}^{j0}; h_n(\bar{W}^*)]$. Поэтому если $w_{nr}^{j0} = h_n(\bar{W}^*)$, то l знает, что он сможет произвести любой объём изделия n , который будет заказан в ответ на его предложения.

Если $w_{nr}^{j0} < h_n(\bar{W}^*)$, то предложения участника l по поставкам изделия n не могут превышать объём w_{nr}^{j0} без риска не выполнить обязательства из-за недостатка комплектующих, и он распределяет по аукционам A_n^j именно этот объём.

В предложении о поставке $y_n^{3,l,j}$ указывается стоимость за единицу продукции $c_n^{l,j}$. Цены поставок участника l образуют матрицу $C^l = (c_n^{l,j})$. Расчёт цен обсуждается в следующем разделе.

Весь корпус предложений о поставках от всех участников вместе со стоимостями обозначим как пару (\hat{Y}^3, \hat{C}) :

$$\hat{Y}^3 = \bigcup_{l \in L} \left(\bigcup_{j \in L} y_n^{3,l,j} \right), \quad \hat{C} = \bigcup_{l \in L} C^l.$$

Этап 4. На этом завершающем этапе по совокупности предложений \hat{Y}^3 , верстается план производства, т.е. распределение поставок между участниками, матрица $\hat{Y}^4(\hat{Y}^3, \hat{C}) = (y_n^{4,l,j})$, $n \in \hat{M}$, $l, j \in \hat{L}$. План будет сформирован, если сумма предло-

жений Центру от производителей изделия t не меньше чем объём заказа, т.е. если $\sum_{l \in L(m)} y_m^{3,l,0} \geq w_m^*$.

Для сформированного плана поставок должны выполняться неравенства

$$y_n^{4,l,j} \leq y_n^{3,l,j} \quad \forall l, j \in \widehat{L}, \forall n \in \widehat{M},$$

т.е. задания участнику l на этапе 4 не могут превышать его предложения на этапе 3.

Процедура выбора значений $y_n^{4,l,j}$ организована в виде координируемой Центром волны, восходящей по рангам узлов технологического графа. Производитель l дожидается получения заданий на поставки изделий $\sum_{j \in L} y_n^{4,j,l}$ от всех участников j , которым он посылал предложения $y_n^{3,l,j}$. После этого участнику l известен весь объём изделий n , который ему необходимо изготовить, $w_n^{4,l} = \sum_{j \in L} y_n^{4,j,l}$. Под этот объём участник l определяет вектор комплектующих $\bar{G}(\bar{w}_n^{4,l})$. Для каждой из ненулевых компонент $g_k(\bar{w}_n^{4,l})$ этого вектора на аукционах третьего этапа A_k^l был создан пул комплектующих объёмом $v(\Pi_k^l)$. Правила формирования предложений на предыдущих этапах гарантируют, что эти пулы достаточны для производства объёма $w_n^{4,l}$.

Участник l выбирает из пула $v(\Pi_k^l)$ предложения, начиная с наиболее дешёвых. Он упорядочивает предложения $y_k^{3,j,l}$ по возрастанию цен за единицу продукции т.е. по величинам $c_k^{j,l}$, формируя пул $\tilde{\Pi}_k^l$ с объёмом $g_k(\bar{w}_n^{4,l})$. Затем l распределяет задания производителям i , чьи предложения $y_k^{3,j,l}$ попали в пул $\tilde{\Pi}_k^l$, т.е. определяет компоненты $y_k^{4,j,l}$. Для всех производителей изделия k из пула $\tilde{\Pi}_k^l$ задания $y_k^{4,j,l} = y_k^{3,j,l}$, кроме последнего по порядку j^* , для которого $y_k^{4,j^*,l} = g_k(\bar{w}_n^{4,l}) - \sum_{i < l} y_k^{3,j^*,l}$.

План производства полностью сгенерирован тогда, когда волна заданий доходит до конечных узлов.

РАСЧЁТ ЦЕН ПОСТАВОК

Если участник l поставил участнику j комплектующие n в объёме $y_n^{4j,j}$, то после исполнения заказа ему от j перечисляется сумма $c_n^{l,j} \cdot y_n^{4j,j}$. Общая сумма перечислений участнику l по всем его поставкам составит величину $\sum_{j \in L} \sum_{n \in M} c_n^{l,j} \cdot y_n^{4j,j}$. Из этих

денег l делает дальнейшие выплаты транспортным компаниям и своим поставщикам комплектующих, после чего его выручка от произведённой продукции составляет величину

$$(3) \quad \Phi^l = \sum_{j \in L} \sum_{n \in M} c_n^{l,j} \cdot y_n^{4j,j} - d^l(\bar{W}^{4j}) - \delta^l(\bar{W}^{4j}) - \sum_{j \in L} \sum_{n \in M} t_n^{l,j} \cdot y_n^{4j,j}.$$

Здесь $d^l(\bar{W}^{4j})$ – аукционная стоимость комплектующих $\bar{G}(\bar{W}^{4j})$, $\delta^l(\bar{W}^{4j})$ – стоимость материалов, необходимых для производства вектора продукции \bar{W}^{4j} и приобретаемых вне ассоциации, $t_n^{l,j}$ – удельные транспортные расходы.

Прибыль θ^l участника l по данному заказу определяется наценкой, установленной участником l на свою продукцию и включённой в цены $c_n^{l,j}$. Она равна разности между выручкой Φ^l и $\psi_1^l(\bar{W}^{4j})$ – собственными финансовыми затратами на сборку произведённых изделий (зарплата, операционные расходы и т.п.):

$$\theta^l = \Phi^l - \psi_1^l(\bar{W}^{4j}).$$

Чем выше назначенные участником l цены продукции C^l , тем выше его прибыль при одинаковом объёме плановых поставок. Но и тем больше риск проиграть в конкуренции за эти поставки. В условиях тотальной неопределённости сложно дать участникам однозначные рекомендации по эффективному выбору значений C^l . Для участника l представляется разумным выбирать цены так, чтобы стоимость производимой им продукции гарантировала его от убытков. Добавив к ценам бонус, участник обеспечит себе положительную прибыль, если только будет привлечён к исполнению заказа.

Производителю l требуется выбрать такие величины $c_n^{l,j}$, чтобы для любых $n \in \bar{M}$, $l, j \in \bar{L}$, для любого вектора $\bar{W}^{4j} \leq \bar{W}^{3,l}$ выполнялось неравенство

$$(4) \quad \psi_1^l(\bar{W}^{4,l}) + d^l(\bar{W}^{4,l}) + \delta^l(\bar{W}^{4,l}) + \sum_{j \in L} \sum_{n \in M} t_n^{l,j} \cdot y_n^{4,l,j} \leq \sum_{j \in L} \sum_{n \in M} c_n^{l,j} \cdot y_n^{4,l,j}$$

Однако цены $c_n^{l,j}$, гарантирующие выполнение неравенства (4), нужно рассчитать на третьем этапе аукционной процедуры, т.е. ещё до формирования плана производства, а это непростая задача. Себестоимость производства $\psi_1^l(\bar{W}^{4,l})$ – сложная, многомерная функция. Стоимость ресурсов, необходимых для производства объёма $w_n^{4,l}$, зависит не только от величины $w_n^{4,l}$, но и от объёмов производства участником l других его изделий, вошедших в свёрстанный план, а также от лимита его ресурсов. Сделаем следующее достаточно естественное допущение.

Условие 1. $\psi_1^l(\bar{W})$ – дифференцируемые функции, такие что если вектор производства $\bar{W}' \geq \bar{W}$, то

$$\frac{\partial \psi_1^l(\bar{W}')}{\partial w_n} \geq \frac{\partial \psi_1^l(\bar{W})}{\partial w_n} \quad \forall n \in \hat{M}.$$

При выполнении условия 1 можно предложить следующее решение. Оценим затраты участника l на производство единичного объёма e_n^l изделия n величиной

$$\rho_n^l = \frac{\psi_1^l(\bar{W}^{3,l}) - \psi_1^l(\bar{W}_{-n}^{3,l})}{w_n^{3,l}},$$

где $\bar{W}_{-n}^{3,l} = \bar{W}^{3,l} - \bar{w}_n^{3,l}$. Иначе говоря, $\bar{W}_{-n}^{3,l}$ – это вектор $\bar{W}^{3,l}$ с обнулённой n -й компонентой.

Из условия 1 нетрудно вывести неравенство

$$\psi_1^l(\bar{W}) \leq \sum_{n \in M} \rho_n^l \cdot w_n = \sum_{n \in M} \rho_n^l \cdot \sum_{j \in L} y_n^{4,l,j} \quad \forall l \quad \forall \bar{W} \leq \bar{W}^{3,l}.$$

Отсюда следует, что

$$(5) \quad \psi_1^l(\bar{W}^{4,l}) \leq \sum_{n \in M} \rho_n^l \cdot w_n^4 = \sum_{n \in M} \rho_n^l \cdot \sum_{j \in L} y_n^{4,l,j}$$

Стоимость комплектующих для производства участником l произвольного вектора продукции \bar{W} равна стоимости всех деталей $g_k(\bar{W})$ по ценам аукционных предложений $c_k^{l,j}$, т.е. $d^l(\bar{W}) = \sum_{k \in M} d^l(\bar{g}_k(\bar{W}))$.

Участник l заинтересован в снижении стоимости используемых комплектующих и стремится минимизировать величины

$d^l(\bar{g}_k(\bar{W}))$, $k \in \bar{M}$. Поэтому на этапе 4 он формирует пулы $\bar{\Pi}_k^l$ так, как описано выше, т.е. делая выбор от более дешёвых предложений к более дорогим. Вследствие этого $d^l(\bar{g}_k(\bar{W}))$ будут кусочно-линейными выпуклыми функциями. Оценим стоимость комплектовочных для производства участником l единицы товара $n \in \bar{M}$ величиной $d^l(\bar{e}_n) = \sum_{k \in \bar{M}} s_k^l \cdot g_k(\bar{e}_n)$, где $s_k^l = d^l(g_k(\bar{W}^{3,l})) / g_k(\bar{W}^{3,l})$.

Тогда для любого вектора $\bar{W} \leq \bar{W}^{3,l}$ для каждого n выполняется неравенство $d^l(\bar{w}_n) \leq d^l(\bar{e}_n) \cdot w_n$ и следовательно

$$(6) \quad d^l(\bar{W}^{4,l}) \leq \sum_{n \in \bar{M}} d^l(\bar{e}_n) \cdot w_n.$$

Положим стоимость материалов, закупаемых участником l для изделий n вне ассоциации, пропорциональной их объёму и обозначим их удельную стоимость через η_n^l .

Тогда если производитель l на этапе 3 аукционной схемы выберет для поставок $y_n^{3,l,j}$ цены

$$c_n^{l,j} = \rho_n^l + s_n^l + \eta_n^l + t_n^{l,j},$$

то, в силу, соотношений (3), (5), будет выполнено неравенство (6). При этом в случаях $\bar{W}^{4,l} = 0$ и $\bar{W}^{4,l} = \bar{W}^{3,l}$ формула (6) трансформируется в равенство.

3. Заключение

Предложенный выше набор аукционных правил базируется на интуитивном представлении о том, что чем большие ресурсы производителей будут задействованы в схеме, тем больше шансов на то, что план производства удастся сформировать. В самом деле, замена любого решения отдельного участника на доминирующее его лишь расширяет множество вариантов формирования общего плана. Если все участники откажутся участвовать в заказе, то он исполнен не будет, но если снять в модели ресурсные ограничения на производство, то легко указать для участников такие совместные решения, при которых план заведомо будет сформирован. Однако можно построить примеры, когда замена участником доминируемого решения доминирующим так повлияет на выборы остальных участников, что в пер-

вом случае план будет сформирован, а во втором – нет. Можно ожидать, что схема окажется эффективной, когда число участников велико, а их ресурсы значительно превосходят те, что необходимы для исполнения заказа. Но зависимость эффективности схемы от параметров её участников слишком сложна для аналитического исследования. Поэтому предложенные правила следует расценивать как эвристические.

Реализация рассмотренного проекта на практике потребовала бы серьёзных затрат по организации взаимодействия участников ассоциации, обучению персонала технологиям такого взаимодействия, информационной поддержки аукционов. Приступать к этому без предварительного исследования было бы нерациональным и рискованным. Представленная работа находится в русле других работ по системному анализу и исследованию операций, в которых предпринимаются попытки решить принципиальные затруднения, возникающие при формализации целевых установок участников конфликтных процессов. Это затруднения, связанные с большой неопределённостью в намерениях участников и многокритериальностью их устремлений. Один из интересных подходов состоит в том, чтобы привлечь непосредственных участников конфликтной ситуации к принятию решений в режиме компьютерной имитации. В этом случае участники принимают решения самостоятельно, а последствия их выборов моделируются в компьютере. Этот подход сродни имитационным играм, но акцент делается не на обучении участников, а на исследовании самой конфликтной ситуации, возможных исходов от совместных действий участников. В качестве примера можно привести работы [1, 6, 8]. Для аукционной схемы представляется интересным и перспективным её исследование в режиме чистой компьютерной имитации, когда решения за участников принимает также компьютер. Это позволило бы провести большую серию экспериментов при разных параметрах технологического графа и предприятий участников производства.

На компьютере моделируется сеть, узлы которой соответствуют участникам ассоциации. Для каждого узла создаётся своя модель производства, конкретизирующая описание, данное в разделе Участники работы [3]. Выбирается множество изделий

\hat{M} и для них строятся производственные графы. Моделируются аукционы и выборы участников согласно правилам (1)–(3). Там, где эти правила предполагают выбор из множества возможных решений, он может быть рандомизирован.

Для каждого узла задаётся начальное состояние ресурсов, и системе предлагается последовательность заказов на производство изделий из набора \hat{M}^1 . Откликом системы на заказ будет сформированный по аукционной схеме план сборки заказа, если такой удаётся построить. Желательно также реализовать решение общей задачи (1)–(3) [3]. В рамках модели это представляется более реальным, чем на практике: в модели легче варьировать число участников и их структуру. Это позволило бы сравнивать решения исходной задачи с решениями, полученными по аукционной схеме.

Исследование схемы в режиме имитации может представлять и теоретический интерес. Как один из примеров конструкции с сетевым принципом управления. В эксперименте можно было бы наблюдать, какова эффективность этого принципа для подобной конструкции: при каких параметрах схема уверенно доставляет допустимое решение, насколько такое решение далеко от оптимального и т.п.

Литература

1. БЕЛОТЕЛОВ Н.В., БРОДСКИЙ Ю.И., ОЛЕНЕВ Н.Н. и др. *Опыт имитационного моделирования при анализе социально-экономических явлений*. – М.: МЗ Пресс, 2005. – 137 с.
2. ВАТЕЛЬ И.А., ЕРЕШКО Ф.И. *Игры с иерархической структурой. Математическая энциклопедия. Т. 2*. – М.: Сов. энциклопедия, 1979. – С. 478–482.
3. ГАСАНОВ И.И. *Проект системы распределения заказа внутри ассоциации производителей посредством множественных аукционов // Управление большими системами*. – 2024. – Вып. 112. – С. 95–108.
4. ГЕРМЕЙЕР Ю.Б. *Игры с противоположными интересами*. – М.: Наука, 1976. – 328 с.

5. ГЕРМЕЙЕР Ю.Б., ВАТЕЛЬ И.А. *Игры с иерархическим вектором интересов* // Техническая кибернетика. – 1974. – №3. – С. 54–69.
6. ЕРЕШКО Ф.И., БЕЛОТЕЛОВ Н.В., БРОДСКИЙ Ю.И. и др. *Имитационные игры как инструмент исследования геополитических проблем* // MLSD'2023. – М. ИПУ РАН – С. 489–494.
7. КАРЛИН С. *Математические методы в теории игр, программировании и экономике*. – М.: Мир, 1964. – 838 с.
8. ОУЭН Г. *Теория игр*. – М.: Мир, 1971. – 229 с.
9. ПАВЛОВСКИЙ Ю.Н. *Имитационные системы и модели*. – М.: Знание, 1990. – 46 с.
10. ПЕТРОСЯН Л.А., ЗЕНКЕВИЧ Н.А., СЁМИНА Е.А. *Теория игр*. – М.: Высшая школа, 1998. – 300 с.

THE PROJECT OF AN ORDER DISTRIBUTION SYSTEM WITHIN THE ASSOCIATION OF MANUFACTURERS THROUGH MULTIPLE AUCTIONS

Igor Gasanov, FRC CSC RAS, Moscow, SRS
(gasanov48@yandex.ru).

Abstract: The description of the project, presented in the article " The project of the order distribution system within the association of manufacturers through multiple auctions " // Upravlenie bol'shimi sistemami. – 2024. – Iss. 112, is ongoing. It explores the network interaction scheme within an association of companies engaged in the assembly production of a fixed set of products. The production process is represented as a directed acyclic graph, with the vertices corresponding to the components produced by the association's members. To assemble lower-rank parts, higher-rank parts are used. Interaction is based on a system of multiple local auctions through which the distribution of production for each new order is carried out among the association's member companies. If the previous work provides a brief overview of the project and outlines the issues related to its implementation, the current paper offers a detailed analysis of the proposed scheme, suggesting rules that participants should follow when making decisions. The paper examines the problem of rational price selection by the participants for the products they manufacture. In conclusion, the risks associated with the project's implementation are highlighted, and it is suggested that the effectiveness of the scheme be studied through computer simulation. The paper also discusses how an appropriate computer model can be built.

Keywords: auction, networking, production planning.

УДК 519.83
ББК 22.18

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

*Поступила в редакцию 24.01.2025.
Опубликована 31.05.2025.*

ПРОГНОЗИРОВАНИЕ ДИНАМИКИ ОБЩЕСТВЕННОГО МНЕНИЯ ПО ЛОНГИТЮДНЫМ ДАНЫМ ВЫСОКОЙ ГРАНУЛЯРНОСТИ: МОДЕЛЬ АБЕЛЬСОНА, РЕГРЕССИОННЫЕ МОДЕЛИ И АНСАМБЛИ МОДЕЛЕЙ¹

Бузиков М. Э.²

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

Петелина Ю. А.³

(Ozon Tech, Москва)

Красоткин С. А.⁴, **Рыжов М. С.**⁵

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

Козицин И. В.⁶

*(ФГБУН Институт проблем управления
им. В.А. Трапезникова РАН, Московский физико-технический
институт, Москва)*

Рассматривается задача прогнозирования динамики общественного мнения на лонгитюдных данных высокой гранулярности, добытых из социальной сети ВКонтакте. Данная задача была предложена участникам Хакатона «УБС Challenge '2024» как одного из досуговых мероприятий XX Всероссийской школы-конференции молодых ученых «Управление большими системами» (УБС), проводившейся в г. Новочеркасске в 2024 году. Настоящая работа посвящена подробному описанию проведенного Хакатона и решений, предложенных его участниками. Для выборки из $N = 1\,648\,829$ пользователей на основании двух снимков их мнений, сделанных с разницей в полгода (в феврале и июле 2018 года), было необходимо построить прогноз распределения общественного мнения в декабре 2018 года. В распоряжении участников также была информация о структуре дружеских связей пользователей. Наибольшую точность прогноза показал ансамбль из двух моделей – модели Абельсона, усиленной при помощи оценки влиятельности по собственному вектору, и модели постоянного тренда.

¹ Авторы выражают признательность С.А. Красновой и Д.А. Новикову.

² Максим Эмонайевич Бузиков, к.ф.-м.н., с.н.с. (me.buzikov@physics.msu.ru).

³ Юлия Александровна Петелина, руководитель группы. (ptlna@yandex.ru).

⁴ Семён Александрович Красоткин, математик. (semen.krassotkin@gmail.com).

⁵ Максим Сергеевич Рыжов, н.с. (ryzhov@phystech.edu).

⁶ Иван Владимирович Козицин, к.ф.-м.н., с.н.с. (kozisin.ivan@mail.ru).

Ключевые слова: модели социального влияния, искусственный интеллект, социальные сети, центральность по собственному вектору, ансамбли моделей.

1. Введение

Моделирование человеческого поведения является актуальной проблемой современной науки, находящейся на стыке таких дисциплин как теория управления, экономика, сетевой анализ и социальная психология [26]. Понимание особенностей формирования мнений людей в социальных группах необходимо для прогнозирования динамики и разработки эффективных стратегий управления социальными системами. Вместе с тем необходимо отметить, что на данный момент для подавляющего большинства моделей, описывающих данные явления (так называемые модели динамики/формирования мнений или модели влияния) еще даже не проведена валидация на эмпирических данных [9]. Здесь можно отметить модель Фридкина – Джонсона, которая была валидирована в рамках лабораторных экспериментов с малыми (~ 4 человека) группами [12]. При этом эмпирические исследования указывают на непредсказуемость динамики мнений людей, когда сам факт изменения мнения спрогнозировать далеко не всегда представляется возможным [18].

В рамках XX Всероссийской школы-конференции молодых ученых «Управление большими системами» (УБС), проводившейся в г. Новочеркасске в 2024 году, было организовано уже традиционное соревнование по искусственному интеллекту и анализу данных – Хакатон «УБС Challenge'2024». В этот раз участникам было предложено разработать модели для прогнозирования динамики политических взглядов пользователей социальной сети ВКонтакте. Имея в распоряжении два снимка мнений выборки $N = 1\,648\,829$ пользователей, сделанных с разницей в полгода – феврале и июле 2018 года, – а также структуру их дружеских связей, было необходимо спрогнозировать, каким будет распределение мнений этих пользователей еще через полгода – в декабре 2018 года.

В соревновании приняло участие три команды (перечислены в порядке, в котором были присланы решения):

– Максим Бузиков (научный сотрудник лаборатории №38 «Управления по неполным данным») ИПУ РАН и Юлия Петелина (team leader, Ozon Tech);

– Семен Красоткин (математик лаборатории №90 «Цифровых систем управления в агропромышленном комплексе» РАН);

– Максим Рыжов (научный сотрудник лаборатории №38 «Управления по неполным данным» РАН).

Для решения поставленной задачи участники Хакатона применили широкий спектр подходов, включая как классические агент-ориентированные модели влияния (модель Абельсона), так и различные линейные модели, а также ансамбли моделей. Настоящая статья посвящена описанию поставленной в рамках Хакатона задачи, предложенных участниками Хакатона решений, а также сравнительному анализу их точности.

Результаты, представленные в статье, ценны в первую очередь тем, что являются одной из немногих попыток построения и валидации прогноза динамики мнений людей на реальных данных высокого уровня детализации, в том числе при помощи аппарата моделей формирования мнений. С другой стороны, наши результаты наглядно демонстрируют необходимость применения модельного подхода для решения такого рода задач: прогнозы, основанные на классических моделях влияния, усиленных методами сетевого анализа (оценка влиятельности вершин при помощи центральности по собственному вектору) оказались более точными, чем линейные data-driven модели, лишенные априорных знаний об устройстве системы. При этом использование ансамблей из моделей (в частности, модели Абельсона и модели постоянного тренда) позволило превзойти по точности нелинейные модели.

Практическая ценность связана с задачами мониторинга и прогнозирования динамики общественного мнения, планированием информационных, маркетинговых и политических кампаний.

Далее в главе 2 будет приведен краткий обзор литературы по прогнозированию динамики общественного мнения, в главе 3 будет сформулирована задача, предложенная участниками, в главе 4 будут введены метрики качества и критерии точности, в главе 5 будут описаны решения, предложенные участниками. Глава 6 посвящена обсуждению полученных результатов и дальнейших направлений исследований.

2. Обзор литературы

Моделирование процессов социального влияния и изменения мнений в обществе играет важную роль в политологии и социологии, а также в смежных областях – в государственном управлении. Начиная примерно с середины 20-го века (с хрестоматийной работы Френча – ДеГроота [8, 10]), в научной литературе было предложено колоссальное количество математических моделей, количественно описывающих те или иные аспекты социального влияния и информационных взаимодействий в социальных сетях [25]. При помощи этих моделей удалось объяснить различные социальные макро- и мезо-явления, например, консенсус, фрагментацию и поляризацию общественного мнения [9], а также образование информационных пузырей и информационных петель [6].

Необходимо отметить, что при всей проработанности теоретической составляющей данных моделей вопрос их применения для описания реальных социальных процессов и, в частности, прогнозирования динамики мнений людей – одной из ключевых прикладных задач в данной предметной области, – является малоизученным, особенно если говорить про классические, агент-ориентированные модели динамики мнений, опирающиеся на признанные в социальной психологии микро-постулаты социального влияния [21].

На данный момент достаточно успешно решаются задачи мониторинга состояния общественного мнения [16], а также определения наиболее влиятельных людей в социальной сети [5]¹.

¹ Здесь под социальной сетью понимается любая социальная сетевая струк-

С одной стороны, причина такого состояния дел кроется в трудоемкости сбора данных, ведь для построения прогноза и проверки его качества необходимо иметь хотя бы два снимка состояния социальной системы с детализацией вплоть до мнений и социальных контактов отдельных людей, так как применение агент-ориентированных моделей социального влияния требует калибровки параметров отдельных агентов [27]. На данный момент работа с эмпирическими данными, как правило, ограничивается анализом макро-состояний социальных систем (к примеру, распределений общественного мнения), которые можно получать путем проведения обезличенных социальных опросов, их темпоральных паттернов (опять же, обезличенных) и ответом на вопрос «какие микро-механизмы социального влияния могли бы привести к таким распределениям мнений и/или их динамике?» [9].

В противовес классическим агент-ориентированным моделям можно поставить data-driven модели [19, 28], в основе которых лежат технологии искусственного интеллекта. Такие модели агностичны к механизмам формирования мнений, не имеют никакой априорной информации о том, как человек должен менять свои взгляды под действием внешних стимулов. Эти паттерны выявляются моделями на стадии обучения и уже после этого строится прогноз.

Вместе с тем, последние исследования предлагают комбинировать методы и подходы из машинного обучения с концептами из социальной психологии – см. работу [21].

3. Описание данных, обозначения и постановка задачи

В распоряжении участников Хакатона был фрагмент данных, добытых из социальной сети ВКонтакте путем повторяющихся во времени измерений крупномасштабной выборки, включающей

тура, в которой вершинам соответствуют люди, а ребрам – различного рода связи между ними [20].

$N = 1\,648\,829$ российских пользователей, которые были получены из генеральной совокупности путем наложения естественных фильтров с последующим случайным отбором – см. [18] для более подробного ознакомления с технологией отбора. В феврале (T_1), июле (T_2) и декабре (T_3) 2018 года для данной выборки пользователей были выгружены подписки на информационные источники (публичные страницы и блогеров¹), на основании которых при помощи алгоритма искусственного интеллекта, описанного в работе [16], были построены оценки мнений относительно политической повестки. Оценка мнения x_i пользователя i является логистической функцией от свертки информационных источников, на которые он подписан, взятых с весами, идентифицированными на стадии обучения модели. Данная оценка $x_i \in [0, 1]$ показывает, насколько оппозиционен данный пользователь по отношению к действующей в Российской Федерации власти: $x_i = 0$ – полная поддержка оппозиции, $x_i = 0,5$ – нейтральная позиция, $x_i = 1$ – приверженность провластным взглядам.

Снимок мнений пользователей в момент времени $t \in \{T_1, T_2, T_3\}$ будем далее обозначать при помощи вектора

$$\mathbf{x}(t) = (x_1(t), \dots, x_N(t)).$$

Помимо этого, в июле 2018 года были выгружены дружеские связи пользователей (связь типа «дружба» предусмотрена функционалом ВКонтакте и образуется, когда пользователь подписывается на другого пользователя путем нажатия клавиши «добавить в друзья», а второй отвечает на эту заявку). Полученный в итоге граф дружеских связей является связным (была проведена предварительная фильтрация вершин, не входящих в гигантскую компоненту связности – доля таких вершин составила менее одного процента). Данный граф можно формализовать при помощи матрицы смежности

$$\mathbf{A} \in \{0, 1\}^{N \times N},$$

¹ Данные аккаунты являются основными продуцентами информации в ВКонтакте. В частности, аккаунты СМИ по классификации ВКонтакте являются публичными страницами.

компоненты которой $a_{i,j}$ удовлетворяют условию $a_{i,j} = a_{j,i}$ (матрица симметрична) и определяются следующим образом: $a_{i,j} = 1$ тогда и только тогда, когда i и j друзья. Далее множество друзей пользователя i будем обозначать V_i :

$$V_i = \{j \in \{1, \dots, N\} \mid a_{i,j} = 1\}.$$

Число пользователей выборки, имеющих мнение из заданного промежутка $\Delta x \subseteq [0, 1]$ в момент времени t , обозначим $y^{\Delta x}(t) \in [0, 1]$:

$$y^{\Delta x}(t) = \frac{\text{card} \{i \in \{1, \dots, N\} \mid x_i(t) \in \Delta x\}}{N},$$

где $\text{card} \{ \dots \}$ обозначает кардинальное число множества $\{ \dots \}$.

Задача, которая была предложена участникам Хакатона, формулировалась следующим образом: на основании первых двух снимков мнений и снимка дружеских связей необходимо построить оценку распределения мнений пользователей в момент времени T_3 . При этом распределение мнений задавалось путем бинирования пространства мнений $[0, 1]$ на 10 промежутков равной длины $1/10$:

$$y^{[0, \frac{1}{10})}(T_3), \quad y^{[\frac{1}{10}, \frac{2}{10})}(T_3), \quad \dots, \quad y^{[\frac{9}{10}, 1]}(T_3).$$

Фокус данной постановки заключается в прогнозировании динамики общественного мнения при помощи атомарной информации на уровне отдельных пользователей. Было принято решение отказаться от задачи предсказания траекторий мнений отдельных пользователей как малоперспективной и труднорешаемой: несмотря на высокий уровень детализации имеющихся в распоряжении данных, в них не содержится информация, позволяющая сделать вывод об индивидуальных поведенческих особенностях пользователей. В качестве такой информации могли бы послужить данные о том, как много времени пользователь проводит онлайн, с кем из своих друзей он коммуницирует более регулярно, а с кем конфликтует [4]. Влиятельность пользователей может зависеть от их статуса¹, а доверие друг к другу может увеличиваться в случае принадлежности к одной социальной группе,

¹ В качестве простейшей прокси статуса может выступать центральность вершины в сети, однако такой (структурный) подход игнорирует другие сферы социальных коммуникаций.

что также не отражено в данных [11]. Полезной могла бы быть информация касательно демографических характеристик пользователей [13, 23, 24]. С другой стороны, эмпирические исследования свидетельствуют о том, что даже сам факт изменения мнения отдельного индивида (не говоря про его количественные аспекты) является достаточно сложным для прогнозирования [18, 22].

4. Метрика качества прогноза и ориентира

Для оценки точности прогноза использовалась суммарная абсолютная ошибка AE . Пусть вектор

$$\mathbf{y}(T_3) = \left(y^{[0, \frac{1}{10}]}(T_3), y^{[\frac{1}{10}, \frac{2}{10}]}(T_3), \dots, y^{[\frac{9}{10}, 1]}(T_3) \right)$$

задает реальное состояние общественного мнения в момент времени T_3 (известное организаторам, но неизвестное участникам), а вектор

$$\mathbf{y}_{pred} = (y_1, \dots, y_{10})$$

– его прогноз, полученный в рамках некоторой прогностической модели. Тогда суммарная абсолютная ошибка AE прогноза \mathbf{y}^{pred} определяется согласно следующему выражению:

$$AE(\mathbf{y}_{pred}) = \|\mathbf{y}(T_3) - \mathbf{y}_{pred}\|_1 = |y^{[0, \frac{1}{10}]}(T_3) - y_1| + \dots + |y^{[\frac{9}{10}, 1]}(T_3) - y_{10}|.$$

В качестве ориентиров рассматривались два базовых прогноза. Первый – это прогноз постоянного тренда, определяемый линейным соотношением

$$\mathbf{y}_{const} = \mathbf{y}(T_2) + \mathbf{y}(T_2) - \mathbf{y}(T_1).$$

Его точность составила

$$AE(\mathbf{y}_{const}) = 40 \times 10^{-4}.$$

Второй базовый прогноз основан на применении агент-ориентированной модели динамики мнений (SCARDO-модели [17]), и его точность составила

$$AE(\mathbf{y}_{SCARDO}) = 37 \times 10^{-4},$$

более подробно – см. работу [1]. Отметим, что данная модель является нелинейной.

Оба эти ориентира были озвучены участникам на установочной лекции, как и их точность.

5. Решения

Далее в данной главе будут приведены решения, предложенные участниками (в порядке, в котором они были получены).

5.1. Решение 1 (Максим Бузиков и Юлия Петелина)

Предположим, что динамика изменения мнений описывается с помощью

$$(1) \quad \dot{x} = \mathbf{W}x,$$

где $\mathbf{W} \in \mathbb{R}^{N \times N}$. Это так называемая модель Абельсона [2] (аналог в непрерывном времени более известной модели ДеГроота [8]). Матрица \mathbf{W} в модели Абельсона отражает структуру влияний в обществе: с какой силой агенты влияют друг на друга. Данную матрицу можно идентифицировать таким образом, чтобы итоговое решение

$$x(t) = e^{\mathbf{W}(t-t_1)}x(t_1)$$

задачи Коши (1) с начальным условием $x(T_1)$ было как можно ближе к $x(T_2)$ в момент $t = T_2$. Таким образом, элементы матрицы \mathbf{W} являются оптимизируемыми величинами. Без дополнительных ограничений оптимизация в пространстве $\mathbb{R}^{N \times N}$ для больших N затруднительна с вычислительной точки зрения, поэтому предпочтительно свести количество неизвестных параметров хотя бы к величине, пропорциональной N .

Сделаем дополнительные предположения относительно вида матрицы \mathbf{W} . Будем характеризовать важность каждого пользователя с помощью центральности по собственному вектору [3]:

$$c = \frac{\mathbf{A}\mathbf{c}}{\lambda_{\max}(\mathbf{A})}.$$

Здесь $\lambda_{\max}(\mathbf{A})$ соответствует максимальному по модулю собственному значению матрицы \mathbf{A} . Так как рассматриваемый граф является ненаправленным и связным, то его матрица смежности неразложима, а значит, в силу теоремы Фробениуса – Перрона собственное число $\lambda_{\max}(\mathbf{A})$ положительно. Соответствующий собственный вектор $\mathbf{c} \in \mathbb{R}^N$ также можно выбрать так, что все его компоненты будут положительны. Именно он определяет

важность вершин в рамках центральности по собственному вектору. Также нам понадобится вектор взвешенных мнений друзей пользователя, где в качестве весов выступают компоненты вектора центральностей \mathbf{c} :

$$\mathbf{w}(\mathbf{x}(t)) = \frac{\mathbf{A}(\mathbf{c} \odot \mathbf{x}(t))}{\mathbf{A}\mathbf{c}} \in \mathbb{R}^N.$$

Здесь \odot означает произведение Адамара (покомпонентное умножение), и деление на вектор понимается также в покомпонентном смысле.

Будем считать, что скорость изменения мнения пользователя пропорциональна разнице мнения самого пользователя и средневзвешенного мнения его друзей:

$$\dot{\mathbf{x}} = \boldsymbol{\beta} \odot (\mathbf{w}(\mathbf{x}) - \mathbf{x}).$$

Здесь $\boldsymbol{\beta} \in \mathbb{R}^N$ характеризует направленность и скорость изменения мнения под влиянием друзей. Используя $\mathbf{A}(\mathbf{c} \odot \mathbf{x}(t)) = (\mathbf{A} \odot \mathbf{c}^\top)\mathbf{x}(t)$, получим:

$$(2) \quad \mathbf{W}(\boldsymbol{\beta}) = \boldsymbol{\beta} \odot \left(\frac{\mathbf{A} \odot \mathbf{c}^\top}{\mathbf{A}\mathbf{c}} - \mathbf{I} \right).$$

Таким образом, в выбранной модели есть всего N неизвестных параметров $\boldsymbol{\beta} \in \mathbb{R}^N$. Эти параметры можно выбрать, решая следующую оптимизационную задачу:

$$f(\boldsymbol{\beta}) = \|\mathbf{x}(T_2) - e^{\mathbf{W}(\boldsymbol{\beta})(T_2-T_1)}\mathbf{x}(T_1)\|^2 \rightarrow \min_{\boldsymbol{\beta}}.$$

Суть данной оптимизации заключается в подборе таких параметров $\boldsymbol{\beta}$, чтобы решение уравнения (1) с $\mathbf{W} = \mathbf{W}(\boldsymbol{\beta})$ проходило в среднем как можно точнее через значение $\mathbf{x}(T_2)$.

Отметим, что градиент функции f имеет следующий вид:

$$\begin{aligned} \nabla f(\boldsymbol{\beta}) = & -2(T_2 - T_1) (\mathbf{w}(\mathbf{x}(T_1)) - \mathbf{x}(T_1)) \\ & \odot \left(e^{\mathbf{W}(\boldsymbol{\beta})(T_2-T_1)}(\mathbf{x}(T_2) - e^{\mathbf{W}(\boldsymbol{\beta})(T_2-T_1)}\mathbf{x}(T_1)) \right). \end{aligned}$$

С вычислительной точки зрения $e^{\mathbf{W}(\boldsymbol{\beta})(T_2-T_1)}\mathbf{x}(t_1)$ легче всего получить, интегрируя систему (1) от T_1 до T_2 с начальным условием $\mathbf{x}(T_1)$ методом Рунге – Кутты. Для вычисления $\nabla f(\boldsymbol{\beta})$ понадобится повторное интегрирование системы (1) с начальным условием $\mathbf{x}(T_2) - e^{\mathbf{W}(\boldsymbol{\beta})(T_2-T_1)}\mathbf{x}(T_1)$.

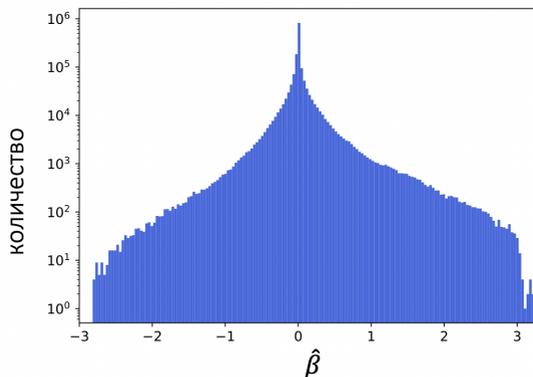


Рис. 1. Гистограмма распределения компонент вектора $\hat{\beta}$.
Вертикальная ось имеет логарифмический масштаб

Оптимизацию будем производить методом градиентного спуска:

$$\beta_{n+1} = \beta_n - \lambda_n \nabla f(\beta_n),$$

где β_0 , λ_0 начальные параметры. В ходе численной оптимизации было выбрано $\beta_0 = \mathbf{0}$, а значения шага λ_n перестраивались вручную, если наблюдалось резкое повышение значения $f(\beta_{n+1})$. Также, если шаг метода становился небольшим, то большие по модулю компоненты вектора β_n обнулялись для выхода из локальных минимумов. С помощью данной ручной настройки удалось подобрать такое $\hat{\beta}$ (см. рис. 1), что $f(\hat{\beta}) \approx 609$. Для сравнения:

$$f(\mathbf{0}) = \|\mathbf{x}(T_2) - \mathbf{x}(T_1)\|^2 \approx 2761.$$

Финальный прогноз мнений каждого из пользователей строился следующим образом:

$$\hat{\mathbf{x}}(T_3) = e^{\mathbf{W}(\hat{\beta})(T_3 - T_2)} \mathbf{x}(T_2).$$

После этого был найден вектор \mathbf{y}_{BP} , для которого величина AE составила

$$AE(\mathbf{y}_{BP}) \approx 142 \times 10^{-4},$$

что превышает величину ошибки прогноза модели постоянного тренда \mathbf{y}_{const} .

Заметим, что прогноз, в котором идентификация модели Абельсона по первым двум снимкам мнений производится без использования центральности по собственному вектору (иными словами, когда вектор \mathbf{c} в формуле (2) является единичным), оказался менее точен, чем с использованием центральности по собственному вектору (200×10^{-4} против 142×10^{-4}). Также необходимо отметить, что без использования центральности по собственному вектору модель Абельсона «хуже» объясняет данные при решении обратной задачи по первым двум снимкам мнений – значение оптимизируемой функции f составило примерно 750, а с использованием центральности по собственному вектору, как было отмечено выше, – 609.

Тот факт, что описанная выше модель и модель постоянного тренда построены «независимо», можно использовать для улучшения прогноза, формируя на их основании ансамбль моделей. Для этого нужно усреднить предсказания по моделям, используя веса $\alpha_{BP} = 1/AE(\mathbf{y}_{BP})$ (для описанной в данном разделе модели¹), $\alpha_{const} = 1/AE(\mathbf{y}_{const})$ (для модели постоянного тренда²):

$$\hat{y}_{BP}^{\Delta x} = \left(\alpha_{const} \cdot y_{const}^{\Delta x} + \alpha_{BP} \cdot \frac{y_{BP}^{\Delta x}}{\sum_k y_{BP}^{\Delta x_k}} \right) / (\alpha_{const} + \alpha_{BP}).$$

Здесь также использована перенормировка $y_{BP}^{\Delta x}$ на $\sum_k y_{BP}^{\Delta x_k}$, так как некоторые компоненты вектора $\hat{\mathbf{x}}(t_3)$ могли выйти за пределы интервала $[0, 1]$. Расчеты показывают, что

$$AE(\hat{\mathbf{y}}_{BP}) \approx 27 \times 10^{-4},$$

т.е. прогноз существенно улучшился и в сравнении с моделью постоянного тренда, и в сравнении со SCARDO-моделью.

5.2. Решение 2 (Семен Красоткин)

Для прогнозирования распределения мнений пользователей в декабре 2018 года (вектора $\mathbf{y}(T_3)$) была использована модель

¹ По правилам Хакатона, участники могли узнать точность прогноза только в момент оглашения итоговых результатов. В связи с этим для построения оценки точности данного прогноза в качестве правильного ответа был взят прогноз модели постоянного тренда.

² Точность данного прогноза была оглашена заранее на установочной лекции.

линейной регрессии. В основе такого выбора лежало предположение о том, что краткосрочные прогнозы можно приближать линейными соотношениями. Кроме того, выбор данной модели обусловлен простотой реализации и малым использованием вычислительных ресурсов. Также предполагалось, что линейная регрессия послужит ориентиром для других решений.

В качестве регрессоров выступали мнение пользователя и мнения его друзей:

$$(3) \quad x_i(T_{k+1}) = \sum_{j \in V_i} \frac{\beta_1}{\text{card } V_i} x_j(T_k) + \beta_2 x_i(T_k) + \beta_3,$$

а коэффициенты регрессии β_1, β_2 и β_3 определялись методом наименьших квадратов. При этом нормировочный множитель $1/\text{card } V_i$ в первом слагаемом правой части позволял учесть тот факт, что пользователи могут иметь разное число друзей: распределение степеней вершин в рассматриваемой социальной сети, как и для большинства такого рода объектов, является степенным [18]. Фактически модель (3) является расширением модели ДеГроота [8] путем включения свободного слагаемого и снятия ограничений на коэффициенты, но с дополнительным ограничением о гомогенности весов влияний.

Модель (3) обучалась на всей совокупности данных и после этого использовалась для прогноза $x_i(T_3)$:

$$x_i(T_3) = \sum_{j \in V_i} \frac{\beta_1}{\text{card } V_i} x_j(T_2) + \beta_2 x_i(T_2) + \beta_3$$

после чего производилось агрегирование по десяти бинам.

Для построенного прогноза $\mathbf{y}_K(T_3)$ точность составила

$$AE(\mathbf{y}_K(T_3)) \approx 171 \times 10^{-4}.$$

5.3. Решение 3 (Максим Рыжов)

В основу предлагаемого решения легла классическая линейная модель ДеГроота [8]. Предположим, что мнение $x_i(t)$ узла i зависит линейно от мнения связанных с ним узлов и его собственного мнения на момент времени $t - 1$:

$$(4) \quad x_i(t) = \sum_{j \in V_i} \beta_{i,j} x_j(t - 1) + \beta_{i,i} x_i(t - 1).$$

В выражении (4) на коэффициенты $\beta_{i,j}$ не накладываются никакие ограничения (в отличие от модели ДеГроота, в которой $\beta_{i,j} \geq 0$ и $\sum_{j \in V_i} \beta_{i,j} = 1$ для любого $i \in \{1, \dots, N\}$).

При помощи первых двух снимков мнений и матрицы смежности для каждого пользователя на основе метода наименьших квадратов были найдены коэффициенты $\beta_{i,j}$. После этого при помощи идентифицированной модели (4) были рассчитаны значения $x_i(T_3)$:

$$x_i(T_3) = \sum_{j \in V_i} \beta_{i,j} x_j(T_2) + \beta_{i,i} x_i(T_2).$$

Далее было произведено агрегирование согласно обозначенным в постановке задачи бинам.

Для построенного прогноза $\mathbf{y}_R(T_3)$ точность составила $AE(\mathbf{y}_R(T_3)) \approx 233 \times 10^{-4}$.

6. Обсуждение результатов участников и выводы

В таблице 1 приведены результаты команд. Все три решения были основаны на построении прогноза изменения мнения каждого пользователя с последующим агрегированием (англ. bottom-up design), а не на использовании каких-либо макро-моделей в духе SIR-модели [15]. При этом все три решения, предложенные участниками, были основаны на линейных моделях. Решение 1 – для случая непрерывного времени, Решения 2 и 3 – для дискретного.

В рамках Решения 1 каждый пользователь-наблюдение характеризовался одним неизвестным параметром, погруженным в контекст модели Абельсона. Гетерогенность влиятельности агентов была достигнута путем вспомогательного вычисления структурной влиятельности, определяемой центральностью по собственному вектору. В Решении 2 предполагалось, что влияние со стороны окружения гомогенно, но отличается от влияния со стороны собственного мнения. Также в этом решении предполагалось наличие свободного слагаемого, что потенциально повышало предиктивную силу модели. В рамках Решения 3 количество параметров модели для каждого пользователя определялось

числом его друзей, что делало модель склонной к переобучению. По-видимому, именно это и случилось в итоге.

Напротив, прогностическая модель с наименьшим числом параметров, учитывающая структурную влияниеемость вершин (Решение 1), показала наилучший результат, если рассматривать одиночные модели (142×10^{-4}). Однако этот прогноз был менее точен, чем полученный при помощи SCARDO-модели в работе [1]. В дальнейшем использование базовой модели из Решения 1 в ансамбле с моделью постоянного тренда позволило превзойти результат SCARDO-модели (27×10^{-4} против 42×10^{-4} для SCARDO-модели). Для этого прогнозы базовой модели из Решения 1 и модели постоянного тренда были взвешены, а веса определялись точностью их прогнозов по отдельности. В то время как точность прогноза модели постоянного тренда была известна участникам Хакатона заранее, точность базовой модели из Решения 1 оставалась неизвестной. Для решения данной проблемы была применена следующая эвристика: оценка точности модели путем сравнения ее не с эталонным значением, а с искаженным, а именно – с прогнозом модели постоянного тренда, который было легко воспроизвести.

Таблица 1. Точность прогнозов моделей

Команда	Показатель (модель)
Максим Бузиков & Юлия Петелина	200×10^{-4} (модель Абельсона) 142×10^{-4} (модель Абельсона + центр. по собств. вekt.) 27×10^{-4} (ансамбль моделей)
Семен Красоткин	171×10^{-4} (линейная модель с гомогенными по популяции коэффициентами и свободным членом)
Максим Рыжов	233×10^{-4} (линейная модель с гетерогенными по популяции коэффициентами и без свободного члена)

Какие выводы можно сделать по итогам Хакатона? Во-первых, «модельный» подход оказался наиболее эффективным: в то время как Решения 2 и 3 были основаны на простейших спецификациях линейной модели без каких-либо априорных предположений о механизмах влияния (в том смысле, что на коэффициенты моделей не накладывались никакие содержательные

ограничения), в основе Решения 1 лежала модель Абельсона [2] – аналог модели ДеГроота [8] в непрерывном времени, усиленная при помощи одного из классических приемов сетевого анализа – оценки влияния вершин сети при помощи Фробениусова вектора. При этом использование центральности по собственному вектору действительно способствовало росту предиктивной силы модели, а также ее способности описывать имеющиеся в распоряжении данные. Без этого модель Абельсона (с гомогенными весами влияний) с точки зрения точности может уступать линейным моделям, лишенным априорной информации о процессе. Открытым остается вопрос, сыграл ли свою роль тот факт, что модель, предложенная в рамках Решения 1 была основана на непрерывном времени, в отличие от Решений 2 и 3, оперирующих дискретным временем.

Второй вывод заключается в том, что применение техник машинного обучения, в частности ансамбля моделей, а также дополнительных эвристик позволяет существенно повысить точность прогноза. Скомбинировав две независимые модели – модель Абельсона и модель постоянного тренда – Максиму Бузикову и Юлии Петелиной удалось значительно улучшить свой прогноз, обойдя оба ориентира (со 142×10^{-4} до 27×10^{-4}). Необходимо учитывать, что второй из указанных ориентиров был получен при помощи SCARDO-модели, которая является нелинейной. Иными словами, ансамбль из двух линейных моделей оказался более успешен с прогностической точки зрения, чем нелинейная модель.

В дальнейшем интересным было бы рассмотреть альтернативные, нелинейные модели социального влияния, основанные, к примеру, на механизме ограниченного доверия [7]. Также перспективным видится тестирование иных метрик центральности [5]. Как было отмечено в работе [1], рассматриваемый набор данных, несмотря на свои внушительные размеры, характеризуется относительно невысоким уровнем изменений состояния общественного мнения. В связи с этим интерес представляет изучение аналогичных постановок на лонгитюдных данных высокой гра-

нулярности, собранных в рамках более длительных временных промежутков, состоящих из более чем трех волн. В этом случае в прогностические модели можно попытаться интегрировать эффект памяти [14] и оценить его влияние на точность прогноза.

Литература

1. КОЗИЦИН И.В. *Построение прогноза динамики общественного мнения при помощи SCARDO-модели* // Управление большими системами. – 2024. – Вып. 108. – С. 124–136. – DOI: <https://doi.org/10.25728/ubs.2024.108.7>.
2. ABELSON R.P. *Mathematical models of the distribution of attitudes under controversy* // Contributions to Mathematical Psychology. – 1964.
3. BONACICH P. *Some unique properties of eigenvector centrality* // Social Networks. – 2007. – Vol. 29, No. 4. – P. 555–564. – DOI: <https://doi.org/10.1016/j.socnet.2007.04.002>.
4. BOND R.M. et al. *A 61-million-person experiment in social influence and political mobilization* // Nature. – 2012. – Vol. 489, No. 7415. – P. 295–298. – DOI: <https://doi.org/10.1038/nature11421>.
5. CHEBOTAREV P., GUBANOV D.A. *How to Choose the Most Appropriate Centrality Measure? A Decision-Tree Approach* // IEEE Trans. on Systems, Man, and Cybernetics: Systems. – 2024. – DOI: <https://doi.org/10.1109/TSMC.2024.3510633>.
6. CINUS F. et al. *The effect of people recommenders on echo chambers and polarization* // Proc. of the Int. AAAI Conf. on Web and Social Media. – 2022. – Vol. 16. – P. 90–101. – DOI: <https://doi.org/10.1609/icwsm.v16i1.19275>.
7. DEFFUANT G. et al. *Mixing beliefs among interacting agents* // Advances in Complex Systems. – 2000. – Vol. 3, No. 01n04. – P. 87–98. – DOI: <https://doi.org/10.1142/S0219525900000078>.
8. DEGROOT M.H. *Reaching a consensus* // Journal of the American Statistical Association. – 1974. – Vol. 69, No. 345. – P. 118–121. – DOI: <https://doi.org/10.1080/01621459.1974.10480137>.

9. FLACHE A. et al. *Models of social influence: Towards the next frontiers* // JASSS – The Journal of Artificial Societies and Social Simulation. – 2017. – Vol. 20, No. 4. – P. 2. – DOI: <https://doi.org/10.18564/jasss.3521>.
10. FRENCH JR J.R.P. *A formal theory of social power* // Psychological Review. – 1956. – Vol. 63, No. 3. – P. 181.
11. FREY V. et al. *Who influences lower-status individuals more: People of higher-status outgroups or people of their lower-status ingroup? Examining the difference between matters of opinion and matters of fact* // Social Science Research. – 2024. – Vol. 123. – P. 103060. – DOI: <https://doi.org/10.1016/j.ssresearch.2024.103060>.
12. FRIEDKIN N.E., JIA P., BULLO F. *A theory of the evolution of social power: Natural trajectories of interpersonal influence systems along issue sequences* // Sociological Science. – 2016. – Vol. 3. – P. 444–472. – DOI: <https://doi.org/10.15195/v3.a20>.
13. 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. – DOI: <https://doi.org/10.3390/computers12060116>.
14. JEDRZEJEWSKI A., SZNAJD-WERON K. *Impact of memory on opinion dynamics* // Physica A: Statistical Mechanics and its Applications. – 2018. – Vol. 505. – P. 306–315. – DOI: <https://doi.org/10.1016/j.physa.2018.03.077>.
15. JI C., JIANG D. *Threshold behaviour of a stochastic SIR model* // Applied Mathematical Modelling. – 2014. – Vol. 38, No. 21–22. – P. 5067–5079. – DOI: <https://doi.org/10.1016/j.apm.2014.03.037>.
16. KOZITSIN I.V. et al. *Modeling political preferences of russian users exemplified by the social network Vkontakte* // Mathematical Models and Computer Simulations. – 2020. – Vol. 12. – P. 185–194. – DOI: <https://doi.org/10.1134/S2070048220020088>.
17. KOZITSIN I.V. *A general framework to link theory and*

- empirics in opinion formation models* // Scientific reports. – 2022. – Vol. 12, No. 1. – P. 5543. – DOI: <https://doi.org/10.1038/s41598-022-09468-3>.
18. 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. – DOI: <https://doi.org/10.1080/0022250X.2021.1956917>.
 19. KULKARNI B. et al. *SLANT+: A nonlinear model for opinion dynamics in social networks* // IEEE Int. Conf. on Data Mining (ICDM). – IEEE, 2017. – P. 931–936. – DOI: <https://doi.org/10.1109/ICDM.2017.117>.
 20. NEWMAN M. *Networks*. – Oxford University Press, 2018.
 21. OKAWA M., IWATA T. *Predicting opinion dynamics via sociologically-informed neural networks* // Proc. of the 28th ACM SIGKDD Conf. on Knowledge Discovery and Data Mining. – 2022. – P. 1306–1316. – DOI: <https://doi.org/10.1145/3534678.35392>.
 22. PANSANELLA V. et al. *Change my mind: Data driven estimate of open-mindedness from political discussions* // Int. Conf. on Complex Networks and Their Applications. – Cham: Springer Int. Publishing, 2022. – P. 86–97. – DOI: https://doi.org/10.1007/978-3-031-21127-0_8.
 23. PESHKOVSKAYA A., BABKINA T., MYAGKOV M. *Social context reveals gender differences in cooperative behavior* // Journal of Bioeconomics. – 2018. – Vol. 20. – P. 213–225. – DOI: <https://doi.org/10.1007/s10818-018-9271-5>.
 24. PESHKOVSKAYA A., BABKINA T., MYAGKOV M. *Gender effects and cooperation in collective action: A laboratory experiment* // Rationality and Society. – 2019. – Vol. 31, No. 3. – P. 337–353. – DOI: <https://doi.org/10.1177/1043463119858788>.
 25. PROSKURNIKOV A.V., TEMPO R. *A tutorial on modeling and analysis of dynamic social networks. Part I* // Annual Reviews in Control. – 2017. – Vol. 43. – P. 65–79. – DOI: <https://doi.org/10.1016/j.arcontrol.2017.03.002>.
 26. 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. – DOI: <https://doi.org/10.1016/j.arcontrol.2018.03.005>.*
27. RAVAZZI C. et al. *Learning hidden influences in large-scale dynamical social networks: A data-driven sparsity-based approach, in memory of Roberto Tempo // IEEE Control Systems Magazine. – 2021. – Vol. 41, No. 5. – P. 61–103. – DOI: <https://doi.org/10.1109/MCS.2021.3092810>.*
28. ZHU L., HE Y., ZHOU D. *Neural opinion dynamics model for the prediction of user-level stance dynamics // Information Processing & Management. – 2020. – Vol. 57, No. 2. – P. 102031. – DOI: <https://doi.org/10.1016/j.ipm.2019.03.010>.*

FORECASTING THE DYNAMICS OF PUBLIC OPINION BASED ON LONGITUDINAL DATA OF HIGH GRANULARITY: THE ABELSON MODEL, REGRESSION MODELS, AND ENSEMBLES OF MODELS

Maksim Buzikov, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, Cand.Sc., Senior Researcher (me.buzikov@physics.msu.ru).

Iuliia Petelina, Ozon Tech, Moscow, Team Leader (ptlna@yandex.ru).

Semen Krassotkin, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, Mathematician (semen.krassotkin@gmail.com).

Maksim Ryzhov, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, researche (ryzhov@phystech.edu).

Ivan Kozitsin, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, PhD, senior researcher, Moscow Institute of Physics and Technology, Moscow, assistant professor (kozitsin.ivan@mail.ru).

Abstract: We consider the problem of forecasting the dynamics of public opinion based on longitudinal data of high granularity gleaned from the social network VKontakte. This problem was suggested to the participants of the Hackathon «UBS Challenge'2024» as one of the leisure events of the XX All-Russian School-Conference of Young Scientists «Management of Large Systems» (UBS), held in Novocherkassk in 2024. This paper is devoted to a detailed description of the Hackathon and the solutions proposed by its participants. For a sample of $N = 1\,648\,829$ users, based on two granular snapshots of their opinions taken six months apart (in February and July 2018), participants have to elaborate on a forecast of the distribution of public opinion in December 2018. The participants also had the information about the structure of friendship ties between users. We report that the highest accuracy was achieved by an ensemble of two models – the Abelson model, enhanced by estimating users' social power via the eigenvector centrality measure, and the constant trend model.

Keywords: social influence models, artificial intelligence, social networks, eigenvector centrality, ensembles of models.

УДК 519.7

ББК 22.18

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

Поступила в редакцию 28.01.2025.

Дата опубликования 31.05.2025.

ИСПОЛЬЗОВАНИЕ ИЕРАРХИЧЕСКОЙ МНОГОФАКТОРНОЙ РЕГРЕССИОННОЙ МОДЕЛИ С ЦЕЛЬЮ АНАЛИЗА И ПРОГНОЗИРОВАНИЯ СЕЛЬСКОХОЗЯЙСТВЕННЫХ ПОКАЗАТЕЛЕЙ

Овчинников П. В.¹, Ткачев А. Н.², Мирославская М. Д.³
(ФГБОУ ВО «ЮРГПУ (НПИ) имени М.И. Платова»,
Новочеркасск)

Представлена методика определения латентных параметров многоуровневой многофакторной регрессионной модели и способы использования модели на примере прогнозирования динамики показателей, отражающих функционирование сельскохозяйственного производства: урожайности зерновых культур и потребления кормовой пшеницы. Актуальность модели обусловлена возможностью применения в целях прогнозирования главного выходного показателя (первый уровень модели) и его промежуточных составляющих (последующие уровни модели). Выполнена обработка статистических данных, применен метод регрессионного анализа информации, выполнено построение графиков результатов моделирования с использованием MS Excel и среды разработки языка программирования Python. Модель строится на основании иерархической зависимости первого и последующих уровней, при которой входные данные выходного показателя группируются как индивидуальные или общие параметры линейных зависимостей промежуточных переменных. В результате апробации модели были получены данные урожайности зерновых культур в зависимости от типа подкормки удобрениями и соответствующие им доли площадей по данным Российской Федерации в целом и Ростовской области в частности; произведен прогноз урожайности на три года, приведено сравнение с результатами применения факторного анализа. В вопросе определения потребления кормовой пшеницы были получены данные объема чистого потребления пшеницы и объема пшеницы, перерабатываемой на комбикорм.

Ключевые слова: регрессионная модель, иерархическая система, урожайность, потребление, зерновые культуры.

1. Введение

Проблема прогнозирования динамики показателей производства и потребления зерновых культур приобретает особую

¹ Петр Вячеславович Овчинников, к.э.н., доцент (pvo78@yandex.ru).

² Александр Николаевич Ткачев, д.т.н., профессор (tkachev.an@mail.ru).

³ Мария Дмитриевна Мирославская, аспирант (miroslavsky.marymir@gmail.com).

актуальность в связи с тем, что именно зерновые стали основой отечественного агропромышленного комплекса. При этом представляет интерес получение прогнозных характеристик как с точки зрения использования зерна (в частности, оценка ожидаемых объемов потребления по видам продукции), так и с точки зрения определения ожидаемого уровня валового сбора, урожайности и т.п. В настоящее время методология и инструментарий для решения задач прогнозирования достаточно хорошо разработаны и разнообразны. Наиболее широко применяются различные регрессионные модели, в практической плоскости применение нашли, прежде всего, одноуровневые модели линейной регрессии. В таких моделях предполагается, что выходная переменная является линейной функцией базисных факторов [23, 28–29]. Линейные одноуровневые модели используются и при оценке функциональных зависимостей и прогнозировании сельскохозяйственных показателей, что представлено, в том числе, в работах Зинченко В.Е., Шубнова М.Г., Бисчокова Р.М. и др. [2, 3, 10, 16, 18, 20, 24]. Значительная часть данных моделей являются локальными и позволяют проводить анализ в рамках определенной предметной области. Представляет интерес задача построения регрессионных моделей для иных частных случаев и, в перспективе, – построения общей модели оценки и прогнозирования показателей сельскохозяйственной деятельности.

Предлагаемая в настоящем исследовании модель разрабатывалась на основе принципов метода моделирования структурных уравнений (SEM) и множественной регрессии (MR) [12, 17, 19, 21, 26–27].

Метод моделирования структурных уравнений является элементом эконометрики, используемым на данный момент для исследования социальных и поведенческих структур [25]. В то же время растет популярность применения модели множественной регрессии, которая используется не только в технических, но и в экономической и сельскохозяйственных исследованиях [1, 6]. Недостатком такой модели является слабый прогнозный потенциал на долгосрочный период, так как колебания любого

входного показателя за пределы существующих в тестовом наборе приводит к выбросам и низкой точности прогноза.

Первым этапом моделирования SEM является выделение латентных (скрытых) переменных, которые не могут быть измерены напрямую или представлены в виде откликов независимых переменных с весовыми коэффициентами. Для определения таких факторов наиболее распространенным является факторный анализ, позволяющий применить целый ряд методов: метод главных компонент, метод наименьших квадратов, метод максимального правдоподобия. Для формирования модели использовалась идея разделения переменных на группы общих и индивидуальных факторов с использованием метода наименьших квадратов.

Использование множественной регрессии обусловлено значимым количеством факторов, которые влияют на выходной параметр модели. При этом выявление скрытых факторов позволяет не только снизить размерность набора входных данных на каждом этапе построения модели, но и определить значения этих факторов как самостоятельных переменных, несущих логический информационный смысл при проведении анализа и построения прогностических моделей.

Предполагается, что предлагаемый тип модели может применяться в случае, когда общая результирующая переменная может быть представлена в виде функциональной зависимости от базисных факторов, каждый из которых, в свою очередь, получает представление как функция от факторов более низкого уровня. В результате модель имеет вид иерархической структуры с множеством входных факторов, см. рис. 1.

Количество уровней зависит от моделируемой системы, при этом высшие уровни подчиняются правилу

$$(1) \quad y = f(y_1, y_2, \dots, y_m) = \sum_{k=1}^m \lambda_k y_k .$$

Возможны следующие ситуации:

1. В случае исследования относительных величин (например урожайности) выполняется условие

$$(2) \sum_{k=1}^m \lambda_k = 1; \quad 0 \leq \lambda_k \leq 1.$$

2. В случае исследования абсолютных показателей (например потребления кормовой пшеницы)

$$(3) \sum_{k=1}^m \lambda_k > 1.$$

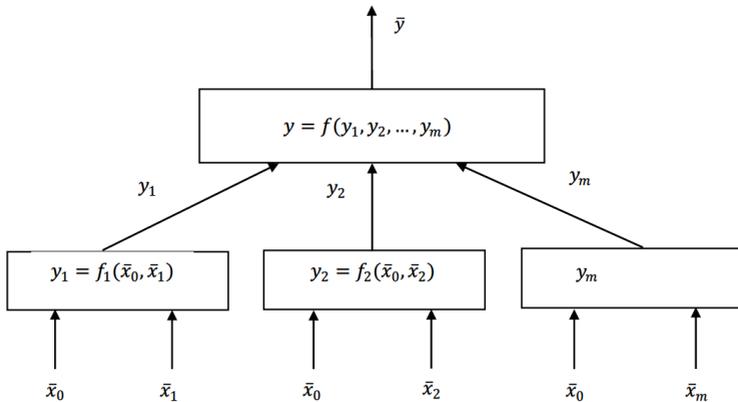


Рис. 1. Структура многоуровневой модели [10]

Математический аппарат, обоснование, валидация представленной модели и идентификация ее коэффициентов была проведена авторами в более ранней публикации [8].

2. Определение параметров модели на различных уровнях

Первый уровень модели представляет из себя множественную регрессию вида:

$$(4) \quad y = a_0 + \sum_{i=1}^n a_i x_i + \varepsilon,$$

где: x_i – входные переменные; a_0, a_i – коэффициенты, подлежащие определению; ε – ошибка оценки регрессии [8].

Промежуточные переменные $y_k, k = 1, 2, \dots, m$, определяются в виде линейной регрессии вида

$$(5) \quad y_k = f_k(\overline{x_o}, \overline{x_k}) = \sum_{i=1}^{n_0} a_{ki} x_{oi} + \sum_{j=1}^{n_k} a_{kj} x_{kj} + \varepsilon_k,$$

где a_{ki} , a_{kj} – подлежащие определению коэффициенты; ε_k – ошибка оценки регрессии [8].

Общая и индивидуальные части промежуточных регрессионных моделей (моделей 2 уровня) определяются по формулам

$$(6) \quad \begin{cases} u_k = \sum a_i x_{ki}, \\ u_o = \sum a_i x_{oi} + a_o; \end{cases}$$

$$(7) \quad \begin{cases} u_k = \sum a_i x_{ki} + \lambda_k a_k, \\ u_o = \sum a_i x_{oi} + \lambda_o a_o; \end{cases}$$

где a_i , a_o – коэффициенты регрессии (4).

Допущение о вариативности определения общей и индивидуальных частей обусловлено зависимостями промежуточных параметров y_k , где возможны явления:

$$(8) \quad y = \sum_{k=1}^m \lambda_k y_k,$$

$$(9) \quad y = \sum_{k=1}^{m-1} \lambda_k y_k + \lambda_o y_o,$$

где y_o – промежуточный параметр, не имеющей индивидуальной части.

Коэффициенты λ_k также определяются в зависимости от видов промежуточных переменных, для (8) и (9) соответственно:

$$(10) \quad \lambda_k = \frac{\sigma_k^2}{\sigma_y^2 + \sum \sigma_k^2},$$

$$(11) \quad \lambda_k = \frac{\sigma_y^2}{\sigma_k^2},$$

где σ_k^2 , σ_y^2 – показатель дисперсии для регрессий y_k , y .

Для каждой модели 2 уровня определяются коэффициенты вхождения в 1 уровень системы:

$$(12) \alpha_k = \frac{w_k}{\sum_{k=1}^m w_k},$$

$$(13) w_k = u_k + \lambda_o u_o,$$

тогда параметры y_k находятся как

$$(14) y_k = \alpha_k y,$$

$$(15) y_k = \frac{\alpha_k y}{\lambda_k},$$

где y – набор выходных переменных.

При наличии более двух уровней в системе расчет выполняется идентично для всех последующих промежуточных факторов.

3. Апробация модели на примере показателя урожайности зерновых культур

3.1. ПОДГОТОВКА ИСХОДНЫХ ПАРАМЕТРОВ ДЛЯ МОДЕЛИРОВАНИЯ

Для реализации и оценки модели использовались сведения об урожайности зерновых культур на территории РФ, содержащиеся в отчете Росстата за период с 2000 по 2021 гг. В качестве выходного (результатирующего) параметра принят y – общая средняя урожайность в стране (ц/га). Факторами множественной регрессии были выбраны: x_1, x_2 – объем вносимых минеральных и органических удобрений (млн т); x_3 – энергообеспеченность сельскохозяйственных организаций (лс/га); x_4, x_5, x_6 – средняя температура по месяцам в вегетативный период (апрель, май, июнь) в градусах; x_7, x_8, x_9 – средняя сумма осадков по месяцам в вегетативный период (в мм).

Задача оценки урожайности предполагает создание трех регрессионных моделей, взаимосвязанных между собой и соответствующих трем случаям категориям обрабатываемых территорий:

– когда в почву вносились минеральные и органические удобрения; обозначим через y_1 среднюю урожайность удобряе-

мых пахотных площадей; y_{11} – урожайность при внесении минеральных удобрений; y_{12} – урожайность при внесении минеральных и органических удобрений;

– когда поля не удобрялись; урожайность обозначена через y_2 [8].

Представленная задача определения урожайности удобряемых и не удобряемых пахотных площадей соответствует модели (1) с подчинением λ_k условию (2) и определением по формуле (10).

Особенностью задачи является влияние факторов $x_4 - x_9$ на выходной показатель. Урожайность любой сельскохозяйственной культуры зависит от климатических условий. Температурный режим и влажность должны находиться в определенном промежутке, так как любые отклонения за пределы оптимальных значений негативно влияют на урожайность [4–5, 16].

Преобразование факторов осуществлялось следующим образом. Выделим из исходных данных те, которые соответствуют наиболее урожайным годам и для них найдем средние значения показателей. Пусть найденные таким образом значения равны $\tilde{x}_i, i = \overline{1, 9}$. Тогда положим

$$(16) \quad e_i(x_i) = x_i - \tilde{x}_i, \quad i = \overline{1, 2}; \quad e_i(x_i) = -|x_i - \tilde{x}_i| + e_{oi}, \quad i = \overline{3, 9}.$$

Урожайность будем искать в виде

$$(17) \quad y_k = \sum_{i=1}^9 a_{ki} e_i(x_i), \quad k = \overline{1, 2},$$

где $a_{21} = a_{22} = 0$, что соответствует случаю, когда поля не удобрялись, $e_i(x_i)$ – базисные факторы.

3.2. ПОСТРОЕНИЕ ИЕРАРХИЧЕСКОЙ МОДЕЛИ

Среднюю урожайность РФ с учетом различий в обработке полей удобрениями можно представить в виде иерархической структуры, см. рис. 2.

Построим данную модель по данным за период 2000–2019 гг.

Построение математической модели происходит в несколько этапов:

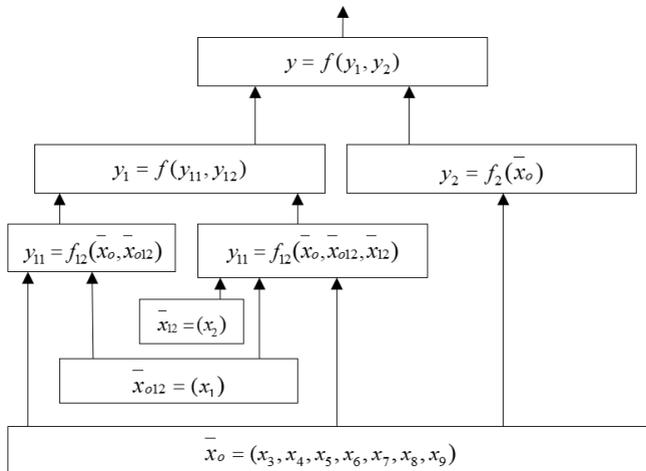


Рис. 2. Иерархическая структура модели урожайности

1 этап. Определение коэффициентов a_o, a_i регрессии 1 уровня (4) с помощью языка программирования Python с использованием библиотеки scikit-learn.

2 этап. Расчет коэффициентов λ_k по формулам (10) или (11) с учетом вариаций (8), (9).

В результате решения задачи получены следующие результаты. Значения λ_k оказались равными: $\lambda_1 = 0,87, \lambda_2 = 0,13$. Заметим, что эти результаты хорошо согласуются со статистическими данными относительно доли удобряемых и неудобряемых площадей. Так, в данных за 2021 год приводятся следующие значения показателей: $\lambda_1 = 0,81, \lambda_2 = 0,19$. При этом имеющиеся расхождения в величине λ_2 можно объяснить погрешностью оценки, так как это значение мало по сравнению с λ_1 и использованием модельных показателей λ_1, λ_2 как среднего значения за тестовый период [8].

3 этап. Определение набора выходных переменных 2 уровня по формулам (12)–(14) и коэффициентов модели. Оценка показателей качества модели: σ – скорректированный коэффициент детерминации модели; MSE – среднеквадратическое отклонение модели.

Для модели y_1 были получены следующие значения: $\sigma = 0,89$, $MSE = 0,98$. Для модели y_2 были получены следующие значения: $\sigma = 0,71$, $MSE = 0,77$.

4 этап. Повторение этапа 2 и 3 для следующих уровней (при их наличии).

Для модели y_{11} были получены следующие значения: $\sigma = 0,89$, $MSE = 0,86$. Для модели y_{12} были получены следующие значения: $\sigma = 0,91$, $MSE = 0,99$. Показатели оценки точности регрессий отражают сильную связь входных и выходных переменных. При этом $\lambda_{11} = 0,36$, $\lambda_{12} = 0,5$.

Построенная модель позволила оценить урожайность зерновых и зернобобовых культур при различных системах внесения удобрений и определить доли площадей для каждой из систем, см. таблицу 1.

Таблица 1. Урожайность зерновых и зернобобовых культур [8]

Год	Урожайность			
	y_{11}	y_{12}	y_2	y
2000	14,15	19,51	13,25	15,60
2001	17,50	20,94	17,86	19,40
2002	16,94	20,84	15,66	19,60
2003	15,78	19,40	16,28	17,80
2004	19,01	20,68	18,52	18,80
2005	16,20	16,77	15,89	18,50
2006	19,29	19,34	17,94	18,90
2007	20,18	20,40	17,73	19,80
2008	23,55	24,60	19,40	23,80
2009	22,78	24,65	19,15	22,70
2010	18,78	20,36	15,15	18,30
2011	21,86	23,38	17,25	22,40
2012	19,79	21,69	15,91	18,30
2013	20,31	22,62	17,37	22,00
2014	21,74	25,91	17,71	24,10
2015	22,64	27,48	18,61	23,70
2016	23,36	28,57	17,41	26,20
2017	26,35	31,94	19,20	29,20
2018	24,69	30,94	17,58	25,40
2019	23,66	30,52	15,26	26,70

Множественная модель 1 уровня имеет высокий прогнозный потенциал при оценки будущих значений средней урожайности, который был проверен на среднесрочный период – 3 года.

С использованием модели был построен прогноз урожайности на 2020–2022 год, прогнозное значение при этом составило: 2020 г. – 28,35 ц/га при фактическом объеме 28,6 ц/га, в 2021 г. – 30,35 ц/га при фактическом 26,7 ц/га, в 2022 г. – 34,87 ц/га, при фактическом объеме в 33,6 ц/га. Отклонение факта от прогноза составляет 0,25 ц/га, 3,65 ц/га и 1,14 ц/га в абсолютных значениях и в относительных – 1%, 14% и 4% соответственно.

При использовании в целях прогнозирования первого уровня модели также можно получить прогнозные значения высокой точности, но при влиянии на фактическое значение урожайности неопределенных внешних факторов данная модель имеет больший процент отклонения прогнозного значения от фактического. Так, например, фактическая урожайности 2021 г. была оценена в 26,7 ц/га, но ее модельное значение представлено в виде 31,8 ц/га, разница составляет 19,1%. В то же время иерархическая модель снизила процент отклонения до 13,7%.

На основании полученного прогнозного значения у можно произвести оценку ее составляющих, а именно, значения показателя урожайности при внесении только минеральных, минеральных и органических удобрений и отсутствия удобрений. Для 2020 г.: $y_{11} - 26,16$ ц/га, $y_{12} - 32,78$ ц/га, $y_2 - 17,27$ ц/га; для 2021 г.: $y_{11} - 28,46$ ц/га, $y_{12} - 34,97$ ц/га, $y_2 - 17,67$ ц/га; для 2022 г.: $y_{11} - 33,0$ ц/га, $y_{12} - 39,78$ ц/га, $y_2 - 21,08$ ц/га.

Данная модель также может эффективно применяться для оценки и прогнозирования урожайности для отдельных регионов с учетом возможности вариации вегетационного периода. При усредненных данных вегетационный период определяется с учетом самой значимой территориально культуры. Так, на территории РФ и, в частности, на территории Ростовской области преобладает пшеница, в результате климатические показатели представлены на основании ее вегетационного периода.

Рассматривая данные для Ростовской области в целях построения модели, было определено отсутствие показателя энер-

гообеспеченности в отчетности Ростовской области. Исключение данного базисного фактора из модели не повлекло существенного изменения качества модели и позволило провести анализ, оценку и прогнозирования показателя урожайности зерновых культур. Таким образом, добавление или изменение факторов в модели возможно, но требует проверки значимости в целях предотвращения перегруженности многофакторной модели.

Иерархическая модель урожайности для Ростовской области идентична построения для РФ, с учетом отсутствия фактора x_3 , и охватывает период с 2000 по 2019 г. в целях настройки модели и период с 2020 по 2022 гг. – в целях проверки качества прогнозирования.

Особенностью данных РО по сравнению с усредненными по РФ являются более динамичные изменения как в сторону увеличения, так и в сторону уменьшения показателя урожайности на период с 2000 г. по 2012 г. и отчетливые тренды роста в периоды 2012 – 2017 гг, 2018 – 2022 гг., что отражает менее явную линейную зависимость по сравнению с данными РФ.

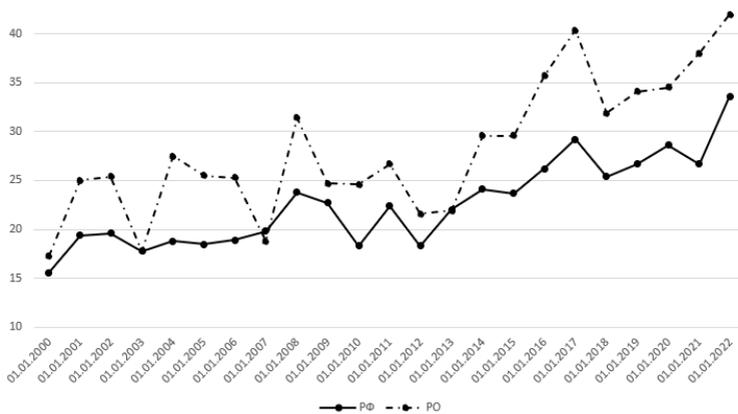


Рис. 3. Изменение показателя урожайности РФ и РО в 2000–2022 гг., ц/га

Характеристики первого уровня иерархической модели РФ: $\sigma = 0,78$, $MSE = 2,18$. При разделении набора на втором уровне

коэффициенты $\lambda_1 = 0,91$, $\lambda_2 = 0,09$ соответствуют удобряемой и неудобряемой территории, средние фактические показатели процента удобряемых территорий РО за 2015–2019 гг.: $\lambda_1 = 0,87$, $\lambda_2 = 0,13$.

Второй уровень модели представлен показателями Y_1 и Y_2 с оценочными параметрами $\sigma = 0,8$, $MSE = 2,27$ и $\sigma = 0,81$, $MSE = 1,35$ для удобряемых и неудобряемых территорий соответственно. Y_1 включает в себя модели y_{11} и y_{12} , характеризующиеся влиянием только минеральных удобрений (y_{11}) и влиянием всех вносимых удобрений (y_{12}). Оценочные характеристики для y_{11} : $\sigma = 0,83$, $MSE = 2,17$; для y_{12} : $\sigma = 0,81$, $MSE = 2,24$; с вкладом в Y_1 $\lambda_1 = 0,45$, $\lambda_2 = 0,46$ соответственно.

Среднее относительное отклонение по всему набору данных 9,02%, при этом с учетом неравномерно изменяющихся показателей урожайности максимальное отклонение составляет 28,7%. Данный показатель отражает более слабую прогнозную точность на среднесрочный период и влияние внешних неучтенных факторов в период 2020–2022 гг., так как среднее относительное отклонение по 20 предыдущим точкам составляет 7,6%.

3.3. СОПОСТАВЛЕНИЕ ИЕРАРХИЧЕСКОЙ МОДЕЛИ И ФАКТОРНОГО АНАЛИЗА

Методика иерархической модели заключается в разделении входных переменных на общую и индивидуальные части. Схожими принципами обладает факторный анализ, который заключается в оценке и группировке входных признаков на скрытые компоненты.

На языке программирования Python определить скрытые компоненты можно с помощью библиотеки `factor_analyzer`, задавая самостоятельно или определяя с помощью «метода локтя» (график каменистой осыпи) оптимальное количество скрытых компонент.

Для набора данных РФ в период 2000–2019 гг. был проведен факторный анализ для выявления скрытых компонент и проверки возможности адекватного разделения переменных

на поднаборы для определения урожайности на удобряемых и неудобряемых площадях.

Исходя из графика (рис. 4) оптимальным количеством компонент является 4, тогда произведем расчет значений факторов. Факторы сгруппировали переменные следующим образом, см. таблицу 2.

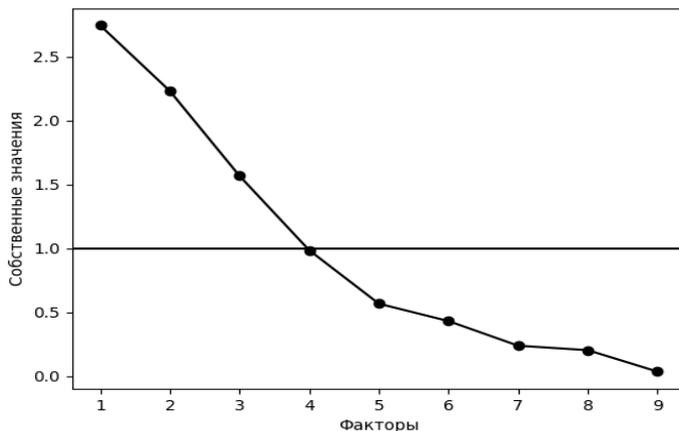


Рис. 4. График каменной осыпи для определения необходимого количества факторов

Таблица 2. Группировка переменных X по 4 факторам

Фактор 1	Фактор 2	Фактор 3	Фактор 4
x_3	x_1	x_5	x_4
x_6	x_2	x_7	
x_8			
x_9			

Очевидно, что фактор 2 описывает влияние удобрений на урожайность, однако в факторе 4 присутствует только одна переменная, что является спорным моментов в оценке правильности выбора указанного количества компонент.

Для проверки выберем в качестве количества формируемых факторов 3, так как именно 3 точки лежат выше уровня собственного значения в размере 1 (таблица 3).

Таблица 3. Группировка переменных X по 3 факторам

Фактор 1	Фактор 2	Фактор 3
x_1	x_5	x_2
x_3	x_7	x_4
x_6	x_9	
x_8		

Фактор 2 полностью представлен показателями осадков, но минеральные и органические удобрения разделились на разные группы факторов. С учетом распределения переменных, построение регрессионной модели урожайности на неудобряемой территории может быть представлено только через фактор 2, то есть с учетом влияния только объема осадков, что противоречит установленному факту зависимости урожайности от колебаний температур [15].

Характеристики регрессионной модели первого уровня на основании 4 факторов: $\sigma = 0,7$, $MSE = 1,75$, что отражает более худшее качество модели по сравнению с общей моделью урожайности по входным переменным X . При этом, используя методику определения доли площадей удобряемых и неудобряемых территорий, результатом являются показатели $\lambda_1 = 0,71$, $\lambda_2 = 0,29$, которые несопоставимы с фактическими представленными ранее показателями.

Определение скрытых факторов для данных урожайности Ростовской области показала схожий результат: «метод локтя» идентичен по числу оптимальных факторов, фактор 4 сформирован одной переменной. Отличие состоит в разделении переменных удобрений по разным компонентам.

Таблица 4. Группировка переменных X по 4 факторам

Фактор 1	Фактор 2	Фактор 3	Фактор 4
x_4	x_3	x_1	x_2
x_5	x_6	x_7	
		x_8	

При использовании трех компонент показатели минеральных и органических удобрений также определяются в разные группы.

Таким образом, в контексте изучения урожайности сельскохозяйственных культур факторный анализ не может полностью удовлетворить запрос на разделение общей урожайности на подтипы в зависимости от существующей подкормки, группировка некоторых переменных к одному фактору не может быть полностью логически обоснована.

4. Потребление кормовой пшеницы

Анализ и оценка объемов потребления зерновых культур являются важным аспектом при производстве зерна. В связи с Распоряжением Правительства РФ «Об утверждении Долгосрочной стратегии развития зернового комплекса Российской Федерации до 2035 года» в период до 2035 г. ожидается увеличение спроса и предложения на продукцию зернового комплекса. На этом основании остро встает вопрос распределения объемов зерна на потребительские и кормовые нужды. Анализ потребления зерновых культур с помощью многофакторных регрессионных моделей представлен авторами в ранней публикации [7]. Проведенный анализ соотношения типов потребления определил, что абсолютное значение объемов продовольственного потребления пшеницы за период 2001–2022 гг. варьируется в пределах 21–24 млн т. В то же время абсолютные и относительные показатели объема кормового потребления растут: так, в 2001 г. доля кормовой пшеницы составляла 35% от общего ее потребления, а к 2022 г. этот показатель вырос до 47%, в абсолютных показателях значения оцениваются в 13 и 21,5 млн т соответственно.

При сравнении объемов производства комбикормов (и доли пшеницы в виде сырья в общей массе) и общего объема кормового потребления было выявлено, что на производство корма для животных уходит только часть пшеницы, так как во многих сельскохозяйственных организациях все еще распространено кормление животных зерновыми культурами без добавления

дополнительных питательных веществ [11]. В связи с этим была поставлена задача оценки объемов пшеницы, направляемых на производство комбикормов и объемов чистого потребления.

В качестве исходных данных выступают показатели, взятые из российских статистических ежегодников Росстата и данных отчета АБ-Центра, а именно: y – потребление кормовой пшеницы (млн т); x_1 – поголовье КРС без учета молочных коров (млн); x_2 – поголовье молочных коров (млн); x_3 – поголовье свиней (млн); x_4 – поголовье овец и коз (млн); x_5 – поголовье птицы (млн).

На основании рекомендаций по питанию сельскохозяйственных животных и птиц была составлена иерархическая схема потребления кормовой пшеницы, см. рис. 5.

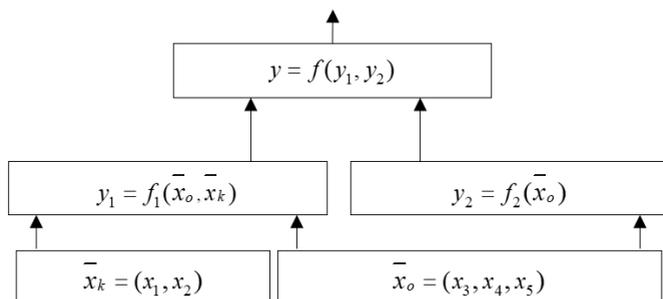


Рис. 5. Иерархическая структура модели потребления кормовой пшеницы

В результате построения множественной регрессионной модели 1 уровня были определены коэффициенты модели, при этом $\sigma = 0,7$, $MSE = 1,33$.

Показатели λ_k определяются по формуле (11) с учетом (9): $\lambda_1 = 0,4$, $\lambda_2 = 0,6$.

Для формирования общей и индивидуальной части использовались формулы (7).

Построенные регрессионные модели 2 уровня определяются показателями оценки: $\sigma = 0,74$, $MSE = 0,84$ для модели y_1

и $\sigma = 0,9$, $MSE = 0,5$ для модели y_2 , что подтверждает значимость результатов моделирования.

Подтверждением достоверности результатов также является среднее значение объема пшеницы в составе произведенных комбикормов по данным с 2015 по 2022 гг. (более ранние значения показателей являются неполными и не могут участвовать в оценке), см. таблицу 5.

Таблица 5. Урожайность зерновых и зернобобовых культур

Период	Объем производства комбикорма, млн т	Объем пшеницы в массе комбикормов, млн т	Модельные значения объема пшеницы на комбикорм, млн т
2015	24,81	9,79	9,42
2016	26,07	10,20	11,72
2017	28,01	10,96	13,64
2018	29,17	11,29	11,69
2019	30,35	11,59	10,54
2020	31,22	11,82	11,78
2021	32,18	12,18	10,38
2022	34,21	12,76	12,56

На основании представленных результатов моделирования можно сделать вывод о применимости модели для анализа и оценки сельскохозяйственных показателей.

5. Вывод

В результате исследования применения иерархической многофакторной регрессионной модели приведена процедура построения регрессионной многоуровневой системы для оценки и прогнозирования урожайности зерновых культур по Российской Федерации в целом и по Ростовской области в частности. Представлена иллюстрация численной процедуры разделения многофакторной модели на уровни и сравнение результатов процедуры с результатами применения факторного анализа для определения скрытых компонент. Применение иерархической модели позволяет не только сгруппировать переменные в фак-

торы, но и логически обосновать эти факторы, представляя их как действительно существующие параметры, а также использовать эти факторы для обособленного анализа.

Для относительных величин (модель урожайности зерновых и зернобобовых культур) построена трехуровневая иерархическая система, проведена оценка качества модели с использованием стандартных методов и специализированного программного обеспечения, построен и подтвержден краткосрочный прогноз на три года вперед.

Для абсолютных показателей (модель потребления кормовой пшеницы) произведено определение состава и объемов входящих в них значений, оценена точность модели.

Оценка предложенной иерархической многофакторной модели на различных показателях агропромышленного комплекса позволяет высоко оценить применимость модели для сельскохозяйственных и агропромышленных нужд.

Литература

1. БЕРГ Н.А., ДЕГТЯРЕВА Н.А. *Принятие управленческих решений в сельском хозяйстве на основе модели множественной регрессии* // Вестник ЧелГУ. – 2023. – №11(481). – С. 167–175.
2. БИСЧОКОВ Р.М. *Анализ, моделирование и прогноз урожайности сельскохозяйственных культур средствами искусственных нейронных сетей* // Вестник Российского университета дружбы народов. Серия: Агрономия и животноводство. – 2022. – Т. 17, №2. – С. 146–157.
3. ЗИНЧЕНКО В.Е. *Прогнозирование урожайности озимых культур и реакции ярового ячменя и озимой пшеницы на различные приемы обработки солонцов в условиях ростовской области*: Автореф. дис. канд. сель.-хоз. наук., Персиановский, 2005. – 22 с.
4. КЛОЧКОВ А.В., СОЛОМКО О.Б., КЛОЧКОВА О.С. *Влияние погодных условий на урожайность сельскохозяйственных культур* // Вестник белорусской государственной сельскохозяйственной академии. – 2019. – №2. – С. 101–105.

5. ЛАМАЖАП Р.Р., ЛИПШИН А.Г. *Влияние климатических условий на урожайность ярового ячменя в республике Тыва* // Вестник КрасГАУ. – 2016. – №12(123). – С. 13–19.
6. МАКСИМОВ Р.А. *Метод определения параметров адаптивной способности с использованием множественного регрессионного анализа взаимосвязи урожайности и ее элементов структуры* // Достижения науки и техники АПК. – 2021. – №6. – С. 4–10.
7. ОВЧИННИКОВ П.В., КОМИССАРОВА М.А., МИРОСЛАВСКАЯ М.Д. *Прогнозирование потребления зерновых культур при помощи многофакторных регрессионных моделей* // XIV Всероссийское совещание по проблемам управления (ВСПУ-2024): сборник научных трудов, 17-20 июня 2024 г., Москва. – С. 2161–2164.
8. ОВЧИННИКОВ П.В., ТКАЧЕВ А.Н., МИРОСЛАВСКАЯ М.Д. и др. *Идентификация смешанных аддитивных регрессионных моделей многоуровневых систем* // Изв. ВУЗов. Сев.-Кавк. регион. Техн. науки. – 2024. – №2. – С. 28–39.
9. ПРОКИНА Л.Н., ПУГАЕВ С.В. *Урожайность озимой пшеницы в зависимости от предшественников, удобрений и известкования* // Аграрная наука Евро-Северо-Востока. – 2022. – №3. – С. 318–326.
10. РОГАЧЕВ А. Ф., МЕЛИХОВА Е. В. *Прогнозирование продуктивности агрокультур на основе ретроспективных данных методом наименьших модулей* // Известия наук. – 2022. – №1(65). – С. 361–369.
11. САВОСТИН Д.С., САВОСТИН С.Д., МАГОМЕДОВ М.Д. и др. *Научное обоснование направлений увеличения объемов производства комбикормов и животноводческой продукции в Российской Федерации* // Экономические системы. – 2022. – №1. – С. 99–109.
12. СЕДЫХ И.А., ИСТОМИН В.А. *Применение иерархических динамических нейро-окрестностных моделей* // XIV Всероссийское совещание по проблемам управления (ВСПУ-2024): Сборник научных трудов, 17-20 июня 2024 г., Москва. – С. 3000–3004

13. СЕМИНЧЕНКО Е.В., СОЛОНКИН А.В. *Влияние климатических факторов на урожай озимой пшеницы и ярового ячменя в условиях сухой степи Нижнего Поволжья* // Siberian Journal of Life Sciences and Agriculture. – 2022. – №3. – С. 58–74.
14. СЕРАЯ Т.М., БОГАТЫРЕВА Е.Н., БИРЮКОВА О.М. и др. *Агроэкономическая эффективность органических удобрений при возделывании озимой пшеницы на дерново-подзолистой легкосуглинистой почве* // Почвоведение и агрохимия. – 2012. – №2(49) – С. 82–96.
15. СОЗИНОВ А.А., ЖЕМЕЛА Г.П. *Улучшение качества зерна озимой пшеницы и кукурузы*. – М.: Колос, 1983. – 270 с.
16. ШУБНОВ М.Г. *Алгоритмы и инструментальные средства нейросетевых технологий моделирования урожайности на основе автокорреляционных функций временных рядов*: Автореф. дис. канд. экон. наук. – Кисловодск, 2013. – 23 с.
17. AFTHANORHAN A., AWANG Z., AIMRAN N. *Five common mistakes for using partial least squares path modeling (pls-pm) in management research* // Contemporary Management Research. – 2020. – Vol. 4, No. 16. – P. 255–278.
18. BHARADIYA J.P., TZENIOS N., REDDY M. *Forecasting of crop yield using remote sensing data, agrarian factors and machine learning approaches* // Journal of Engineering Research and Reports. – 2023. – Vol. 24, No. 12. – P. 29–44.
19. CHENG E.W. *SEM being more effective than multiple regression in parsimonious model testing for management development research* // Journal of Management Development. – 2001. – Vol. 7, No. 20. – P. 650–667.
20. DISSANAYAKE D.M.P.W., RATHNAYAKE R.M.K.T., CHATHURANGA G. *Crop yield forecasting using machine learning techniques - a systematic literature review* // KDU Journal of Multidisciplinary Studies (KJMS). – 2023. – Vol. 5, No. 1. – P. 54–65.
21. ELANGO VAN N., RAJENDRAN R. *Structural equation modeling-a second-generation multivariate analysis* // National Conference on Indian Business Management «Emerging Man-

- agement Paradigm in Indian Business» at: Sri Ramakrishna Institute of Technology, Coimbatore Volume. – 2015. – P. 33–54.
22. HAYFIELD T., RACINE J.S. *Nonparametric econometrics: the NP package* // Journal of statistical software. – 2008. – Vol. 5, No. 27. – P. 1–32.
 23. ITIDAL A.A., MOHAMMAD A.A., WALEED B.A. *Application of stepwise multiple regression to supersaturated designs data of water pollution in Saudi Arabia* // JP Journal of Biostatistics. – 2024. – Vol. 3, No. 24. – P. 487–515.
 24. MERONI M., WALDNER F., SEGUINI L. et al. *Yield forecasting with machine learning and small data: what gains for grains?* // Agricultural and Forest Meteorolog. – 2021. – Vol. 308–309. – P. 1–13.
 25. NUNKOO R., RAMKISSOON H. *Structural equation modeling and regression analysis in tourism research* // Current Issues in Tourism. – 2011. – No. 15(8) – P. 1–26.
 26. PUTRA W.B.T.S. *Modul Pembelajaran SEM-PLS: Permasalahan, Kepercayaan Umum, Tahapan Spesifikasi dan Evaluasi Model PLS-SEM, HCMs, serta PLS-MGA* // Indonesian School of Research. – 2024. – 85 p.
 27. SADIKAJ G., WRIGHT A.G.C., DUNKLEY D.M. et al. *Multi-level structural equation modeling for intensive longitudinal data: a practical guide for personality researchers* // The Handbook of Personality Dynamics and Processes. – 2021. – P. 855–885.
 28. WANG W., et al. *Forecasting elections with non-representative polls* // Int. Journal of Forecasting. – 2015. – P. 980–981.
 29. ZOU H., HASTIE T. *Regularization and variable selection via the elastic NET* // Journal of the Royal Statistical Society: Series B (Statistical Methodology). – 2005. – Vol. 67, No. 2. – P. 301–320.

USING A HIERARCHICAL MULTI-VARIABLE REGRESSION MODEL FOR THE PURPOSE OF ANALYSIS AND FORECASTING AGRICULTURAL INDICATORS

Petr Ovchinnikov, Platov South-Russian State Polytechnic University (NPI), Novocherkassk, Cand.Sc., Associate Professor (pvo78@yandex.ru).

Alexander Tkachev, Platov South-Russian State Polytechnic University (NPI), Novocherkassk, Doctor of Science, Professor (tkachev.an@mail.ru).

Maria Miroslavskaya, Platov South-Russian State Polytechnic University (NPI), Novocherkassk, Graduate Student (miroslavsky.marymir@gmail.com).

Abstract: The article presents a methodology for determining the latent parameters of a multi-level multifactor regression model and ways of using the model using the example of forecasting the dynamics of indicators reflecting the functioning of agricultural production: grain crop yield and feed wheat consumption. The relevance of the model is due to the possibility of using it for forecasting the main output indicator (the first level of the model) and its intermediate components (subsequent levels of the model). The article processes statistical data, applies the method of regression analysis of information, and plots the results of modeling using MS Excel and the Python programming language development environment. The model is built on the basis of a hierarchical dependence of the first and subsequent levels, in which the input data of the output indicator are grouped as individual or common parameters of linear dependencies of intermediate variables. As a result of testing the model, data on grain crop yields were obtained depending on the type of fertilizer application and the corresponding shares of areas according to the Russian Federation as a whole and the Rostov Region, in particular, a yield forecast for three years was made, and a comparison with the results of applying factor analysis was given. In determining the consumption of feed wheat, data on the volume of net wheat consumption and the volume of wheat processed into compound feed were obtained.

Keywords: regression model, hierarchical system, yield, consumption, grain crops.

УДК 519.6 +330.4

ББК 22.19

Статья представлена к публикации членом редакционной коллегии Р.М. Нижегородцевым.

Поступила в редакцию 31.10.2024.

Опубликована 31.05.2025.

АЛГОРИТМ ИДЕНТИФИКАЦИИ ЗАГРЯЗНЕНИЙ ПОВЕРХНОСТИ ВОДОЕМОВ ДЛЯ ГЕОИНФОРМАЦИОННЫХ СИСТЕМ

Вытовтов К. А.¹, Гладких Т. Я.²
(ФГБУН Институт проблем управления
им. В.А. Трапезникова РАН, Москва)

Рассматривается актуальная проблема обнаружения нефти и нефтепродуктов на поверхности воды, имеющая решающее значение для экологического мониторинга водных ресурсов. Рост загрязнения водных объектов нефтепродуктами обуславливает необходимость разработки эффективных методов их обнаружения и оценки состояния экосистем. В данной работе представлена методика, основанная на исследовании спектров отраженного солнечного света, которая позволяет с высокой точностью обнаруживать нефтяные загрязнения. Авторами получены аналитические выражения для индексов коэффициентов спектральной яркости, которые в дальнейшем использованы в геоинформационных системах (ГИС). Эти показатели помогают в обработке и интерпретации данных, полученных с помощью дистанционного зондирования, значительно упрощают процесс мониторинга и анализа состояния водных объектов. Методика показала свою эффективность в различных условиях и может быть адаптирована для использования в режиме реального времени. Результаты исследования способствуют совершенствованию методов экологического контроля и обеспечивают более надежную защиту водных ресурсов от загрязнения. Кроме того, в исследовании рассматривается потенциальная интеграция этого метода с другими системами экологического мониторинга для повышения точности и эффективности обнаружения загрязнения. Благодаря этому усовершенствованиям, предлагаемая методика позволяет не только выявлять разливы нефти, но и помогает в управлении ими и снижении негативных последствий, связанных с разливами нефти.

Ключевые слова: нефтепродукты, экологический мониторинг, мультиспектральные изображения, спектрометрия, обработка информации, геоинформационная система.

1. Введение

В последние годы актуальность проблемы загрязнения водных ресурсов нефтью и нефтепродуктами значительно возросла

¹ Константин Анатольевич Вытовтов, д.т.н., в.н.с. (vytovtov_konstan@mail.ru).

² Татьяна Яновна Гладких, н.с. (golnikt@yandex.ru).

в связи с увеличением объемов добычи, транспортировки и переработки углеводородного сырья [7, 18, 26]. По критериям экологической опасности, таких как токсичность, канцерогенность, распространенность и частота встречаемости, нефтепродукты относятся к приоритетным загрязняющим веществам [2]. Все эти факторы увеличивают риски возникновения чрезвычайных ситуаций, в частности аварий [3], где эффективные методы обнаружения загрязнений играют ключевую роль. В связи со сложившейся ситуацией быстрое и точное обнаружение и идентификация нефтяных пятен на водной поверхности становится важнейшей задачей, требующей разработки и применения эффективных методов мониторинга.

В имеющихся публикациях, посвященных способам идентификации нефти на воде, рассматриваются разнообразные подходы к выбору метода и средств мониторинга. Так, авторами статьи [1] были проанализированы преимущества и недостатки дистанционного, а конкретно спутникового и авиационного мониторинга загрязнений, и был описан комплексный, многоуровневый мониторинг, где немаловажную роль играет обработка и анализ поступившей информации с помощью геоинформационных систем. Использование ГИС-технологий становится ключевым инструментом, который обеспечивает экологическую безопасность [5]. Данная технология позволяет извлекать ключевую информацию, такую как прогнозирование распространения разливов нефти, зависимость от внешних факторов, оценку интенсивности и определение источника загрязнений [16]. Возможность картирования нефтяных разливов [11, 14] позволяет оценить их пространственное распределение по водной поверхности и тенденции распространения. На основе ГИС возможно создание систем поддержки принятия решений, обеспечивающих ввод, хранение и обработку данных для решения проблем планирования и управления [8]. При этом авторы [10] полагают, что ГИС-технологии целесообразно рассматривать как средство создания комплексной платформы, применимой в экологическом мониторинге для идентификации нефтяных разливов и прогнозирования их распространения. Все это требует соответствующего совершенствования существующих методов [13, 23] и создания новых под-

ходов спектрального анализа, спутниковой [19] и воздушной разведки с применением беспилотных летательных аппаратов (БПЛА) [15, 21, 25], а также лабораторных исследований, которые должны быть как можно более точными и оперативными для эффективного принятия мер по предотвращению экологических последствий нефтяных аварий. Проблема использования ГИС в области идентификации загрязнений водной поверхности нефтяными веществами заключается в необходимости точного и быстрого обнаружения и мониторинга таких загрязнений с использованием геопространственных данных. Недостаточная точность данных, а также несовершенство методов их анализа может затруднять эффективное применение геоинформационных систем для решения данной проблемы. Дополнительной сложностью является интеграция разнообразных исходных данных из многочисленных источников, включая спутниковые снимки, данные от БПЛА и наземные измерения. Важной задачей становится не только сбор данных, но и их стандартизация и нормализация для создания обобщенной и эффективной модели мониторинга. По этой причине разработка универсальных протоколов для обработки и анализа экологических данных, совместимых с ГИС, становится критически важной для обеспечения оперативности реагирования на инциденты. В этом контексте ученые и разработчики [6, 9, 20] стремятся к созданию мощных алгоритмов машинного обучения и искусственного интеллекта, способных автоматически анализировать и интерпретировать данные о загрязнениях, что является следующим шагом к повышению эффективности экологического мониторинга.

В настоящей работе предлагаются математическая модель и алгоритм обработки мультиспектральных изображений для идентификации нефтяных загрязнений на водных поверхностях для решения задач экологического мониторинга.

2. Формулировка проблемы

Рассматривается система экологического мониторинга, основанная на анализе мультиспектральных изображений нефтяных загрязнений водной поверхности, полученных с использова-

нием БПЛА. Основная сложность поставленной задачи заключается в вариабельности условий распространения нефтяных пятен, которые зависят от множества факторов, включая тип нефти, погодные условия, течения и температуру воды [22].

Таким образом, целью данной работы является разработка математической модели и алгоритма идентификации нефтяных загрязнений на водной поверхности с использованием мультиспектральной съемки, учитывающих спектры отраженных от нефтяных пленок сигналов при их разной толщине, добавлении лодочного масла. Для проверки корректности математической модели и соответствующего алгоритма анализируются данные натуральных экспериментов, которые позволят в реальных условиях оценить эффективность полученных результатов.

3. Математическая модель и алгоритм идентификации загрязнений

В данном разделе предложены математическая модель и алгоритм идентификации нефтяных и масляных загрязнений водной поверхности с использованием типовых мультиспектральных камер БПЛА. Использовался квадрокоптер с мультиспектральной камерой DJI Phantom 4. Параметры каналов камеры: центральная длина волны «синего канала» равна $450 \text{ нм} \pm 16 \text{ нм}$, центральная длина волны «зеленого канала» равна $560 \text{ нм} \pm 16 \text{ нм}$, центральная длина волны «красного канала» равна $650 \text{ нм} \pm 16 \text{ нм}$, центральная длина волны канала «красный край» равна $730 \text{ нм} \pm 16 \text{ нм}$, центральная длина волны «инфракрасного канала» равна $840 \text{ нм} \pm 16 \text{ нм}$. Число эффективных пикселей каждой матрицы: 2,08 млн. Угол обзора: $62,7^\circ$. Фокусное расстояние: 5,74мм. Кроме того, в экспериментах использовалась камера MicaSense RedEdge-MX с аналогичными параметрами. Особенностью предлагаемой модели является учет интенсивности солнечного излучения в различное время суток и время года, коэффициентов отражения от воды и нефтяной пленки и состояния атмосферы. В результате производится сравнение интенсивностей отраженного солнечного света от воды и нефти по каналам мультиспектральной камеры.

Коэффициент отражения от полубесконечной водной среды вычисляется на основе классических законов электродинамики и равен [4]

$$(1) R_{\text{вода}}(\lambda) = (\rho_{\text{воздух}} \cos \alpha_{\text{пад}} - \rho_{\text{вода}} \cos \alpha_{\text{прел}}) / (\rho_{\text{воздух}} \cos \alpha_{\text{пад}} + \rho_{\text{вода}} \cos \alpha_{\text{прел}}),$$

где $\rho_{\text{воздух}}$ – волновое сопротивление воздуха; $\rho_{\text{вода}}$ – волновое сопротивление воды; $\alpha_{\text{пад}}$ – угол падения; $\alpha_{\text{прел}}$ – угол преломления.

Коэффициент отражения от нефтяной пленки на водной поверхности рассчитывается с использованием так называемой матрицы преобразования [4, 23] как

$$(2) R_{\text{нефть}}(\lambda) = [(m_{11} + m_{12}p_{\text{вода}})p_{\text{воздух}} - (m_{21} + m_{22}p_{\text{вода}})] / [(m_{11} + m_{12}p_{\text{вода}})p_{\text{воздух}} + (m_{21} + m_{22}p_{\text{вода}})],$$

где m_{ij} – элементы матрицы преобразования для изотропной среды:

$$(3) M(\lambda) = \begin{pmatrix} \cos(\varphi(\lambda, \theta, d)) & -(j/p) \sin(\varphi(\lambda, \theta, d)) \\ -jp \sin(\varphi(\lambda, \theta, d)) & \cos(\varphi(\lambda, \theta, d)) \end{pmatrix}.$$

Здесь $\varphi(\lambda, \theta, d) = k_0 n(\lambda) d \cos(\theta)$; $p_{\text{воздух}} = \sqrt{\varepsilon_{\text{воздух}}(\lambda) / \mu} \cos \theta$ для случая, когда вектор напряженности электрического поля лежит в плоскости падения (ТЕ-волна); $p_{\text{воздух}} = \sqrt{\mu / \varepsilon_{\text{воздух}}(\lambda)} \cos \theta$ – когда вектор напряженности электрического поля перпендикулярен плоскости падения (ТМ-волна); $p_{\text{вода}} = \sqrt{\varepsilon_{\text{вода}}(\lambda) / \mu} \cos \theta$ для ТЕ-волны; $p_{\text{вода}} = \sqrt{\mu / \varepsilon_{\text{вода}}(\lambda)} \cos \theta$ – для ТМ-волны; θ – угол падения, зависящий от времени суток и времени года; d – толщина слоя нефти; λ – длина волны в свободном пространстве; $k_0 = 2\pi / \lambda$ – волновое число в свободном пространстве.

Тогда распределение интенсивности в спектре отраженного света может быть найдено как

$$(4) S(\lambda) = S_{\text{солнца}}(\lambda) R(\lambda),$$

где $S_{\text{солнца}}(\lambda)$ – распределение поля в видимом спектре излучения Солнца, определяемое экспериментально (рис. 1).

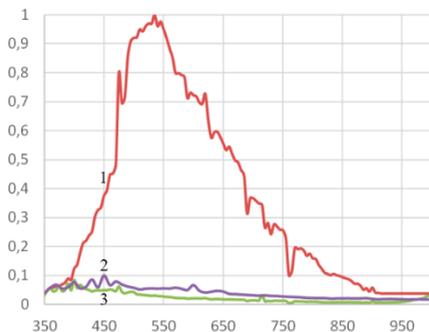


Рис. 1. Приведенные нормированные спектры: 1 – спектр солнечного света; 2 – спектр пленки толщиной 100мкм; 3 – спектр пленки толщиной 1мм

В предлагаемой математической модели также учитываются центральная длина волны λ_0 и ширина полосы пропускания $\Delta\lambda$ каждого канала соответствующей мультиспектральной камеры. Тогда суммарная интенсивность в i -м канале камеры может быть найдена следующим образом:

$$(5) \quad S_i = \int_{\lambda_0 - \Delta\lambda}^{\lambda_0 + \Delta\lambda} S_{\text{солнца}}(\lambda) R(\lambda) d\lambda.$$

Индексы коэффициентов спектральной яркости (КСЯ) для N -канальной камеры в простейшей форме рассчитываются по формуле

$$(6) \quad K_i = (S_i - S_N) / (S_i + S_N), \quad i = \overline{1, N}.$$

С учетом этого результирующий индекс КСЯ:

$$(7) \quad K = \sum_{i=1}^N \sum_{j=1}^N [\alpha_{ij} (s_i - s_j) / (s_i + s_j)], \quad i \neq j,$$

и проверяется соответствие его значения заданному диапазону. Здесь α_{ij} ($0 \leq \alpha_{ij} \leq 1$, $\sum_i \alpha_{ij} = 1$) – весовые коэффициенты, которые находятся экспериментально.

На основании анализа выражения (7) принимается решение о наличии или отсутствии загрязнения.

Таким образом, в соответствии с математической моделью алгоритм включает в себя следующие этапы (рис. 2):

- Расчет коэффициентов отражения от воды и нефтяной пленки по формулам (1) и (2).

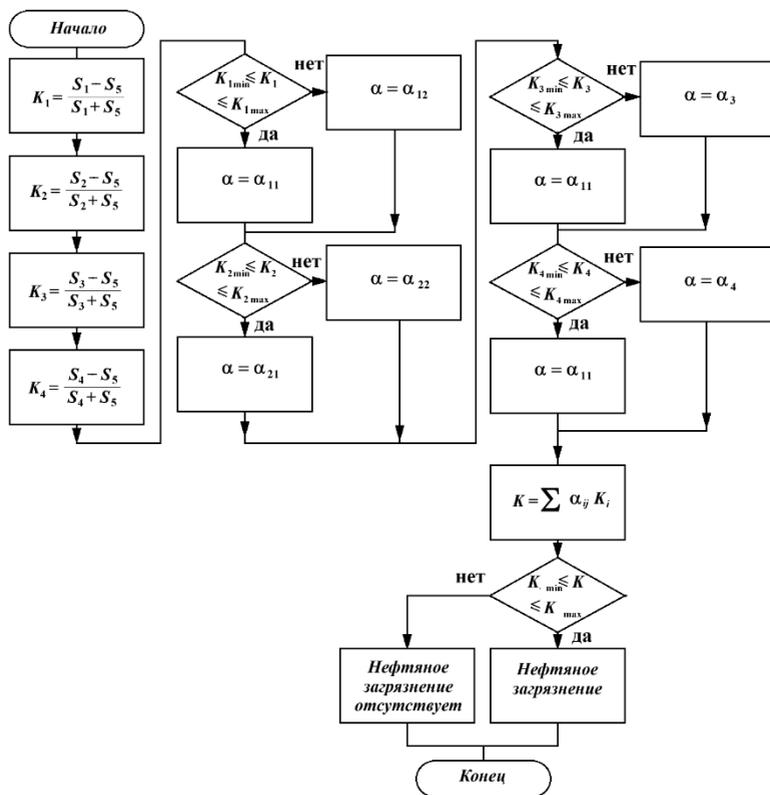


Рис.2. Блок-схема алгоритма определения загрязнений водной поверхности

- Расчет суммарной интенсивности в i -м канале камеры по формуле (5).
- Расчет простейших индексов коэффициентов спектральной яркости для чистой водной поверхности и нефтяной пленки на поверхности воды по формуле (6).
- Для каждого индекса проверяется нахождение значений в пределах соответствующих диапазонов: $K_{i\ min} \leq K_i \leq K_{i\ max}$ и в зависимости от этого весовым коэффициентам α присваиваются определенные значения α_{ij} ($0 \leq \alpha_{ij} \leq 1$, $\sum_i \alpha_{ij} = 1$).

– Расчет результирующего индекса коэффициента спектральной яркости по формуле (7).

– Принятие решения о наличии или отсутствии загрязнения.

Блок-схема алгоритма идентификации нефтяных загрязнений с использованием построенной математической модели представлена на рис. 2.

4. Лабораторный эксперимент

Для идентификации нефтяных загрязнений на водной поверхности необходимо исследовать как спектр отраженного света от нефтяной пленки, так и спектр отраженного света от самой воды, поскольку именно различия в интенсивности отраженного света по каналам мультиспектральной камеры позволяют проводить эффективный мониторинг загрязнений.

Измерение спектров отраженного солнечного света от пресной воды и нефтяной пленки проводилось в лабораторных условиях. Схема эксперимента представлена на рис.3.

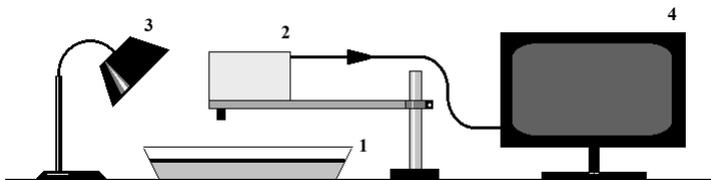


Рис. 3. Схема эксперимента: 1 – кювета с исследуемой жидкостью; 2 – оптический спектрометр «Flame» Ocean Insight; 3 – галогенная лампа; 4 – компьютер

На первом этапе были проведены измерения оптическим спектрометром «Flame» Ocean Insight спектра чистой воды в кюветах без нефтяных пленок для получения фоновых данных. Затем была проведена серия измерений спектральных характеристик нефтяных пленок марки Urals и пленок двухтактного лодочного масла на воде с фиксацией полученных спектральных данных.

Результаты измерений отраженного спектра тонкой нефтяной пленки различной толщины представлены на рис. 4. Здесь

кривая 1 соответствует толщине слоя 10 мкм; кривая 2 – толщине слоя 20 мкм; кривая 3 – толщине слоя 25 мкм; кривая 4 – толщине слоя 40 мкм; кривая 5 – толщине слоя 50 мкм. Из результатов измерений видно, что увеличение толщины нефтяной пленки приводит к изменению интенсивности отраженного света, но не изменяет форму спектра.

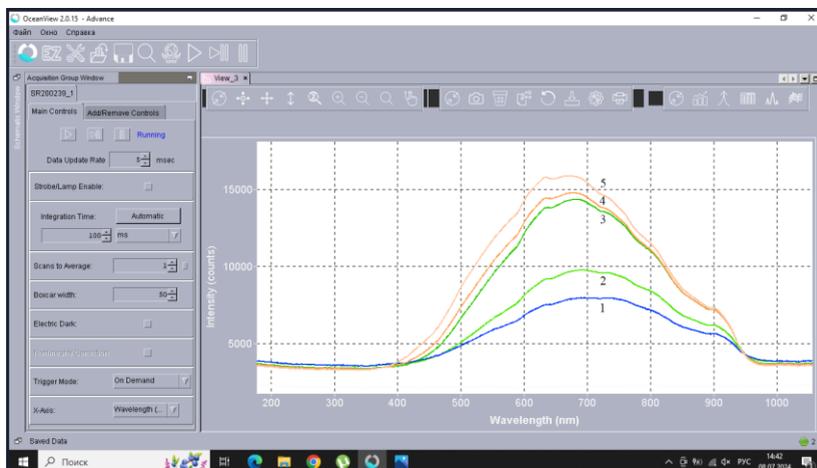


Рис. 4. Спектры отраженного нефтяной пленкой света от галогенной лампы: 1 – толщина слоя 10 мкм; 2 – толщина слоя 20 мкм; 3 – толщина слоя 25 мкм; 4 – толщина слоя 40 мкм; 5 – толщина слоя 50 мкм

В данном случае учитывается, что в полевом эксперименте будет использоваться пятиканальная камера с параметрами: для «синего канала» $\lambda_0 = 475$ нм и $\Delta\lambda = 20$ нм, для «зеленого канала» $\lambda_0 = 560$ нм и $\Delta\lambda = 20$ нм, для «красного канала» $\lambda_0 = 668$ нм и $\Delta\lambda = 20$ нм, для канала «красный край» $\lambda_0 = 717$ нм и $\Delta\lambda = 10$ нм, центральная длина волны «инфракрасного канала» равна 840 нм и $\Delta\lambda = 10$ нм.

Например, для нефтяной пленки толщиной 50 мкм (рис.4) имеем $K_1 \approx 0,600$, $K_2 \approx 0,487$, $K_3 \approx 0,298$, $K_4 \approx 0,2$. Тогда $K = \alpha_1 0,6 + \alpha_2 0,487 + \alpha_3 0,298 + \alpha_4 0,2$. При $\alpha_1 = \alpha_2 = 0,5$ и $\alpha_3 = \alpha_4 = 0$ имеем $K = 0,5435$.

Для корректного выбора весовых коэффициентов α_i в дальнейшем необходимо провести аналогичные вычисления для водной поверхности, на которой исследуется загрязнение. Коэффициенты выбираются таким образом, чтобы разница между индексами коэффициентов спектральной яркости загрязнения и поверхности была максимальна.

5. Верификация алгоритма. Полевой эксперимент

После успешного завершения лабораторных испытаний и разработки алгоритма идентификации нефти авторами статьи были проведены полевые эксперименты, чтобы проверить эффективность своего решения в реальных условиях. В первую очередь была определена цель эксперимента – оценить точность и надежность алгоритма, а также его способность выявлять нефтяные загрязнения на мультиспектральных изображениях, полученных с камер, установленных на беспилотных летательных аппаратах (БПЛА). Применение БПЛА для получения мультиспектральных снимков позволяет проводить идентификацию малоразмерных загрязнений, что дает преимущество перед аэрофотосъемкой и съемкой с космических аппаратов. Полученные коэффициенты были проанализированы с помощью геоинформационной системы для выявления характерных спектральных признаков и отличий между различными типами нефтяных пленок и подбора соотношений для более точной идентификации нефтяных загрязнений на водной поверхности.

Предложенный алгоритм может быть внедрен в системы мониторинга в реальном времени, что позволяет оперативно отслеживать и анализировать изменения в состоянии водных поверхностей. Использование разработанного алгоритма в ГИС позволяет получить изображение мест загрязнений с точными границами пятна. Пример изображения нефтяного пятна, полученного в результате контролируемого разлива нефти марки Urals на водном полигоне во время проведения натуральных испытаний. Съёмка была проведена 27 января 2023 года в утреннее время в малооблачную погоду. Съёмка проводилась мультиспектральной камерой с высоты 50 м. На водную поверхность в проруби было нанесено небольшое количество сырой нефти для получения эффекта

тонкой нефтяной пленки на водной поверхности. В соответствии с [12] толщина пленки составила 2–5 мкм. При рассмотрении с малого расстояния (с края проруби) пленка имела радужную расцветку. После завершения испытаний нефтяная пленка была локализована и удалена механическим способом. Проверка качества очистки проведена визуально. Полученные в результате съемки изображения были обработаны в ГИС (Рис. 5). На RGB-снимке (рис. 5а) отчетливо видны контуры проруби и отсутствие в ней посторонних предметов. Тонкое нефтяное пятно можно распознать только если заранее известно о его существовании и на малых расстояниях. На обработанном изображении (рис. 5б) цветовая гамма выбрана таким образом, что снежный покров отображен желтым цветом, водная поверхность – насыщенным синим цветом, нефтяные пятна в зависимости от толщины – оттенками голубого. Отметим, что съемка, например, в ясную погоду или в другое время года может потребовать дополнительных исследований, поскольку при этом изменится спектр отраженного от воды света.

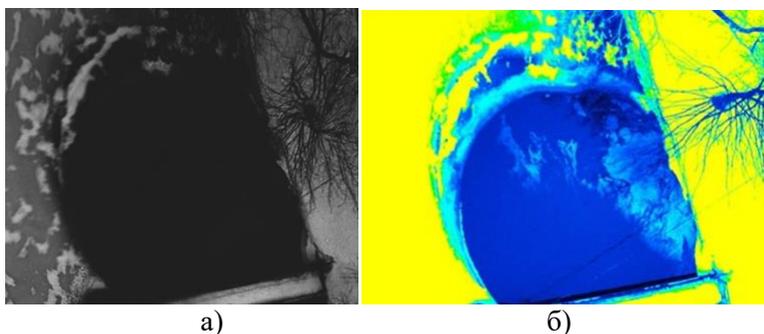


Рис. 5. а) Изображение нефтяной пленки на воде на RGB- снимке; б) Изображение нефтяной пленки на воде после обработки данных в ГИС

Внедрение данного алгоритма в ГИС обеспечивает возможность интеграции с другими источниками данных, такими как данные с беспилотных летательных аппаратов. Дистанционное зондирование в сочетании с применением ГИС-технологий улучшают идентификацию и классификацию разливов нефти, что

должно привести к созданию конечного продукта - карты распространения разливов нефти [17].

Практическое применение разработанного метода показало его эффективность в различных условиях, что подтверждается рядом выполненных полевых испытаний.

6. Заключение

В работе представлены математическая модель для нахождения индексов коэффициентов спектральной яркости и алгоритм идентификации загрязнений на водных поверхностях, основанный на анализе спектров отраженного солнечного света и выполнена его интеграция с геоинформационными системами. Предложенная математическая модель является аналитической и позволяет находить требуемые индексы коэффициентов спектральной яркости для различных условий окружающей среды еще до полевых испытаний, что существенно снижает затраты на экологический мониторинг.

Экспериментальные исследования показали, что предложенная математическая модель и алгоритм, основанный на анализе спектральных признаков, позволяют не только точно идентифицировать виды нефтяных пленок, но и оперативно отслеживать изменения в реальном времени.

Литература

1. АНТОНЕЦ К.В. *Комплексный мониторинг нефтегазовых загрязнений* // International Agricultural Journal. – 2021. – №1. – С. 49–54.
2. БАРАБАШИН Т.О., КОРАБЛИНА И.В., ПАВЛЕНКО Л.Ф. и др. *Методическое обеспечение мониторинга загрязнения водных объектов азово-черноморского бассейна* // Водные биоресурсы и среда обитания. – 2018. – Т. 1. – №3-4. – С. 9–27.
3. БЕСКИД П.П., БОГДАНОВ П.Ю., МИКЛУШ В.А. и др. *Результаты исследований в области дистанционных методов обнаружения нефтяных загрязнений на водной поверхности, проводимых в РГГМУ* // Гидрометеорология и экология. – 2020. – №60. – С. 371–391.

4. БОРН М., ВОЛЬФ Э. *Основы оптики*. М.: Изд-во Наука, 1973. – С. 24–81.
5. БУЗМАКОВ С.А., САННИКОВ П.Ю., СИВКОВ Д.Е. и др. *Разработка геоинформационных систем для управления окружающей средой и экологической безопасностью в районах эксплуатируемых нефтяных месторождений // Антропогенная трансформация природной среды*. – 2021. – Т. 1. – С. 102–127.
6. БУКИН О.А., ПРОЩЕНКО Д.Ю., ЧЕХЛЕНКО А.А. и др. *Методы оптического мониторинга нефтяного загрязнения морских акваторий с использованием беспилотных летательных аппаратов // Оптика атмосферы и океана*. – 2019. – Т. 32(4). – С. 324–328.
7. БЫКОВ Р.Е., ИГНАТЬЕВА Н.В., СЛАДКОВ А.Ю. *Технические средства контроля водных поверхностей // Человек и море*. – 1994. – С. 74–79.
8. ЕПИФАНОВ А.В., ХУРШУДЯН Н.С., ШАРЕНКОВ Д.В. и др. *Оценка техногенного воздействия на водные объекты с применением геоинформационных систем // Учебно-методическое пособие*. – 2010. – С. 36–62.
9. КАЗАНЦЕВ А.Ю. *Применение искусственного интеллекта в предотвращении и минимизации сбросов нефтесодержащих вод с судов // Каспийский научный журнал*. – 2024. – Т. 2. – №1. – С. 9–20.
10. КАЛАНТАЕВСКИЙ Ю.С., АЛЕКСАНДРОВА А.В., КСАНДОПУЛО С.Ю. и др. *Применение географических информационных систем для экологического мониторинга в нефтегазовом комплексе // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета*. – 2013. – Т. 86. – С. 92–101.
11. ЛУНЕВ Б.С., КРАСИЛЬНИКОВ П. А., ИЛАРИОНОВ С. А. и др. *Картирование территории при проведении геоэкологического мониторинга средствами ГИС // Фундаментальные исследования*. – 2014. – №11–1. – С. 89–93.
12. МОЛЬКОВ А.А., КАПУСТИН И.А., ЕРМОШКИН А.В. и др. *Дистанционные методы определения толщины плёнок*

- нефти и нефтепродуктов на морской поверхности // Современные проблемы дистанционного зондирования земли из космоса. – 2020. – Т. 17. – №3. – С. 9–27.*
13. ФЕДОТОВ Ю.В., БЕЛОВ М.Л., МАТРОСОВА О.А. и др. *Метод обнаружения нефтяных загрязнений на водной поверхности, основанный на регистрации флуоресцентного излучения в двух узких спектральных диапазонах. // Вестник Московского государственного технического университета им. Н.Э. Баумана. Серия «Приборостроение». – 2010. – Т. 2. – С. 39–47.*
 14. ЯЩЕНКО И.Г., АЛЕКСЕЕВА М.Н., СВАРОВСКАЯ Л.И. *Геоинформационные технологии для анализа нефтяных загрязнений рек // Интерэкспо Гео-Сибирь. – 2014. – Т. 7. – С. 38–43.*
 15. BAYIRHAN İ., GAZIOGLU C. *Use of unmanned aerial vehicles (UAV) and marine environment simulator in oil pollution investigations // Baltic journal of modern computing. – 2020. – Vol.8. – P. 327–336.*
 16. BARABANOVA E., VYTOVTOV K., GLADKIKH T. et al. *Environmental monitoring of water surface pollution in the visible range by using UAVs // Journal of Communications Technology and Electronics. – 2023. – Vol. 68. – P. 388–392.*
 17. ELJABRI A., GALLAGHER C. *Developing integrated remote sensing and GIS procedures for oil spills monitoring at Libyan coast // WIT Transactions on Ecology and the Environment. – 2012. – Vol. 44. – P. 17–20.*
 18. FERREIRA N.M., COUTINHO R., DE OLIVEIRA L.S. *Emerging studies on oil pollution biomonitoring: A systematic review // Marine Pollution Bulletin. – 2023. – Vol. 192. – P. 115081.*
 19. FINGAS M., BROWN C.E. *A review of oil spill remote sensing sensors // Sensors (Basel). – 2017. – Vol.18.*
 20. GHORBANI Z., BEHZADAN A.H. *Monitoring offshore oil pollution using multi-class convolutional neural networks // Environmental Pollution. – 2021. – Vol. 289. – P. 117884.*
 21. GÓMEZ C., GREEN D.R. *Small unmanned airborne systems to support oil and gas pipeline monitoring and mapping // Arab. J Geosci. – 2017. – Vol. 10. – DOI: 10.1007/s12517-017-2989-x.*

22. LUKYANOV A., HOZAN H., SIALOUNAS G. et al. *Dynamics of oil slicks on wavy water surfaces* // Environmental Fluid Mechanics. – 2024. – P. 1–20.
23. MOHAMMADI M., SHARIFI A., HOSSEINGHOLIZADEH M. et al. *Detection of oil pollution using SAR and optical remote sensing imagery: A case study of the Persian Gulf* // Journal of the Indian Society of Remote Sensing. – 2021. – Vol. 49. – No. 10. – P. 2377–2385.
24. VYTOVTOV K. *Analytical investigation of stratified isotropic media* // Journal of the Optical Society of America A. – 2005. – Vol. 22(4). – P. 689–696.
25. WANASINGHE T., GOSINE R., DE SILVA O. et al. *Unmanned aerial systems for the oil and gas industry: overview, applications and challenges* // IEEE Access. – 2020. – Vol. 8. – P. 166980–166997.
26. YANG M., ZHANG B., XIN X. et al. *Microplastic and oil pollution in oceans: Interactions and environmental impacts* // Science of the Total Environment. – 2022. – Vol. 838. – P. 156142.

ALGORITHM FOR IDENTIFICATION OF WATER SURFACE POLLUTION FOR GEOINFORMATION SYSTEMS

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

Tatiana Gladkikh, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, Doctor of Science, Research Associate (golnikt@yandex.ru).

Abstract: The article considers the urgent problem of detection of oil and oil products on the water surface, which is of crucial importance for ecological monitoring of water resources. The growth of water bodies pollution with oil products necessitates the development of effective methods for their detection and assessment of the ecosystem state. This paper presents a new analytical technique based on the study of the spectra of reflected sunlight, which allows for the high-precision detection of oil pollution. The authors obtained analytical expressions for the indices of spectral brightness coefficients, which were subsequently used in geographic information systems (GIS). These indicators help in processing and interpreting data obtained by remote sensing, significantly simplifying the process of monitoring and analyzing the state of water

bodies. The method has proven its effectiveness in various conditions and can be adapted for use in real time. The research results contribute to the improvement of environmental monitoring methods and provide more reliable protection of water resources from pollution. In addition, the study considers the potential integration of this method with other environmental monitoring systems to improve the accuracy and efficiency of pollution detection. Thanks to these improvements, the proposed methodology not only allows for the detection of oil spills, but also helps in their management and the reduction of negative consequences associated with oil spills.

Keywords: petroleum products, environmental monitoring, multispectral images, spectrometry, information processing, geoinformation system.

УДК 504.064.36

ББК 26.192

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

Поступила в редакцию 31.10.2024.

Опубликована 31.05.2025.

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ГИБРИДНЫХ ЭНЕРГЕТИЧЕСКИХ СИСТЕМ С ПОМОЩЬЮ ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ¹

Фёдорова Е. А.², Тюленина Я. А.³, Мелошников Н. П.⁴
(Национальный исследовательский Томский государственный
университет, Томск)

Предлагается новый математический подход к оценке эффективности гибридных электроэнергетических установок, использующих альтернативный источник энергии совместно с центральной сетью электроснабжения, основанный на методах теории массового обслуживания и позволяющий учитывать стохастический характер природных процессов. В качестве математической модели гибридной энергетической системы, использующей альтернативную (непостоянную) энергию как один из источников электроэнергии, предлагается система массового обслуживания с двумя блоками приборов, один из которых имеет случайное время работы и ограниченную мощность. Исследуется трехмерный марковский случайный процесс. С помощью метода начальных моментов получены формулы стационарных характеристик системы. Выполнен численный анализ зависимости характеристик системы от входных параметров, а также представлен пример расчёта эффективности использования источника солнечной электроэнергии для бытовых нужд на основе реальных данных.

Ключевые слова: математическое моделирование, гибридные энергетические системы, теория массового обслуживания.

1. Введение

Переход к возобновляемым источникам энергии является одним из приоритетных направлений развития страны и мирового сообщества. Альтернативная энергетика (солнечная, ветровая, гидроэнергетика) способствует экологической устойчивости региона, позволяет избежать зависимости бюджета от объемов до-

¹ Работа выполнена при финансовой поддержке РФФ, грант №24-71-00022, <https://rscf.ru/project/24-71-00022/>.

² Екатерина Александровна Фёдорова, к.ф.-м.н., доцент (ekat_fedorova@mail.ru).

³ Яна Андреевна Тюленина, м.н.с. (yancch@yandex.ru).

⁴ Наталья Павловна Мелошниковна, м.н.с. (meloshnikovana@gmail.com).

бычи полезных ископаемых и их цены на мировом рынке, кроме того, исследования в этой области могут стимулировать инновации и развитие новых технологий в сфере энергетики, привлекая внимание инвесторов и ускоряя внедрение новых энергетических решений [4]. Однако при очевидных достоинствах на сегодняшний день такая энергетика не находит всеобщего распространения. Так, в европейских странах чистая энергетика в среднем занимает 30% всей вырабатываемой энергии, а в России доля возобновляемых источников составляет менее 1% [11]. Основной причиной этого является меньшая вырабатываемая мощность (по сравнению с традиционными источниками электроэнергии) и, конечно, сильная зависимость от климатических условий [12]. Компромиссным решением является внедрение гибридных энергетических систем, сочетающих несколько источников энергии (например, центральную сеть электроснабжения и собственные солнечные панели) [8].

Теоретических работ, посвященных анализу гибридных энергетических систем, не так много [15, 17, 19], в основном это экономические расчеты инвестиционных проектов перехода на альтернативную энергетику [9, 20] или расчеты технических возможностей проектируемой энергосистемы, базирующиеся на уравнениях математической физики и электротехнике [2, 5, 6, 13]. В большинстве работ используются среднестатистические значения, не учитывается стохастический характер природных явлений и объема потребляемой электроэнергии в течение периода наблюдения [16, 18]. Это может приводить к достаточно большой погрешности в полученных оценках показателей проектируемой станции. Имитационное моделирование также применяется для расчета характеристик конкретных энергетических систем [3].

Методы стохастического моделирования, в частности теория массового обслуживания (ТМО), используются для энергетических систем крайне редко [14, 21]. Это объясняется высокой степенью детерминированности традиционных энергетических объектов, что не характерно для альтернативной энергетики, поэтому стохастическое моделирование здесь оправдано.

Теория массового обслуживания – это научная дисциплина, являющаяся разделом теории вероятностей и занимающаяся анализом процессов в системах обслуживания, в которых однородные действия (события, задачи) повторяются многократно в одних и тех же условиях [7]. В исследуемых гибридных энергетических системах такими действиями являются запросы на электроэнергию от потребителей. Наиболее близкое по математическому подходу исследование представлено в статье [1], где авторы применяют аппарат теории массового обслуживания для анализа гибридных сетей связи, сочетающих радиоканал с оптическим и таким образом обеспечивающих более высокую скорость передачи информации и надежность системы.

В рамках данной работы предлагается применить аппарат теории массового обслуживания к моделированию гибридных энергетических систем, использующих источник альтернативной энергии в дополнение к центральному энергоснабжению. В разделе 2 описывается рассматриваемая техническая система и формулируется постановка задачи. В разделе 3 представлено подробное описание предлагаемой математической модели в терминах теории массового обслуживания. В разделе 4 выполняется исследование математической модели методом начальных моментов. В разделах 5 и 6 представлены численные примеры расчетов характеристик системы, анализ их зависимости от входных параметров и оценка экономической эффективности использования альтернативного источника энергии.

2. Постановка задачи

Для демонстрации предлагаемого подхода рассмотрим следующую систему: некоторый объект (производственный цех, «умный дом» и т.п.) потребляет электрическую энергию от двух источников: центральной сети и собственной установки альтернативной энергии (например, солнечные панели). Альтернативная энергия не вырабатывается постоянно, а имеет стохастический характер, зависящий от природных условий конкретной местности, и используется для дополнительных целей, напрямую

не влияющих на производство, которые могут быть реализованы в отложенный момент времени (например, заряд аккумуляторных устройств). Ставится задача оценки экономической эффективности установки альтернативной энергии в такой системе. На решение об использовании дополнительного источника электроэнергии влияют ряд факторов, главным образом, климатические особенности местности нахождения объекта и объемы потребления электроэнергии. Аппарат математического моделирования позволит нам выявить характер зависимости конечной стоимости электроэнергии от этих факторов, а также сделать выводы о наиболее оптимальных характеристиках оборудования альтернативного источника.

3. Математическая модель

Для того чтобы учитывать стохастический характер природных условий и объем потребления в течение суток, мы предлагаем использовать аппарат теории случайных процессов, а именно теорию массового обслуживания.

Предлагается следующая модель. В систему поступает простейший поток заявок (запросов на электроэнергию) с интенсивностью λ . Имеется два блока обслуживания: в первом блоке (альтернативной электроэнергии) ограниченное число приборов, равное K . Во втором блоке (городской электроэнергии) количество приборов не ограничено. Данные блоки могут работать совместно, второй блок работает всегда, а первый включается и выключается с некоторой периодичностью. Время работы первого блока распределено по экспоненциальному закону с параметром для нахождения во включенном состоянии γ_1 , а в выключенном – γ_0 . Когда накопленная энергия в первом блоке заканчивается, использующий её блок выключается и остается работать только второй блок с городской электрической сетью. При поступлении заявки в систему приоритетным считается обслуживание на первом блоке, если он включен и не переполнен. Если обслуживание на первом блоке невозможно, то заявка начинает обслуживание на втором блоке. Время обслуживания заявки в блоках распре-

делено по экспоненциальному закону с параметром для первого блока μ_1 , а для второго – μ_2 . При включении и выключении первого блока заявки не переходят с одного блока на другой, а продолжают обслуживание во время нового периода включения. Схематически предлагаемая модель представлена на рис. 1.

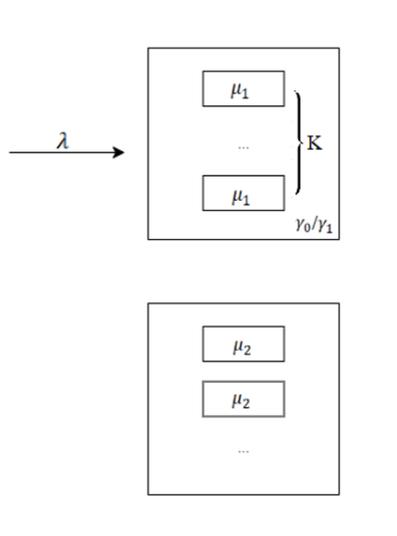


Рис. 1. Математическая модель

Обозначим случайные процессы: $i_1(t)$ – число заявок в первом блоке в момент времени t , $0 \leq i_1 \leq K$; $i_2(t)$ – число заявок во втором блоке в момент времени t , $i_2 \geq 0$; $k(t)$ – состояние первого блока, $k(t) = \{0, 1\}$, где 1 – первый блок включен, 0 – первый блок выключен.

Введём следующие обозначения для вероятностей: $P(k, i_1, i_2, t)$ – вероятность того, что в момент времени t система находится в состоянии k и в ней находится i_1 заявок на первом блоке и i_2 на втором.

Трёхмерный случайный процесс $\{k(t), i_1(t), i_2(t)\}$ является марковским, запишем систему дифференциальных уравнений

Колмогорова для распределения $P(k, i_1, i_2, t)$:

$$(1) \left\{ \begin{array}{l} \frac{\partial P(0, i_1, i_2, t)}{\partial t} = -P(0, i_1, i_2, t)(\lambda + i_2\mu_2 + \gamma_0) + \\ \quad + P(1, i_1, i_2, t)\gamma_1 + P(0, i_1, i_2 + 1, t)(i_2 + 1)\mu_2 + \\ \quad + P(0, i_1, i_2 - 1, t)\lambda, \quad i_1 < K, i_2 \geq 0, \\ \frac{\partial P(1, i_1, i_2, t)}{\partial t} = -P(1, i_1, i_2, t)(\lambda + i_1\mu_1 + i_2\mu_2 + \gamma_1) + \\ \quad + P(0, i_1, i_2, t)\gamma_0 + P(1, i_1, i_2 + 1, t)(i_2 + 1)\mu_2 + \\ \quad + P(1, i_1 + 1, i_2, t)(i_1 + 1)\mu_1 + \\ \quad + P(1, i_1 - 1, i_2, t)\lambda, \quad i_1 < K, i_2 \geq 0, \\ \frac{\partial P(0, K, i_2, t)}{\partial t} = -P(0, K, i_2, t)(\lambda + i_2\mu_2 + \gamma_0) + \\ \quad + P(1, K, i_2, t)\gamma_1 + P(0, K, i_2 + 1, t)(i_2 + 1)\mu_2 + \\ \quad + P(0, K, i_2 - 1, t)\lambda, \quad i_2 \geq 0, \\ \frac{\partial P(1, K, i_2, t)}{\partial t} = -P(1, K, i_2, t)(\lambda + K\mu_1 + i_2\mu_2 + \gamma_1) + \\ \quad + P(0, K, i_2, t)\gamma_0 + P(1, K, i_2 + 1, t)(i_2 + 1)\mu_2 + \\ \quad + P(1, K - 1, i_2, t)\lambda + P(1, K, i_2 - 1, t)\lambda, \quad i_2 \geq 0. \end{array} \right.$$

Запишем систему уравнений для стационарных вероятностей $P(k, i_1, i_2)$:

$$(2) \left\{ \begin{array}{l} -P(0, i_1, i_2)(\lambda + i_2\mu_2 + \gamma_0) + P(1, i_1, i_2)\gamma_1 + \\ \quad + P(0, i_1, i_2 + 1)(i_2 + 1)\mu_2 + P(0, i_1, i_2 - 1)\lambda = 0, \\ -P(1, i_1, i_2)(\lambda + i_1\mu_1 + i_2\mu_2 + \gamma_1) + P(0, i_1, i_2)\gamma_0 + \\ \quad + P(1, i_1, i_2 + 1)(i_2 + 1)\mu_2 + P(1, i_1 + 1, i_2)(i_1 + 1)\mu_1 + \\ \quad + P(1, i_1 - 1, i_2)\lambda = 0, \\ -P(0, K, i_2)(\lambda + i_2\mu_2 + \gamma_0) + P(1, K, i_2)\gamma_1 + \\ \quad + P(0, K, i_2 + 1)(i_2 + 1)\mu_2 + P(0, K, i_2 - 1)\lambda = 0, \\ -P(1, K, i_2)(\lambda + K\mu_1 + i_2\mu_2 + \gamma_1) + P(0, K, i_2)\gamma_0 + \\ \quad + P(1, K, i_2 + 1)(i_2 + 1)\mu_2 + P(1, K - 1, i_2)\lambda + \\ \quad + P(1, K, i_2 - 1)\lambda = 0. \end{array} \right.$$

4. Метод моментов

Для нахождения стационарных характеристик исследуемого процесса применим метод начальных моментов [7, 10] к системе уравнений (2).

Введём характеристические функции следующего вида:

$$H_k(i_1, u) = \sum_{i_2=0}^{\infty} e^{ju i_2} P(k, i_1, i_2), \quad j = \sqrt{-1}.$$

Тогда из системы уравнений (2), можно получить следующую систему дифференциальных уравнений для частичных характеристических функций:

$$(3) \left\{ \begin{array}{l} -H_0(i_1, u)(\lambda + \gamma_0) - \frac{1}{j}H'_0(i_1, u)\mu_2 + H_1(i_1, u)\gamma_1 + \\ + \frac{1}{j}H'_0(i_1, u)\mu_2e^{-ju} + H_0(i_1, u)\lambda e^{ju} = 0, \quad i_1 < K, \\ -H_1(i_1, u)(\lambda + \gamma_1) - H_1(i_1, u)i_1\mu_1 - \frac{1}{j}H'_1(i_1, u)\mu_2 + \\ + H_0(i_1, u)\gamma_0 + \frac{1}{j}H'_1(i_1, u)\mu_2e^{-ju} + \\ + H_1(i_1 + 1, u)(i_1 + 1)\mu_1 + H_1(i_1 - 1, u)\lambda = 0, \quad i_1 < K, \\ -H_0(K, u)(\lambda + \gamma_0) - \frac{1}{j}H'_0(K, u)\mu_2 + H_1(K, u)\gamma_1 + \\ + \frac{1}{j}H'_0(K, u)\mu_2e^{-ju} + H_0(K, u)\lambda e^{ju} = 0, \\ -H_1(K, u)(\lambda + \gamma_1) - H_1(K, u)K\mu_1 - \frac{1}{j}H'_1(K, u)\mu_2 + \\ + H_0(K, u)\gamma_0 + \frac{1}{j}H'_1(K, u)\mu_2e^{-ju} + H_1(K - 1, u)\lambda + \\ + H_1(K, u)\lambda e^{ju} = 0, \end{array} \right.$$

где $H'_k(i_1, u) = \frac{\partial H_k(i_1, u)}{\partial u}$.

Обозначим стационарные вероятности состояний первого блока (включен или выключен и сколько на нем заявок):

$$R_k(i_1) = H_k(i_1, 0).$$

Тогда вероятности того, что первый блок включен или выключен, определяются как

$$R_k = \sum_{i_1=0}^K H_k(i_1, 0) = \sum_{i_1=0}^K \sum_{i_2=0}^{\infty} P(k, i_1, i_2), \quad k = 0, 1.$$

Полагая в системе (3) $u = 0$ и добавляя условия нормировки, имеем следующие уравнения:

$$(4) \left\{ \begin{array}{l} -R_0(i_1)\gamma_0 + R_1(i_1)\gamma_1 = 0, \quad i_1 < K, \\ -R_0(i_1)(\lambda + \gamma_1 + i_1\mu_1) + R_0(i_1)\gamma_0 + \\ + R_1(i_1 + 1)(i_1 + 1)\mu_1 + R_1(i_1 - 1)\lambda = 0, \quad i_1 < K, \\ -R_0(K)\gamma_0 + R_1(K)\gamma_1 = 0, \\ -R_1(K)(\gamma_1 + K\mu_1) + R_0(K)\gamma_0 + R_1(K - 1)\lambda = 0, \\ \sum_{i_1=0}^K (R_0(i_1) + R_1(i_1)) = 1. \end{array} \right.$$

Откуда нетрудно получить

$$(5) \quad R_1 = \sum_{i_1=0}^K R_1(i_1) = \frac{\gamma_0}{\gamma_1 + \gamma_0}, \quad R_0 = \sum_{i_1=0}^K R_0(i_1) = \frac{\gamma_1}{\gamma_1 + \gamma_0},$$

а значения $R_k(i_1)$ вычисляются через соотношения

$$(6) \quad R_0(i_1) = R_1(i_1) \frac{\gamma_1}{\gamma_0},$$

$$R_1(i_1) = R_1(K) \left(\frac{\mu_1}{\lambda} \right)^{K-i_1} \prod_{m=0}^{K-i_1-1} (K-m),$$

где $R_1(K)$ находится из условия нормировки:

$$(7) \quad R_1(K) = \frac{\gamma_0}{\gamma_0 + \gamma_1} \left(1 + \sum_{i=0}^{K-1} \left(\frac{\mu_1}{\lambda} \right)^{K-i} \prod_{m=0}^{K-i-1} (K-m) \right)^{-1}.$$

Из свойств характеристической функции:

$$\left. \frac{\partial H_0(i_1, u)}{\partial u} \right|_{u=0} = jm_0(i_1), \quad \left. \frac{\partial H_1(i_1, u)}{\partial u} \right|_{u=0} = jm_1(i_1),$$

где $m_0(i_1)$ и $m_1(i_1)$ – частичные моменты первого порядка, т.е. математические ожидания числа заявок на втором блоке при нахождении первого блока в состоянии k и i_1 заявок на нем.

Дифференцируя систему уравнений (3) по u и принимая $u = 0$, имеем следующую систему уравнений:

$$(8) \quad \begin{cases} -m_0(i_1)(\gamma_0 + \mu_2) + m_1(i_1)\gamma_1 + R_0(i_1)\lambda = 0, \\ -m_1(i_1)(\lambda + \gamma_1 + i_1\mu_1 + \mu_2) + m_0(i_1)\gamma_0 + \\ \quad + m_1(i_1 + 1)(i_1 + 1)\mu_1 + m_1(i_1 - 1)\lambda = 0, \\ -m_0(K)(\gamma_0 + \mu_2) + m_1(K)\gamma_1 + R_0(K)\lambda = 0, \\ -m_1(K)(\gamma_1 + K\mu_1 + \mu_2) + m_0(K)\gamma_0 + \\ \quad + m_1(K - 1)\lambda + R_1(K)\lambda = 0. \end{cases}$$

Путем суммирования соответствующих уравнений по всем $i_1 = \overline{0, K}$ получаем формулы для частичных первых моментов числа заявок во втором блоке:

$$(9) \quad m_1 = \sum_{i_1=0}^K m_1(i_1) = \frac{\gamma_0 \lambda R_0 + R_1(K) \lambda (\gamma_0 + \mu_2)}{(\gamma_1 + \mu_2)(\gamma_0 + \mu_2) - \gamma_0 \gamma_1},$$

$$m_0 = \sum_{i_1=0}^K m_0(i_1) = \frac{\gamma_1 m_1 + R_0 \lambda}{\mu_2 + \gamma_0}.$$

Таким образом, математическое ожидание числа заявок во втором блоке вычисляется по формуле

$$(10) \quad m = (R_0 + R_1(K)) \cdot \lambda / \mu_2,$$

т.е. оно равно отношению интенсивности входящего потока на интенсивность обслуживания на втором блоке, умноженной на вероятность того, что первый блок не доступен (по любой из причин).

Применяя дальше метод моментов к системе (3), можно при желании получить формулы для частичных моментов более высоких порядков.

5. Численный анализ

Выполним численный анализ чувствительности характеристик системы к входным параметрам. Рассмотрим следующий пример. Пусть

$$K = 3, \mu_1 = 0,5, \mu_2 = 1, \gamma_1 = 0,4, \gamma_0 = 0,6.$$

Продемонстрируем зависимость полученных характеристик от интенсивности входящего потока λ . На рис. 2 представлены значения частичных первых моментов числа заявок на втором блоке при включенном ($m1$) и выключенном ($m0$) первом блоке приборов для различных значений λ .

На рис. 3 представлена зависимость среднего числа заявок на первом блоке для разных значений интенсивности входящего

потока, вычисляемая как
$$M = \sum_{k=0}^K k \cdot (R_0(k) + R_1(k)).$$

Очевидно, что с ростом числа поступающих заявок в единицу времени растет и среднее число заявок, находящихся в системе, при этом для процесса $i_2(t)$ этот рост линейный, тогда как для процесса $i_1(t)$ имеет нелинейный характер. Рассмотрим процесс $i_1(t)$ подробнее. На рис. 4 представлены значения стационарных вероятностей числа заявок на первом блоке $R(i_1(t)) = P\{i_1(t) = i_1\}$ при различных λ . В рассматриваемом примере $i_1(t)$ принимает целые значения $0 \leq i_1 \leq 3$.

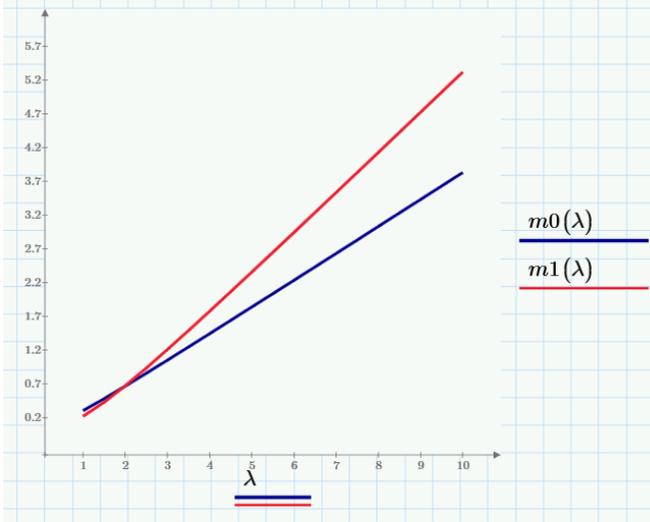


Рис. 2. Математическое ожидание числа заявок на втором блоке при различных λ

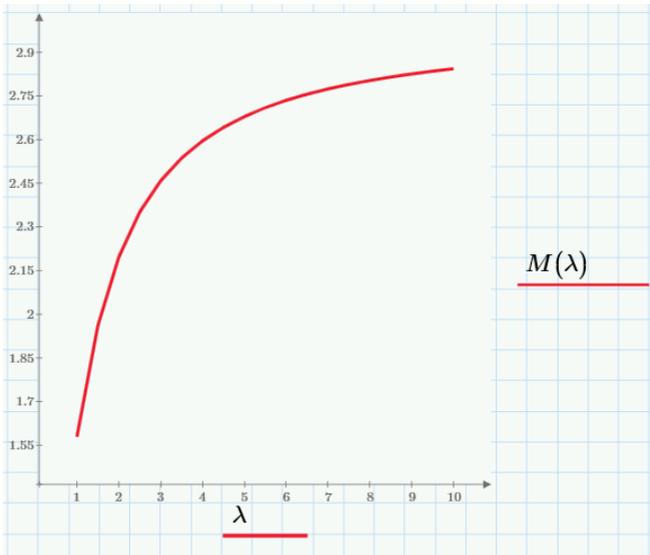


Рис. 3. Изменение среднего числа заявок на первом блоке от интенсивности входящего потока

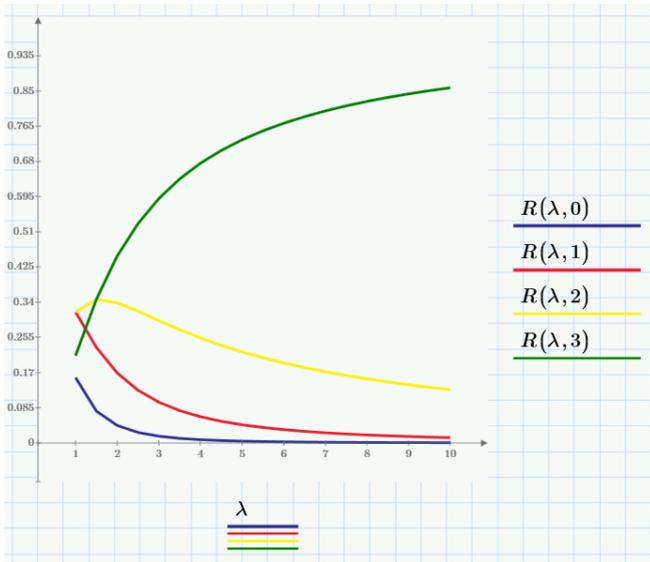


Рис. 4. Стационарные вероятности числа заявок на первом блоке при различных λ

Из графика можно сделать вывод о том, что при увеличении интенсивности входящего потока заявок значительно увеличивается вероятность $P(i_1(t)) = K$, т.е. того, что все обслуживающие приборы на первом блоке будут заняты. Поэтому и математическое ожидание числа заявок при увеличении интенсивности входящего потока стремится к числу количества приборов на первом блоке K .

Двумерное стационарное распределение вероятностей числа заявок на первом блоке и состояния самого блока для фиксированного значения $\lambda = 2$ представлено в таблице 1.

Выполним анализ зависимости параметров системы от времени работы первого блока. Введем параметр среднего времени работы первого блока t в часах ($t = \overline{1, 16}$), где 16 часов – максимальная долгота дня на территории России. Тогда интенсивности включения и выключения первого блока определяются как

$$\gamma_1 = 1/t, \quad \gamma_0 = 1/(24 - t).$$

Таблица 1. Распределение вероятностей состояний первого блока при $\lambda = 2$

	$i_1 = 0$	$i_1 = 1$	$i_1 = 2$	$i_1 = 3$
$k = 0$	0,017	0,068	0,135	0,180
$k = 1$	0,025	0,101	0,203	0,270

В таблице 2 представлены значения стационарных вероятностей включенного и выключенного состояния первого блока для разных t .

Таблица 2. Значения вероятностей состояния первого блока для разных t

	$t = 1$	$t = 2$	$t = 5$	$t = 10$
R_0	0,958	0,917	0,792	0,583
R_1	0,042	0,083	0,208	0,417

На рис. 5 представлено изменение значений частичных первых моментов числа заявок на втором блоке при включенном (m_1) и выключенном (m_0) первом блоке приборов в зависимости от t . Среднее же значение числа заявок на первом блоке на всем интервале принимает постоянное значение $M = 2,923$.

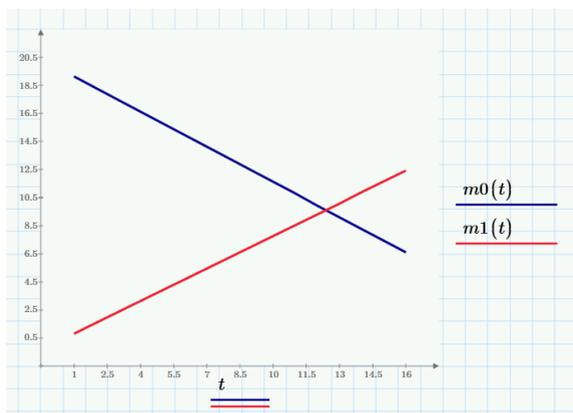


Рис. 5. Математическое ожидание числа заявок на втором блоке при различных t

6. Расчет эффективности

Решим задачу оценки экономической эффективности установки солнечной панели для бытовых нужд на примере г. Москвы. Введем функцию затрат на электроэнергию на основе предлагаемой модели в следующем виде:

$$L_1 = c_1(R_0m_0 + R_1m_1) + c_2,$$

где c_1 – стоимость электроэнергии из городской сети за кВт·ч, c_2 – «стоимость» производимой альтернативной электроэнергии.

Средняя стоимость электроэнергии в Москве составляет $c_1 = 7,33$ руб./кВт·ч. Стоимость альтернативной электроэнергии c_2 за кВт·ч будем рассчитывать как дисконтированную на весь период использования полную стоимость устанавливаемого оборудования и его обслуживания. Бытовые солнечные станции бывают различных характеристик, для примера выберем сетевую солнечную электростанцию СЗ-ДН мощностью 5 кВт. Стоимость такой солнечной панели составляет примерно 46 4120 руб., а срок службы – 25 лет. Дисконтированная полная стоимость установки с учетом ставки дисконтирования (например, 8%) составит $cost = 1\,168\,464$ руб. Примем расходы на обслуживание этой станции в размере $\Delta = 5\%$ от стоимости в год. Тогда

$$c_2 = \frac{cost \cdot (1 + \Delta)}{T} \approx 5,6 \text{ руб./кВт·ч, где } T = 25 \cdot 365 \cdot 24 \text{ ч.}$$

Средняя длительность работы солнечной панели на территории г. Москвы в сутки составляет в зимние месяцы 4 часа, а в летний период – 8 часов [12]. Тогда параметры работы первого блока зададим следующими: $\gamma_0 = \frac{1}{20}$, $\gamma_1 = \frac{1}{4}$ и $\gamma_0 = \frac{1}{16}$, $\gamma_1 = \frac{1}{8}$ в зимний и летний период соответственно. Параметры обслуживания примем $\mu_1 = \mu_2 = 1$, а $K = 5$.

Параметр входящего потока в модели λ имеет физический смысл объема потребления электроэнергии в единицу времени (т.е. среднему значению запрашиваемой мощности). Для среднестатистического потребителя он составляет около $\lambda = 1,33$ кВт. Стоит отметить, что весь объем потребления можно разложить на постоянную (постоянно работающее электрооборудование, например, охранная и противопожарная системы) и переменную

часть (например, освещение). В рамках данного исследования анализируется переменная составляющая.

Для оценки эффективности описанной гибридной энергетической системы будем сравнивать затраты L_1 с затратами в системе без использования альтернативной электроэнергии. Математической моделью негибридной системы является классическая бесконечнолинейная система массового обслуживания $M|M|\infty$ с известным математическим ожиданием числа заявок в системе $M = \lambda/\mu_2$ и функцией затрат

$$L_2 = c_1 M.$$

На рис. 6, 7 представлено сравнение значений L_1 и L_2 в зависимости от параметра λ в летний и зимний периоды.

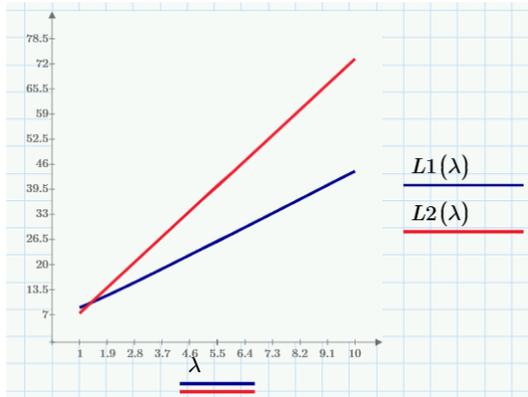


Рис. 6. Зависимость затрат от потребления в летний период

По точкам пересечения прямых L_1 и L_2 можно сделать вывод, при каком объеме потребляемой электроэнергии в час установка солнечной панели будет иметь положительный экономический эффект. В рассмотренном примере такими значениями является $\lambda^* = 1,374$ кВт для летнего периода (рис. 6) и $\lambda^* = 2,439$ кВт для зимнего (рис. 7).

Заметим, что в приведённых примерах был сделан упор на анализ зависимости конечной стоимости электроэнергии в гибридной системе от объемов потребления электроэнергии в еди-

ницу времени (т.е. от λ в обозначениях математической модели). Однако результаты моделирования можно также использовать для анализа экономической эффективности гибридной системы в зависимости от климатических условий (т.е. среднего времени работы установки альтернативной энергии γ_1) или выбора оптимальных характеристик приобретаемого оборудования – заявленной мощности среднесуточной выработки (задавая значения K и μ_1), срока службы T и др.

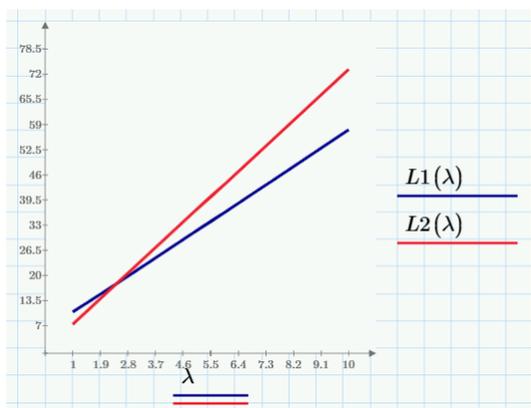


Рис. 7. Зависимость затрат от потребления в зимний период

7. Заключение

В работе была предложена математическая модель гибридной энергетической системы в виде системы массового обслуживания с дополнительным непостоянным блоком приборов. С помощью метода начальных моментов найдены стационарные вероятности состояний непостоянного блока и среднее число заявок на втором блоке. Проведён численный анализ результатов моделирования. Представлен пример оценки экономической эффективности гибридной энергетической системы при использовании источника солнечной энергии для бытовых нужд на среднестатистических данных в выбранной местности.

Результаты аналитического исследования могут быть применены для анализа гибридных энергетических систем различных конфигураций и географических особенностей местности путем задания нужных значений параметров модели, соответствующих техническим и климатическим данным.

Отметим, что данная работа представляет первый этап исследования. В будущем планируется учесть, что объем потребления электроэнергии имеет неэкспоненциальный вид распределения. В частности, интенсивности запросов в разное время суток имеют разные значения, что в дальнейшем предполагается описать входящим ММРР-потокм заявок.

Литература

1. ВИШНЕВСКИЙ В.М., СЕМЕНОВА О.В., ШАРОВ С.Ю. *Моделирование и анализ гибридного канала связи на базе лазерной и радио технологий* // Управление большими системами. – 2011. – Вып. 35. – С. 237–249.
2. ГРИГОРАШ О.В. КРИВОШЕЙ А.А., СМЫК В.В. *Автономные гибридные электростанции* // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. – 2016. – №124(10). – С. 1–12.
3. ГОРДИЕВСКИЙ Е.М., МИРОШНИЧЕНКО А.А., КУЛГАНАТОВ А.З. и др. *Разработка имитационной модели мобильного энергокомплекса на базе ВИЭ в программе MatLab* // Вестник ПНИПУ. Электротехника, информационные технологии, системы управления. – 2019. – №31. – С. 51–71.
4. ИВАНОВА А.Ю. *Альтернативные источники в энергетике: виды и принципы функционирования* // Int. Scientific Review. – 2016. – С. 29–32.
5. МАРЧЕНКО О.В. *Математическая модель энергосистемы с возобновляемыми источниками энергии* // Известия РАН. Энергетика. – 2006. – №3. – С. 154–161.

6. МИТРОФАНОВ С.В., ПЕРЕПЕЛКИН К.А. *Математическое моделирование гибридной ветро-солнечной станции для электроснабжения собственных нужд* // Вестник ЮУрГУ. Серия «Энергетика». – 2022. – №22(3). – С. 18–26.
7. НАЗАРОВ А.А., ТЕРПУГОВ А.Ф. *Теория массового обслуживания*. – Томск: НТЛ, 2004. – 228 с.
8. НОВОКРЕЩЕНОВ О.В., ОТМАХОВ Г.С., ХУАДЕ М.Ю. *Комбинированные системы электроснабжения на возобновляемых источниках энергии* // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. – 2017. – №132(08). – С. 1–12.
9. РЕЗИНЬКОВ А.А., ЩЕКЛЕИН С.Е. *Сравнительный анализ эффективности гибридных (солнечно-дизельных) электростанций для регионов России* // Энерго- и ресурсосбережение. Энергообеспечение. Нетрадиционные и возобновляемые источники энергии. Атомная энергетика : сборник научных трудов. – Екатеринбург : Издательство Уральского университета. – 2021. – С. 385–388.
10. СИНЯКОВА И.А., МОИСЕЕВА С.П. *Метод моментов для исследования математической модели параллельного обслуживания кратных заявок потока Марковского восстановления* // Известия Томского политехнического университета. – 2012. – №321(5). – С. 24–28.
11. *Enerdata* [Электронный ресурс]. – Режим доступа: <https://yearbook.enerdata.net/renewables/wind-solar-share-electricity-production.html>.
12. *Global Solar Atlas* [Электронный ресурс]. – Режим доступа: <https://globalsolaratlas.info/map>.
13. HERBST A., TORO F., REITZE F., JOCHEM E. *Introduction to Energy Systems Modelling* // Swiss Journal of Economics and Statistics. – 2012. – Vol. 148(2). –P. 111–135.

14. FERRETTI I., CAMPARADA M., ZAVANELLA L.E. *Queuing Theory-Based Design Methods for the Definition of Power Requirements in Manufacturing Systems* // Energies. – 2022. – No. 15. – 7621.
15. KING M., DACHENG LI., DOONER M. et al. *Mathematical Modelling of a System for Solar PV Efficiency Improvement Using Compressed Air for Panel Cleaning and Cooling* // Energies. – 2021. – No. 14.
16. KUMAR A., SOOD Y., RAM N.M. *Hybrid Energy System Simulation and Modelling Incorporating Wind and Solar Power* // IEEE Int. Students' Conf. on Electrical, Electronics and Computer Science (SCEECS-2024). – 2024.
17. MOHAMMED I., ABUASHOUR M.W., THA'ER O.S. et al. *Modelling, simulations and operational performance of a stand-alone hybrid wind/PV energy system supplying induction motor for pumping applications* // Int. Journal of Engineering Systems Modelling and Simulation (IJESMS). – 2018. – Vol.10, No. 1. – P. 12–25.
18. PERVEEN G., RIZWAN M., GOEL N. *Comparison of intelligent modelling techniques for forecasting solar energy and its application in solar PV based energy system* // IET Energy Systems Integration. – 2019. – No. 11. – P. 34–51.
19. TYAGI S., GOYAL N., KUMAR A. et al. *Stochastic hybrid energy system modelling with component failure and repair* // Int. Journal of System Assurance Engineering and Management. – 2021. – Vol. 13. – P. 842–852.
20. VARSHNEY S., SHEKHAR C. *Optimal management strategies of renewable energy systems with hyperexponential service provisioning: an economic investigation* // Frontiers. Energy Research. – 2023. – No. 11. – P. 1–22.
21. WATERMEYER M., THOMAS M., GROTHE O. et al. *A hybrid model for day-ahead electricity price forecasting: Combining fundamental and stochastic modelling* // Engineering, Economics. – 2023. – P. 1–38.

MATHEMATICAL MODELING OF HYBRID ENERGY SYSTEMS USING QUEUEING THEORY

Ekaterina Fedorova, National Research Tomsk State University, Tomsk, Associate Professor, (ekat_fedorova@mail.ru).

Yana Tyulenina, National Research Tomsk State University, Tomsk, Junior Researcher (yancch@yandex.ru).

Natalia Meloshnikova, National Research Tomsk State University, Tomsk, Junior Researcher (meloshnikovana@gmail.com).

Abstract: In the paper, a new mathematical approach for assessing the efficiency of hybrid energy system, using alternative energy sources (solar panels, wind turbines) based on queueing theory is proposed, which allows to consider into account the stochastic nature of natural processes. As a mathematical model of a hybrid energy system using an alternative (non-permanent) energy as one of the source of electricity, we propose a queueing system with two units of servers, one of which is non-permanent and limited. A three-dimensional Markov random process is studied. Using the initial moments method, formulas for stationary characteristics are derived. A numerical analysis of the system characteristics dependency on input parameters is performed. An example of assessment of the efficiency of using solar energy for personal needs based on statistical data is presented.

Keywords: mathematical modeling, hybrid energy systems, queueing theory.

УДК 519.872

ББК 31.6

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

Поступила в редакцию 27.01.2025.

Дата опубликования 31.05.2025.

РАЗРАБОТКА АДАПТИВНОЙ СИСТЕМЫ УПРАВЛЕНИЯ СВЕТОФОРНЫМИ ОБЪЕКТАМИ С ИСПОЛЬЗОВАНИЕМ МАРКОВСКИХ ПРОЦЕССОВ ПРИНЯТИЯ РЕШЕНИЙ¹

Тисленко Т. И.², Семенова Д. В.³

(Сибирский федеральный университет, г. Красноярск)

Описан программный комплекс MARLIN24, предназначенный для реализации адаптивного управления светофорными объектами. Цель разработки — оптимизация работы светофоров для сокращения суммарного времени пребывания транспортных средств в зоне действия оптических детекторов. Архитектура комплекса включает три ключевых модуля: модуль адаптивного управления, модуль имитационного моделирования транспортных потоков и модуль валидации, а также дополнительный модуль визуализации. В модуле адаптивного управления реализованы четыре подхода: фиксированный план; некоординированное подкрепленное обучение; координированное мультиагентное обучение с подкреплением; управление длительностью фаз с помощью MISO контроллера нечёткой логики (Multiple Input Single Output, MISO). Имитационное моделирование транспортных потоков для проверки результатов управления осуществляется посредством модуля микросимуляции на основе модели «разумный водитель» (Intelligent Driver Model, IDM). В модуле валидации для моделирования реалистичных показаний оптических датчиков, отражающих дорожную обстановку, используются копула-функции, принимающие на вход маргинальные распределения. Маргинальные распределения были получены на основе исторических данных об интенсивности движения за период 2019–2020 гг. Разработанный комплекс программ имитационного моделирования MARLIN24 позволяет исследовать и сравнивать между собой ряд моделей управления реальными участками дорожной сети г. Красноярска при различных условиях.

Ключевые слова: имитационное моделирование транспортных потоков, модель IDM, модель MOBIL, копулы Маршалла – Олкина, машинное обучение с подкреплением.

1. Введение

Одной из задач, решаемых в ходе реализации транспортной стратегии России на 2035 год [6], является увеличение про-

¹ Работа поддержана Красноярским математическим центром, финансируемым Минобрнауки РФ (Соглашение № 075-02-2025-1790).

² Тимофей Иванович Тисленко, аспирант (timtisko@mail.ru).

³ Дарья Владиславовна Семенова, к.ф.-м.н., доцент (DVSeменова@sfu-kras.ru).

пусковой способности и скоростных параметров дорожной инфраструктуры. При этом необходимо учитывать общесоциальные целевые ориентиры транспортной стратегии: подвижность населения, снижение аварийности, рисков и угроз безопасности по видам транспорта, снижение доли транспорта в загрязнении окружающей среды. Таким образом, актуальна проблема моделирования транспортных потоков и оценки эффективности управления светофорными объектами для наиболее нагруженных участков дорожной сети.

Системы, управляющие светофорными объектами, подразделяют на те, которые корректируют сигналы светофоров в реальном времени и реагируют на текущую дорожную обстановку – АСУДД (адаптивные системы управления дорожным движением), и те, которые работают согласно фиксированному плану управления – неадаптивные. В таблице 1 представлены наиболее известные АСУДД, которые были разработаны в различные временные периоды и для различных условий движения.

Таблица 1. Распространенные модели АСУДД

Критерий	UTCS-1	SCOOT	OPAC	АСУДД «Микро»	MARLIN
Город	Вашингтон	Лондон	Арлингтон, Тускон	Красноярск	Торонто
Временной период	1970е	1995	1983,1989	1993	2010
Длительность фаз	фиксированная		переменная		
Оптимизация	офлайн	онлайн			
Предсказание	нет		есть		
Устройство	централизованная		децентрализованная		
Основные ограничения	постоянный сбор данных	сенсоры далеко	только для 8 фаз	находится в разработке	«ПРОКЛЯТИЕ РАЗМЕРНОСТИ»

Одна из ранних систем управления городским дорожным движением была UTCS-1 (Urban Traffic Control System) [9]. Система UTCS-1 использовала фиксированные сигнальные планы, которые менялись в зависимости от времени суток (утрен-

ний, дневной, вечерний планы). Для стабильной работы системы UTCS-1 необходимо регулярно вручную корректировать сигнальные планы, что является существенным недостатком. На данный момент считается устаревшей, так как более современные системы UTCS-2, UTCS-3, SCOOT позволяют гибко управлять дорожным движением в реальном времени.

Система SCOOT (Split Cycle Offset Optimization Technique) [10] анализирует данные о дорожной обстановке и корректирует светофорные сигналы, чтобы предотвратить образование заторов до их появления. Для работы данной системы требуется установка плотной сети индукционных петель, камер и других датчиков движения на расстоянии не менее сорока метров и не более двухста метров до регулируемых перекрёстков. Централизованное управление SCOOT направлено на устранение «эффекта волны». Следует заметить, что продолжительность активной фазы (зеленого сигнала) ограничена и не продлевается динамически, так как параметр Split задаёт её максимальный процент от времени в цикле светофорного объекта.

Система OPAC (Optimized Policies for Adaptive Control) [13] – это адаптивная система управления светофорами, схожая по назначению с системой SCOOT, но использующая другой подход для оптимизации транспортных потоков. Система OPAC разработана в США и предназначена для улучшения дорожной ситуации в реальном времени путем адаптации фаз светофоров в зависимости от условий трафика. В системе OPAC существует ограничение в восемь фаз для каждого светофора. Это ограничение связано с практическими соображениями, поскольку каждая фаза представляет собой отдельное направление движения или определённую комбинацию разрешённых манёвров на перекрёстке (например, движение прямо, поворот налево или направо). Эти ограничения заложены аппаратно, т.е. на уровне контроллеров, и не могут быть изменены конечным пользователем.

Автоматизированная система управления дорожным движением (АСУДД) «Микро» [1] – наиболее широко используемая в России система, успешно применяемая в следующих регионах:

Красноярский край, Иркутская область, Белгородская область, Воронежская область, Хабаровский край, Московская область. АСУДД «Микро» является децентрализованной системой и поддерживает до шести GPRS-серверов, которые позволяют подключить до 250 перекрёстков. Детекторы АСУДД «Микро» работают на расстоянии до ста метров. Основным недостатком системы является тот факт, что реализация адаптивных алгоритмов находится в стадии разработки.

В работах [4, 5, 11, 19, 20] для адаптивного управления светофорными объектами было предложено использовать метод мультиагентного обучения с подкреплением. Данный подход получил название MARLIN (Multiagent Reinforcement Learning for Integrated Network) и был успешно применен в современной АСУДД в Канаде. Целью обучения с подкреплением является сокращение времени проезда транспортных средств через выбранные участки дорожной сети. Управление светофорными объектами считается эффективным, если транспортные средства находятся на детектируемых участках менее двух циклов. Для работы требуется установка камер на расстоянии менее ста метров от стоп-линий. Агенты (светофоры) могут работать без информации о полной дорожной обстановке и управлять движением децентрализованно. Перечисленные особенности можно отнести к достоинствам подхода на основе MARLIN. Существенным ограничением является рост вычислительной сложности при увеличении обзора агента, известный как «проклятие размерности».

Целью работы является разработка программных средств для имитационного моделирования и сравнительного анализа различных показателей эффективности для моделей управления дорожным движением. Структура статьи следующая. В разделе 2 приводится общее описание работы четырех модулей комплекса MARLIN24: симуляции, визуализации, валидации и модуля адаптивного управления светофорными объектами.

В разделе 3 приводится описание работы модуля симуляции. Данный модуль предназначен для имитации показаний оптических детекторов. В параграфе 3.2 предлагается описание особен-

ностей программной реализации модуля симуляции. Включение графического интерфейса в модуле визуализации является опциональным и позволяет производить отладку применяемого управления. Данный модуль отображает транспортные средства, дорожную сеть, дополнительную информацию о содержании названий дорог, их координатах. Также в модуле реализована функция ускорения и замедления модельного времени, выделения маршрутов и масштабирования.

В разделе 4 приводится описание работы модуля валидации. В параграфе 4.1 сформулирована задача моделирования интенсивностей транспортных потоков с использованием копул Маршалла – Олкина. В параграфе 4.2 представлено описание работы модуля на примере исторических данных, полученных с оптических детекторов.

В разделе 5 приводится описание модуля управления светофорными объектами. В параграфе 5.1 проблема управления светофорными объектами сводится к задаче мультиагентного обучения с подкреплением для одного и нескольких агентов. В параграфе 5.2 приводится схема взаимодействия модуля управления светофорными объектами с модулями симуляции и валидации.

В разделе 6 представлены вычислительные эксперименты, включающие в себя моделирование перемещения транспорта по дорожной сети, моделирование интенсивности транспортных потоков, построение эффективного управления двумя светофорами.

2. *Общее описание программного комплекса MARLIN24*

Фокус наших исследований сосредоточен на разработке программных и математических инструментов для адаптивного управления сетью светофорных объектов участков дорожной сети города Красноярска. В настоящей работе представлен новый программный комплекс MARLIN24, реализующий часть методов новейших АСУДД. Промежуточные результаты по разработке комплекса и эксперименты представлены в [4, 5, 19, 20]. В статье в качестве примера приведены две модели участков дорожной

сети города Красноярска: модель двух перекрестков (пр. Свободный – ул. Лесопарковая, пр. Свободный – ул. Годенко) и модель перекрестков микрорайона Покровский. Оценка эффективности применяемого управления и валидация полученных результатов осуществлялись с помощью симуляционных экспериментов на основе статистических данных с детекторов за 2018–2019 гг.

Общая структура комплекса MARLIN24 приведена на рис. 1. Комплекс состоит из четырех модулей: симуляции, валидации, адаптивного управления светофорными объектами, визуализации. Модуль визуализации является опциональным. Его функционал и используемые компоненты будут описаны ниже.

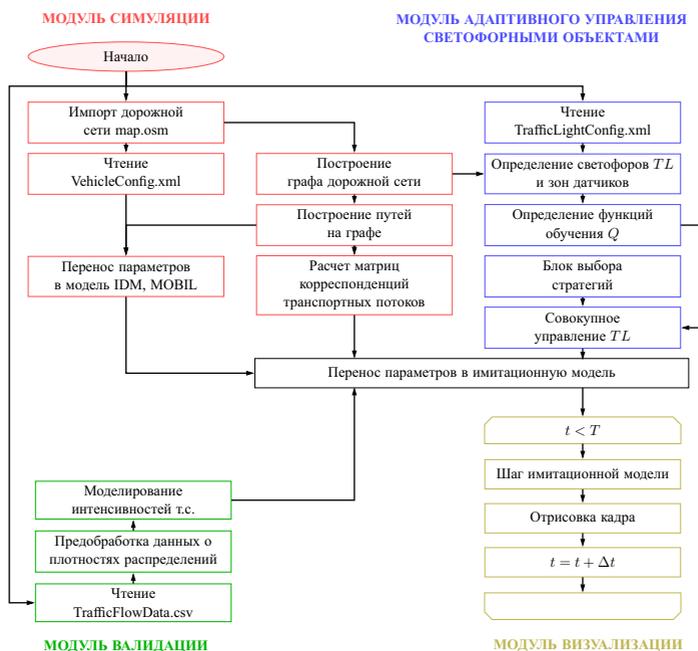


Рис. 1. Структура программного комплекса MARLIN24

3. Модуль симуляции

Модуль симуляции предназначен для оценки эффективности выбранного управления светофорными объектами. В модуле реализованы две микросимуляционные модели движения, описывающие перемещения транспортных средств по дорожной сети. Первая модель IDM (Intelligent Driver Model) реализует движение транспортных средств по однополосной прямой дороге [21]. Вторая модель – MOBIL (Microscopic Optimally Balanced Intersection Lanes) – описывает перестроение на многополосной дороге [22]. В процессе симуляции движения транспортных средств учитывается управление фазами и циклами светофорных объектов, а также предусмотрено имитирование показаний оптических датчиков и работы светофоров.

3.1. Модели движения транспортных средств

Для описания движения отдельных транспортных средств используется модель IDM [21]. Данная модель позволяет учитывать следующие параметры: минимальная безопасная дистанция, максимальная разрешенная скорость и коэффициент торможения транспортных средств. В таблице 2 представлены значения параметров для модели IDM, используемые далее в статье.

Таблица 2. Основные обозначения для модели IDM

Символ	Значение
i	номер транспортного средства
$s_{0,i}$	минимальная безопасная дистанция для т.с. i
$v_{0,i}$	максимальная желательная скорость i
δ	компонента «гладкости» ускорения
T_i	время реакции i -го водителя
a_i	ускорение i
b_i	коэффициент торможения i
s^*	желаемое расстояние между i и $i - 1$

Модель «разумный водитель» (Intelligent Driver Model, IDM) относится к классу моделей движения за лидером. Все транспортные средства рассматриваются как индивидуальные сущности, обладающие характеристиками и поведением. Водители транспортных средств регулируют скорость в зависимости от расстояния до впереди идущих и их скорости. На рис. 2 представлено взаимное расположение и характеристики текущего автомобиля, расположенного в i -й позиции, и $(i - 1)$ -го автомобиля, находящегося перед ним.

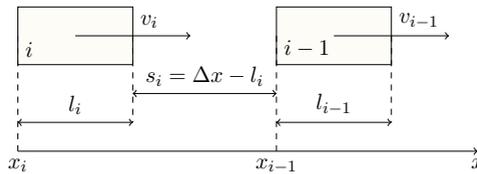


Рис. 2. Иллюстрация модели IDM

Система уравнений, описывающих текущую скорость i -го автомобиля и расстояние до $(i - 1)$ -го автомобиля в модели IDM, в классических обозначениях имеет вид

$$(1) \quad \begin{cases} \frac{dv_i}{dt} = \underbrace{a_i \left(1 - \left(\frac{v_i}{v_{0,i}} \right)^\delta \right)}_{a_{\text{free_road}}} - \underbrace{a_i \left(\frac{s^*(v_i, \Delta v_i)}{s_i} \right)^2}_{a_{\text{deceleration}}}, \\ s^*(v_i, \Delta v_i) = s_{0,i} + v_i T_i + \frac{v_i \Delta v_i}{2\sqrt{a_i b_i}}, \end{cases}$$

где $a_{\text{free_road}}$ обозначает ускорение, возникающее на свободной дороге, $a_{\text{deceleration}}$ – замедление, возникающее при сближении транспортных средств.

При имитационном моделировании для нахождения значений скорости и ускорения использовались формулы, вытекающие из численного метода «пристрелки» [3]:

$$\begin{cases} \frac{dv}{dt}(t) = a_{\text{free_road}}(t) + a_{\text{deceleration}}(t), \\ v(t + \Delta t) = v(t) + \frac{dv}{dt}(t)\Delta t, \\ x(t + \Delta t) = x(t) + v(t)\Delta t + \frac{1}{2} \frac{dv}{dt}(t)(\Delta t)^2, \\ s(t + \Delta t) = x_i(t + \Delta t) - x(t + \Delta t) - l_i. \end{cases}$$

Шаг симуляции Δt выбирается как шаг по времени при численном решении системы (1).

Существенным ограничением модели IDM является ее применимость только к однополосному движению. Одним из способов расширить ее применимость к многополосным дорожным сетям является введение алгоритмов, описывающих перестроение транспортных средств. В работе используется модель перестроения MOBIL [22].

В основе модели MOBIL лежит идея о том, что водители принимают решения о перестроении и изменении скорости движения из соображений проходимости и безопасности. Конкретное изменение полосы движения, например с правой полосы движения на левую полосу, как показано на рис. 3, зависит, как правило, от двух следующих транспортных средств на текущей полосе движения и, соответственно, на целевой полосе движения.

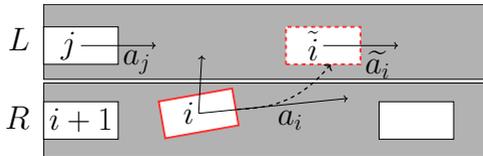


Рис. 3. Модель смены полосы MOBIL

Стимул для перестроения есть, если после первого фиктивного перестроения с правой полосы R на левую полосу L сумма собственного ускорения согласно модели IDM и ускорения соседних транспортных средств выше на порог изменения δ :

$$R \rightarrow L : (\tilde{a}_i - a_i) + m((\tilde{a}_{i+1} - a_{i+1}) + (\tilde{a}_j - a_j)) \geq \delta,$$

где $m \in [-\infty; \frac{1}{2}] \cup [1; +\infty]$ – вручную задаваемый коэффициент вежливости, символ « $\widetilde{\bullet}$ » обозначает измененные характеристики. В [22] предложены следующие интерпретации коэффициента m :

$m \geq 1$, *альтруистичное поведение*: автомобиль не перестраивается, если он ухудшит общую дорожную ситуацию;

$0 \leq m \leq 0,5$, *реалистичное поведение*: движение остальных транспортных средств менее приоритетно;

$m < 0$, *вредительское поведение*: автомобиль перестраивается, если он замедлит остальные.

Заметим, что в модели IDM скорости, а следовательно, и ускорения связаны формулой (1) и изменяются последовательно от лидирующего транспортного средства к последующему. Также следует учитывать, что если транспортное средство i перестраивается на соседнюю полосу перед транспортным средством j , то значение коэффициента торможения транспортного средства j по модулю не должно превышать значения коэффициента комфортного торможения b_j . Поэтому ограничение на ускорение \tilde{a}_j , можно выразить следующим образом: $\tilde{a}_j \geq -b_j$.

3.2. Описание работы модуля

Модуль симуляции трафика на вход получает конфигурационный файл `VehicleConfig.xml`, в котором содержатся такие параметры как максимально разрешенная скорость, коэффициент торможения (покрытие дороги), количество полос. Дополнительно в модуль симуляции поступает информация о дорожной сети в виде мультиграфа `map.osm`, предобработанная библиотекой `osmnx` [8]. Для построения маршрутов транспортных средств в дорожной сети используется библиотека `NetworkX` [17].

Каждый шаг по времени t в модуле симуляции может быть отображен в модуле визуализации с использованием библиотеки `pygame`. Модуль визуализации отрисовывает дорожную сеть, транспортные средства и отладочную информацию, а также позволяет изменять модельное время.

4. Модуль валидации

В модуле валидации решается задача имитационного моделирования интенсивности движения транспортного потока. В данной работе под интенсивностью транспортного потока будем понимать число автомобилей, проезжающих через поперечное сечение участка дорожной сети в единицу времени. В модуле предусмотрены два способа подсчета интенсивности: первый способ, учитывающий количество машин за фиксированный временной интервал Δt [4], второй – величину временного интервала между проездом двух автомобилей [5].

4.1. Моделирование интенсивностей транспортных потоков

На практике оптические датчики закреплены за соответствующей отдельной полосой и движение по полосам, как правило, является зависимым. Обозначим множество индексов детектируемых участков на полосах дорожной сети $Zones = \{1, 2, \dots, Z\}$.

Пусть случайная величина X_z с непрерывной функцией распределения $F_z(x)$ есть временной интервал между проездом двух автомобилей через сечение детектируемого участка в зоне $z \in Zones$. Тогда система зависимых случайных величин (X_1, \dots, X_Z) с совместной функцией распределения $H(x_1 \dots x_Z)$ описывает интенсивность движения транспортных потоков через детектируемые участки дорожной сети $1, 2, \dots, Z$.

По теореме Склера [16] совместная функция распределения представима копулой C

$$(2) \quad H_{X_1 \dots X_Z}(x_1, \dots, x_Z) = C(F_1(x_1), \dots, F_Z(x_Z)).$$

Для описания совместного распределения временных интервалов проезда автомобилей будем использовать копулу Маршалла – Олкина [14, 16, 23]. Обозначим $\tilde{z} = \arg \min(u_1^{-\theta_1}, \dots, u_Z^{-\theta_Z})$. Тогда копула Маршалла – Олкина примет вид

$$(3) \quad C(u_1, \dots, u_Z; \theta_1, \dots, \theta_Z) = u_{\tilde{z}}^{-\theta_{\tilde{z}}} \prod_{z \in Zones} u_z,$$

для всех $(u_1, \dots, u_Z) \in [0; 1]^Z$, параметры $\theta_z \in [0; 1]$, $z \in Zones$.

Используя формулы (2)–(3), нетрудно показать, что совместная плотность h_{X_1, \dots, X_Z} примет вид

$$(4) \quad h(x_1, \dots, x_Z) = \frac{\partial^Z H}{\partial x_1 \dots \partial x_Z} = \frac{\partial^Z C}{\partial u_1 \dots \partial u_Z} \prod_{z \in Zones} \frac{dF_z(x_z)}{dx_z} = (1 - \theta_{\bar{z}}) f_{\bar{z}}^{-\theta_{\bar{z}}} \prod_{z \in Zones} f_z(x_z).$$

4.2. Описание модуля валидации

В комплексе MARLIN24 предусмотрена имитация показаний оптических датчиков с помощью модуля валидации. Данный модуль предназначен для численного моделирования случайных величин, описывающих интенсивность движения транспортных средств. Использование модуля позволяет увеличить объем выборки при проведении вычислительных экспериментов, а варьирование параметров копулы Маршалла – Олкина позволяет моделировать интенсивность потоков в разное время суток.

Иллюстрация работы оптических датчиков приведена на рис. 4а, зоны работы оптических датчиков – на рис. 4б. Опишем совместное распределение интенсивностей движения на примере полос 1 и 2 рис. 4б.

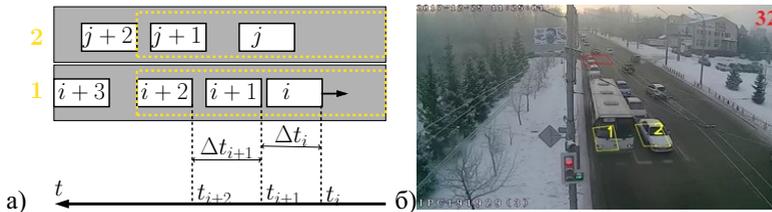


Рис. 4. Принцип работы оптических датчиков: а) скриншот окна комплекса MARLIN24; б) зоны оптических датчиков (жёлтый цвет)

Пусть случайная величина X с функцией распределения $F(x)$ описывает длину временного интервала Δt_i между проездом автомобилей i и $i + 1$ через сечение детектируемого участка на полосах 1. Случайная величина Y с функцией распределения $G(y)$ – длину временного интервала Δt_j на второй

полосе соответственно. Процедура вычисления длин интервалов Δt_i следующая. В таблице фиксируется дискретная временная отметка с частотой обновления детектора 300 с, например t_i . Когда транспортное средство i перемещается в зону обзора оптического датчика, счетчик транспортных средств M увеличивается на 1. Длина временного интервала между проездом двух автомобилей определяется как $\Delta t_i = \frac{300}{M}$. Далее аналогично ведется подсчет транспортных средств для отметки $t_{i+1} = t_i + 300$. Упорядоченные по дискретному разбиению t совокупности интервалов $\{(\Delta t_i, \Delta t_j)\}$ образуют выборку, которая используется для анализа совместного распределения интенсивности движения на разных полосах.

Этап предобработки состоит в оценивании плотности распределений случайных величин X и Y , описывающих число детектируемых транспортных средств на полосах 1 и 2 соответственно на основе данных, полученных с оптических детекторов города Красноярска с 2019 по 2020 год. На первом шаге строятся ядерные оценки плотности с ядром Епанечникова [12]. Далее формулируется упрощающее предположение о том, что каждая из рассматриваемых случайных величин представима в виде смеси нормальных распределений. С использованием EM-алгоритма [7], на вход которого подавались значения ядерной оценки плотности, определяются параметры смесей. Результаты этапа предобработки на 10-й итерации EM алгоритма и вид маргинальных плотностей распределений случайных величин X и Y приведены на рис. 5а и рис. 5б.

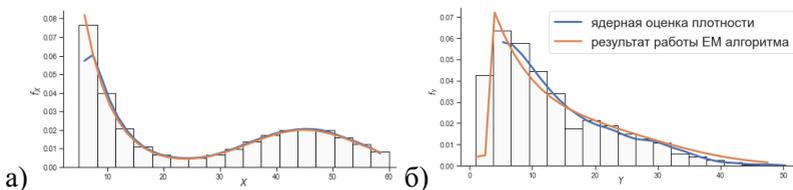


Рис. 5. Результаты этапа предобработки: оценка плотностей маргинальных плотностей: а) X ; б) Y

Для этапа имитационного моделирования среднесуточных интенсивностей была разработана модификация метода дискретной суперпозиции Монте-Карло для генерации значений случайной величины (X, Y) [2, 5].

Гистограмма выборки, полученной моделированием копулой Маршалла – Олкина с параметрами $\bar{\theta} = (\frac{9}{10}, \frac{1}{4})$ и функция плотности распределения (4) представлены на рис. 6.

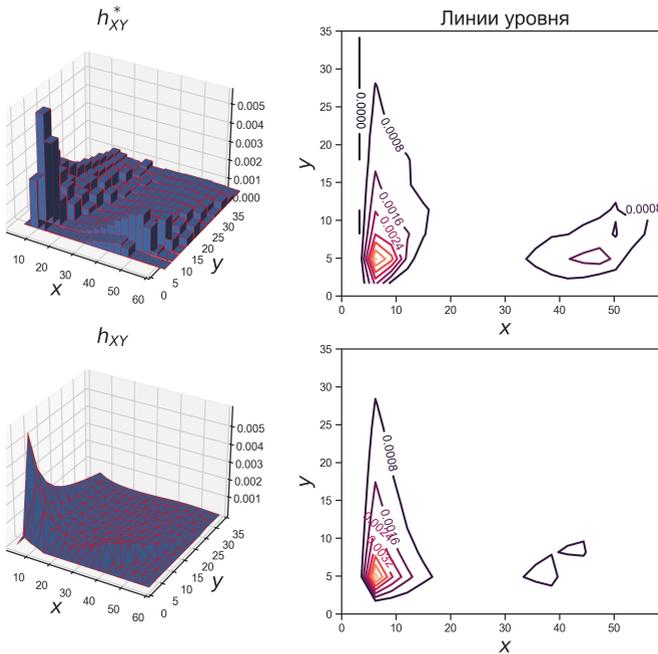


Рис. 6. Сравнение эмпирической h_{XY}^* с параметрами $\bar{\theta} = (\frac{9}{10}, \frac{1}{4})$ и теоретической h_{XY} функции плотности с.в. (X, Y)

Сравнение значений выборки с моделируемыми значениями (X, Y) приведено на рис. 7. Полученная выборка (точки на графике справа) используется далее в комплексе MARLIN24 при генерации виртуальных транспортных средств.

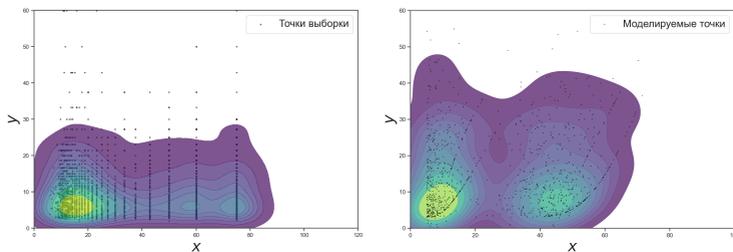


Рис. 7. Сравнение моделируемых значений и выборки (X, Y)

5. Модуль адаптивного управления светофорными объектами

Опишем задачу управления светофорным объектом как задачу управления агентом в стохастической среде. Агент (светофорный объект) не располагает ресурсами и решает задачу целесообразности активации той или иной фазы. Обозначим множество всех действий агента символом \mathcal{A} . Среда – детектируемые перекрестки с оптическими датчиками, которые распознают машины на отрезках дорог за сто метров до стоп-линий. Состояние среды отражает активность фаз светофорных объектов и время, которое машины находятся в детектируемой зоне. Обозначим множество всех состояний символом \mathcal{S} .

В качестве математической модели сети светофоров в работе рассматривается управляемый марковский процесс с конечным числом действий и состояний. Таким образом, проблема управления светофорными объектами сводится к задаче мультиагентного обучения с подкреплением (Multiagent Reinforcement Learning).

5.1. Задача мультиагентного обучения с подкреплением для светофорных объектов

Опишем поведение светофорных объектов (агентов) с помощью марковского процесса принятия решений $\langle \mathcal{S}, \mathcal{A}, \mathbb{P}, r \rangle$ [18]. Процесс принятия решений для агента будет выглядеть следующим образом. В момент времени t агент наблюдает состояние среды $s_t \in \mathcal{S}$ и выбирает действие $a_t \in \mathcal{A}$. Среда отвечает гене-

рацией наград $R_t = r(s_t, a_t)$ и переходит в состояние $s_{t+1} = s'$ с вероятностью $p(s' | s_t, a_t)$ согласно матрице переходов \mathbb{P} .

Функция оценки эффективности применяемого управления $\delta = \{a_t, t \in \mathbb{N}\}$, составляющая траекторию процесса $\mathcal{T} = \{s_0, a_0, s_1, a_1, \dots, s_T, a_T\}$, получается как функция

$$(5) \quad V = \sum_{t=0}^{\infty} \gamma^t r(s_{t+1} | s_t, \delta_t) = \lim_{T \rightarrow \infty} \mathbb{E}_{\mathcal{T}} \sum_{t=0}^T \gamma^t R_t,$$

где величина $0 < \gamma < 1$ называется коэффициентом переоценки и показывает, во сколько раз уменьшается отложенное вознаграждение за один временной шаг. Переоценка задает приоритет получения награды в ближайшее время перед получением той же награды через некоторое время. Математический смысл условия $0 < \gamma < 1$ состоит в том, чтобы гарантировать ограниченность функционала V .

Формальная постановка задачи вычисления оценки эффективности управления светофорным объектом представлена ниже.

- Дано: марковский процесс принятия решения $\langle \mathcal{S}, \mathcal{A}, \mathbb{P}, r \rangle$ для управления светофорным объектом, активная в начальный момент времени фаза светофорного объекта s_0 .
- Найти: управление светофорного объекта $\delta^* = \{a_t^* \}_{0 \leq t < \infty}$, которое доставит максимум функции оценки его эффективности (5).

Решение задачи поиска оптимального совокупного управления светофорными объектами дорожной сети ищется методом динамического программирования согласно принципу оптимальности Вальда – Беллмана. В задаче управления фазами светофорного объекта уравнение Вальда – Беллмана имеет вид [20]

$$(6) \quad V^* = \max_{a \in \mathcal{A}} \sum_{s' \in \mathcal{S}} p(s' | s, a) (r(s, a) + \gamma V^*(s')).$$

Формула (6) может быть переписана в итерационной записи, называемой Q -обучением [18]. Функция суммарных вознаграждений при оптимальном управлении на шаге t имеет вид

$$V^* (\{s_{t'}, \delta\}_{t' \in \mathbb{N}, t' \leq t}) = \max_{a \in \mathcal{A}} Q_t(s_t, a).$$

Считаем, что нам известно состояние среды s_{t+1} и оптимальное управление a_{t+1} на шаге $t + 1$, соответствующий итерации l , и условимся, что итерация Q идет по индексу l , тогда функция Q для агента представима в рекурсивном виде:

$$\begin{aligned} Q_{l+1}(s, a) &= \underbrace{p(s_{t+1}|s, a)}_{\alpha_l} \left(r_{t+1} + \gamma V^*(s_{t+1}) \right) + \\ &+ \underbrace{\sum_{s' \in \mathcal{S}/s_{t+1}} p(s'|s, a)}_{1-\alpha_l} \left(r(s'|s, a) + \gamma V(s') \right) = \\ &= \alpha_l \left(r_l + \gamma \max_{a' \in \mathcal{A}} Q_l(s_{t+1}, a') \right) + (1 - \alpha_l) Q_l(s, a). \end{aligned}$$

Цель поиска оптимального совокупного управления светофорными объектами дорожной сети заключается в увеличении максимального совокупного вознаграждения, определяемого функцией Q . Для задачи поиска оптимального управления светофорным объектом справедливы следующие утверждения [20]:

- существует единственное точное решение;
- оценка точности приближенного решения на n -м шаге имеет вид $\rho(Q_n, Q_0) \leq \frac{\gamma^n \rho(Q_1, Q_0)}{1 - \gamma}$, где $Q_l \in \mathbb{R}_{\infty}^{|\mathcal{A}|+|\mathcal{S}|}$ - векторы значений $Q(s, a)$ на шаге l ; $\forall q, w \in \mathbb{R}_{\infty}^{|\mathcal{A}|+|\mathcal{S}|}$ расстояние $\rho(q, w) = \max_{j \in \mathbb{N}, j \leq |\mathcal{A}|+|\mathcal{S}|} |q_j - w_j|$;

- приближенное решение находится согласно формулам

$$\begin{aligned} V^*(s) &= \max_{a \in \mathcal{A}} \lim_{l \rightarrow +\infty} Q_l(s, a), \\ a_t(s) &= \arg \max_{a' \in \mathcal{A}} Q_l(s, a'). \end{aligned}$$

5.2. Описание модуля управления светофорными объектами

Схема подсчета функции оценки эффективности управления управления светофорными объектами представлена на рис. 8.

В имитационной среде (SIMULATION) моделируются транспортные потоки с интенсивностями, полученными в модуле валидации, автомобили перемещаются в имитационной среде (SIMULATION), пока не выйдут из ее зоны покрытия.

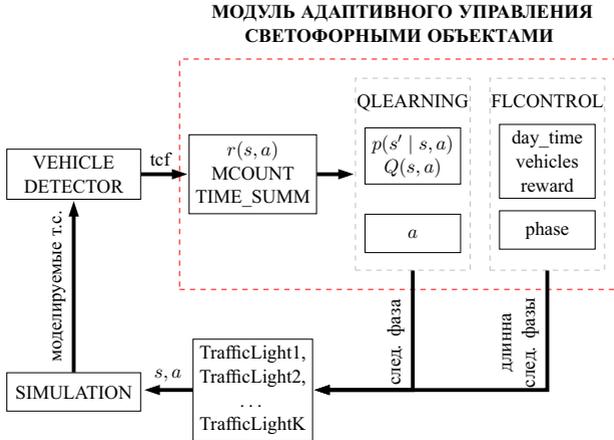


Рис. 8. Схема подсчета функции оценки эффективности управления управления светофорными объектами в комплексе MARLIN24

При попадании машины на детектируемый участок дорожной сети z во вспомогательном модуле, имитирующем поступление информации с оптических датчиков (VEHICLE DETECTOR), пары, состоящие из указателей на объект машины и текущего времени модели, добавляются в одну из коллекций tcf (time collection forward) для выбранного вручную множества светофорных объектов TrafficLight1, ..., TrafficLightK. На следующем шаге симуляции машины удаляются из коллекции tcf при проезде через зону. В течении периода времени Δt во вспомогательном модуле выбора управления вызывается процесс переключения фаз (QLEARNING), реализующий управление согласно выбранной стратегии совокупного управления.

В качестве альтернативы процессу переключения фаз светофора (QLEARNING), используемому по умолчанию в модуле адаптивного управления светофорными объектами, можно выбрать изменение длительности фазы в следующем цикле. В работе [20] данный подход был успешно реализован с помощью контроллера нечеткой логики [24] (FLCONTROL).

6. Вычислительные эксперименты

Для исследования представленных в работе моделей были проведены серии вычислительных экспериментов. Эксперименты проводились на ПК с процессором Intel Core i7-10510U CPU@1.80ГГц и оперативной памятью объемом 8ГБ.

Опишем подробнее процесс подсчета политики управления $\hat{\pi}(a|s)$, опираясь на структурную схему комплекса MARLIN24 на рис. 1. Модуль адаптивного управления светофорными объектами загружает управляющий конфигурационный файл `trafficLightConfig.xml`. В конфигурационном файле содержится информация о возможных направлениях движения, количестве фаз и циклах светофорных объектов. Далее комплекс программных средств MARLIN24 связывает показания датчика в имитационном модуле и рассчитывает оптимальное управление для светофорных объектов.

Оптические датчики (VEHICLE DETECTOR) в имитационной среде (SIMULATION) записывают момент появления t_i пронумерованного транспортного средства $i \in I \subset \mathbb{N}$ в зоне $z \in Zones = \{z^{(0)}, z^{(1)}, \dots, z^{(m)}\}$, $m \in \mathbb{N}$. Отметим, что при имитационном моделировании псевдослучайная интенсивность движения транспортных средств будет задана алгоритмически. Это означает, что мы можем сконструировать множество пар (i, z) таких, что автомобиль i находится в детектируемой зоне z в момент времени t . Определим данное множество как отношение $\psi_t \subset \mathbb{N} \times Zones$, для которого $i\psi_t z$. Введем также отношение $\phi \subset Zones \times \mathcal{S}$, описывающее зоны z , в которых состояние s' разрешает движение. Сгруппируем автомобили в зонах в соответствии с фазой светофорного объекта s' , которая разрешает движение транспортных средств в этих зонах, и обозначим $I(s', t) = \{i \mid t_i < t, i\psi_t z, z\phi s'\}$.

Приведем рассуждения, исходя из которых считается функция вознаграждения. Для каждой полосы определено число машин на отрезке дороги, начинающемся с детектора и заканчивающемся стоп-линией перекрестка. Пусть $r : \mathcal{S} \times \mathcal{A} \mapsto \mathbb{R}$ – функция

вознаграждения агента при изменении наблюдаемого состояния s_t при действии $a_t = \delta(s_t)$. В момент времени t для каждого состояния s_t значение функции $r(s_t, a_t) = R_t$ определяется как суммарное время, затраченное всеми машинами из $I(s', t)$ на ожидание активации состояния s' действием a_t . Поскольку цель агента – минимизация задержки трафика автомобилей, то награда агента за активацию состояния s' , которое приводит к разгрузке детектируемой зоны из $z\phi s'$, равна длительности нахождения машин на ней, т.е. $R_t = \sum_{i \in I(s', t)} (t - t_i)$.

Далее для построенного множества светофорных объектов TL и зон детекции z вычисляются функция наград $r(s, a)$, число проехавших машин (MCOUNT), суммарное время проезда через детектируемые участки дорожной сети (TIME_SUMM) и обучающие функции Q . При симуляции, формируется двумерная выборка $\mathcal{X} = \{(s_i, a_i)\}_{i=1}^T$ объемом T порядка 10^6 . В результате управления δ^* , принятого из соображений увеличения значения функции оценки эффективности Q , рассчитывается несмещенная оценка распределения $\mathcal{P} = \{p(s, a) \mid s \in \mathcal{S}, a \in \mathcal{A}\}$ двумерной случайной величины (s, a) , где функция $p(s, a)$ – вероятность того, что в состоянии s агент принял решение a . На основании выборочных вероятностей $\hat{p}(s, a)$ вычисляются оценки политики агента $\hat{\pi}(a|s)$ для каждого $s \in \mathcal{S}$

$$\hat{\pi}(a|s) = \frac{\hat{p}(s, a)}{\sum_{a \in \mathbb{A}} \hat{p}(s, a)} = \frac{\hat{p}(s, a)}{\hat{p}(s)}.$$

В ходе серии из 1000 симуляций были построены усредненные кривые обучения функции оценки эффективности управления \hat{Q} при равновесной по Нэшу стратегии [15]. Кривые обучения Q^0 и Q^1 для первого и второго светофорных объектов соответственно приведены на рис. 9. Синим цветом отмечены графики принимаемых значений Q для каждой эпохи по модельному времени t , красным цветом – их усредненные значения \hat{Q} . В результате эффективного управления время ожидания транспортного средства в среднем не превышает длины цикла светофорно-

го объекта. Также было продемонстрировано, что значительного улучшения управления при расширении покрытия дорожной сети может и не быть. Таким образом, координированное управление светофорными объектами в целях ускорения вычислений может быть рассмотрено только в тех участках, где его применение дает ощутимое улучшение в управлении. В остальных случаях может быть рассмотрен некоординированный подход, и, следовательно, «проклятие размерности», возникающее с ростом размерности матриц при вычислениях, не является серьезной проблемой.

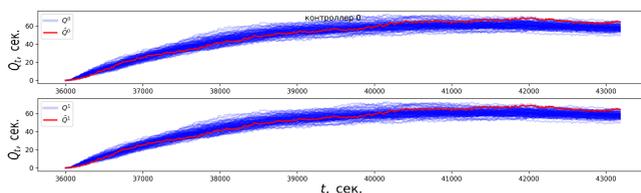


Рис. 9. Пример сходимости кривых обучения Q^0 и Q^1 для двух светофорных объектов

С помощью разработанного программного комплекса было проведено сравнение четырех подходов [19]: координированного управления (MARLIN), некоординированного управления (MARL), фиксированного плана (FIXED), управления с переменной длительностью фаз (FUZZY). Координированное управление (MARLIN) продемонстрировало улучшение среднего времени проезда через детектируемые участки дорожной сети на 30%.

Сравнительный анализ эффективности управления на исторических данных в [20] показал, что среднее время ожидания участников движения при использовании MARLIN24 сопоставимо с результатами системы АСУДД24.

В таблице 3 приведено сравнение различных показателей эффективности, рассчитанных на модельных и исторических данных, для управления светофорными объектами участка дорожной сети г. Красноярска. Показатели эффективности рассмотрены для различных режимов управления: координированного адаптивного управления светофорными объектами участка дорожной

сети (MARLIN), некоординированного адаптивного управления (MARL), для светофорных объектов с фиксированным координационным планом (FIXED) и с нечетким контроллером (FUZZY).

Таблица 3. Сравнение показателей эффективности управления на моделируемых и исторических данных

Модель	Среднее ожидание, ($\frac{c}{\text{маш.}}$)		M COUNT, (маш.)		TIME_SUMM, (c)	
	Мод.	Истор.	Мод.	Истор.	Мод.	Истор.
FIXED	12,7	13,0	5412	5394	68771	69880
MARL	12,04	12,8	5415	5402	64911	69347
MARLIN	11,8	12,4	5415	5400	63817	66804
FUZZY	12,8	12,9	5413	5396	69089	69637

7. Заключение

В статье описан программный комплекс MARLIN24 для адаптивного управления сетью светофорных объектов. Комплекс имеет модульную архитектуру, включающую следующие компоненты: адаптивное управление, симуляция, валидация и визуализация. Предложенные модели MARL и MARLIN превосходят фиксированный план управления светофорными объектами по всем показателям производительности и являются одинаково эффективными при отсутствии совместных решений. Это подтверждает конкурентоспособность MARLIN24 для разгрузки сложных участков дорожной сети.

В данной работе отличительной чертой является применение копул Маршалла – Олкина при моделировании транспортных потоков. Модуль валидации адаптирован для эмпирических данных города Красноярска. В условиях ограниченной эмпирической информации это позволило, во-первых, параметрически описать зависимость случайных величин, во-вторых, упростить моделирование случайных величин, описывающих интенсивность движения по полосам, и, в-третьих, увеличить объём выборки моделируемых данных.

Литература

1. АСУДД «МИКРО–М».[Электронный ресурс]. – Режим доступа: <http://asud55.ru/archives/1346>.
2. ВОЙТИШЕК А.В. *Основы метода Монте-Карло: Учебное пособие* / Под ред. В.Н. Задорожного. Новосибирск: НГУ, 2010. – 108 с.
3. ЗАЛИЗНЯК В.Е. *Численные методы. Основы научных вычислений: учебное пособие для бакалавров.* – М.: ЮРАЙТ, 2012. – С. 356.
4. ТИСЛЕНКО Т.И. *Моделирование интенсивностей транспортных потоков при помощи копул Маршала – Олкина* // ИТ. НАУКА. КРЕАТИВ. Т. 5. Системы управления, информационные технологии и математическое моделирование: материалы I Междунар. форума (Омск, 14–16 мая 2024 г.): в 5 ч. / Под науч. ред. П.С. Ложникова, отв. ред. И.Г. Ольгина. – Омск: Изд-во ОмГТУ, 2024. – С. 70.
5. ТИСЛЕНКО Т.И. *Моделирование интенсивностей транспортных потоков для модуля валидации комплекса MARLIN24* // Информационные технологии и математическое моделирование (ИТММ-2024): Материалы XXIII Международной конференции имени А.Ф. Терпугова. – Томск: Изд-во ТГУ, 2024. – С. 336–341.
6. *Транспортная стратегия Российской Федерации.* Утверждена распоряжением Правительства Российской Федерации от 22 ноября 2008 года №1734-р. 2008 [Электронный ресурс]. – Режим доступа: <http://mintrans.gov.ru>.
7. BISHOP С.М. *Pattern Recognition and Machine Learning.* – Springer, 2007.
8. BOEING G. *OSMnx: Python for Street Networks.* [Электронный ресурс]. – Режим доступа: <https://osmnx.readthedocs.io/>.

9. CARINI RAYMOND N. *Application of the UTCS-1 Network Simulation Model to Select Optimal Signal Timings in a Multi-Linear Street System* // Interim Report. Publication for Urban Traffic Control System. Joint Highway Research Project. – 1977. – P. 164.
10. CHANDLER M.J.H. *SCOOT and Bus Detection. OTRC Proc. Annual Summer Meeting* // Traffic Control Studies in London. – 1990. – Vol. P269. – P. 111–128.
11. EL-TANTAWY S., ABDULHAI B., ABDELGAWAD H. *Multiagent Reinforcement Learning for Integrated Network of Adaptive Traffic Signal Controllers (MARLIN-ATSC): Methodology and Large-Scale Application on Downtown Toronto* // IEEE Trans. on Intelligent Transportation Systems. – 2013. – Vol. 14, No. 3. – P. 1140–1150.
12. EPANECHNIKOV V.A. *Non-Parametric Estimation of a Multivariate Probability Density* // Theory of Probability and Its Applications. – 1969. – P. 153–158.
13. GARTNER N.H. *OPAC: Strategy for Demand-responsive Decentralized Traffic Signal Control* // IFAC Proc. Volumes. – 1990. – Vol. 23. – P. 241–244.
14. MARSHALL A.W., OLKIN I. *Families of Multivariate Distributions* // Journal of the American Statistical Association. – 1988. – Vol. 83, No. 403. – P. 834–841.
15. NASH J. *Equilibrium Points in N-Person Games.* // Proc. of the National Academy of Sciences. – 1950. – Vol. 36, No. 1. – P. 48–49.
16. NELSEN R. B. *An Introduction to Copulas.* – Springer, 2006.– 270 p.
17. NETWORKX DEVELOPERS. *NetworkX.* [Электронный ресурс]. – <https://networkx.org/>.
18. SUTTON R.S., BARTO A.G. *Reinforcement Learning: An Introduction.* – Cambridge, MA: MIT Press, 2015.–P. 145–147.
19. TISLENKO T.I., SEMENOVA D.V., SERGEEVA N.A. et al. *Multiagent Reinforcement Learning for Integrated Network: Applying to a Part of the Road Network of Krasnoyarsk*

- City* // IEEE 16th Int. Conf. on Application of Information and Communication Technologies (AICT-2022). – 2022. – P. 1–5.
20. TISLENKO T.I., SEMENOVA D.V., SOLDATENKO A.A. *Modeling and Comparison of Different Management Approaches on the Intersections Network* // IEEE 26th Int. Conf. Distributed Computer and Communication Networks: Control, Computation, Communications (DCCN-2023). – 2023. – P. 25–29.
21. TREIBER M., HENNECKE A., HELBING D. *Congested traffic states in empirical observations and microscopic simulations* // Transportation Physics Reviews E. – 2000. – Vol. 62. – P. 1805–1824.
22. TREIBER M., HELBING D. *Realistische Mikrosimulation von StraBenverkehr mit einem einfachen Modell* // Symposium "Simulationstechnik ASIM-2002 Rostock. – 2002. – P. 514–520.
23. QUESADA-MOLINA J.J., RODRIGUEZ-LALLENA J.A. *Bivariate copulas with quadratic sections* // Journal of Nonparametric Statistics. – 1995. – Vol. 5, No. 4. – P. 323–337.
24. ZADEH L.A. *Fuzzy Sets, Fuzzy Logic, Fuzzy Systems*. – World Scientific Press, 1996.

DEVELOPMENT OF AN ADAPTIVE TRAFFIC LIGHT CONTROL SYSTEM USING MARKOV DECISION PROCESSES

Timofey Tislenko, Siberian Federal University, Krasnoyarsk, postgraduate student (timtisko@mail.ru).

Darya Semenova, Siberian Federal University, Krasnoyarsk, Cand.Sc. (Physics and Mathematics), assistant professor (DVSemenova@sfu-kras.ru).

Abstract: The paper presents the results of development of the MARLIN24 software suite, designed to implement adaptive control of traffic light systems. The primary objective of the development is to optimize the operation of traffic signals in order to minimize the total time vehicles spend within the detection zones of optical sensors. The architecture of the software suite comprises three principal modules: the adaptive control module, the traffic flow simulation module, and the validation module, complemented by an additional visualization module. The adaptive control module integrates four control approaches: fixed-time planning, uncoordinated reinforcement learning, coordinated multi-agent reinforcement learning, and phase duration control based on a MISO fuzzy logic controller (Multiple Input Single Output, MISO). Traffic flow simulation for performance evaluation is conducted via a microsimulation module utilizing the "Intelligent Driver Model" (IDM). The validation module employs copula functions to generate realistic optical sensor data reflecting actual traffic conditions, with marginal distributions derived from historical traffic intensity data collected during 2019–2020. The MARLIN24 software suite facilitates the analysis and comparison of multiple traffic control models on real-world sections of the road network in Krasnoyarsk under various conditions.

Keywords: traffic flow simulation, IDM model, MOBIL model, Marshall – Olkin copulas, reinforcement learning.

УДК 519.1 + 519.2 + 519.6 + 519.8(075) + 004.42 + 519.85
ББК 221.7

*Статья представлена к публикации
членом редакционной коллегии Б.Р. Андриевским.*

*Поступила в редакцию 24.01.2025.
Дата опубликования 31.05.2025.*